BOUNDS FOR LATTICE DISTRIBUTIONS HAVING MONOTONE HAZARD RATE WITH APPLICATIONS

by

Gamanlal P. Shah

OPERATIONS RESEARCH CENTER
COLLEGE OF ENGINEERING

EVALUATION COPY
UNIVERSITY OF CALIFORNIA - BERKELEY
Bounds for Lattice Distributions Having Monotone Hazard Rate With Applications

by

Gamanlal P. Shah
Operations Research Center
University of California, Berkeley

June 1965

This research has been partially supported by the Office of Naval Research under Contract Nonr-3656(18) with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.
Summary

Using methods which have been employed for obtaining bounds on continuous failure distributions, bounds on lattice distributions are derived under the increasing (decreasing) failure rate assumptions. The discrete bounds are convenient in a number of applications such as life table analysis and reliability theory.
1. Introduction

A discrete distribution of a random variable \( X \) is called a lattice distribution if there exist real numbers \( a \) and \( h > 0 \) such that every possible value of \( X \) can be represented in the form \( a + kh \), where \( k \) runs through integer values. We call \( h \) the span of the distribution. The geometric, binomial and Poisson distributions are important examples of lattice distributions.

It seems that there is essentially no literature on Chebyshev type bounds for lattice distributions. Of course, the classical Chebyshev bounds are attained by discrete distributions. However, these need not be lattice. The general method used by Karlin and Studden and others [12, 11] for obtaining Chebyshev type bounds can, of course, in theory be applied to lattice distributions. However, we will restrict ourselves to a special class of lattice distributions where these methods do not provide sharp bounds.

For convenience assume that \( a = 0 \), \( h = 1 \) and let \( p_k = P[X = k] \); \( k = 0, 1, 2, \ldots \). Define the failure rate \( r(k) \) for lattice distributions as

\[
r(k) = \frac{p_k}{\sum_{j=k}^\infty p_j} \quad \text{if} \quad \sum_{j=k}^\infty p_j > 0
\]

and note that \( 0 \leq r(k) \leq 1 \). Let \( F(k) = \sum_{j=0}^{k-1} p_j \) and \( F(k) = \sum_{j=k}^\infty p_j \). We say that \( F \) is lattice IFR (lattice DFR)
If and only if \( r(k) \) is non-decreasing (non-increasing) in \( k \).

Note that it is shown in [3] that \( F \) is lattice IFR (lattice DFR)
iff \( \log F(k) \) is concave (convex). Further equivalent definitions
of lattice IFR (lattice DFR) distributions in terms of Pólya
frequency function of order two and total positivity of order two have
been discussed in [1,2,3,10]. Note that the negative binomial
distribution

\[
p_k = \binom{k-1}{\alpha-1} p^\alpha (1-p)^{k}\alpha ; \quad \alpha > 0 \\
k = 0,1,2,\ldots
\]

is lattice IFR (increasing failure rate) for \( \alpha > 1 \) and lattice DFR
(decreasing failure rate) for \( \alpha < 1 \). For \( \alpha = 1 \), it is the
geometric distribution with a constant failure rate and hence it
belongs to both IFR and DFR classes. The geometric distribution
is the only distribution with this property in the discrete case.
The binomial and Poisson distributions are lattice IFR.

Lattice distributions occur naturally in many applications.
Fatigue failure data obtained under dynamic loading is commonly
recorded in terms of the number of cycles to failure. For many
structures subject to fatigue failure, a lattice IFR distribution
would seem appropriate. A mathematical model describing such types
of failure has been discussed in [5]. As another example we mention
human mortality data where it is common to record time of death
to the nearest year or the nearest month. For certain types of
cancer mortality data a decreasing failure rate seems appropriate,[4].

In any situation where grouped data is derived from a continuous
IFR(DFR) distribution, a lattice IFR (lattice DFR) distribution will be
appropriate.
2. Bounds on Lattice Distributions

2.1 IFR Bound:

Let $F$ denote a left continuous distribution function such that $F(0^-) = 0$. Bounds on $1 - F$ under the IFR (DFR) assumption in the continuous case when the first moment $\mu_1$ is given can be found in [2]. In the IFR case, the lower bound is given by Theorem 1.

**Theorem 1:** Let $F$ be a left continuous distribution. Let $F$ be IFR with mean $\mu_1$, then

$$F(t) \geq \begin{cases} e^{-\frac{t}{\mu_1}} ; & t < \mu_1 \\ 0 ; & t \geq \mu_1 \end{cases}$$

The inequality is sharp.

A similar bound exists for lattice distributions and we shall discuss this after proving the following lemma.

**Lemma:** Let $F$ be a lattice distribution with mean $\mu_1$. Assume that $F$ is IFR and let $s_k = \sum_{i=1}^{k} F(i)$ then for $t \leq \mu_1$

$$F(t) \geq F(t-k) \left( 1 - \frac{F(t-k)}{\mu_1 - s_{t-k}} \right)^k ; k = 1, 2, \ldots, t-1.$$

**Proof:** We shall prove (1) by induction. By definition,

$$\bar{F}(k) = \sum_{i=k}^{\infty} P_i,$$

so that $\log \bar{F}(k) = \log \left( \sum_{i=k}^{\infty} P_i \right)$. Note that the function $\log \bar{F}(k)$ is defined only for integer values. By linear interpolation we can define $\log \bar{F}(k)$ for any real value of $k$ as shown in Figure 2.1.
Since we assume that \( F \) is IFR, it follows that \( \log F(k) \) is concave. Hence there exists a supporting line to \( \log F(k) \) as shown in Figure 2.1.
FIGURE 2.1 SHOWING SUPPORTING LINE FOR $\log \bar{F}(k)$
Thus by log concavity, we have

\[
\log \bar{F}(k) \leq (L_{i+1} - L_i)k - iL_{i+1} + (i+1)L_i \quad k = i, i+1, \ldots
\]

But \( L_i = \log \bar{F}(i) \) and \( L_{i+1} = \log \bar{F}(i+1) \), therefore

\[
\bar{F}(k) \leq \left[ \frac{\bar{F}(i+1)}{\bar{F}(i)} \right]^k \frac{[\bar{F}(i)]^{i+1}}{[\bar{F}(i+1)]^i} \quad \text{for} \quad k = i, i+1, \ldots
\]

Let \( s_k = \bar{F}(1) + \bar{F}(2) + \ldots + \bar{F}(k) \) where \( s_0 = 0 \). Then we have

\[
\sum_{k=1}^{\infty} \bar{F}(k) = \sum_{k=1}^{i-1} \bar{F}(k) + \sum_{k=i}^{\infty} \bar{F}(k)
\]

\[
\leq s_{i-1} + \sum_{k=1}^{\infty} \left[ \frac{\bar{F}(i+1)}{\bar{F}(i)} \right]^k \frac{[\bar{F}(i)]^{i+1}}{[\bar{F}(i+1)]^i}
\]

\[
= s_{i-1} + \frac{\bar{F}(1)}{\bar{F}(i+1)} \sum_{k=1}^{\infty} \left[ \frac{\bar{F}(i+1)}{\bar{F}(i)} \right]^{k-1}
\]

Thus \( \mu_1 \leq s_{i-1} + \frac{\bar{F}(1)}{\bar{F}(i+1)} \). Solving for \( \bar{F}(i+1) \), we obtain

\[
(2) \quad \bar{F}(i+1) \geq \bar{F}(1) \cdot \frac{\bar{F}(1)}{\mu_1 s_{i-1}}
\]

From (2) with \( i = t-1 \), it follows that (1) holds for \( k = 1 \). Now assume that (1) is true for \( k = n-1 \); \( n < t \) i.e.;

\[
(3) \quad \bar{F}(t) \geq \bar{F}(t-n+1) \left( 1 - \frac{\bar{F}(t-n+1)}{\mu_1 s_{t-n}} \right)^{n-1} = \varphi(\bar{F}(t-n+1))
\]

where \( \varphi(x) = x(1 - \frac{x}{\mu_1 \varepsilon})^{n-1} \).

-6-
Note that \( \phi'(x) = \left(1 - \frac{x}{\mu_1 - s_{t-n}}\right)^{n-2} \left(1 - \frac{nx}{\mu_1 - s_{t-n}}\right) \)

\[ \geq 0 \quad \text{if} \quad x \leq \frac{\mu_1 - s_{t-n}}{n} \]

**To Show:** \( x \leq \frac{\mu_1 - s_{t-n}}{n} \)

By definition \( x = \bar{F}(t-n+1) \leq 1 \), hence we need only show \( \frac{\mu_1 - s_{t-n}}{n} \geq 1 \).

But \( s_{t-n} \leq t-n \) implies \( \frac{\mu_1 - s_{t-n}}{n} \geq \frac{\mu_1 - t+n}{n} \geq 1 \) since \( t \leq \mu_1 \)

by assumption.

Thus we can substitute for \( \bar{F}(t-n+1) \) in (3) the bound given

by (2) with \( i=t-n \) and obtain

\[
\bar{F}(t) \geq \bar{F}(t-n) \left(1 - \frac{\bar{F}(t-n)}{\mu_1 - s_{t-n-1}}\right) \left[1 - \frac{\bar{F}(t-n) \left(1 - \frac{\bar{F}(t-n)}{\mu_1 - s_{t-n-1}}\right)}{\mu_1 - s_{t-n}}\right]^{n-1}
\]

Noting \( s_{t-n} = s_{t-n-1} + \bar{F}(t-n) \) and letting \( c = 1 - \frac{\bar{F}(t-n)}{\mu_1 - s_{t-n-1}} \),

the above inequality becomes

\[
\bar{F}(t) \geq \bar{F}(t-n) c \left[\frac{\left(1 - \frac{\bar{F}(t-n)}{\mu_1 - s_{t-n-1}}\right)^2}{\mu_1 - s_{t-n-1}}\right]^{n-1}
\]

\[ = \bar{F}(t-n) c \left[\frac{c^2}{c}\right]^{n-1} = c^n \bar{F}(t-n) .
\]

This is (l) for \( k = n \), and the induction is complete.

**Theorem 2:** If \( F(t) \) is a lattice IF distribution with mean

\( \mu_1 \geq 1 \) placing mass \( p_k \) at \( k = 0,1,2,\ldots \) and \( \bar{F}(t) = \Sigma_{k=t}^{\infty} p_k \); then
F(t) ≥ \begin{cases} 
\left( \frac{\mu_1}{1+\mu_1} \right)^t & \text{for } t \leq \mu_1 \\
\frac{\mu_1 - [\mu_1]}{1+\mu_1 - [\mu_1]} & \text{for } t = [\mu_1] + 1 \\
0 & \text{for } t > [\mu_1] + 1 
\end{cases}

Where \([\mu_1]\) denotes the greatest integer contained in \(\mu_1\). The inequality is sharp.

**Proof**: Since \(\log \overline{F}(k)\) is concave in \(k\) and noting figure 2.1 for \(i=0\), we see that

\[
\log \overline{F}(k) < L_1 k
\]
or
\[
\overline{F}(k) < [\overline{F}(1)]^k
\]

Thus
\[
\mu_1 = \sum_{k=1}^{\infty} \overline{F}(k) < \sum_{k=1}^{\infty} [\overline{F}(1)]^k = \frac{\overline{F}(1)}{1-\overline{F}(1)}
\]

(4) or \(\overline{F}(1) \geq \frac{\mu_1}{1+\mu_1}\)

Now consider (1) with \(k = t-1\); this gives

\[
\overline{F}(t) \geq \overline{F}(1) \left[ 1 - \frac{\overline{F}(1)}{\mu_1} \right]^{t-1} = \psi(\overline{F}(1))
\]

Note that \(\psi'(x) = (1 - \frac{x}{\mu_1})^{t-2}(1 - \frac{tx}{\mu_1}) ≥ 0\) if \(t ≤ \mu_1\) and \(\mu_1 ≥ 1\) (since \(x ≤ 1\)), and thus from (4), we have

\[
\overline{F}(t) ≥ \psi(\frac{\mu_1}{1+\mu_1}) = \left( \frac{\mu_1}{1+\mu_1} \right)^t \text{ for } t ≤ \mu_1
\]
This bound for \( t \leq \mu_1 \) is attained by the geometric distribution; i.e.,

\[
p_k = \frac{1}{1+\mu_1} \left( \frac{\mu_1}{1+\mu_1} \right)^k ; \quad k = 0, 1, 2, \ldots
\]

We conjecture that

\[
\bar{F}(t) \geq \begin{cases} 
\frac{\mu_1 - [\mu_1]}{1+\mu_1 - [\mu_1]} & \text{for } t = [\mu_1] + 1 \\
0 & \text{for } t > [\mu_1] + 1
\end{cases}
\]

where \([\mu_1]\) denotes the greatest integer contained in \(\mu_1\).

For \( t > [\mu_1] + 1 \), it is easily seen that the bound is attained by

\[
\bar{G}(t) = \begin{cases} 
1 & ; \quad t \leq [\mu_1] \\
\frac{\mu_1 - [\mu_1]}{1+\mu_1} & ; \quad t = [\mu_1] + 1 \\
0 & ; \quad t > [\mu_1] + 1
\end{cases}
\]

The upper bound on \( 1 - F \) for the continuous distribution when \( F \) is IFR is given by the following theorem.

**Theorem 3:** Let \( F \) denote a left continuous distribution. Assume that \( F \) is IFR with mean \( \mu_1 \), then

\[
\bar{F}(t) \leq \begin{cases} 
1 & ; \quad t \leq \mu_1 \\
e^{-wt} & ; \quad t > \mu_1
\end{cases}
\]

where \( w \) depends on \( t \) and satisfies \( 1 - \mu_1 = e^{-wt} \). The inequality is sharp [2].
An analogous bound exists for lattice distributions and is given by

**Theorem 4:** Let $F$ denote a lattice distribution. Assume that $F$ is IFR with mean $\mu_1$, then

$$\bar{F}(t) \leq \begin{cases} 1 & ; \; t \leq \mu_1 \\ q^t & ; \; t > \mu_1 \end{cases}$$

where $q$ depends on $t$ and satisfies $q\left(\frac{1-q^t}{1-q}\right) = \mu_1$. The inequality is sharp.

**Proof:** Let us define,

$$G(k) = \begin{cases} q^k & ; \; k = 0, 1, 2, \ldots, t \\ 0 & ; \; k > t \end{cases}$$
FIGURE 2.2 LOGARITHMIC CURVE OF THE FUNCTIONS $\bar{F}(k)$ AND $\bar{G}(k)$
Note that $\log G(k)$ is linear and $\log F(k)$ is concave, hence $F(k)$ can cross $G(k)$ at most once for $k < t$.

If $t > \mu_1$, we can always determine $q$ so that

$$
\sum_{k=1}^{\infty} G(k) = \mu_1 \\
or \\
\sum_{k=1}^{t} q^k = \mu_1
$$

or

$$
q \left( \frac{1-q^t}{1-q} \right) = \mu_1 \text{ which is solvable for } q.
$$

Since, $F(k)$ and $G(k)$ are distribution functions with the same mean $\mu_1$, they must cross at least once for $0 < k \leq t$.

Thus, for this choice of $q$, $F(k)$ necessarily crosses $k$ exactly once from above for $0 < k < t$. Therefore, $F(t) < q^t$ unless $F$ coincides identically with $G$. The degenerate distribution concentrating at $\mu_1$ provides the upper bound for $t \leq \mu_1$ when

$\mu_1$ is an integer.

Since $G$ is also IFR, the bound is sharp.

2.2 DFR Bound:

It is shown in [2] that the sharp lower bound on $1-F$ where $F$ is a left continuous DFR distribution with known mean, say $\mu_1$, is zero and the sharp upper bound is given by the following theorem.

**Theorem 5:** Let $F$ be left continuous. Assume that $F$ is DFR with mean $\mu_1$, then...
\[
\overline{F}(t) \leq \begin{cases} 
\frac{t}{e^{\mu_1}} & ; \quad t < \mu_1 \\
\frac{\mu_1 e^{-1}}{t} & ; \quad t \geq \mu_1 
\end{cases}
\]

The inequality is sharp.

Similarly, for lattice distributions the sharp lower bound on

\[1 - F \] where \( F \) is a lattice DFR distribution with known mean \( \mu_1 \) is zero. For example, consider

\[ \overline{G}(k) = \begin{cases} 
0 & , \quad k < 0 \\
\frac{a}{q}^{-k} & , \quad k = 0, 1, 2, \ldots 
\end{cases} \]

where \( a \) is arbitrarily small and \( 0 \leq q \leq 1 \). Choose \( q \) such that \( G \) has mean \( \mu_1 \). Then \( G \) is DFR with a jump at the origin and \( \overline{G}(k) < a \) for all \( k = 0, 1, 2, \ldots \). Letting \( a \to 0 \), we see that the sharp lower bound is zero.

Thus, for lattice DFR distributions only the upper bound on \( \overline{F}(t) \) is given.

**Theorem 6:** Let \( F \) be a lattice distribution. Assume that \( F \) is DFR with mean \( \mu_1 \), then

\[
\overline{F}(t) \leq \begin{cases} 
\left(\frac{\mu_1}{\mu_1 t}\right)^t & ; \quad t < \mu_1 \\
\left(\frac{t-1}{t}\right)^t & ; \quad t \geq \mu_1 
\end{cases}
\]

The inequality is sharp.
Proof: By definition, \( \overline{F}(k) = \sum_{j=1}^{\infty} p_j \).

We remark that the function \( \log \overline{F}(k) \) is defined only for integer values. By linear interpolation, we can define \( \log \overline{F}(k) \) for any real value of \( k \) as shown in figure 2.3.

Since we assume that \( F \) is DFR, it follows that \( \log \overline{F}(k) \) is convex, and therefore, there exists a supporting line to \( \log \overline{F}(k) \), say, at \( k = t \).
FIGURE 2.3 SHOWING SUPPORTING LINE FOR DFR DISTRIBUTION
Let $\alpha$ denote the intersection of the supporting line with $y$-axis
and let $L = \log F(t)$. Obviously, $L < \alpha \leq 0$ and the equation of
the supporting line is $y = \left(\frac{L-\alpha}{t}\right)^k + \alpha$. Thus we have,

\[
\log F(k) \geq \left(\frac{L-\alpha}{t}\right)^k + \alpha \quad ; \quad k \geq 0
\]

or

\[
F(k) \geq \frac{L-\alpha}{t}^k + \alpha
\]

Hence,

\[
\mu_1 \geq \sum_{k=1}^{\infty} e^{ak+\alpha} = e^{\alpha+a} \sum_{k=1}^{\infty} e^{a(k-1)} = \frac{e^{\alpha+a}}{1-e^a}
\]

since $|e^a| < 1$.

That is

\[
\mu_1 \geq \frac{e^a}{e^{-a}-1}
\]

Solving for $L$, we obtain

\[
L < \alpha - t \log(1 + \frac{e^a}{\mu_1})
\]

Let $\phi(\alpha) = \alpha - t \log(1 + \frac{e^a}{\mu_1})$ so that

\[
L \leq \phi(\alpha) \quad \text{for some } \alpha, \; L < \alpha \leq 0,
\]

or

\[
L \leq \max_{L<\alpha \leq 0} \phi(\alpha)
\]
Note that

\[ \varphi'(\alpha) = 1 - \frac{t}{e^{-\alpha} + 1} \]

and

\[ \varphi''(\alpha) = \frac{-t \mu_1 e^{-\alpha}}{(e^{-\alpha} + 1)^2} < 0. \]

Since \( \varphi''(\alpha) < 0 \), we know that \( \varphi(\alpha) \) is strictly concave and hence it has a maxima, say, at \( \alpha = \alpha^* \).

Thus, from (6) we have,

(7) \[ L \leq \varphi(\alpha^*) \]

so it remains to obtain \( \alpha^* \). \( \alpha^* \) can be found by equating \( \varphi'(\alpha) \) to zero; that is

\[ 1 - \frac{t}{e^{-\alpha} + 1} = 0. \]

This gives us,

\[ \alpha^* = \log\left(\frac{\mu_1}{t-1}\right) \quad \text{for} \quad t \geq \mu_1. \]

Note that for \( t < \mu_1 \),

\[ \varphi'(0) = 1 - \frac{t}{\mu_1 + 1} > 0. \]

Also we have

\[ \varphi''(\alpha) < 0 \quad \text{for any} \quad \alpha \quad \text{where} \quad L < \alpha \leq 0. \]
Thus, for \( t < \mu_1 \), we have \( \phi'(\alpha) > 0 \) and \( \phi''(\alpha) < 0 \) which shows that \( \phi \) is strictly increasing concave function and hence the maximum of \( \phi(\alpha) \) is achieved at the origin.

That is

\[
\alpha^* = \begin{cases} 
0 & \text{if } t < \mu_1 \\
\log\left(\frac{\mu_1}{t-1}\right) & \text{if } t \geq \mu_1 
\end{cases}
\]

Using (8), the expression in (7) becomes,

\[
L \leq \begin{cases} 
\phi(0) & \text{if } t < \mu_1 \\
\phi[\log(\frac{\mu_1}{t-1})] & \text{if } t \geq \mu_1 
\end{cases}
\]

or

\[
L \leq \begin{cases} 
\log\left(\frac{\mu_1}{\mu_1+1}\right)^t & \text{if } t < \mu_1 \\
\log[\left(\frac{\mu_1}{t-1}\right)\left(\frac{t-1}{t}\right)^t] & \text{if } t \geq \mu_1 
\end{cases}
\]

But \( L = \log F(t) \), so finally we have

\[
F(t) \leq \begin{cases} 
\left(\frac{\mu_1}{\mu_1+1}\right)^t & \text{if } t < \mu_1 \\
\left(\frac{\mu_1}{t-1}\right)\left(\frac{t-1}{t}\right)^t & \text{if } t \geq \mu_1 
\end{cases}
\]

We observe that equality in (9) for \( t < \mu_1 \) is attained by the geometric distribution \( P_k = \frac{1}{1+\mu_1} \left(\frac{\mu_1}{1+\mu_1}\right)^k \); \( k = 0,1,2,... \) and for
t \geq \mu_1$ equality in (9) is attained by the geometric distribution $G$ where $G(k) = \left(\frac{\mu_1}{t-1}\right)^k; k = 1, 2, \ldots$. Hence, the inequalities given by (9) are sharp.

2.3 **Comparison of the Continuous and Lattice Bounds:**

Comparing Theorem 1 and Theorem 2, we remark that for $\mu_1$ fixed, the continuous upper bound is everywhere smaller than the lattice upper bound. Similar comparisons hold for Theorem 5 and Theorem 6.
3. Application of Bounds

3.1 Motivation to Life Table Analysis:

We note that the failure rate function, \( r(t) \) is familiar in other branches of science and is known by a variety of names. In the life sciences it is used by actuaries under the name of "Force of Mortality" to compute mortality tables [15]. The important work of life table analysis is to calculate survival rate, i.e., the proportion of persons surviving a specified interval of time. Various methods for calculating survival rate have been discussed in several research papers by Ederer [7]; Ederer, Cutler and Axtell [8], Berkson and Gage [4], etc. Some functions that have been used to define the force of mortality are Gompertz's formula and Makeham's formula. A detailed discussion concerning the estimation of the parameters in Makeham's law of mortality can be found in Grenander's paper [9].

Grenander remarks in his paper that we should specify carefully the probability assumptions to be used in mortality studies. The basic assumption is that our data forms a sample of independent observations of stochastic variables, with the same probability distribution except when sampling errors are serious. Further Grenander remarks in his paper that mortality estimates will be used to compute premiums, reserves and so on, and it is the future mortality that is of interest for making a prognosis for several decades forward in time. Since the development of the mortality depends upon factors that are very difficult to handle, it is likely that only a rough prognosis can be made.
We are interested in giving upper bounds on survival probability which will be useful in the life sciences in order to compare mortality rates in different populations. To give such bounds, which can be used for all ages, we need to have some a priori information about the failure rate function such as IFR or DFR, as we shall see in a cancer study.

3.2 Application of bounds to a cancer study:

Patient survival is generally accepted as the principal criterion for measuring the effectiveness of treatment in cancer. The American College of Surgeons requires the maintenance of a cancer registration and follow up program for approval of a hospital cancer program. We shall investigate a particular type of cancer from a large number of cancer studies given in [6] in order to apply the bounds.

We are interested in giving upper bounds for the survival probability (also called survival rate). Data of table 1 is taken from [6, p. 51f], which represents the observed survival rates for breast and genito-urinary organs cancer for the period 1942-56 of both sexes. For this particular type of cancer there were 41,157 patients at the beginning of the year 1942. The observed survival rates were calculated by the usual formula and is shown in Table 1.

Our aim is to use this data to give an upper bound for the survival rate and to extend it for all ages.
# TABLE - 1

<table>
<thead>
<tr>
<th>Years after diagnosis,</th>
<th>Observed Survival rates in percentage.</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{F}(t)$</td>
<td>$\text{Log}_e \hat{F}(t)$</td>
</tr>
<tr>
<td>0</td>
<td>100.0</td>
<td>4.60517</td>
</tr>
<tr>
<td>1</td>
<td>71.6</td>
<td>4.27110</td>
</tr>
<tr>
<td>2</td>
<td>59.3</td>
<td>4.08261</td>
</tr>
<tr>
<td>3</td>
<td>51.7</td>
<td>3.94546</td>
</tr>
<tr>
<td>4</td>
<td>46.0</td>
<td>3.82864</td>
</tr>
<tr>
<td>5</td>
<td>41.7</td>
<td>3.73050</td>
</tr>
<tr>
<td>6</td>
<td>38.0</td>
<td>3.63759</td>
</tr>
<tr>
<td>7</td>
<td>35.0</td>
<td>3.55535</td>
</tr>
<tr>
<td>8</td>
<td>32.4</td>
<td>3.47816</td>
</tr>
<tr>
<td>9</td>
<td>30.0</td>
<td>3.40120</td>
</tr>
<tr>
<td>10</td>
<td>27.8</td>
<td>3.32504</td>
</tr>
<tr>
<td>11</td>
<td>25.9</td>
<td>3.25424</td>
</tr>
<tr>
<td>12</td>
<td>24.2</td>
<td>3.18635</td>
</tr>
<tr>
<td>13</td>
<td>22.5</td>
<td>3.11352</td>
</tr>
<tr>
<td>14</td>
<td>21.1</td>
<td>3.04927</td>
</tr>
<tr>
<td>15</td>
<td>20.0</td>
<td>2.99573</td>
</tr>
</tbody>
</table>

Since successive differences of the function $\text{Log}_e \hat{F}(t)$ are increasing in $t$ except at $t = 12$, we can assume that the successive differences of the function $\text{Log}_e \hat{F}(t)$ are increasing in $t$ for all values of $t$. The graph of the function $\text{Log}_e \hat{F}(t)$ is shown in Figure 3.1.
YEARS AFTER DIAGNOSIS,
From successive differences of the function \( \log e F(t) \) and from the graph it is clear that \( \log e F(t) \) is convex and therefore we claim that \( F \) is DFR.

Now we shall obtain upper bounds for \( 1 - F \) using Theorem 5 and Theorem 6 and compare them.

We need to estimate \( \mu_l \), the mean of \( F \). Let \( \hat{\mu}_l \) denote an estimate of \( \mu_l \). Thus,

\[
\hat{\mu}_l = \sum_{t=1}^{\infty} \frac{1}{t} F(t)
\]

\[
= \sum_{t=1}^{14} \frac{1}{t} F(t) + \sum_{t=1}^{\infty} \frac{1}{t} F(t)
\]

\[
= \mu_1' + \mu_1'' , \text{ say.}
\]

From the data of Table 1, we have

\[
\mu_l = \sum_{t=1}^{14} \frac{1}{t} F(t) = 527.2 .
\]

Hence, it remains to calculate \( \mu_1'' \). From the graph of the function \( \log e F(t) \) it is clear that for \( t \geq 8 \), \( F(t) \) has an exponential form and therefore, \( \mu_1'' \) can be approximated by

\[
\mu_1'' = \sum_{t=1}^{\infty} \frac{1}{t} F(t)
\]

\[
= \int_{14}^{\infty} g(x)dx \quad \text{where } g(x) \text{ is an exponential curve.}
\]

Let \( g(x) = ce^{-a(x-14)} \) for \( x \geq 14 \).

So, \( \log e g(x) = \log e c - a(x-14) \) implies \( \log e c \) is equal to the height at
$x = 14$ and $-a$ is the slope of the curve $\log_e g(x)$. This gives $c = 21.1$ and an estimate of $-a$ is given by

$$-a = \frac{\log_e \hat{F}(9) - \log_e \hat{F}(12)}{9 - 12} = -0.07162.$$

Hence,

$$\mu_1'' = \int_{14}^{\infty} (21.1)e^{0.07162(x-14)} \, dx$$

$$= (21.1)\int_{0}^{\infty} e^{0.07162} \, dy$$

$$= 294.6.$$

Finally, we obtain

$$\hat{\mu}_1 = 527.2 + 294.6$$

$$= 821.8$$

or $$\hat{\mu}_1 = 8.218 \text{ years }.$$

Using an estimate of the mean of $F$ as $8.218$, we shall calculate upper bounds for $1-F$ in Table 2.
### TABLE - 2

Upper Bounds on Survival Rate.

<table>
<thead>
<tr>
<th>Years after diagnosis</th>
<th>Observed survival rates in %</th>
<th>Upper bounds on $\hat{F}(t)$ using Th. 6 in %</th>
<th>Upper bounds on $\hat{F}(t)$ using Th. 5 in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t$</td>
<td>$\hat{F}(t)$</td>
<td>$\hat{F}(t)$</td>
<td>$\hat{F}(t)$</td>
</tr>
<tr>
<td>0</td>
<td>100.0</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>1</td>
<td>71.6</td>
<td>89.15</td>
<td>88.54</td>
</tr>
<tr>
<td>2</td>
<td>59.3</td>
<td>79.47</td>
<td>78.37</td>
</tr>
<tr>
<td>3</td>
<td>51.7</td>
<td>70.85</td>
<td>69.42</td>
</tr>
<tr>
<td>4</td>
<td>46.0</td>
<td>63.16</td>
<td>61.47</td>
</tr>
<tr>
<td>5</td>
<td>41.7</td>
<td>56.31</td>
<td>54.42</td>
</tr>
<tr>
<td>6</td>
<td>38.0</td>
<td>50.20</td>
<td>48.19</td>
</tr>
<tr>
<td>7</td>
<td>35.0</td>
<td>44.75</td>
<td>42.66</td>
</tr>
<tr>
<td>8</td>
<td>32.4</td>
<td>39.89</td>
<td>37.78</td>
</tr>
<tr>
<td>8.218</td>
<td>31.9</td>
<td>38.94</td>
<td>36.79</td>
</tr>
</tbody>
</table>

(By Interpolation)

| 9                     | 30.0                        | 35.55                                       | 33.58                                       |
| 10                    | 27.8                        | 31.80                                       | 30.23                                       |
| 11                    | 25.9                        | 28.79                                       | 27.48                                       |
| 12                    | 24.2                        | 26.29                                       | 25.18                                       |
| 13                    | 22.5                        | 24.16                                       | 23.24                                       |
| 14                    | 21.1                        | 22.39                                       | 21.58                                       |
| 15                    | 20.0                        | 20.82                                       | 20.13                                       |

From the last two columns we observe that the upper bounds on $\hat{F}(t)$ obtained by Theorem 6 are everywhere larger than those obtained by Theorem 5. These bounds can be compared graphically as shown in Figure 3.2 and Figure 3.3.
Figure 3.2

Years after diagnosis, x

Survival rate in percentage, y

Discrete case

Survival rate observed
Upper bounds on survival rate
**SURVIVAL RATE IN PERCENTAGE, Y**

- **Observed Survival Rate**
- **Upper Bounds on Survival Rate**

**Figure 3.3**

Years after diagnosis, X

Continuos Case
These results on bounds are of particular interest in comparing the mortality in different populations, when \textit{a priori} information about the mortality function is known.

Suppose we are interested in comparing the survival rates of two types of cancer, say, stomach and lung. Let us assume that both have decreasing failure rate with mean \( \mu_1 \) and \( \mu_2 \) respectively. We note that one of the means is greater than the other say, \( \mu_1 > \mu_2 \). Also note that, the estimate of the means can be found by studying sample information. Using Theorem 6, we can find upper bounds on survival rates for both stomach and lung cancer. Then, by comparing the upper bounds on stomach and lung cancer, we can estimate the greatest difference in survival rate for different periods.

2.3 Application of bounds to reliability theory:

In reliability theory, a decreasing failure rate would seem to correspond to some physical mechanism of improvement with age, so that the longer the unit survives the less the chance of failure in the next unit of time. We will discuss the reliability problem given in [13] by establishing the upper bound.

Proschan was interested in obtaining information as to the distribution of failure intervals for the air conditioning system of each member of a fleet of Boeing 720 jet airplanes [13]. Records were kept for the time of successive failures (to the nearest hour) of the air conditioning system [13, p. 316].

First we note that it was proved in [13] that the failure distribution formed by pooling failure intervals for a given component operating in different types of equipment would be DFR. Thus, we know that the data follows a DFR distribution and hence we will give an
upper bound on survival probability using Theorem 6. We remark that an estimate of the parent mean can be estimated by sample information and it was found in [13] that the mean of all the failure intervals was 93.14. Since the sample size is large, this estimate should be fairly accurate. Using 93.14 as an estimate of the parent mean, we give theoretical upper bound for the pooled data.

Figure 3.4 shows the upper bounds on survival probability. The cross-over in the beginning for the observed survival probability may be due to random fluctuation.
Acknowledgement

The author would like to gratefully acknowledge the assistance and encouragement of his thesis advisor Richard Barlow and also Dr. Albert Marshall of the Boeing Scientific Research Laboratories.
Bibliography


