CHAPTER I

INTRODUCTION AND SUMMARY

1.1 Individuals in any particular population are exposed to various risks of death, which may be grouped into \( k \) (\( k \geq 2 \)) mutually exclusive and exhaustive classes. All these \( k \) risks are said to compete for the life of each member of the population.

The various risks to which an individual is subject must be considered in a cause specific mortality study. Every individual is exposed to various risks of death - cancer, tuberculosis, heart failure, kidney failure, etc. A patient who is under study for cancer may be involved in a fatal accident or he may succumb to a heart attack. These other risks as well as the risk of death due to diseases under study are called competing risks.

However, death is not a repetitive event and is usually attributed to a single cause. The theory of competing risks deals with the assessment of a specific risk in the complicating presence of other risks.

A question of interest may be: 'are people suffering from arteriosclerotic heart disease more likely to die from pneumonia than people with another heart disease?' A meaningful comparison between the two groups with respect to their susceptibility to pneumonia would have to evaluate the effect of arteriosclerosis as a competing risk.
Similarly, in a life testing situation in the physical sciences a system of k components is arranged in a series. The system fails when the first component fails. Thus, the different components of the system 'compete' for the life of the system. The theory of competing risks helps in answering questions, such as, if the performance of a certain component is improved, what will be its effect on the performance of the system? How can we estimate the failure rate due to a particular risk in the competing presence of other risks of failure? Does one component have a stochastically smaller lifetime than the other?

If a patient dies during the course of the study, the cause of death and the time to death are noted. For survivors at the end of the study, the length of time in the experiment are recorded. These values are said to be censored since the observations are stopped prior to death. Such censoring may be regarded as a competing risk, since censoring at a certain time x prevents an individual from dying from a particular cause during the experiment, just as death at time x prevents the individual from dying from some other cause. Thus, in particular, competing risks analysis is applicable to data in which there are no competing risks in the ordinary sense, but only a single risk with censoring. However, censoring can be treated separately, because censoring need not always be random and the interest of the experimenter may not lie in the study of the censoring pattern.
Cox (1959, 1962) preferred to call this multi-risk model and, for the case \( k = 2 \), distinguished it from the mixture of distributions or the single risk model in which the individuals of type I are subject only to risk of failure of the first type and individuals of type II are subject only to the risk of failure of second type. However, in the multi-risk model all individuals are subject to both the risks; the actual failure is due to only one of the two risks.

The terms 'risk' and 'cause' refer to the same condition, but are distinguished by their position in time relative to the occurrence of death. Prior to death the condition referred to is a risk, after death the same condition is the cause. For example, tuberculosis is a risk of dying to which an individual is exposed, but tuberculosis is the cause of death if it is the disease from which the individual eventually dies.

The concept of competing risks has been found useful in the interpretation of data from a variety of scientific investigations. There is a population of objects, a random sample of which is observed over a specified period of time. The information about the individuals is restricted to vital state - living or dead and, if dead, the time of death and an assignable cause of death. Most models postulate the existence of a number of independently operating risks of death, that is, risk of death from one cause is assumed to be independent of, and is also assumed to be unaffected by, changes in the risk of death from other causes.
The terms ‘death’ and ‘failure’ essentially mean the same thing – death referring to the living organisms and failure referring to the physical mechanisms, components, etc. We will use the term failure to describe the breakdown of an individual – which may be a living organism or an inanimate object such as light bulbs, electric or physical equipments, etc.

1.2 The beginning of the theory of competing risks goes back to the eighteenth century. Bernoulli (1760) was interested in the effect of eradication of smallpox on the mortality structure of the population. He constructed a hypothetical life table after the elimination of smallpox. The basic assumption was that the individuals saved from smallpox were subject to other causes of failure in exactly the same manner as the rest of the population. D’ Alembert (1761) and Laplace (1812) found Bernoulli’s assumptions too rigid and modified his approach to a new one with the required flexibility. Other significant early contributors were Makeham (1874) and Todhunter (1949). Karn (1931) analyzed massive lifetable data. She considered the effect on mortality tables after, in turn, eliminating cancer, pulmonary tuberculosis and heart diseases.

The term ‘competing risks’ was used by Neyman (1950). Berkson and Elvaback (1960) applied the theory of competing risks to some questions that arose in the study of relation of smoking to lung cancer. Altschuler (1976) considered the occurrence of tumour in animals in the competing presence of the possibility of a tumourless death. For an account of historical background of competing risks see Gail (1975),

1.3 Chiang (1961a, 1961b, 1968) defined three types of probability of failure from a specific cause, viz., crude, net and partial crude probability. Crude probability is the probability of failure in the presence of all other risks in the population; net probability is the probability of failure when \( \ell \text{th} \) (say) risk is the only risk acting on the population; and partial crude probability is the probability of failure from the \( \ell \text{th} \) (say) risk if a particular risk is eliminated as a risk of failure. The concept of population exposed to only one cause of failure has little value in preventive medicine. But the partial crude probability is of interest in predicting demographic changes that might occur in a particular population if a disease were conquered. Chiang obtained mathematical relationship between these probabilities and derived their estimates under the assumption that the observed number of deaths classified in a two-way table according to cause of failure and age at failure follow a multinomial distribution.

The basic assumption underlying Chiang’s work was that the relative force of mortality (ratio of force of mortality due to a specific cause to the total force of mortality) is independent of exact time of failure in the interval of interest. David (1970) showed that Chiang’s proportionality assumption is satisfied whenever the underlying life distribution has one of the three possible forms of the extreme valued distribution.
of the minimum. Kimball (1969) pointed out an anomaly in this assumption. But Chiang (1970), Pike (1970), and Mantel and Blair (1970) showed that the proportionality assumption is sound and led to a model with desirable properties. Gupta (1981) examined the various statements made by Kimball and Chiang, modified them, and supplied mathematical proofs under more general assumptions. Johnson (1976) extended the above results of Chiang without assuming independence of failure times.

Actuaries and demographers use the expression multiple decrement instead of competing risks. A multiple decrement model may be described as a time continuous Markov chain with one transition state labelled 0, and k absorbing states numbered from 1 to k. \( P_i(t) \), \( i = 0, 1, 2, \ldots, k \), is the probability that the process is in state i at time \( t \), given it started in state 0 at time 0. \( a_i(t) \) is the infinitesimal transition probability from state 0 to state i at time \( t \). Aalen (1976, 1978) obtained nonparametric estimates for the cumulative hazard rates \( \beta_i(t) = \int_0^t a_i(s) \, ds \). These estimates are based on minimal sufficient and complete statistics. They are generalizations of Kaplan – Meier (1958) estimators. Aalen studied the bias of the estimates, their strong consistency and asymptotic normality.

1.4 Let \( X_1, X_2, \ldots, X_k \) denote the latent times to failure of an individual subject to k risks, where \( X_i \) represents the age at death if cause i were the only cause of failure. What is actually observed is \( T = \min(X_1, X_2, \ldots, X_k) \) and the cause of failure C,
where \( C = j \) if \( X_j = \min(X_1, X_2, \ldots, X_k) \). A question of interest is, given the minimum and the cause of failure, can we say something about the joint distribution of \( X_1, X_2, \ldots, X_k \)? This is, in fact, the identifiability problem in competing risks. The pair \((T, C)\) is called the identified minimum.

If \( X_1, X_2, \ldots, X_k \) are independent and identically distributed random variables with a common distribution function \( F(x) \), then it is uniquely determined by the distribution of the minimum, that is, \( P[T < x] = 1 - [1 - F(x)]^k \). Berman (1963) considered the case when \( X_i \)'s are independent with distribution function \( F_i(x) \), where the \( F_i \)'s are not all the same. Then the joint distribution of \((T, C)\) uniquely determines \( F_i(x), i = 1, 2, \ldots, k \).

In fact,

\[
F_i(x) = 1 - \exp\left\{ \int_{-\infty}^{x} \left[ 1 - \sum_{j=1}^{k} H_j(t) \right]^{-1} dH_i(t) \right\}, \quad i = 1, 2, \ldots, k.
\]

where

\[
H_j(x) = P[C = j, T < x], \quad j = 1, 2, \ldots, k.
\]

In general, it is not possible on the basis of \((T, C)\) alone, to determine whether \( X_i \)'s are independent (see, for example, Johnson (1990)). The knowledge of the Cox (1962), Tsiatis (1975), Johnson and identified minimum can only determine the derivative of the joint survival function \( S(X_1, X_2, \ldots, X_k) \) along the diagonal \( X_1 = X_2 = \ldots = X_k = t \), and generally this is not sufficient to determine \( S(X_1, X_2, \ldots, X_k) \) uniquely.

Miller (1977), Johnson and Johnson (1980, 1981) showed that the distribution of the identified minimum \((T, C)\) determines
an infinite class of equivalent models for \((X_1, X_2, \ldots, X_k)\). Each such class contains a unique model for which the latent failure times are independent, and at least one marginal survival distribution function is proper, assuming that the distribution of \(T\) is proper.

Langberg, Proschan and Quinzi (1978) gave necessary and sufficient conditions under which a system with dependent components is equivalent in life length and failure pattern to another system with independent components. They (1981) obtained strongly consistent estimators for the unobservable marginal distribution of interest. These estimators are analogous to those of Kaplan – Meier (1958).

Thus, the general dependent model cannot be identified uniquely. The next simplifying assumption would be to assume some kind of parametric form for the joint survival function. Nádas (1971) showed that the bivariate normal distribution is identifiable. Basu and Ghosh (1978) estimated the parameters for bivariate normal distribution under competing risks set up and made comments on the identifiability of the general multivariate normal distribution. They (1980) considered the identifiability of various bivariate exponential distributions. Arnold (1983) introduced some parametric classes and obtained explicit expressions for the parametric estimates.

1.5 In this dissertation, we consider the problem of comparing the two risks to which individuals in a particular population are exposed. Let \(X\) and \(Y\) be two random variables, with distribution
functions $F$ and $G$, respectively, denoting the hypothetical times to failure of the individual due to the two risks. We assume that both $F$ and $G$ belong to $\mathcal{F}$, the class of all absolutely continuous distribution functions $H(x)$, with $H(x) = 0$ for $x \leq 0$. We draw independent random samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ from these two distributions. For the $i$th individual in the sample both $X_i$ and $Y_i$ are not observable. What we observe is the time to failure $T_i = \min(X_i, Y_i)$ and the cause of failure $\delta_i$, that is,

$$T_i = \min(X_i, Y_i)$$

$$\delta_i = \begin{cases} 1 & \text{if } X_i > Y_i, \\ 0 & \text{if } X_i \leq Y_i. \end{cases}$$

On the basis of the times to failure $T_1, T_2, \ldots, T_n$ and the causes of failure $\delta_1, \delta_2, \ldots, \delta_n$ we wish to test the hypothesis of equality of two distribution functions $F$ and $G$, that is,

$$H_0 : F(x) = G(x), \text{ for every } x \geq 0.$$ 

Armitage (1959) and Allen (1963) proved the following result.

**Theorem 1.1** The pair $(X, Y)$ follows proportional hazards model (see, for example, Cox (1972)), i.e., $1 - G = (1 - F)^\beta$, for a positive
constant \( p \), iff the random variables \( T = \min(X, Y) \) and 
\( \delta = I(X > Y) \) are independent, where,

\[
I(a > b) = \begin{cases} 
1, & \text{if } a > b \\
0, & \text{otherwise}.
\end{cases}
\]

As a consequence of the above result we have the following theorems and their corollaries.

**Theorem 1.2** Under \( H_0 \), \( \delta_i \) and \( T_i \) are independent for every 
\( i = 1, 2, \ldots, n \).

Let \( R_i \) be the rank of \( T_i \) among \( T_1, T_2, \ldots, T_n \). Then we 
have the following results.

**Corollary 1.1** Under \( H_0 \), \( \delta_1, \delta_2, \ldots, \delta_n \) and the vector 
\( R = (R_1, R_2, \ldots, R_n) \) are mutually independent.

**Corollary 1.2** Each \( \delta_i \) is a Bernoulli random variable with 
probability of success \( p = P[X > Y] \). Under \( H_0 \), \( p = \frac{1}{2} \) and the 
vector \( R \) is uniformly distributed over \( Q \), the set of all 
permutations of the integers \( (1, 2, \ldots, n) \).

**Theorem 1.3** Under \( H_0 \),

\[
P(R_1 = r_1, R_2 = r_2, \ldots, R_n = r_n; \delta_1 = a_1, \delta_2 = a_2, \ldots, \delta_n = a_n) = \frac{1}{2^n n!},
\]

where each \( a_i \) is either 0 or 1.

**Proof:** As a consequence of above results we have, under \( H_0 \),

\[
P(R_1 = r_1, R_2 = r_2, \ldots, R_n = r_n; \delta_1 = a_1, \delta_2 = a_2, \ldots, \delta_n = a_n)
= P(R_1 = r_1, R_2 = r_2, \ldots, R_n = r_n) \times
P(\delta_1 = a_1, \delta_2 = a_2, \ldots, \delta_n = a_n)
\]
\[
\frac{1}{n!} \prod_{i=1}^{n} P(S_i = a_i) = \frac{1}{2^n n!}
\]

Let \(T_1 < T(2) < \ldots < T(n)\) denote the ordered \(T_i\)s. Define,

\[
W_j = \begin{cases} 
1, & \text{if } T(j) \text{ corresponds to a } Y \text{ observation,} \\
0, & \text{otherwise.}
\end{cases}
\]

Thus, \(W_1, W_2, \ldots, W_n\) are indicator functions corresponding to ordered minima \(T_1, T(2), \ldots, T(n)\). Trivially,

\[
\sum_{i=1}^{n} S_i = \sum_{j=1}^{n} W_j
\]

**Theorem 1.4** Under \(H_0\),

\[
P[ \sum_{i=1}^{n} R_i S_i = w] = P[ \sum_{j=1}^{n} R_j W_j = w]
\]

**Proof:** From corollary 1.1, it follows that

\[
P[ \sum_{i=1}^{n} R_i S_i = w] = \sum_{R \in \mathcal{R}} P[ \sum_{i=1}^{n} R_i S_i = w \mid R = r] P[R = r]
\]

where, \(\mathcal{R} = \{r \mid r \text{ is a permutation of } (1, 2, \ldots, n)\}\)

Thus,

\[
P[ \sum_{i=1}^{n} R_i S_i = w] = \sum_{R \in \mathcal{R}} P[ \sum_{i=1}^{n} R_i S_i = w \mid R = r] P[R = r]
\]

\[
= \sum_{R \in \mathcal{R}} P[ \sum_{j=1}^{n} R_j W_j = w] P[R = r]
\]

\[
= \sum_{R \in \mathcal{R}} P[ \sum_{j=1}^{n} R_j W_j = w]
\]

\[
= P[ \sum_{j=1}^{n} R_j W_j = w]
\]
Let $D_1, D_2, \ldots, D_n$ be the antiranks of the times to failure $T_1, T_2, \ldots, T_n$, that is, $D_k = j$ if the $k^{th}$ ordered $T$ observation $T(k) = T_j$, the $j^{th}$ unordered observation. Thus $D_k$ labels the $T$ which corresponds to the $k^{th}$ ordered minimum value. Therefore,

$$W_k = \delta D_k$$

**Theorem 1.5** Under $H_0$, $W_1, W_2, \ldots, W_n$ are independent and identically distributed with $P[W_1 = 1] = P[W_1 = 0] = \frac{1}{2}$, $i = 1, 2, \ldots, n$.

**Proof:**

$$P[W_1 = 1] = P[\delta D_1 = 1].$$

$$= \sum_{k=1}^{n} P[\delta D_1 = 1|D_1 = k] P[D_1 = k]$$

Just as, under $H_0$, $\delta_1, \delta_2, \ldots, \delta_n$ are independent of ranks, they are also independent of the antiranks. Hence, we have that $\delta_1, \delta_2, \ldots, \delta_n$ and $(D_1, D_2, \ldots, D_n)$ are mutually independent under $H_0$. Therefore, under $H_0$,

$$P[W_1 = 1] = \sum_{k=1}^{n} P[\delta = 1] P[D_1 = k]$$

$$= \sum_{k=1}^{n} P[X_k > Y_k] P[D_1 = k]$$

$$= \frac{1}{2} \sum_{k=1}^{n} P[D_1 = k]$$

$$= \frac{1}{2}$$

$$= P[W_1 = 0]$$
And, under $H_0$,
\[
P[W_1 = w_1, \ldots, W_n = w_n] = P[\delta_{D_1} = w_1, \ldots, \delta_{D_n} = w_n]
\]
\[
= \sum_{d \in \mathcal{P}} P[\delta_{D_1} = w_1, \ldots, \delta_{D_n} = w_n | D = d] P[D = d]
\]
where, $\mathcal{P} = \{d | d \text{ is a permutation of } (1, 2, \ldots, n)\}$
\[
= \sum_{d \in \mathcal{P}} P[\delta_{d_1} = w_1, \ldots, \delta_{d_n} = w_n] P[D = d]
\]
\[
= \frac{1}{2^n} \sum_{d \in \mathcal{P}} P[D = d]
\]
\[
= \frac{1}{2^n} \sum_{i=1}^{n} P[W_i = w_i].
\]

Hence, we have the required result.

**Theorem 1.6** Under $H_0$, $W = \sum_{i=1}^{n} R_i \delta_i$ is symmetric about $\frac{n(n+1)}{4}$.

**Proof:** From Corollaries 1.1 and 1.2 it follows that, under $H_0$, $1 - \delta_i$, $i = 1, 2, \ldots, n$, is a Bernoulli random variable with probability of success $p = \frac{1}{2}$. Then,
\[
(\delta_1, \ldots, \delta_n; R) \overset{d}{=} (1 - \delta_1, \ldots, 1 - \delta_n; R) \tag{1.1}
\]
where $\overset{d}{=}$ means that the expressions on the two sides have the same distribution.

Computing $W$ on each side of the equation (1.1), we get,
\[
\sum_{i=1}^{n} R_i \delta_i \overset{d}{=} \sum_{i=1}^{n} R_i (1 - \delta_i)
\]
\[
\Rightarrow \sum_{i=1}^{n} R_i \delta_i = \frac{n(n+1)}{2} - \sum_{i=1}^{n} R_i \delta_i
\]

\[
\Rightarrow \sum_{i=1}^{n} \frac{n(n+1)}{4} - \sum_{i=1}^{n} \frac{n(n+1)}{4} = w
\]

Hence the result.

1.6 The second chapter of the thesis is based on likelihood considerations. The likelihood function, under the competing risks set up, is given by

\[
L(\xi, \delta) = \prod_{i=1}^{n} \left[ g(t_i) \bar{F}(t_i) \right]^{\delta_i} \left[ f(t_i) \bar{G}(t_i) \right]^{-\delta_i},
\]

where \( F \) and \( G \) are the distribution functions of the hypothetical lifetimes \( X \) and \( Y \) due to risks I and II, respectively, \( \bar{F} = 1-F \) and \( \bar{G} = 1-G \) are the corresponding survival functions and \( f \) and \( g \) are the corresponding probability density functions.

We assume that the two lifetimes are exponentially distributed random variables and derive a likelihood ratio statistic for testing the equality of parameters of the exponential distributions.

Then we derive the locally most powerful tests and the locally most powerful rank tests for testing

\( H_0 : F(x) = G(x), \) for every \( x \geq 0, \)

against the location alternative

\( H_{A_1} : G(x) = F(x+\theta), \) for every \( x \geq 0, \theta \geq 0, \)

and the scale alternative

\( H_1 : G(x) = F[(\theta+1)x], \) for every \( x \geq 0, \theta \geq 0. \)
1.7 In the third chapter some tests based on heuristic grounds have been developed for testing $H_0$ against $H_{A_2}$: $F(x) \lessgtr G(x)$, for every $x \geq 0$, with strict inequality over a set of positive probability.

The first test is based on the information about the causes of failure only. The proposed test statistic is the sign statistic

$$U_1 = \frac{1}{n} \sum_{i=1}^{n} \delta_i$$

where $\delta_i = \begin{cases} 1 & \text{if } X_i > Y_i, X_j > Y_j, \\ 0 & \text{if } X_i > Y_i, Y_j > X_j \text{ and } \min(X_j, Y_j) > \min(X_i, Y_i), \\ -1 & \text{if } Y_i > X_i, X_j > Y_j \text{ and } \min(X_i, Y_i) > \min(X_j, Y_j). \end{cases}$

$U_1$ is the proportion of failures due to the second cause.

The above statistic does not make use of the times to failure $T_1, T_2, \ldots, T_n$. We can simultaneously look at the times to failure and the causes of failure of the $i^{th}$ and $j^{th}$ individuals in the sample. There are eight mutually exclusive arrangements of the observed pairs $(T_i, \delta_i)$ and $(T_j, \delta_j)$. The test is based on the generalized U-statistic (see, for example, Bhapkar (1961) and Lehmann (1963)).

$$U_2 = \frac{1}{\binom{n}{2}} \sum_{1 \leq i < j \leq n} \phi_2(X_i, Y_i, X_j, Y_j)$$

where

$$\phi_2(X_i, Y_i, X_j, Y_j) = \begin{cases} 3 & \text{if } X_i > Y_i, X_j > Y_j, \\ 1 & \text{if } X_i > Y_i, Y_j > X_j \text{ and } \min(X_j, Y_j) > \min(X_i, Y_i), \\ 0 & \text{if } Y_i > X_i, X_j > Y_j \text{ and } \min(X_i, Y_i) > \min(X_j, Y_j). \end{cases}$$
For testing $H_0$ against $H_{A_1}$ we propose a test based on the U-statistic

$$U_j = \frac{1}{(n-2)} \sum_{i<j<n} \phi_j(X_i, Y_i, X_j, Y_j)$$

where,

$$\phi_j(X_i, Y_i, X_j, Y_j) = \begin{cases} 1, & \text{if } X_i > Y_i, X_j > Y_j, \\
& \text{or } X_i > Y_i, X_j \geq X_j, \text{ and } \\
& \min(X_j, Y_j) > \min(X_i, Y_i), \\
& \text{or } X_j > Y_j, Y_i \geq X_i, \text{ and } \\
& \min(X_i, Y_i) > \min(X_j, Y_j), \\
& 0, \text{ otherwise.} \end{cases}$$

The exact and asymptotic distributions of the proposed statistics have been studied. The critical points for $S_2 = \binom{n}{2} U_2$ have been derived. In order to facilitate the computation of the statistics, they have been expressed in terms of the ranks of the observed minima. Consistency and unbiasedness properties of the tests based on these statistics have been discussed.
1.8 In Chapter 4 we first compare the tests for stochastic dominance based on $U_1$ and $U_2$ in the Pitman asymptotic relative efficiency (ARE) sense for some parametric alternatives in $H_{A_2}$. Then the various tests proposed for location alternative $H_{A_1}$ are compared in the Pitman ARE sense.

1.9 Let us define failure rate of an individual and an increasing failure rate distribution (see, for example, Barlow and Froschan (1975)).

(i) **Failure Rate**: The failure rate (also called the hazard rate) $r_F(t)$ corresponding to a survival function $\overline{F}(t)$ of a random variable $X$ is defined by

$$r_F(t) = \lim_{\Delta_t \to 0} \frac{1}{\Delta_t} \Pr[X \leq t + \Delta_t | X > t]$$

for $t > 0$ such that $\overline{F}(t) > 0$, provided the density function $f(t)$ exists.

(ii) **Increasing Failure Rate (IFR) Distribution Function**: A life distribution $F \in \mathcal{F}$ is called an increasing failure rate distribution if $r_F(t)$ is a nondecreasing function of $t$ for every $t \geq 0$.

Let $r_F(x)$ and $r_G(x)$ denote the failure rates corresponding to $F$ and $G$, respectively.

In the fifth chapter we propose a test based on a $U$-statistic for testing,

$$H_0 : F(x) = G(x), \text{ for every } x \geq 0,$$

or equivalently,

$$r_F(x) = r_G(x), \text{ for every } x \geq 0.$$
against the alternative  
\[ H_{A_3} : r_F(x) \leq r_G(x), \text{ for every } x \geq 0 , \]
with strict inequality over a set of nonzero probability.

The test is based on the following U-statistic.

\[ U_4 = \frac{1}{n \choose 2} \sum \sum \phi_4(X_i, Y_i, X_j, Y_j) \]

where,

\[ \phi_4(X_i, Y_i, X_j, Y_j) = \begin{cases} 
1, & \text{if } X_i > Y_i, X_j > Y_j, \\
\text{or } X_i > Y_i, Y_j \geq X_j \text{ and } \min(X_i, Y_i) > \min(X_j, Y_j), \\
\text{or } Y_i \geq X_i, X_j > Y_j \text{ and } \min(X_j, Y_j) > \min(X_i, Y_i), \\
0, & \text{otherwise.} 
\end{cases} \]

The exact and the asymptotic distributions of the test statistic have been considered. The statistic has been expressed as a function of the ranks of \( T_1, T_2, \ldots, T_n \). Consistency and unbiasedness properties of the test have been discussed. Finally, we compare the \( U_4 \) test, in the Pitman ARE sense, with the \( U_1 \) and \( U_2 \) tests for some specific parametric alternatives belonging to \( H_{A_3} \).

Some practical examples involving competing risks are also provided and the use of some of the above tests is illustrated.

1.10 In the sixth chapter we consider the problem of comparing
simultaneously two populations exposed to two different risks. Let $X$ and $U$ ($Y$ and $V$) be random variables denoting the hypothetical lifetimes of individuals exposed to Risk I and Risk II, respectively, in the first (second) population, with cumulative distribution functions $F_1$ and $F_2$ ($G_1$ and $G_2$), $F_1$, $F_2$, $G_1$, $G_2 \in \mathcal{F}$. We assume that the two risks act independently in both the populations. Let $(X_1, U_1), (X_2, U_2), \ldots, (X_n, U_n)$ be a random sample from the first population and $(Y_1, V_1), (Y_2, V_2), \ldots, (Y_m, V_m)$ be an independent random sample from the second population. Since the failure of each individual is assumed to occur due to only one cause, we can neither observe the pair $(X_i, U_i)$, nor the pair $(Y_j, V_j)$. What is actually observed are the times to failure $T_1, T_2, \ldots, T_n$ and $S_1, S_2, \ldots, S_m$, and the corresponding causes of failure $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m$, where,

$$T_i = \min.(X_i, U_i)$$

$$= \begin{cases} U_i, & \text{if } X_i > U_i, \\ X_i, & \text{if } X_i \leq U_i. \end{cases}$$

and,

$$\varepsilon_i = \begin{cases} 1, & \text{if } X_i > U_i, \\ 0, & \text{if } X_i \leq U_i, \end{cases} \quad i = 1, 2, \ldots, n.$$

$$S_j = \min.(Y_j, V_j)$$

$$= \begin{cases} V_j, & \text{if } Y_j > V_j, \\ Y_j, & \text{if } Y_j \leq V_j. \end{cases}$$
and, $\epsilon_j = \begin{cases} 1, & \text{if } Y_j > V_j, \\ 0, & \text{if } Y_j \leq V_j \end{cases}, \quad j = 1, 2, \ldots, m$

On the basis of the observed information $(T_1, S_1)$, $(T_2, S_2)$, ..., $(T_n, S_n)$ from the first sample and $(S_1, \epsilon_1)$, $(S_2, \epsilon_2)$, ..., $(S_m, \epsilon_m)$ from the second sample we wish to test the hypothesis

$H_0 : F_1(x) = G_1(x), F_2(x) = G_2(x)$, for every $x \geq 0$,

against the alternative

$H_A : F_1(x) < G_1(x), F_2(x) < G_2(x)$, for every $x \geq 0$, with strict inequality over a set of nonzero probability.

We propose two tests for the above problem. The first test is based on the Wilcoxon–Mann–Whitney statistic based on observed times to failure $T_1, T_2, \ldots, T_n$ and $S_1, S_2, \ldots, S_m$. The statistic is

$W_1 = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} \phi(T_i, S_j)$,

where,

$\phi(T_i, S_j) = \begin{cases} 1, & \text{if } T_i > S_j, \\ 0, & \text{otherwise}. \end{cases}$

The second test makes use of the times to failure as well as the causes of failure of individuals in the two samples. The test is based on the statistic,

$W_2 = (B_1, B_2) \sum_{H_0}^{-1} (B_1, B_2)'$
where,  \( B_1 = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} \phi_5(X_i, U_i, Y_j, V_j) \),
\[ B_2 = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} \phi_6(X_i, U_i, Y_j, V_j) \],

\[ \phi_5(X_i, U_i, Y_j, V_j) = \begin{cases} 
1, & \text{if } \min(X_i, U_i) > \min(Y_j, V_j), \quad V_j > Y_j, \\
-1, & \text{if } \min(X_i, U_i) < \min(Y_j, V_j), \quad U_i > X_i, \\
0, & \text{otherwise}.
\end{cases} \]

\[ \phi_6(X_i, U_i, Y_j, V_j) = \begin{cases} 
1, & \text{if } \min(X_i, U_i) > \min(Y_j, V_j), \quad Y_j > V_j, \\
-1, & \text{if } \min(X_i, U_i) < \min(Y_j, V_j), \quad X_i > U_i, \\
0, & \text{otherwise}.
\end{cases} \]

and,
\[ \Sigma^{-1} \text{ is the inverse, under } H_0, \text{ of the dispersion matrix } \Sigma \text{ corresponding to } (B_1, B_2). \]

We study the asymptotic distributions of \( W_1 \) and \( W_2 \) and compare their performance, in the Pitman ARE sense, using the approach of Rothe (1981), which allows the computation of the Pitman ARE when one test statistic has asymptotic normal distribution and the other has asymptotic chi-square distribution.

Finally, we provide a list of relevant references.

Before we conclude this chapter, it is worthwhile to remark that although we are restricting ourselves to non-negative random variables because of applications mainly in the survival analysis and reliability situations, yet all the procedures suggested in the subsequent chapters are applicable even for random variables defined on the entire real line or a part thereof.