Cure Rate Estimation Based on Uncensored and Censored Data

Mohammed Taj Uddin

A thesis
in
The Department
of
Mathematics and Statistics

Presented in partial fulfillment of requirements
For the degree of Master of Science at
Concordia University
Montreal, Quebec, Canada.

August, 2004

Your file Votre référence
ISBN: 0-612-94675-4
Our file Notre référence
ISBN: 0-612-94675-4

Cure Rate estimation based on uncensored and censored data

Mohammed Taj Uddin

Abstract

This thesis deals with the analysis of cure model under uncensored and censored data. First we provide related concepts and second, we develop the cure rate model. Thirdly, we study the non-parametric maximum likelihood method to estimate the parameter of cure model. We consider both uncensored and censored data as well as Type 1 and Type 2 censoring to estimate the parameter of the model. Finally, we develop some parametric and non-parametric estimating equations to estimate the parameter of the cure rate model.

Acknowledgement

All praises and glory to Almighty Creator, who provides me with an opportunity for acquiring the knowledge and experience for building my career.

I am grateful to my supervisor **Dr. Arusharka Sen** for accepting me as his student and suggesting to work in this field. His continuous encouragement, constructive discussion, intellectual suggestions and necessary corrections of my research work have made this thesis possible. I also acknowledge his cooperation in providing me with all his valuable reference materials in this area.

I am thankful to Dr. Yogendra P. Chaubey for his valuable suggestions and encouragement. I am sure that without the help of my supervisor, encouragement and financial support, I could have never completed this thesis.

I am grateful to the Department of Mathematics and Statistics at Concordia University, Canada for its financial support as well as every professor and secretary who helped me to overcome any type of technical and theoretical difficulty throughout the preparation of this thesis.

I wish to thank my wife Hasiba Begum for her great sacrifice and understanding during the period of my study at Concordia University in Canada. Finally I am grateful to my parents and all of my friends and relatives who constantly encouraged for my success.

Contents

Chapter 1

Survival analysis and related concepts

1.1 Introduction

Survival analysis is a loosely defined statistical term encompassing a variety of statistical techniques for analyzing positive valued random variables. Typically the value of the random variable is the time to failure of a physical component (say, mechanical or electrical or biological) or the time to the death of a biological unit (say, patient, animal, cell etc.). However it could be the time to the learning of a skill or it may not even be a time at all. For example, it could be the number of dollars that a health insurance company pays in a particular case.

While the origins of survival analysis might be attributed to the early work on mortality tables centuries ago, the more modern era started about a half century ago with applications to engineering and medical science. World War II stimulated interest in the reliability of military equipment and this interest in reliability carried over into the past war era for military and commercial products. Most of the statistical research for engineering and medical applications was concentrated on parametric survival models. Within the past few decades there has been an increase in the number of clinical trials in medical research and this has shifted the statistical focus to non- parametric approaches. The analysis of survival experiment is complicated by issues of censoring, where an individual life length is known to occur only in a certain period of time and by truncation, where individuals enter the study only if they survive a sufficient length of time or individuals are included in the study only if the event has occurred by a given date. In recent years, the use of counting process methodology has resulted in developing theory

to account for censoring and truncation in survival experiments. Anderson *et al.* (1993) give an excellent survey of the mathematics of this theory.

1.2 Some basic quantities related to the estimation of Cure Model

The problem of analyzing time to event data arises in a number of applied fields such as medicine, biology, public health, epidemiology, engineering, economics and demography etc. A common feature of this data set is that they contain either censored or truncated observations. Censored data arise when an individual's life length is known to occur only in a certain period of time. Possible censoring schemes are right censoring, where all that is known is that the individual is still alive at a given time, left censoring when all that is known is that the individual has not experienced the event of interest prior to the start of the study or interval censoring, where the only information is that the event occurs within some interval. Truncation schemes are left truncation, where only individuals who survive a sufficient time are included in the sample and right truncation, where only individuals who have experienced the event by a specified time are included in the sample.

Let X be the time until some specified event. This event may be death, the appearance of a tumor, the development of disease, recurrence of a disease, equipment break down, conception, cessation of smoking, cessation of breast feeding and so forth. More precisely, here X is a non-negative random variable from a homogeneous population. Basically the following four quantities (functions) characterize the distribution of X, namely, the survival function $S(x)$, which is the probability of an individual surviving beyond time x, the hazard rate, which is the chance an individual of age x experiences the event in the next instant, the survival density function, which is the unconditional

probability of the event occurring at time x and the mean residual life at time x, which is the mean time to the event of interest, given that the event has not occurred at time x. If we know any one of these four quantities, then the other three can be uniquely determined. In practice, these four functions along with the following quantity known as the cumulative hazard function, are used to illustrate the different aspects of the distribution of X.

1.3 The Survival Function

The most basic quantity used to describe time to event data is the survival function, which is the probability of an individual surviving beyond time x. It is denoted by $S(x)$ and defined as $S(x) = \Pr(X > x)$.

In some contexts involving systems or lifetimes of manufactured itimes, $S(x)$ is considered as the reliability function. Note that the Survival function is non-negative and non- increasing function with $S(0) = 1$ and $S(\infty) = \lim_{x \to \infty} S(x) = 0$. In some special cases, we consider $S(\infty) > 0$ where some individuals never fail. For example, to estimate the cure model (mixture model), $S(\infty) > 0$. On the other hand, if X is a continuous random variable, the survival function, $S(x)$ is the complement of the Cumulative distribution function, that is $S(x) = 1 - F(x)$, where $F(x) = \Pr(X \leq x)$. Also the survival function is the integral of the probability density function, $f(x)$, that is

$$S(x) = \Pr(X > x) = \int_x^{\infty} f(t)dt$$

Therefore,

$S'(X) = -F'(X)$; where prime denotes the derivative,

and hence, $f(x) = -S'(X)$

In Survival analysis, sometimes discrete random variables arise due to rounding off measurements, grouping of failure times (life times) into intervals, or when life times refer to an integral number of units. Suppose that X can assume on values $x_j, j = 1,2,...$ with probability mass function $P(x_j) = \Pr(X = x_j), j = 1,2,...$ and $x_1 < x_2 < .....$ The Survival function of a discrete random variable X is given by

$$S(x) = \Pr(X > x) = \sum_{x_j > x} P(x_j)$$

Note that, in practice many types of survival curves are used but they have the same fundamental properties. They are monotone, non-increasing functions equal to one at origin and zero as the survival time tends to infinity. Their rate of decline varies according to the risk of experiencing the event. Nevertheless, this quantity continues to be a proper description of survival in the applied literature and can be very useful in comparing two or more mortality patterns.

1.4 The Hazard Function

A very important and fundamental concept in survival analysis is the hazard function. This function is sometimes given other names, such as 'The forces of mortality' in Demography, the inverse of 'The Mill's ratio' or 'the hazard rate' in Economics, 'The conditional failure rate' in reliability, 'The intensity function' in Stochastic process, and 'The age specific failure rate' in Epidemiology. The hazard rate is defined by

$$h(x) = \lim_{h \to 0} \frac{P(x \le X < x + h \mid X \ge x)}{h}$$

$$= \lim_{h \to 0} \frac{S(x) - S(x + h)}{hS(x)}$$

$$= -\lim_{h\to 0} \frac{S(x+h)-S(x)}{hS(x)}$$

$$= -\frac{S'(x)}{S(x)}, \tag{1.4.1}$$

Thus, $h(x) = \frac{f(x)}{S(x)}$, (1.4.2)

The equation (1.4.1) can be written in a nice form involving $S(x)$ only as,

$$h(x) = -\frac{d}{dx}\log[S(x)] \tag{1.4.3}$$

Now integrating (1.4.3) with respect to x and using $S(0) = 1$, we obtain,

$$\ln[S(x)] = -\int_0^x h(x)dx$$

Hence, $S(x) = \exp\left[-\int_0^x h(x)dx\right]$, (1.4.4)

A very important and related quantity is the cumulative hazard function $H(x)$, defined by $H(x) = \int_0^x h(t)dt = \ln[S(x)]$

Therefore, for continuous life times,

$$S(x) = \exp[-H(x)] = \exp\left[-\int_0^x h(t)dt\right], \tag{1.4.5}$$

Since, $S(\infty) = 0$, then $H(\infty) = \infty$ and thus it follows from (1.4.2)

$$f(x) = h(x)\exp\left[-\int_0^x h(t)dt\right], \tag{1.4.6}$$

The hazard function is useful in determining the appropriate failure distributions utilizing qualitative information about the mechanism of failure and for describing the way in which the chance of experiencing the event changes with time. Information about the nature of the hazard rate is very helpful in selecting a model.

When X is a discrete random variable, the hazard function is given by

$$h(x_j) = \Pr\left\langle X = x_j \,\middle|\, X \geq x_j \right\rangle = \frac{P(x_j)}{S(x_{j-1})} \tag{1.4.7}$$

Where, $j = 1,2,....$ and $S(x_0) = 1$. Note that, $P(x_j) = S(x_{j-1}) - S(x_j)$ and therefore,

$$h(x_j) = 1 - \frac{S(x_j)}{S(x_{j-1})}, \tag{1.4.8}$$

Also the survival function can be written as the product of conditional probabilities

$$S(x) = P\left\langle X > x \,\middle|\, X \geq x_j \right\rangle = \prod_{x_j \leq x} S(x_j)/S(x_{j-1}), \tag{1.4.9}$$

Thus, the survival function can be written in terms of hazard function as

$$S(x) = \prod_{i=1}^{n} \{1 - h(x_j)\}, \tag{1.4.10}$$

For discrete lifetime, the cumulative hazard function is defined as

$$H(x) = \sum_{x_j \leq x} h(x_j), \tag{1.4.11}$$

Cox and Oakes (1984) prefer to define the cumulative hazard function for discrete lifetime as

$$H(x) = \sum_{x_j \leq x} \ln\{1 - h(x_j)\}, \tag{1.4.12}$$

1.5 The Mean Residual Life Function

The next basic quantity of interest in survival analysis is the mean residual life at time x. This quantity measures the expected lifetime of an individual given the survival until time x. It is defined as

$$mrl(x) = E(X - x \mid X > x) \tag{1.5.1}$$

For a continuous random variable,

$$mrl(x) = E(X - x \mid X > x)$$

$$= \frac{\int_x^\infty (t-x) f(t)dt}{S(x)}$$

$$= \frac{\int_x^\infty (t-x)(-dS(t))}{S(x)}$$

$$= \frac{-(t-x)S(t)\Big|_x^\infty + \int_x^\infty S(t)dt}{S(x)}$$

$$= \frac{\int_x^\infty S(t))dt}{S(x)}$$

Thus, $mrl(x) = \frac{\int_x^\infty S(t)dt}{S(x)}$, (1.5.2)

Note that, the mean life, $\mu = mrl(0)$ is the total area under the survival curve and

$$\mu = E(x) = \int_0^\infty tf(t)dt$$

$$= \int_0^\infty S(t)dt, \quad (1.5.3)$$

Also the variance of X in terms of survival function is given by the following expression

$$Var(X) = 2\int_0^\infty tf(t)dt - [\int_0^\infty tf(t)dt\ ]^2$$

1.6. Different types of Censoring and Truncation

Right censoring: Failure time data often include some individuals who do not fail during their observation period, the data on these individuals are said to be right censored. In some situations, right censoring arises simply, because some individuals are still surviving at the time that the study is terminated and the analysis is done. In other

instances, individuals may move away from the study area for reasons unconnected with the failure time endpoint, so contact is lost. In yet other instances, individuals may decide to withdraw or may be withdrawn from the study because of worsening or improving prognosis. A right censoring mechanism is said to be independent if the failure rates that apply to individuals on trial at each time are the same as those that would have applied had there been no censoring.

Type1 censoring: Censored data are called Type1 censoring if observations occur only at a specified fixed time. For example, in life testing when all unit are put on test at the same time and the data are collected and analyzed at a specified point in time. For life data it is also called time censored if the censoring time is fixed and the number of failures in the fixed time is random. Basically Type1 censoring is a special kind of right censoring. But in a censoring scheme if the individuals enter the study at different times and the termination point of the study is predetermined by the investigator so that the censoring times are known when an individual is entered in to the study. In such studies, individuals have their own specific, fixed censoring time. This kind of censoring is termed as generalized Type1 censoring (*c.f.* David and Moeschberger, 1978).

Type2 censoring: A second type of right censoring is called Type 2 censoring in which the study continues until the failure of the first 'r' individuals, where r is some predetermined integer ($r < n$). Experiment involving Type 2 censoring are often used in testing of equipment life. Here all items are put on test at the same time and the test is terminated when 'r'of the 'n' items have failed. Such an experiment may have time and money because it could take long time for all items to fail. The statistical treatment of Type 2 censored data is simple because the data consists of the 'r' smallest life time in a

random sample of 'n' life times, so that the theory of order statistics is directly applicable to determining the likelihood function and any inferential technique employed.

Random censoring: Sometimes, individuals will experience some other competing events which cause them to be removed from the study. In such cases, the event of interest is not observable. This situation has been studied in details in competing risk theory (*c.f.* David and Moeschberger, 1978) and is termed as random censoring. Some events which cause the individuals to be randomly censored, with respect to the event of interest, are accidental deaths, migration of human population, death due to some cause other than the one of interest, patients withdrawn from a clinical trial, and so forth. If the distribution of random censoring times contains no parameters common with survival function, S(t), then the estimates of such parameters may be obtained in the usual behavior for generalized Type1 censoring.

Some comments on Type1 censoring: Type1 censoring can be considered as a special case of random censoring if we allow the censoring time to have degenerate distributions, each with mass at one fixed point. Though it may be desirable to make inferences conditional on the censoring time in any given situation, the properties of the procedures averaged over the distribution of the censoring time may be of interest when planning studies, and in some applications. Though the individual random censorship model is often reasonable, in many situations the censoring process is linked to the failure time process. Suppose, for example, that the termination date for a medical trial is not fixed before the study commences, but is chosen later, with the choice influenced by the results of the study up to that time. In such instances it may be difficult to write down a model

that fully represents the process under study. Hopefully, the likelihood function of Type1 censoring is still applicable in many such complicated situations.

Left censoring: Individuals can also be subject to left censoring, which occurs if the individual is observed to fail prior to some time, but actual time of failure is otherwise unknown. In other words, a life time associated with a specific individual in a study is considered to be left censoring if it is less than a censoring time, i.e., the event of interest has already occurred for the individual before that person is observed in the study at time. For such individuals, we know that they have experienced the event sometimes before censoring time, but their exact event time is unknown. The exact life time will be known if and only if the life time is greater than or equal to censoring time.

In the next chapter, we will discuss some methods of estimation in survival analysis.

Chapter 2

Some methods of estimation in survival analysis

2.1 Non-parametric maximum likelihood method

The Method of maximum likelihood is the most commonly employed estimation procedure. From the theoretical and practical point of view, it is the most general method of estimation. This method was initially formulated by C.F. Gauss but as a general method of estimation was first introduced by R.A. Fisher and later on developed by him in a series of papers.

For some models, explicit solution of the estimating equations may be obtained. However, for other models, the solutions can not be obtained explicitly. In this situation some computational methods of finding estimators such as Newton-Raphson method, Iterative method, Scoring method, EM method and Lagrange Multiplier method have been employed.

Suppose that X is a random variable with probability density function $f(x;\theta)$, θ to be estimated and $x_1, x_2, ...x_n$ is a random sample of size n. The joint probability density function of the random variable comprising the sample is called the likelihood function of the sample and is given by

$$L(x_1, x_2, ...x_n;\theta) = f(x_1;\theta)f(x_2;\theta)..f(x_n;\theta)$$

$$= \prod_{i=1}^{n} f(x_i;\theta), \qquad (2.1.1)$$

If there exists a value θ^* such that $L(x_1, x_2, ...x_n;\theta^*) \geq L(x_1, x_2, ...x_n;\theta)$ for all possible choices of θ, then based upon the meaning of likelihood function, θ^* maximizes the equation (2.1.1) and this value θ^* is considered as the maximum likelihood estimate.

Choosing the value of θ that makes it most likely that the data would be as obtained is certainly a reasonable approach. Therefore, if a value θ^* can be found such that θ^* maximizes the likelihood function (2.1.1) for a given set of sample values $x_1, x_2, ...x_n$, then θ^* is called the maximum likelihood estimate for the given set of sample values. Since the likelihood function is a function of parameter, θ under the sample information $(x_1, x_2, ...x_n)$, the maximum likelihood estimate will be a function of the sample values. If the sample functional relationship between the estimate and the sample information $(x_1, x_2, ...x_n)$ holds for all possible choices of the x_i, then that functional relationship can be taken as an estimation rule and the result will be an estimator $\hat{\theta}$ of θ, this estimator being known as maximum likelihood estimator or the maximum likelihood filter.

In Non-parametric maximum likelihood method, we write the non-parametric likelihood function as

$$L_n(F) = \prod_{i=1}^{n} (\Delta F(x_i)), \tag{2.1.2}$$

where, $F(.)$ is the common distribution function of $X_i, 1 \le i \le n$ and $\Delta F(x_i) = F(x_i) - F(x_{i-})$ is the jump of $F(.)$ at $X_i, 1 \le i \le n$. Putting $\Delta F(x_i) = p_i$ in (2.1.2), we can write

$$L_n(p_1, p_2, ..., p_n) = \prod_{i=1}^{n} p_i, \tag{2.1.3}$$

Now we maximize (2.1.3) subject to condition

$$\sum_{i=1}^{n} p_i = 1, \; p_i \ge 0, 1 \le i \le n, \tag{2.1.4}$$

The log-likelihood function becomes

$$\log L_n(p_1, p_2, \ldots p_n) = \sum_{i=1}^{n} \log p_i \,, \tag{2.1.5}$$

By using Lagrange multiplier method we can maximize (2.1.5). By adding a Lagrange multiplier λ, (2.1.5) becomes

$$\log L_n(p_1, p_2, \ldots p_n) = \sum_{i=1}^{n} \log p_i - \lambda(\sum_{i=1}^{n} p_i - 1)\,, \tag{2.1.6}$$

Thus, the non-parametric maximum likelihood estimator of p_i is obtained by the solution of the following equations

$$\frac{\partial \log L}{\partial p_i} = 0, \frac{\partial \log L}{\partial \lambda} = 0\,,\ i = 1,2,..,n, \tag{2.1.7}$$

Now, $\frac{\partial \log L}{\partial p_i} = 0$ gives

$$\frac{1}{p_i} - \lambda = 0$$

Therefore, $p_i = \frac{1}{\lambda}$, i = 1,2,…,n , (2.1.8)

Similarly, $\frac{\partial \log L}{\partial \lambda} = 0$ gives

$$\sum_{i=1}^{n} p_i = 1, \tag{2.1.9}$$

From (2.1.8), we can write

$$\sum_{i=1}^{n} p_i = \frac{n}{\lambda}\,, \tag{2.1.10}$$

Using (2.1.9) in (2.1.10) we obtain

$$\hat{\lambda} = n\,, \tag{2.1.11}$$

Therefore, the non-parametric maximum likelihood estimator of p_i is

$$\hat{p}_i = \frac{1}{n}, \text{ i} = 1,2,\ldots\ldots,\text{n}$$

For Type 1 censoring the non-parametric likelihood function can be written as follows

$$L_n(F) = \prod_{i=1}^{n} (\Delta F(x_i))^{\delta_i} (1 - F(T)^{1-\delta_i} \quad , \tag{2.1.12}$$

where, $\delta_i = 1_{\{x_i \le T\}}$

Suppose $x_1 \le x_2 \le \ldots. \le x_n$ are order statistics and let $k = \max\{ i \mid X_{(i)} \le T \}$.Putting $p_i = \Delta F(x_i)$, the non-parametric likelihood function becomes

$$L_n(p_1, p_2, .. p_n) = (\prod_{i=1}^{k} p_i)(1 - \sum_{j=1}^{k} p_j)^{n-k} \quad , \tag{2.1.13}$$

By using Lagrange multiplier method we can obtain the non-parametric maximum likelihood estimate of the parameter.

The log-likelihood function can be written as

$$\log L_n(p_1, p_2, \ldots, p_n) = \sum_{i=1}^{k} \log p_i + (n-k) \log(1 - \sum_{i=i}^{k} p_i) \quad , \tag{2.1.14}$$

Now, we maximize (2.1.14) subject to condition $\sum_{i=1}^{k} p_i \le 1$, (2.1.15)

Adding a non-negative slack variable S in (2.1.15) we get $\sum_{i=1}^{k} p_i + S = 1$, (2.1.16)

Adding Lagrange multiplier λ in (2.1.14) we get the following equation

$$\log L_n(p_1, p_2, \ldots, p_n) = \sum_{i=1}^{k} \log p_i + (n-k) \log(S) - \lambda(\sum_{i=1}^{k} p_i + S - 1), \tag{2.1.17}$$

Therefore the non-parametric maximum likelihood estimator of p_i is obtained by the solution of the following simultaneous equations

$$\frac{\partial \log L_n(p_1, p_2, \ldots p_k)}{\partial p_i} = 0, \frac{\partial \log(p_1, \ldots p_k)}{\partial \lambda} = 0, \frac{\partial \log(p_1, \ldots p_k)}{\partial S} = 0, \tag{2.1.18}$$

Now $\frac{\partial \log L_n(p_1, p_2, \ldots p_k)}{\partial p_i} = 0$, gives

$$\frac{1}{p_i} - \lambda = 0 , \tag{2.1.19}$$

$\frac{\partial \log L_n(p_1, p_2, \ldots p_k)}{\partial \lambda} = 0$, gives

$$\sum_{i=1}^{k} p_i + S - 1 = 0 , \tag{2.1.20}$$

$\frac{\partial \log L_n(p_1, p_2, \ldots p_k)}{\partial S} = 0$ gives

$$\frac{n-k}{S} - \lambda = 0 , \tag{2.1.21}$$

Solving the equations (2.1.19), (2.1.20) and (2.1.21), we obtain the following estimate of p_i and S respectively,

$$p_i = \frac{1}{n}, S = 1 - \frac{k}{n}, 1 \le i \le k$$

2.2 Expectation-Maximization Algorithm (EM-algorithm)

The Expectation–Maximization (EM) algorithm is most useful in much more complicated problems and often provides a simple and intuitively appealing algorithm. It has been used widely in various contexts to fit models with incomplete data (censored data). In particular, interval censored data is analyzed by EM-algorithm. It can be very slow to converge and may not converge to the Maximum Likelihood Estimator (MLE).

Basically, the EM-algorithm (Dempster *et al.* 1977) is a convenient and widely used tool to compute Maximum Likelihood Estimates for incomplete data. Suppose that the complete data are denoted by x and that the incomplete data are denoted by $y = g(x)$, where $g(.)$ is one to one function that corresponds to some grouping or incomplete observation of the complete data. Let the probability density function of x depends on an unknown parameter vector θ and that the log-likelihood function for θ given x, the complete data log-likelihood, is $l_x(\theta)$. The log-likelihood from the incomplete data is denoted by $l_y(\theta)$. In many cases, $l_y(\theta)$ is of complicated form, whereas $l_x(\theta)$ is relatively much simpler and more easily maximized. In this situation, the EM-algorithm often provides a simple approach to finding the MLE of parameter vector, θ based on y.

Let $\theta^{(0)}$ be an initial estimate of θ. The algorithm proceeds through an expectation (E) step and then a maximization (M) step:

E-step: Calculate $Q(\theta, \theta^{(0)}) = E[l_x(\theta) \mid y; \theta^{(0)}]$, (2.2.1)

M-step: Find $\theta^{(1)}$ to maximize $Q(\theta, \theta^{(0)})$, (2.2.2)

In the next iteration, replace $\theta^{(0)}$ with $\theta^{(1)}$ and then repeat step (i) and step (ii) to convergence. This algorithm converges to the value of θ that maximizes the log-likelihood function $l_y(\theta)$ based on the incomplete data. If $l_y(\theta)$ is multi-modal, however the algorithm may converge to a local maximum or it can converge to a saddle point. In some fairly simple problems, such as the interval censoring, the algorithm can be shown to be globally convergent. In most applications, however, one would wish to

explore various initial values to be assured that the MLE has been obtained. It should be noted that the estimation of the variance of $\hat{\theta}$ generally involves a separate calculation of the observed information based on the incomplete likelihood, $l_y(\theta)$ and can not be computed directly by taking second derivative of Q. Louis (1982) gives some alternative approaches to estimating the variance-covariance matrix for $\hat{\theta}$.

It can be seen that the algorithm is monotone in that, at each step, the incomplete log-likelihood $l_y(\theta)$ can not decrease. To show this we observe that

$$l_y(\theta) = l_x(\theta) - l_{x|y}(\theta), \tag{2.2.3}$$

where, the second term on the right hand is the log-likelihood function coming from the conditional density of x given y. Taking an expectation conditional on y at $\theta = \theta^{(0)}$ gives

$$l_y(\theta) = Q(\theta, \theta^{(0)}) - E[l_{x|y}(\theta) \mid y; \theta^{(0)}], \tag{2.2.4}$$

which follows that

$$l_y(\theta^{(1)}) - l_y(\theta^{(0)}) = A + B, \tag{2.2.5}$$

where,

$$A = Q(\theta^{(1)}, \theta^{(0)}) - Q(\theta^{(0)}, \theta^{(0)}) \geq 0, \tag{2.2.6}$$

and

$$B = E[l_{x|y}(\theta^{(0)}) \mid \theta^{(0)}] - E[l_{x|y}(\theta^{(1)}) \mid \theta^{(0)}], \tag{2.2.7}$$

Since $l_{x|y}(\theta)$ is the log of the conditional density of x given y, it follows from well-known Jenson's Inequality that $B \geq 0$, and the monotonic follows. From the above discussion, it is easily seen that the MLE $\hat{\theta}$ coming from incomplete likelihood, $l_y(\theta)$ is a fixed point of the algorithm and under differentiability conditions, is a solution to the self-consistency equation

$$\frac{d}{d\theta_*} Q(\theta_*, \theta) \mid \theta_{*=\theta} = 0 \ , \tag{2.2.8}$$

In chapter 3, we will discuss the cure model along with some basic properties.

Chapter 3

Cure Model

3.1 Introduction

In some clinical trials, a substantial proportion of patients who respond favorably to treatment subsequently appear to be free of any signs or symptoms of the diseases and may be considered as cured, while the remaining patients may eventually relapse. The objective of these clinical trials is on estimating the proportion of patients who are cured (cure rate or cure fraction) and the failure time distribution of proportion of patients who are not cured (uncured patients).

The survival models incorporating a cure fraction are called cure rate models. Nowadays these models are becoming popular in analyzing data from cancer (or other diseases) clinical trials. The cure rate model has been widely used for modeling time to event data for various types of cancers for which a large (significant) proportion of patients are cured. For example, breast cancer, non-Hodgkins lymphoma, leukemia, prostate cancer, melanoma, head and neck cancer etc. Actually a cure model is a survival model where a fraction of population is not exposed to the hazard of interest. For example, a treatment for a fatal disease prevents death (from that disease) in a proportion of those treated (the cure rate or cure fraction). It may or may not delay death in the remaining individuals (vulnerable fraction). In fact cure model is a special case of frailty model, whether or not a patient is cured is an 'unobserved explanatory variable'.

3.2 Mixture model:

The mixture model was introduced by Berkson and Gage (1952). In this model assume that T be a non-negative random variable denoting the failure time. A certain proportion

π of population is 'cured'(cured patients) and the remaining 1- π are not cured (uncured patients). For this model, $S(t)$ denote the survival function for the whole population (cured and uncured patients) and $S^*(t)$ denote the survival function for the uncured patients in the population. The value of the survival function for the cured patients is 1. Thus the mixture model that has long been used for cure rate estimation can then be written as follows

$$S(t)=\pi + (1-\pi) S^*(t), \qquad (3.2.1)$$

Note that the survival function of the failure time distribution of cured patients is set equal to one for all finite values of the survival time because it is assumed that cured patients will never experience a relapse or death to the disease under investigation.

To estimate $S^*(t)$ in this model, we must specify the failure time distribution of uncured patients. The specification can be parametric or non parametric, which leads to parametric and semi-parametric cure model respectively. In parametric cure models, we assume a particular distribution for the failure time distribution of uncured patients. The most typical distribution that have been used for this purpose are: the Exponential distribution (Jones *et al.* 1981; Goldman, 1984; Ghitany *et al.* 1994), the Weibull distribution (Farewell, 1982; 1986), the Lognormal distribution (Boag,1949; Gamel *et al.* 1990), the Gompertz distribution(Gordon, 1990a, b; Cantor and Shuster, 1992) the Log-logistic distribution, the Gamma distribution, the Normal distribution, the Exponential power distribution, the Inverse Gaussian distribution and the Pareto distribution. More comprehensive distribution families such as the Extended Generalised Gamma (EGG) distribution and the Generalised F (GF) distribution (Yamaguchi,1992; Peng *et al.* 1998) have also been proposed recently to accommodate other forms of failure time

distributions for uncured patients. The model (3.2.1) is called Standard Cure Rate Model. It has been extensively used in the statistical literature by many authors, including Ewell and Ibrahim (1997), Farewell (1982, 1986), Goldman (1984), Gray and Tsiatis (1989), Greenhouse and Wolfe (1984), Halpern and Brown (1987a, 1987b), Kuk and Chen (1992), Laska and Meisner (1992), Sposto, Sather, and Baker (1992), Stangl and Greenhouse (1998), Taylor (1995), Yamaguchi (1992), Cantor and Shuster (1992), Peng and Carriere (2002), Chen, Ibrahim and Sinha (1999), Peng (2003), Yakovlev (1994), Spoto (2002), Peng, Dear and Denham (1998), Wienke *et al.* (2003), Li and Taylor (2002), Brown and Ibrahim (2003), Frankel and Longmate (2002), Tsodikov (2002), Tsodikov (2001). The model (3.2.1) is also called Mixture Cure Rate Model by several authors. In this model, the fraction of non-cured of patients ultimately will experience recurrence or other treatment failure according to a statistical distribution function, $F(t)$. Note that in model (3.2.1),

$$S^*(t) = 1 - F(t), \tag{3.2.2}$$

and $F(0) = 0, F(\infty) = 1$, so that $S^*(0) = 1, S^*(\infty) = 0$ and $S_1(\infty) = \pi$, the plateau value. The hazard function for this model is

$$h(t) = \frac{(1-\pi) f(t)}{S_1(t)}, \tag{3.2.3}$$

$$= \frac{f_1(t)}{S_1(t)}, \tag{3.2.4}$$

Where, $f(t)$ is the density corresponding to $F(t)$ and $f_1(t) = (1-\pi) f(t)$.

Originally, when treatments for cancer comprised surgery or radiation therapy alone administered over a short period, Mixture Cure Models has a practical interpretation

(Boag, 1949, Berkson and Gage, 1952). Patients either were cured by treatment (i.e., the disease was eradicated) or they were not, and the later would experience a recurrence after some time. Nowadays with combined modality treatment that can last up to three years in some children cancers, this interpretation does not apply, since eradication of disease, if it occurs, can occur at any time during treatment. The time to cure can not reliably be observed using current technology. Hence even though mixture cure models often fit cancer data well, they usually can not be viewed literally as describing a mixture of cured and uncured patients. The literal interpretation of the cure model is meaningful also in some non-cancer applications (Hauck *et al.* 1997). Cure models were first proposed 50 years ago (Boag, 1949, Berkson *et al.* 1952 and haybittle 1959) and have since received regular attention in the statistical literature (Haybittle 1965, Mould *et al.* 1975, O'Neill 1979, Farewell 1982, Goldman 1984, Sposto 1985, Farewell 1986, Halpern 1987, Goldman 1991, Sposto *et al.* 1992, Kuk *et al.* 1992, Maller *et al.*1992, Cantor *et al.* 1992, Ghitany *et al.* 1994, Yakovlev 1994, Cantor *et al.* 1994, Ghitany *et al.* 1995. Lee *et al.* 1995, Zhou *et al.* 1995, Ghitany *et al.* 1995, Laska *et al.* 1992, Tsodikov *et al.* 1998, Peng *et al.* 1998 Gieser *et al.* 1998, Tsodikov 1998, Hauck *et al.* 1997, de Angelis *et al.* 1999, Sy *et al.* 2000). However, they have not attained wide use or acceptance in the medical research literature, perhaps in part because of their reliance on particular parametric forms. However, parametric cure models provide a good empirical description of outcome in paediatric cancer data (Sposto *et al.* 1985, Sposto *et al.* 1992, Geiser *et al.* 1998) and have been used in analysis of adult (Gamel *et al.* 1993, Gamel amd Jones 1993, Gamel and Vogel 1993, Gamel and McLean 1994, Gamel and Vogel 1994, Gamel *et al.*, 1995) and childhood cancer (Nesbit *et al.* 1994). Most importantly, they provide a

single analytic method within which the effect of treatments and prognostic factors on the proportion cured can be assessed separately from their effect on the time to failure.

Although the mixture model appears to be attractive and is widely used, it has some drawbacks. Chen *et al.* (1999) have identified the following drawbacks of model (3.2.1).

Firstly, in the presence of covariates, it can not have a proportional hazard structure which is a desirable property for any survival model, because many asymptotic and computational results require a proportional hazard structure.

Secondly, when including covariates through the parameter π via a standard regression model, (3.2.1) yields improper posterior distributions for many types of non-informative improper priors, including the uniform prior for the regression coefficients. This is the most crucial drawback of model (3.2.1), because Bayesian inference with the model (3.2.1) essentially requires a proper prior.

Thirdly, the model (3.2.1) does not appear to describe the underlying biological process generating the failure time, at least in the context of cancer relapse, where cure rate model are frequently used.

3.3 Frailty model:

Chen *et al.* (1999) and Yakovlev *et al.* (1993) have proposed a different type of cure rate model, which is considered as a frailty model. The model they propose overcomes the drawbacks of mixture model. The model is quite attractive in several respects.

i) The model is derived from a natural biological motivation.

ii) It has proportional hazard structure through the cure rate parameter π, and thus has an appealing interpretation.

iii) It is very computationally attractive.

iv) The proposed model has a mathematical relationship with the standard cure rate model. Specifically one can show that any standard cure rate model can be written as the proposed model and vice versa.

v) The model yields proper posterior distribution under a wide class of non-informative improper prior for the regression coefficients, including an improper uniform prior. This is an especially solid feature of the model.

Suppose that for an individual in the population, N denotes the total number of carcinogenic cells (often called clonogens) for that individual left active after the initial treatment, and then assume that N has a Poisson distribution with mean θ. Also let Z_i denote the random time for the ith clonogenic cell to produce a detectable cancer mass. That Z_i can be viewed as an incubation time for the ith clonogenic cell. The variable Z_i, i = 1, 2, ..., are assumed to be identically independently distributed (i.i.d.) with a common distribution function $F(t) = 1 - S(t)$ and are independent of N. Where $S(t)$ is the survival function. The time to relapse of cancer can be defined by the random variable $T = \min\{ Z_i,\ 0 \le i \le N \}$, $P(Z_0 = \infty) = 1$ and N is independent of the sequence Z_1, Z_2, ...The survival function for T and hence the survival function for the population is given by

$S_p(t) = P(T > t) = P(\text{ no cancer by time } t)$

$= P(N = 0) + P(Z_1 > t, ...Z_N > t, N \ge 1)$

$= \exp(-\theta) + P(Z_1 > t)P(Z_2).....P(Z_N)P(N \ge 1)$

$= \exp(-\theta) + S(t)S(t)....S(t)\{P(N = 1) + P(N = 2) + ...\}$

$= \exp(-\theta) + S(t)^k \sum_{k=1}^{\infty} P(N = k)$

$$= \exp(-\theta) + \sum_{k=1}^{\infty} S(t)^k \frac{e^{-\theta}\theta^k}{k!}$$

$$= \exp(-\theta) + \exp(-\theta)\sum_{k=1}^{\infty} \frac{\{S(t)\theta\}^k}{k!}$$

$$= \exp(-\theta)(1 + \sum_{k=1}^{\infty} \frac{\{S(t)\theta\}^k}{k!})$$

$$= \exp(-\theta)\sum_{k=0}^{\infty} \frac{\{S(t)\theta\}^k}{k!}$$

$$= \exp(-\theta)\exp(\theta\, S(t))$$

$$= \exp(-\theta + \theta\, S(t))$$

$$= \exp(-\theta + \theta(1 - F(t)), \text{ since } S(t) = 1 - F(t)$$

$$= \exp(-\theta\, F(t)), \qquad (3.3.1)$$

The model (3.3.1) is not a proper survival function. Because $S_p(\infty) = \exp(-\theta)$. Note that $F(0) = 0$ and $F(\infty) = 1$. We observe that model (3.3.1) shows explicitly the contribution to the failure time of two distinct characteristics of tumor growth: the initial number of carcinogenic cells and the rate of their progression. Thus the model incorporates parameters bearing clear biological meaning. The model (3.3.1) is suitable for any type of failure data that has a surviving fraction (cure rate). Thus the model can be useful for modeling various types of failure time data, including time to relapse, time to death, time to first infection and so forth.

We also observe that the cure rate π is given by

$$S_p(\infty) = P(N = 0) = \exp(-\theta), \qquad (3.3.2)$$

As $\theta \to \infty$, the cure rate tends to zero, whereas as $\theta \to 0$, the cure rate tends to 1.*i.e.*, the cure rate lies between 0 and 1.

Note that by taking first derivative of (3.3.1), we get,

$S_p'(t) = \exp(-\theta F(t))(-\theta F'(t)$

$= - \theta f(t) \exp(-\theta F(t))$

Where $S_p'(t)$ and $F'(t)$ denote the first derivative of $S_p(t)$ and $F(t)$ respectively and

$F'(t) = f(t)$

Or, $- S_p'(t) = \theta f(t) \exp(-\theta F(t))$

Since, $- S_p'(t) = f_p(t)$

The density function corresponding to model (3.3.1) is given by

$$f_p(t) = \theta f(t) \exp(-\theta F(t)), \tag{3.3.3}$$

We observe that $S_p(t)$ is not a proper survival function, because $S_p(\infty) \neq 0$. Therefore $f_p(t)$ is not a proper probability density function. But $f(t)$ in model (3.3.3) is a proper density function. The hazard function is given by

$$h_p(t) = \frac{f_p(t)}{S_p(t)} = \frac{\theta f(t) \exp(-\theta F(t))}{\exp(-\theta F(t))}$$

$$= \theta f(t), \tag{3.3.4}$$

We observe that $h_p(t)$ is not a proper hazard function corresponding to a probability distribution, because $S_p(t)$ is not a proper survival function. The cure rate model (3.3.1) yields an attractive form for the hazard function in (3.3.4). Basically, we observe that $h_p(t)$ is multiplicative in θ and $f(t)$ and thus has the proportional hazard structure, with

the covariates modeled through θ. This form of hazard is more appealing than the one from standard cure rate model in (3.2.1), which does not have the proportional hazard structure if $\pi = (e^{-\theta})$ is modeled as function of covariates. The proportional hazard property in (3.3.4) is also computationally attractive. The survival function for the 'non-cured' group is given by

$$S^*(t) = P(T > t \mid N \geq 1)$$

$$= \frac{P(Z_1 > t, Z_2 > t, \ldots Z_N > t, N \geq 1)}{P(N \geq 1)}$$

$$= \frac{\sum_{k=1}^{\infty} S^k(t) \frac{e^{-\theta}\theta^k}{k!}}{1 - P(N < 1)}$$

$$= \frac{e^{-\theta} \sum_{k=1}^{\infty} \frac{(\theta S(t))^k}{k!}}{1 - P(N = 0)}$$

$$= \frac{e^{-\theta} (\sum_{k=0}^{\infty} \frac{(\theta S(t))^k}{k!} - 1)}{1 - e^{-\theta}}$$

$$= \frac{e^{-\theta} \sum_{k=0}^{\infty} \frac{(\theta S(t))^k}{k!} - e^{-\theta}}{1 - e^{-\theta}}$$

$$= \frac{e^{-\theta} e^{\theta S(t)} - e^{-\theta}}{1 - e^{-\theta}}$$

$$= \frac{e^{-\theta + \theta S(t)} - e^{-\theta}}{1 - e^{-\theta}}$$

$$= \frac{e^{-\theta(1 - S(t))} - e^{-\theta}}{1 - e^{-\theta}}$$

$$= \frac{e^{-\theta F(t)} - e^{-\theta}}{1 - e^{-\theta}}$$

$$= \frac{\exp(-\theta F(t)) - \exp(-\theta)}{1 - \exp(-\theta)}, \tag{3.3.5}$$

Note that, in (3.3.5), $S^*(0) = 1$ and $S^*(\infty) = 0$, so that $S^*(t)$ is a proper survival function. Thus the survival density corresponding to (3.3.5) is given by

$$f^*(t) = -\frac{d}{dt} S^*(t) = -\left(\frac{e^{-\theta F(t)}(-\theta F'(t))}{1 - e^{-\theta}}\right)$$

$$= \frac{e^{-\theta F(t)} \theta f(t)}{1 - e^{-\theta}}$$

$$= \left(\frac{\exp(-\theta F(t))}{1 - \exp(-\theta)}\right) \theta f(t), \tag{3.3.6}$$

Here, $f^*(t)$ is a proper density function. Since $S^*(t)$ is a proper survival function. The hazard function for the non-cured group (population) is given by

$$h^*(t) = \frac{f^*(t)}{S^*(t)} = \left(\frac{\exp(-\theta F(t))}{1 - \exp(-\theta)}\right) \theta f(t) \div \frac{\exp(-\theta F(t)) - \exp(-\theta)}{1 - \exp(-\theta)}$$

$$= \frac{\exp(-\theta F(t)) h_p(t)}{\exp(-\theta F(t)) - \exp(-\theta)}, \quad [\text{ by using (3.3.4) }]$$

$$= \left(\frac{\exp(-\theta F(t))}{\exp(-\theta F(t)) - \exp(-\theta)}\right) h_p(t), \tag{3.3.7}$$

We observe that the model (3.3.7) does not have a proportional hazard structure, because $\frac{\exp(-\theta F(t))}{\exp(-\theta F(t)) - \exp(-\theta)}$ can never be free of t for any $f(t)$. Also $h^*(t) \to 0$ as $t \to \infty$ and thus $h^*(t)$ converges to the hazard function of the incubation time random variable Z as $t \to \infty$. At last it can be shown that the hazard function $h^*(t)$ is an increasing function

of θ. There is an attractive mathematical relationship between the models in (3.2.1) and (3.3.1). From (3.3.5), we can write

$$\exp(-\theta F(t)) - \exp(-\theta) = (1 - \exp(-\theta))S^*(t)$$

$$\text{or, } \exp(-\theta F(t)) = (1 - \exp(-\theta))S^*(t) + \exp(-\theta), \tag{3.3.8}$$

Using (3.3.8) in (3.3.1), we obtain the following model as

$$S_p(t) = \exp(-\theta) + (1 - \exp(-\theta))S^*(t), \tag{3.3.9}$$

Where, $S^*(t)$ is given by (3.3.5). Therefore, $S_p(t)$ is a standard cure rate model with cure rate equal to $\pi = e^{-\theta}$ and the survival function for the non-cured population given by $S^*(t)$. This shows that every model defined by (3.3.1) can be written as a standard cure rate model. This results also means that every standard cure model corresponds to some model of the form (3.3.1) for some θ and distribution function $F(.)$. We want to mention here that if the covariates enter through θ, then $S_p(t)$ can be taken to have a Cox proportional hazards structure, but $h^*(t)$ will not have a proportional hazard structure. In the proposed model, we model the entire population as a proportional hazard model, whereas in the standard cure rate model, only the non-cured group can be modeled with a proportional hazard structure.

3.4 Review of cure rate estimation

The following sections present brief review of the work done in cure rate estimation.

3.5 Work of Peng and Carriere

Peng and Carriere(2002) have proposed parametric and semi-parametric cured models for cure rate estimation. In their study, several parametric and semi-parametric models are compared and their estimation methods are discussed within the framework of EM algorithm. They showed that semi-parametric cure models can achieve efficiency levels similar to those of parametric cure models, provided that the failure time distribution is well specified and non-cured patients have an increasing hazard rate. Therefore they recommended that the semi-parametric model is a viable alternative to parametric cure models. They have proposed mixture models for these analyses.

Let T denote the failure time and $S(t \mid x, z)$ be the survival function of T, where x and z are two covariate vectors on which the distribution of T may depend. A mixture models for cure rate estimation can then be written as follows

$$S(t \mid x, z) = \pi(z) S_u(t \mid x) + 1 - \pi(z) \qquad (3.5.1)$$

Where, $\pi(z)$ is the proportion of non-cured patients depending on z with unknown parameter vector γ by a logistic form $\log[\pi(z)/\{1-\pi(z)\}] = \gamma' z$ and $S_u(t \mid x)$ is the survival function of the failure time distribution of non-cured patients depending on x with unknown parameter vector β. The survival function of the failure time distribution of cured patients is set to be equal to 1 because it is assumed that cured patients will never experience a relapse or death due to disease under investigation. To estimate β, one should specify the failure time distribution of non-cured patients. The specification can be parametric and non-parametric. In parametric cure models we assume a particular distribution for the failure time of non-cured patients. Semi-parametric cure models in

cure rate estimation have gained attention only in recent years. Taylor (1995) used Kaplan-Mier survival estimator to model the failure time distribution of uncured patients and the EM algorithm to obtain estimates of the parameter of interest. Peng and Dear (2000) and Sy and Taylor (2000) proposed to use EM algorithm to estimate the parameters which simplifies the method used by Kuk and Chen (1992). To facilitate the application of these models, they compared the performance of parametric and semi-parametric cure models. They reviewed the existing estimation method and used simulation to compare the mean squared error (MSE) and bias of the estimates from the models. They perform their analysis as follows:

Cure model and the EM algorithm

Considering the data in the form $(t_i, \delta_i, x_i, z_i), i = 1,2,..n,$ where t_i, is the observed survival time for the ith patient; δ_i, is the censoring indicator with 1 if t_i is uncensored and 0 otherwise; and x_i, and z_i are observed values of two covariate vectors. Let $\theta' = (\beta', \gamma')$ that relates to x_i,and $z_i$respectively. To use the EM algorithm, define $c' = (c_1, c_2, ..., c_n)$, where c_i is an indicator of the status of being cured for ith patient, taking a value 1 if the patient is cured and 0 otherwise. Given c, the complete log-likelihood function is

$$l(\theta|c) = \log \prod_{i=1}^{n} [\{\pi(z_i) f_u(t_i|x_i)\}^{1-c_i}]^{\delta_i} [\{1-\pi(z_i)\}^{c_i} \{\pi(z_i) S_u(t_i|x_i)\}^{1-c_i}]^{1-\delta_i}, \qquad (3.5.2)$$

The EM algorithm starts with an initial value $\theta^{(0)}$. The E-step in the $(r+1)$th iteration calculates the expectation of the complete log-likelihood function $l(\theta|c)$ with respect to

C, conditional on the observed data and $\theta^{(r)}$, the estimate of θ at the rth iteration. This is equivalent to calculating the following conditional expectation

$$g_i^{(r)} = E(1-c_i)\left|\theta^{(r)}\right) = \delta_i + (1-\delta_i)\frac{\pi(z_i)S_u(t_i|x_i)}{1-\pi(z_i)+\pi(z_i)S_u(t_i|x_i)}, \tag{3.5.3}$$

That is, (3.5.3) is the rth estimator of the probability of the ith patient being uncured. Let $g^{(r)'} = (g_1^{(r)}, g_2^{(r)}, ..., g_n^{(r)})$. The M-step in the $(r+1)$th iteration maximizes the expected complete log-likelihood function with respect to θ to obtain $\theta^{(r+1)}$, where the expected log-likelihood function is the sum of the two following functions:

$$l_1(\gamma\left|g^{(r)}\right) = \sum_{i=1}^{n}[g_i^{(r)}\log\pi(z_i) + (1-g_i^{(r)})\log\{1-\pi(z_i)\}], \tag{3.5.4}$$

$$l_2(\beta\left|g^{(r)}\right) = \sum_{i=1}^{n}[g_i^{(r)}\log s_u(t_i|x_i) + \delta_i \log h_u(t_i|x_i)] , \tag{3.5.5}$$

where $h_u(.) = f_u(.)/s_u(.)$ is the hazard function of the failure time distribution of uncured patients. The algorithm is stopped until $| \, |\theta^{(r+1)} - \theta^{(r)}| \, |$ is sufficiently small. Equation (3.5.4) can be maximized by the Newton-Raphson method and maximizing equation (3.5.5) will depend on how the failure time distribution of uncured patients is specified.

Parametric cure model

In parametric cure models, a specific distribution is assumed for the failure time distribution of uncured patients. The most commonly used distribution includes the Exponential, Gamma, Weibull and Lognormal distribution. Yamaguchi(1992) and Peng *et al.*(1998) proposed the use of Extended Generalized Gamma(EGG) Distribution and the Generalized F(GF) Distribution. If the GF distribution is used in parametric cure models, the density function of the failure time distribution of uncured patients is given by

$$f_u(t|x) = p\lambda(\lambda t)^{ps_1-1}[1+(\lambda t)^p]^{-s_1-s_2} B^{-1}(s_1,s_2) \quad , \tag{3.5.6}$$

where, $\lambda = (s_1/s_2)e^{-\beta' x}$ is a scale parameter with positive shape parameters s_1, s_2, p, and $B(s_1,s_2)$ is the beta function evaluated at s_1, s_2.The advantages of using the GF distribution is its flexibility: it include all the distributions mentioned above as special cases. For example, when $s_1 \to \infty$ or $s_2 \to \infty$, the GF distribution reduces to the EGG distribution with the following density

$$f_u(t|x) = \frac{|q|(q^{-2})^{q^{-2}} p\lambda(\lambda t)^{p/q-1}\exp[-q^{-2}(\lambda t)^{qp}]}{\Gamma q^{-2}}, q \neq 0$$

$$= p(\sqrt{2\pi t})^{-1}\exp(-p^2[\log(\lambda t)]^2/2), q = 0$$

where, $\lambda = \exp(-\beta' x)$ and $q = 1/\sqrt{s_1}$ if $s_2 \to \infty$ and $q = -1/\sqrt{s_2}$ if $s_1 \to \infty$. The EGG distribution further reduces to the lognormal when $q = 0$, the Weibull distribution when $q = 1$, and the exponential when $q = 0$ and $p = 1$. Detailed discussion of the relationship to other distributions can be found in Peng *et al.*(1998). Once the distribution of failure time

for the uncured patients is specified $\beta^{(r+1)}$ can be obtained by maximizing (3.5.5). When the GF distribution is reduced to Weibull distribution, the hazard function of uncured patients becomes $h_u(t|x) = pt^{p-1}\exp(-\beta' x)$, where x also acts on the hazard function multiplicatively, which leads to a PH model.

Semi-parametric cure model

Considering that the PH assumptions are appropriate for the effect of x on the failure time distribution of uncured patients (that is, $h_u(t|x) = h_{u0}(t)\exp(\beta' x)$, where $h_{u0}(t)$ is the base line hazard function of uncured patients, Sy and Taylor (2000) and Peng and Dear (2000) recently proposed Cox's partial-likelihood type method to estimate β semi-parametrically without specifying $h_{u0}(t)$.This method can viewed as an alternative to the parametric approaches above. For this method, a Cox's partial-likelihood-type function can be obtained from (3.5.5):

$$\sum_{j=1}^{k}[\beta' S_{(j)} - d_j \log(\sum_{i\in R_j}^{n} g_i^{(r)} \exp(\beta' x)] \ , \tag{3.5.7}$$

where k is the number of distinct uncensored failure times $\tau_1 \le \tau_2 \le ..... \le \tau_k$, d_j is the number of uncensored observations at τ_j, $S_{(j)}$ is the sum of covariates associated with the uncensored observations at τ_j, and R_j is the risk set at τ_j. Therefore in the M-step, β can be separated from $h_{u0}(t)$ and $\beta^{(r+1)}$ can be obtained directly from (3.5.7).

A non-parametric estimate of $h_{u0}(t)$ or the baseline cumulative hazard function $H_{u0}(t) = \int_0^t h_{u0}(w)dw$ can be obtained following the method similar to those proposed for the Cox's PH model (Kalbfleisch and prentice, 1980, p. 85). That is, a non-parametric maximum likelihood estimator of $H_{u0}(t)$ in the $(r+1)$th iteration of the EM algorithm is given by $H_{u0}^{(r+1)}(t) = \sum_{j:\tau_j \le t} \frac{d_j}{\sum_{i \in R_j} g_i^{(r)} \exp(\beta^{(r+1)'} x_1)}$.

Finally, Peng and Carriere (2000) conducted a simulation study to compare the performance of the parametric and semi-parametric cure models in studying cure proportions and the failure time distribution of uncured patients. Performance is measured by the MSE and bias of estimated survival function and cure proportion. They found that when the hazard function of uncured patients is decreasing, all cure models suffer from large MSEs and biases. The poor performances of the models can be explained in part by the fact that when the hazard rate of uncured patients is deceasing, it is difficult to distinguish a cured patient from an uncured patient who has a small hazard rate after a long term follow up. Even the Weibull cure model has rather large MSE and biases in estimating the parameters. Semi-parametric cure models especially suffer from a large bias in cure proportion because the estimated survival and hazard function are forced to be zero and infinite, respectively, after the largest observed failure time. However when the hazard function is not decreasing, the PH cure model becomes viable alternative for cure proportion. The findings reported here suggested that its performance is comparable to that of an appropriately specified parametric cure model. Nonetheless,

the semi-parametric cure model is useful in situations where determining an appropriate distribution for the failure time of uncured patients is difficult.

3.6 Work of Cantor and Shuster

When analyzing survival data from clinical trials, it is sometimes clear that a non-zero proportion of patients can be considered as cured. Cantor and Shuster (1992) make a constructive discussion about parametric and non-parametric methods for estimating cure rate based on censored survival data. They use Kaplan-Meier method for non-parametric estimation of cure proportion. On the other hand for parametric estimation of cure rates, they assume a survival function S(t) for which $\lim_{t\to\infty} S(t) = S(\infty) > 0$ i.e., in a proportion of patients the event never occurs, using MLE one can estimate $S(\infty)$, which is considered as cure rate. Two such survival functions are:

$$S_1(t) = \pi + (1-\pi)\exp(-\lambda t), 0 < \pi < 1, \lambda > 0 \ , \tag{3.6.1}$$

$$S_2(t) = \exp[\beta^{-1}\alpha\{1 - \exp\beta(t)\}], \alpha > 0, \beta < 0 \ , \tag{3.6.2}$$

The model (3.6.1) represents the case where a proportion π of patients are cured, while the remaining 1-π have an exponential failure rate. The model (3.6.2) is a modified Gompertz survival distribution, for which $S_2(t)$ approaches $\exp(\alpha/\beta)$ asymptotically. The model (3.6.1) based on the exponential distribution has been extensively discussed by Goldman (1984), who studied the performance of Maximum likelihood estimates of the parameters of $S_1(t)$ by Monte Carlo method. The model (3.6.2) was developed by Gompertz (1825) and was motivated by observed population life tables. Garg *et al.*,

(1970) discuss the Gompertz distribution and the maximum likelihood estimation of its parameters. They do not consider the situation in which the hazard is decreasing, causing $S(\infty)$, the 'cure rate' to be positive. Gehan and Siddiqui (1973) present a least squares procedure for estimating the parameters of the Gompertz distribution. Given a set of survival data, either of the above survival functions can be used as the basis of likelihood function whose logarithm can be maximized to find parameter estimates. Estimates of cure rates are given by $\hat{\pi}$ or $\exp(\hat{\alpha}/\hat{\beta})$. The maximization of the likelihood function based on $S_1(t)$ creates problems. Newton-type methods often fail or require a large number of iterations. Convergence of values outsides the domain of the parameter space occurs frequently for data that do not support a non-zero cure rate, for example where the Kaplan Meier curve does not have a plateau. On the other hand, the likelihood function based on Gompertz distribution is maximized readily using Newton- Raphson method. Here $\text{var}(\hat{\alpha})$ may be estimated from the information matrix and $\text{var}[\exp(\hat{\alpha}/\hat{\beta})]$ using delta method. Finally they recommended that Gompertz model is more suitable than exponential model in parametric estimation of cure rate.

Since the exponential distribution, in many cases, clearly does not fit the survival data well, the use of non-parametric methods such as Kaplan-Meier is attractive. However, the Kaplan-Meier method is not free of problems and limitations. Cantor and Shuster (1992) have demonstrated the potential for misleading or counter-intuitive results. Miller (1983) shows that the asymptotic efficiency of the Kaplan–Meier estimator tends to be low relative to the MLE for a given distribution. In fact, the asymptotic efficiency approaches 0 as t tends to infinity. The Gompertz distribution may provide a useful alternative to

non-parametric method. Estimation of cure rate using this distribution may be preferable to the use of the plateau of the Kaplan-Meier curve.

3.7 Work of Tsodikov

Tsodikov (2001) studied the estimation of survival function based on proportional hazard model with cure. He proposed an algorithm to fit the proportional hazard model restricted by the fixed survival rates at the end of observation period. He used parametric cure model to estimate the proportion of long term survivors. To combine the stability of the parametric method, the survival function is estimated non-parametrically conditional on the cure rates provided by the parametric analysis. Time to relapse or freedom from treatment failure is an endpoint used in many cancer clinical trials when evaluation of the efficiency of primary therapy of cancer is of interest. Usually the semi parametric proportional hazard model [D.R. Cox , 1972] is used to estimate the treatment effects. The population survival function is given by

$G(t) = \exp\{-H(t)\}$, where, $H(t) = \int h(x)dx$ is the cumulative hazard function and $h(x)$ is the hazard function. The survival function $G(t)$ is improper which means that the cumulative hazard is bounded

$$H(t) \quad \leq \theta \ \lim_{t\to\infty} H(t) = \theta \ , \tag{3.7.1}$$

By considering $H(t) = \theta F(t)$, we can adjust the above property. Where $F(t)$ is the proper distribution function. An alternative form of Cox's model is given by

$$G(t) = \exp\{-\theta F(t)\} , \tag{3.7.2}$$

Estimation of cure rates based on continuous model

Along with model (3.7.2) with parametrically specified $F(.)$, [Yakovlev and Tsodikov,1996] there exists a variety of other parametric cure model. Most of them are of mixture type. The simplest model of this type would assume that the population is divided into two subpopulations due to some unobserved prognostic factor, so that an individual is either a long term survivor with some probability or has a proper survival function otherwise [Farewell, V.T., 1982]. Note that the model (3.7.2) can be simultaneously interpreted both as a mixture model, as well as simple model of carcinogenesis. According to the carcinogenic interpretation, the observed tumor originates from clonogens that have a random progression time X with the distribution function F (.). It is assumed that the number ν of such clonogens in a patients is Poisson distributed with parameter θ. The observed tumor onset is associated with completing of the progression by one of the competing clonogens, so that the tumor onset time U is given by U = min $\{X_i$, i = 1,2,..., $\nu\}$. Where X_i, is the time of ith clonogens to produce detectable tumor. Let X_i, is i.i.d random variables with distribution function F (.). Thus the $S(t;\nu) = [1 - F(t)]^{\nu}$. If the number of clonogens in a patients could be measured, we would have a PH model with the baseline cumulative hazard as

$$H_0(t) = -\log(1 - F(t)), H_0(\infty) = \infty.$$

Since ν is unobservable, we take an expectation over the frailty and obtain (3.7.2) for the population survival function. Suppose that $t_1 \le t_2 \le .... \le t_n$ are the ordered failure times with $t_o = 0, t_{n+1} = \infty$. Let m_{ij} be the number of failures at t_i in the treatment group j

=1,2,…,k, and n_{ij} be the number of censored observations. Let θ_j be the value of θ in the treatment group j. The cure rates are given by exp(-θ_j). The log-likelihood of data based on a continuous model can be written as

$$l = \sum_{i=1}^{n}\sum_{j=1}^{k}\{m_{ij}\log[\theta_j f(t_i)] - (m_{ij} + n_{ij})\theta_j F(t_i)\}, \tag{3.7.3}$$

Where $f(.)$ is the probability density function corresponding to $F(.)$ The estimates of θ_j are obtained by the solution of following equation

$$\frac{\partial l}{\partial \theta_j} = 0, \text{j} = 1,2,\ldots,\text{k}\ , \tag{3.7.4}$$

Thus the MLE of θ_j is given by

$$\hat{\theta}_j = \frac{M_j}{\sum_{i=1}^{n}(m_{ij} + n_{ij})F(t_i)} = \frac{M_j}{\sum_{i=1}^{n} R_{ij}\Delta F_i}\ , \tag{3.7.5}$$

Where M_j is the number of failures in the treatment group j, R_{ij} is the number of patients at risk at t_i –0 in group j.

Restricted estimate of survival function under the cure rate

In order to estimate $F(.)$ non-parametrically, Tsodikov (2001) consider it as a step function and use θ as estimated from a continuous model. The generalized log-likelihood function of the data on the class of step functions can be written as

$$l = \sum_{i=1}^{n}\sum_{j=1}^{k}\{m_{ij}\log[\Delta G_{ij}] - n_{ij}\theta_j F_i\}, \tag{3.7.6}$$

Where, $\Delta G_{ij} = G_j(t_{i-1}) - G_j(t_i)$ and F_n=1. Suppose that l as a function of ΔF_i, i =1,2,..,n-1, where $\Delta F_n = 1 - \sum_{i=1}^{n-1} \Delta F_i$. To get the MLE of F, we maximize (3.7.6) with respect to ΔF_i, i =1,2,...,n-1. This is performed by the following numerical algorithm.

Define the functions $z_i = \varphi_i(x)$ as solutions to the following equations

$$\sum_{j=1}^{k} \frac{\theta_j m_{ij}}{1-\exp\{-\theta_j z_i\}} = \sum_{j=1}^{k} \{R_{ij} - m_{nj} - n_{nj}\} + \sum_{j=1}^{k} \frac{m_{nj}\theta_j}{1-\exp\{-\theta_j x\}}, \qquad (3.7.7)$$

$i = 1,2,..,n-1$

Then, solve the equation $x + \sum_{i=1}^{n-1} \varphi_i(x) = 1$, (3.7.8)

Let x^* be the solution of (3.7.8). As a result of the above estimation procedure, we have

$\Delta \hat{F}_i = \varphi(x^*)$, i =1,2,...,n-1, $\Delta \hat{F}_n = x^*$.

He showed that the above algorithm indeed finds the maximum likelihood estimator for F.

Chapter 4

Estimation of Cure rate

4.1 Introduction:

Suppose that X is the life time of a patient. Then $P(X=\infty)=\lim_{t\to\infty}P(X>t)=e^{-\theta}$, which is considered as cure rate. On the other hand, $P(X<\infty)=1-e^{-\theta}, 0\le t\le\infty$, which is the probability of non-cured. Since $X\in[0,\infty)\cup\{"cured"\}$, hence the probability density function of X can be written as

$$f_\theta(x)=\theta f_0(x)e^{-\theta F_0(x)}1_{\{x<\infty\}}+e^{-\theta}1_{\{x="cured"\}} \tag{4.1.1}$$

with respect to the measure

$$\mu(A)=\int_{A\cap[0,\infty)}dx+1_{\{"cured"\in A\}} \quad \text{for } A\subseteq[0,\infty)\cup\{"cured"\}$$

Thus, $P\{X\in A\}=\int_{A\cap[0,\infty)}\theta f_0(x)e^{-\theta F_0(x)}dx+e^{-\theta}1_{\{"cured"\in A\}}$

By using non-parametric maximum likelihood method we can estimate the parameter of the cure models [(3.2.1) and (3.3.1)]. We are interested to estimate the parameter by using uncensored and censored data.

4.2 Estimation of cure rate from mixture model

Uncensored case: $F_0(.), f_0(.)$ and θ are unknown. Here we observe both cured and non-cured group.

Suppose that we have the data in the form $(x_i,\in_i), i=1,2,..,n$, where x_i denotes the survival time for the ith patient, $\in_i$ is the cured indicator with 1 if x_i is not cured and 0 otherwise i.e., $\in_i=1_{\{X_i<\infty\}}$.

From the mixture model (3.2.1), we can write

$$1-F(x)=\pi+(1-\pi)(1-F^{*}(x))$$

or, $F(x)=(1-\pi)F^{*}(x)$

Therefore, $\Delta F(x)=(1-\pi)\Delta F^{*}(x)$ (4.2.1)

Likelihood function: The non-parametric likelihood function is given by

$$L_n(\pi,F)=\prod_{i=1}^{n}(\Delta F(x_i))^{\epsilon_i}(\pi)^{1-\epsilon_i}, \quad (4.2.2)$$

Where, $\Delta F(x_i)$ = jump of F(.) at x_i

Using (4.2.1) in (4.2.2)) we get the following likelihood as

$$L_n(\pi,F)=\prod_{i=1}^{n}\{(1-\pi)\Delta F^{*}(x_i)\}^{\epsilon_i}(\pi)^{1-\epsilon_i}$$

$$=\prod_{i=1}^{n}\{(1-\pi)p_i\}^{\epsilon_i}(\pi)^{1-\epsilon_i}$$

Where, $p_i=\Delta F^{*}(x_i)$

Therefore, the log likelihood function becomes

$$\log L_n(\pi,F)=\sum_{i=1}^{n}\epsilon_i\log(1-\pi)+\sum_{i=1}^{n}\epsilon_i\log p_i+\sum_{i=1}^{n}(1-\epsilon_i)\log\pi \quad (4.2.3)$$

We want to maximize (4.2.3) subject to condition $\sum_{i=1}^{n}p_i=1$

Adding a Lagrange multiplier λ, we obtain the following equation

$$\log L_n(\pi,F)=\sum_{i=1}^{n}\epsilon_i\log(1-\pi)+\sum_{i=1}^{n}\epsilon_i\log p_i+\sum_{i=1}^{n}(1-\epsilon_i)\log\pi-\lambda(\sum_{i=1}^{n}p_i-1) \quad (4.2.4)$$

Therefore, the non-parametric maximum likelihood estimators of p_i and π are obtained by the solution of the following equations

$$\frac{\partial \log L_n(\pi, p_i)}{\partial p_i} = 0, \frac{\partial \log L_n(\pi, p_i)}{\partial \pi} = 0 \tag{4.2.5}$$

Thus, $\frac{\partial \log L_n(\pi, p_i)}{\partial p_i} = 0$ gives

$$\frac{\epsilon_i}{p_i} - \lambda = 0, i = 1,2,..,n \tag{4.2.6}$$

$\frac{\partial \log L_n(\pi, p_i)}{\partial \lambda} = 0$ gives

$$\frac{\sum_{i=1}^{n}(1-\epsilon_i)}{\pi} - \frac{\sum_{i=1}^{n}\epsilon_i}{1-\pi} = 0 \tag{4.2.7}$$

Solving (4.2.6) and (4.2.7), we get

$$\hat{p}_i = \frac{\epsilon_i}{\sum_{i=1}^{n}\epsilon_i} \text{ and } \hat{\pi} = \frac{\sum_{i=1}^{n}(1-\epsilon_i)}{\sum_{i=1}^{n}(1-\epsilon_i) + \sum_{i=1}^{n}\epsilon_i}$$

Censored case: $F_0(.), f_0(.)$ and θ **are unknown.** Suppose that we have data in the form $(Z_i, \delta_i), i = 1,2,..n.$ Where $Z_i = \min\{X_i, T\}$, δ_i is the censoring indicator and $\delta_i = 1_{\{X_i \le T\}}$, T is fixed censoring time.

Likelihood function: The non-parametric likelihood function based on Type 1 censoring can be written as follows

$$L_n(F, \pi) = \prod_{i=1}^{n}\{(1-\pi)p_i\}^{\delta_i}\{\pi + (1-\pi)S^*(T)\}^{1-\delta_i} \tag{4.2.8}$$

$$= L_n(F, \pi) = [\prod_{i=1}^{k}\{(1-\pi)p_i\}]\ \{\pi + (1-\pi)S^*(T)\}^{n-k}$$

Where, $k = \max\{i | X_i \le T\}$

The log-likelihood function becomes

$$\log L_n(F,\pi) = k\log(1-\pi) + \sum_{i=1}^{k}\log p_i + (n-k)\log\{\pi + (1-\pi)(1-\sum_{i=1}^{k} p_i)\} \qquad (4.2.9)$$

Note that the above equation (4.2.9) is very difficult to maximize subject to condition

$\sum_{i=1}^{k} p_i \le 1$.

4.3 Estimation of cure rate from frailty model

In uncensored data, we consider the following cases:

Case –(a1): $F_0(.), f_0(.)$ are known and θ **unknown and also both cured and non-cured observed.**

This case is fully parametric. Suppose that we have the data in the form $(x_i, \in_i), i = 1,2,..,n$, where x_i denotes the survival time for the ith patient, $\in_i$ is the cured indicator with 1 if x_i is not cured and 0 otherwise. Also, $x \in [0,\infty)\cup\{\text{"cured"}\}$, $1 \le i \le n$.

Likelihood function:

The likelihood function is given by

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i) = \prod_{i=1}^{n} (\theta f_0(x_i) e^{-\theta F_0(x_i)})^{\in_i} (e^{-\theta})^{1-\in_i}, \qquad (4.3.1)$$

Where, $\in_i = 1_{\{X_i<\infty\}}$ and $1-\in_i = 1_{\{X_i=\text{"cured"}\}}$

$$= \theta^{\sum_{i=1}^{n}\in_i} \{f_0(x_i)\}^{\sum_{i=1}^{n}\in_i} e^{-\theta\sum_{i=1}^{n}\in_i F_0(x_i)} e^{-\theta\sum_{i=1}^{n}(1-\in_i)}$$

The log-likelihood function is given by

$$\log L(\theta) = (\sum_{i=1}^{n}\in_i)\log\theta + \sum_{i=1}^{n}\in_i \log f_0(x_i) - \theta\sum_{i=1}^{n}\in_i F_o(x_i) - \theta\sum_{i=1}^{n}(1-\in_i), \qquad (4.3.2)$$

Differentiating (4.3.2) with respect to θ and equating to zero i.e., $\frac{\partial \ln L(\theta)}{\partial \theta} = 0$ gives

$$\frac{\sum_{i=1}^{n} \in_i}{\theta} - \sum_{i=1}^{n} \in_i F_o(x_i) - \sum_{i=1}^{n} (1 - \in_i) = 0$$

$$\text{or, } \hat{\theta} = \frac{\sum_{i=}^{n} \in_i}{\sum_{i=1}^{n} \in_i F_o(x_i) + \sum_{i=1}^{n} (1 - \in_i)}, \tag{4.3.3}$$

$$\text{Therefore, } \hat{\theta} = \frac{\frac{1}{n}\sum_{i=}^{n} \in_i}{\frac{1}{n}\sum_{i=1}^{n} \in_i F_o(x_i) + \frac{1}{n}\sum_{i=1}^{n} (1 - \in_i)}, \tag{4.3.4}$$

Which is the required estimate of θ. Therefore, the estimate of cure rate is $\pi = e^{-\theta}$. Now, we will show that $\hat{\theta}$ converges to the true parameter θ.

Convergence of $\hat{\theta}$:

From the law of convergence, we can write,

$$\frac{1}{n}\sum_{i=1}^{n} \in_i \xrightarrow{a.s} E(\in) = P(X < \infty) = 1 - e^{-\theta}, \tag{4.3.5}$$

$$\frac{1}{n}\sum_{i=1}^{n} (1 - \in_i) \xrightarrow{a.s} E(1 - \in) = P(X > \infty) = e^{-\theta}, \tag{4.3.6}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} \in_i F_0(x_i) \xrightarrow{a.s} E(\in F_o(x))$$

$$= E(1_{\{X<\infty\}} F_0(x))$$

$$= \int_0^{\infty} \theta f_0(t) F_0(t) e^{-\theta F_o(t)} dt$$

$$= \int_0^1 \theta z e^{-\theta z} dz \quad [\text{ putting } F_0(t) = z]$$

$$= \int_0^1 \theta z e^{-\theta z} dz + \int_0^1 e^{-\theta z} dz$$

$$= -[z e^{-\theta z}]_0^1 - [\frac{1}{\theta} e^{-\theta z}]_0^1$$

$$= -e^{-\theta} - \frac{1}{\theta}(e^{-\theta} - 1)$$

$$= \frac{1-e^{-\theta}}{\theta} - e^{-\theta}, \tag{4.3.7}$$

Using (4.3.5), (4.3.6) and (4.3.7) in (4.3.4), we obtain the following expression

$$\hat{\theta} = \frac{1-e^{-\theta}}{(\frac{1-e^{-\theta}}{\theta}) - e^{-\theta} + e^{-\theta}}$$

Therefore, $\hat{\theta} \to \theta$

Case-(a2): $F_0(.), f_0(.)$ are known and θ **unknown and only non-cured are observed.**

This case is also fully parametric. We also consider here only non-cured group i.e., we observe X if $X < \infty$ i.e., $X_i \in [0, \infty)$. The p.d.f of X given $X < \infty$ can be written as

$$f_\theta^{nc}(x) = \frac{\theta f_o(x) e^{-\theta F_0(x)}}{1-e^{-\theta}}, \; X < \infty \tag{4.3.8}$$

Likelihood function: The likelihood function is given by

$$L^{nc}(\theta) = \prod_{i=1}^{n} f_\theta^{nc}(x)$$

$$= \prod_{i=1}^{n} \frac{\theta f_0(x_i) e^{-\theta F_0(x_i)}}{1-e^{-\theta}}$$

$$= \frac{\theta^n \prod_{i=1}^{n} f_0(x_i) e^{-\theta \sum_{i=1}^{n} F_0(x_i)}}{(1-e^{-\theta})^n}$$

Thus the log-likelihood function becomes

$$\log L^{nc}(\theta) = n\log\theta + \sum_{i=1}^{n} \log f_0(x_i) - \theta \sum_{i=1}^{n} F_0(x_i) - n\log(1-e^{-\theta}), \tag{4.3.9}$$

Differentiating (4.3.9) with respect to θ and equating to zero i.e., $\frac{\partial \log L(\theta)}{\partial \theta} = 0$ gives

$$\frac{n}{\theta} - \sum_{i=1}^{n} F_0(x_i) - \frac{ne^{-\theta}}{1-e^{-\theta}} = 0$$

$$\text{or, } \frac{1}{\theta} - \frac{1}{n}\sum_{i=1}^{n} F_0(x_i) - \frac{e^{-\theta}}{1-e^{-\theta}} = 0$$

Therefore, the estimate of θ can be obtained by the following equation

$$\frac{1}{\hat{\theta}} - \frac{1}{n}\sum_{i=1}^{n} F_0(x_i) - \frac{e^{-\hat{\theta}}}{1-e^{-\hat{\theta}}} = 0, \tag{4.3.10}$$

$$\frac{1}{\hat{\theta}} - \frac{e^{-\hat{\theta}}}{1-e^{-\hat{\theta}}} = \frac{1}{n}\sum_{i=1}^{n} F_0(x_i), \tag{4.3.11}$$

Also we can write (4.3.10) as

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} F_0(x_i) + \frac{e^{-\hat{\theta}}}{1-e^{-\hat{\theta}}}}, \tag{4.3.12}$$

Comment: The equation (4.3.10) is an estimating equation of θ and it can not be solved analytically but it may be solved numerically. Therefore, the numerical solution of this equation is the required estimate of θ.

Convergence of θ:

From the law of convergence we can write,

$$\frac{1}{n}\sum_{i=1}^{n} F_0(x_i) \xrightarrow{a.s} E\left\{F_0(x) \middle| x < \infty\right\}$$

$$= \frac{\int_0^{\infty} \theta F_0(t) f_0(t) e^{-\theta F_0(t)} dt}{P(x < \infty)}, \tag{4.3.13}$$

Using (4.3.5) and (4.3.7) in (4.3.13), we obtain the following expression as

$$\frac{1}{n}\sum_{i=1}^{n} F_0(x_i) = \frac{\frac{1-e^{-\theta}}{\theta} - e^{-\theta}}{1-e^{-\theta}}$$

$$= \frac{1}{\theta} - \frac{e^{-\theta}}{1-e^{-\theta}}, \tag{4.3.14}$$

Therefore,

$$\frac{1}{\hat{\theta}} - \frac{e^{-\hat{\theta}}}{1-e^{-\hat{\theta}}} = \frac{1}{n}\sum_{i=1}^{n} F_0(x_i) \xrightarrow{a.s} \frac{1}{\theta} - \frac{e^{-\theta}}{1-e^{-\theta}}$$

And hence,

$$\frac{1}{\hat{\theta}} - \frac{e^{-\hat{\theta}}}{1-e^{-\hat{\theta}}} = \xrightarrow{a.s} \frac{1}{\theta} - \frac{e^{-\theta}}{1-e^{-\theta}}, \tag{4.3.15}$$

Case-b1: $F_0(.), f_0(.)$ and θ **are unknown. Here we observe both cured and non-cured group.**

Suppose that we have the data in the form $(x_i, \in_i)$, $i = 1,2,..,n$, where x_i denotes the survival time for the ith patient, $\in_i$ is the cured indicator with 1 if x_i is not cured and 0 otherwise i.e., $\in_i = 1_{\{X_i < \infty\}}$.

Let $F(x) = \int_0^x \theta f_0(t) e^{-\theta F_0(t)} dt$

$= \int_0^{F_0(x)} \theta e^{-\theta z} dz$ [by putting $F_0(t) = z$]

$= \theta [\frac{e^{-\theta z}}{-\theta}]_0^{F(x)}$

$= 1 - e^{-\theta F_0(x)}$, (4.3.16)

Therefore, the non-parametric likelihood function is given by

$$L_n(\theta, F) = \prod_{i=1}^{n} (\Delta F(x_i))^{\epsilon_i} (e^{-\theta})^{1-\epsilon_i} , \quad (4.3.17)$$

where, $\Delta F(x_i)$ = jump of $F(.)$ at x_i

$= 1 - e^{-\theta F_0(x_i)} - (1 - e^{-\theta F_0(x_{i-})})$

$= e^{-\theta F_0(x_{i-})} - e^{-\theta F_0(x_i)}$

$= e^{-\theta F_0(x_{i-})} (1 - \frac{e^{-\theta F_0(x_i)}}{e^{-\theta F_0(x_{i-})}})$

$= e^{-\theta F_0(x_{i-})} (1 - e^{-\theta [F_0(x_i) - F_0(x_{i-})]})$

$= e^{-\theta (p_1 + p_2 + ... + p_{i-1})} (1 - e^{-\theta p_i})$, (4.3.18)

where, $p_i = F_0(x_i) - F_0(x_{i-})$

Therefore, the above likelihood function (4.3.17) can be written as

$$L_n(\theta, p_1, p_2, ... p_n) = \prod_{i=1}^{n} (e^{-\theta (p_1 + p_2 + ... + p_{i-1})} (1 - e^{-\theta p_i}))^{\epsilon_i} (e^{-\theta})^{1-\epsilon_i} , \quad (4.3.19)$$

We want to maximize (4.3.19) subject to condition

$$p_1 + p_2 + ... p_n = 1, p_i \geq 0 , \quad (4.3.20)$$

The log-likelihood function becomes

$$\log L_n(\theta, p_1, p_2, ..., p_n) = -\theta \sum_{i=1}^{n} \in_i (p_1 + p_2 + ... + p_{i-1}) + \sum_{i=1}^{n} \in_i \log(1 - e^{-\theta p_i}) - \theta \sum_{i=1}^{n} (1 - \in_i)$$

Or,

$$\log L_n(\theta, p_1, p_2, ..., p_n) = -\theta \sum_{i=1}^{n-1} p_i (\sum_{j=i+1}^{n} \in_i) + \sum_{i=1}^{n} \in_i \log(1 - e^{-\theta p_i}) - \theta \sum_{i=1}^{n} (1 - \in_i), \quad (4.3.21)$$

By using Lagrange multiplier method we can maximize (4.3.21). Adding Lagrange multiplier λ, we can write (4.3.21) as follows

$$\log L_n(\theta, p_1, p_2, ..., p_n) = -\theta \sum_{i=1}^{n-1} p_i (\sum_{j=i+1}^{n} \in_j) + \sum_{i=1}^{n} \in_i \log(1 - e^{-\theta p_i}) - \theta \sum_{i=1}^{n} (1 - \in_i) - \lambda (\sum_{i=1}^{n} p_i - 1) \quad (4.3.22)$$

Therefore, the non-parametric maximum likelihood estimators of θ and p_i are obtained by the solution of the following equations

$$\frac{\partial \log L}{\partial \theta} = 0, \frac{\partial \log L}{\partial p_i} = 0, \quad (4.3.24)$$

Now $\frac{\partial \log L}{\partial \theta} = 0$ gives,

$$-\sum_{i=1}^{n-1} p_i (\sum_{j=i+1}^{n} \in_j) + \sum_{i=1}^{n} \in_i \frac{p_i e^{-\theta p_i}}{(1 - e^{-\theta p_i})} - \sum_{i=1}^{n} (1 - \in_i) = 0, \quad (4.3.25)$$

Similarly, $\frac{\partial \log L}{\partial p_i} = 0$ gives, $i = 1,2,...,n-1$

$$-\theta \sum_{j=i+1}^{n} \in_j + \in_i \frac{\theta e^{-\theta p_i}}{(1 - e^{-\theta p_i})} - \lambda = 0, \quad (4.3.26)$$

And $\frac{\partial \log L}{\partial p_n} = 0$ gives,

$$\in_n \frac{\theta e^{-\theta p_n}}{(1 - e^{-\theta p_n})} - \lambda = 0, \quad (4.3.27)$$

Multiplying (4.3.26) by p_i and summing over i from 1 to n, we obtain the following equation

$$-\theta\sum_{i=1}^{n} p_i\left(\sum_{j=i+1}^{n}\in_j\right)+\sum_{i=1}^{n} p_i \in_i \frac{\theta e^{-\theta p_i}}{(1-e^{-\theta p_i})}-\lambda\sum_{i=}^{n} p_i = 0, \qquad (4.3.28)$$

Since, $\sum_{i=1}^{n} p_i = 1$, so the equation (4.1.28) becomes

$$-\theta\sum_{i=1}^{n} p_i \sum_{j=i+1}^{n}\in_j+\sum_{i=1}^{n} p_i \in_i \frac{\theta e^{-\theta p_i}}{(1-e^{-\theta p_i})}-\lambda = 0\ , \qquad (4.3.29)$$

From (4.3.25) and (4.3.29), we obtain

$$\hat{\lambda} = \theta\sum_{i=1}^{n}(1-\in_i), \qquad (4.3.30)$$

Now using the estimate of λ in (4.3.26), we get

$$-\theta\sum_{j=i+1}^{n}\in_j+\in_i \frac{\theta e^{-\theta p_i}}{(1-e^{-\theta p_i})}-\theta\sum_{i=1}^{n}(1-\in_i) = 0, \text{ i =1,2, ...,n},$$

(4.3.31)

Therefore, $\sum_{j=i+1}^{n}\in_j-\in_i \frac{e^{-\theta p_i}}{(1-e^{-\theta p_i})}+\sum_{i=1}^{n}(1-\in_i) = 0\ , \qquad (4.3.32)$

Comment: This equation (4.3.32) can be considered as an estimating equation of p_i which can not be solved analytically but may be solved numerically. So the solution of this equation is our desired estimate of p_i.

Again using (4.3.30) in (4.3.27), we may obtain the estimate of p_n

$$\in_n \frac{\theta e^{-\theta p_n}}{(1-e^{-\theta p_n})}-\theta\sum_{i=1}^{n}(1-\in_i) = 0$$

Comment: The above equation also can not be solved analytically but may be solved numerically.

Finally the estimate of θ may be obtained from the numerical solution of equation (4.3.25).

Case-b2: $F_0(.), f_0(.)$ and θ **are unknown and only non-cured group are observed.** Suppose that we have the data in the form $(x_i, \in_i)$, $i = 1,2,..,n$, where x_i denotes the survival time for the ith patient, $\in_i$ is the cured indicator with 1 if x_i is not cured and 0 otherwise i.e., $\in_i = 1_{\{X_i<\infty\}}$. The non-parametric likelihood function can be written as

$$L_n^{nc}(\theta, p_1, p_2, ..., p_n) = \prod_{i=1}^{n} \frac{e^{-\theta(p_1+p_2+....p_{i-1}}(1-e^{-\theta p_i})}{1-e^{-\theta}}$$

$$= \frac{e^{-\theta\sum_{i=1}^{n}(p_1+p_2+....p_{i-1})} \prod_{i=1}^{n}(1-e^{-\theta p_i})}{(1-e^{-\theta})^n}$$

The log-likelihood function is given by

$$\log L_n^{nc}(\theta, p_1, p_2, ...p_n) = -\theta\sum_{i=1}^{n}(p_1 + p_2 + ...p_{i-1}) + \sum_{i=1}^{n}\log(1-e^{-\theta p_i}) - n\log(1-e^{-\theta})$$

$$= -\theta\sum_{i=1}^{n-1}(n-i)p_i) + \sum_{i=1}^{n}\log(1-e^{-\theta p_i}) - n\log(1-e^{-\theta}), \qquad (4.3.34)$$

We want to maximize (4.3.34) subject to condition $\sum_{i=1}^{n} p_i = 1$,

By using Lagrange multiplier method we can maximize (4.3.34). Adding Lagrange multiplier λ in (4.3.34), we can write

$$\log L_n^{nc}(\theta, p_1, p_2, ...p_n) = -\theta\sum_{i=1}^{n-1}(n-i)p_i) + \sum_{i=1}^{n}\log(1-e^{-\theta p_i}) - n\log(1-e^{-\theta}) - \lambda(\sum_{i=1}^{n} p_i - 1)$$

(4.3.35)

Therefore, the non-parametric maximum likelihood estimators of θ and p_i are obtained by the solution of following equations

$$\frac{\partial \ln L}{\partial \theta} = 0, \frac{\partial \ln L}{\partial p_i} = 0, \tag{4.3.36}$$

Now $\frac{\partial \ln L}{\partial \theta} = 0$ gives

$$-\sum_{i=1}^{n-1}(n-i)p_i + \sum_{i=1}^{n}\frac{p_i e^{-\theta p_i}}{1-e^{-\theta p_i}} - \frac{ne^{-\theta}}{1-e^{-\theta}} = 0, \tag{4.3.37}$$

And $\frac{\partial \ln L}{\partial p_i} = 0$ gives, i =1,2,…,n-1

$$-(n-i)\theta + \frac{\theta e^{-\theta p_i}}{1-e^{-\theta p_i}} - \lambda = 0, \tag{4.3.38}$$

$\frac{\partial \log L}{\partial p_n} = 0$ gives

$$\frac{\theta e^{-\theta p_n}}{1-e^{-\theta p_n}} - \lambda = 0 , \tag{4.3.39}$$

Multiplying (4.3.38) by p_i and summing over i from1 to n we obtain

$$-\theta\sum_{i=1}^{n} p_i(n-i) + \sum_{i=1}^{n} p_i \frac{\theta e^{-\theta p_i}}{1-e^{-\theta p_i}} - \lambda\sum_{i=1}^{n} p_i = 0 , \tag{4.3.40}$$

Since $\sum_{i=1}^{n} p_i = 1$, so the above equation becomes

$$-\theta\sum_{i=1}^{n} p_i(n-i) + \sum_{i=1}^{n} p_i \frac{\theta e^{-\theta p_i}}{1-e^{-\theta p_i}} - \lambda = 0, \tag{4.3.41}$$

Multiplying equation (4.3.37) by θ and subtracting from (4.3.41), we obtain

$$-\lambda + \frac{n\theta e^{-\theta}}{1-e^{-\theta}} = 0$$

Therefore, $\hat{\lambda} = \frac{n\theta e^{-\theta}}{1-e^{-\theta}}$, (4.3.42)

Using the estimate of λ in (4.3.38), we obtain

$$-(n-i)\theta + \frac{\theta e^{-\theta p_i}}{1-e^{-\theta p_i}} - \frac{n\theta e^{-\theta}}{1-e^{-\theta}} = 0$$

or, $$-(n-i) + \frac{e^{-\theta p_i}}{1-e^{-\theta p_i}} - \frac{ne^{-\theta}}{1-e^{-\theta}} = 0$$

or, $$\frac{e^{-\theta p_i}}{1-e^{-\theta p_i}} = (n-i) + \frac{ne^{-\theta}}{1-e^{-\theta}}$$

Thus, $p_i = -\frac{1}{\theta}[\log\{(n-i) + \frac{ne^{-\theta}}{1-e^{-\theta}})\} + \log(1-e^{-\theta p_i})]$, (4.3.43)

Comment: This is an estimating equation of p_i, i =1,2,..,n-1. and it can not be solved analytically but it may be solved numerically and the solution of this equation is the desired estimate of p_i.

Again, using (4.3.42) in (4.3.39) we obtain

$$\frac{\theta e^{-\theta p_n}}{1-e^{-\theta p_n}} - \frac{n\theta e^{-\theta}}{1-e^{-\theta}} = 0$$

$$\frac{e^{-\theta p_n}}{1-e^{-\theta p_n}} = \frac{ne^{-\theta}}{1-e^{-\theta}}$$

$$p_n = -\frac{1}{\theta}\{\log(\frac{ne^{-\theta}}{1-e^{-\theta}}) + \log(1-e^{-\theta p_n})\}, \quad (4.3.44)$$

We observe that the estimate of p_n may be obtained from the numerical solution of the equation (4.3.44)

Finally the estimate of θ can be obtained from the numerical solution of the equation (4.3.37).

In censored data, we consider the following cases:

Case –(a): $F_0(.), f_0(.)$ **are known** and θ **is unknown.** So, this case is fully parametric. We are going to estimate the parameter based on Type 1 censoring and Type 2 censoring.

4.4 Estimation of parameter based on Type 1 censoring for known distribution

A Type 1 censoring is said to apply when each individual has a fixed potential censoring time C_i such that T_i is observed if $T_i \leq C_i$; otherwise we only know that $T_i > C_i$. Type 1 censoring often arises when a study is conducted over a specified time period. Let us consider the following notation for type 1 censoring.

$$t_i = \min(T_i, C_i),\ \delta_i = 1\{T_i \leq C_i\}, \tag{4.4.1}$$

The likelihood function of Type 1 censoring is based on the probability distribution of (t_i, δ_i), i =1,2,..,n. Here t_i and δ_i both are random variables and their joint probability density function is

$$f(t_i)^{\delta_i} p(T_i > C_i)^{1-\delta_i}, \tag{4.4.2}$$

Assuming that $T_1, T_2, ..., T_n$ are independent. We obtain the likelihood function from (4.4.2) as

$$L = \prod_{i=1}^{n} f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}, \tag{4.4.3}$$

Using the cure model mentioned above we can write the likelihood function as follows

$$L(\theta) = \prod_{i=1}^{n} (\theta f_0(t_i) e^{-\theta F_0(t_i)})^{\delta_i} (e^{-\theta F_0(C)})^{1-\delta_i}, \tag{4.4.4}$$

$$= \theta^{\sum_{i=1}^{n}\delta_i} (f_0(t_i))^{\sum_{i=1}^{n}\delta_i} e^{-\theta\sum_{i=1}^{n}\delta_i F_0(t_i)} e^{-\theta\sum_{i=1}^{n} F_0(C)(1-\delta_i)}$$

Therefore the log-likelihood function becomes

$$\log L(\theta) = \sum_{i=1}^{n} \delta_i \log\theta + \sum_{i=1}^{n} \delta_i \log f_0(t_i) - \theta\sum_{i=1}^{n} \delta_i F_0(t_i) - \theta\sum_{i=1}^{n}(1-\delta_i)F_0(C), \qquad (4.4.5)$$

Thus the estimate of θ is obtained by the solution of the equation $\frac{\partial \log L}{\partial\theta} = 0$, which gives

$$\frac{\sum_{i=1}^{n}\delta_i}{\theta} - \sum_{i=1}^{n}\delta_i F_0(t_i) - \sum_{i=1}^{n}(1-\delta_i)F_0(C) = 0$$

$$\hat{\theta} = \frac{\sum_{i=1}^{n}\delta_i}{\sum_{i=1}^{n}\delta_i F_0(t_i) + \sum_{i=1}^{n}(1-\delta_i)F_0(C)}, \qquad (4.4.6)$$

Therefore, $\hat{\theta} = \frac{\frac{1}{n}\sum_{i=1}^{n}\delta_i}{\frac{1}{n}\sum_{i=1}^{n}\delta_i F_0(t_i) + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)F_0(C)}$, (4.4.7)

4.5 Estimation of parameter based on Type 2 censoring for known distribution

In Type 2 censoring, the data consist of r smallest life times $T_{(1)} < T_{(2)} < .... < T_{(r)}$ out of a random sample of n life times $T_1, T_2, ..., T_n$. If T_i, i =1,2,…,n, has a probability density function $f(t)$ and survival function $S(t)$, then the joint probability density function of $T_{(1)} < T_{(2)} < .... < T_{(r)}$ is (*c.f.* David,1981) given by

$$L = \frac{n!}{(n-r)!}\{\prod_{i=1}^{r} f(t_i)\}S(t_{(r)})^{n-r}, \qquad (4.5.1)$$

Using cure rate model in (4.5.1), the likelihood function can be written as

$$L(\theta) = \frac{n!}{(n-r)!}\{\prod_{i=1}^{r} \theta f_0(t_i)e^{-\theta F_0(t_i)}\}\{e^{-\theta F(t_r)}\}^{n-r}$$

The log-likelihood function becomes

$$\log L(\theta) = \log(\frac{n!}{(n-r)!}) + r\log\theta + \sum_{i=1}^{r}\log f_0(t_i) - \theta\sum_{i=1}^{r}F_0(t_i) - (n-r)\theta F(t_r), \qquad (4.5.2)$$

Thus, $\frac{\partial \log L(\theta)}{\partial\theta} = 0$ gives

$$\frac{r}{\theta} - \sum_{i=1}^{r}F_0(t_i) - (n-r)F(t_r) = 0$$

Therefore, the estimate of θ is given by the following expression as

$$\hat{\theta} = \frac{r}{\sum_{i=1}^{r}F_0(t_i) + (n-r)F(t_r)}, \qquad (4.5.3)$$

Case-b: $f_0(.), F_0(.)$ **and** θ **are unknown.** So this case is fully non-parametric. We are going to estimate the parameter based on Type 1 censoring and Type 2 censoring.

4.6 Estimation of parameter under type 1 censoring for unknown distribution

Suppose that we have data in the form (Z_i, δ_i), i =1,2,..,n. Where $Z_i = \min\{X_i, T\}$, δ_i is the censoring indicator and $\delta_i = 1_{\{X_i \le T\}}$, T is fixed censoring time. Therefore the non-parametric likelihood function can be written as

$$L(\theta, p_1, p_2, ..p_n | z_1, z_2, .., z_n)$$

$$= \prod_{i=1}^{n}((e^{-\theta(p_1+p_2+....p_{i-1})})(1-e^{-\theta p_i}))^{\delta_i}\ (e^{-\theta F_0(T)})^{1-\delta_i}, \qquad (4.6.1)$$

$$= \prod_{i=1}^{n}((e^{-\theta(p_1+p_2+....p_{i-1})})(1-e^{-\theta p_i}))^{\delta_i}\ (e^{-\theta(p_1+p_2+...p_k)})^{1-\delta_i}$$

$$= e^{-\theta\sum_{i=1}^{n}\delta_i(p_1+p_2+..p_{i-1})}\prod_{i=1}^{n}(1-e^{-\theta p_i})^{\delta_i}e^{-\theta\sum_{i=1}^{n}(1-\delta_i)(p_1+p_2...+p_k)}, \qquad (4.6.2)$$

Where $p_i = F_0(X_i) - F_0(X_{i-})$ and $k =$ largest i such that $X_i \le T$. Since there are kth observations are uncensored and $(n-k)$th are censored. So the above (4.6.2) non-parametric likelihood function becomes

$$L(\theta, p_1, p_2,.., p_n | z_1, z_2,.., z_n) = e^{-\theta\sum_{i=1}^{k}(p_1+p_2+..p_{i-1})}\prod_{i=1}^{k}(1-e^{-\theta p_i})\ e^{-\theta(n-k)(p_1+p_2+...p_k)}, \qquad (4.6.3)$$

Hence the log-likelihood function becomes

$$\log L = -\theta\sum_{i=1}^{k}(p_1 + p_2 + ...p_{i-1}) + \sum_{i=1}^{k}\log(1-e^{-\theta p_i}) - \theta(n-k)(p_1 + p_2 + ..p_k)$$

$$= -\theta\sum_{i=1}^{k-1}(k-i)p_i + \sum_{i=1}^{k}\log(1-e^{-\theta p_i}) - \theta(n-k)\sum_{i=1}^{k}p_i, \qquad (4.6.4)$$

We want to maximize (4.6.4) subject to condition

$p_1 + p_2 + ...p_k \le 1$

i.e. $p_1 + p_2 + ...p_k + S = 1,$ (4.6.5)

where, S is a non-negative slack variable.

By adding a Lagrange multiplier λ in (4.6.4), we get

$$\log L = -\theta\sum_{i=1}^{k-1}(k-i)p_i + \sum_{i=1}^{k}\log(1-e^{-\theta p_i}) - \theta(n-k)\sum_{i=1}^{k}p_i - \lambda(\sum_{i=1}^{k}p_i + S - 1), \qquad (4.6.6)$$

The non-parametric maximum-likelihood estimator of θ, p_i are obtained by the solution of the following equations

$$\frac{\partial \log L}{\partial \theta} = 0, \frac{\partial \log L}{\partial p_i} = 0, \frac{\partial \log L}{\partial \lambda} = 0, \frac{\partial \log L}{\partial S} = 0, \qquad (4.6.7)$$

Thus, $\frac{\partial \log L}{\partial \theta} = 0$ gives

$$-\sum_{i=1}^{k-1}(k-i)p_i + \sum_{i=1}^{k}\frac{p_i e^{-\theta p_i}}{1-e^{-\theta p_i}} - (n-k)\sum_{i=1}^{k} p_i = 0, \tag{4.6.8}$$

Similarly, $\frac{\partial \log L}{\partial p_i} = 0$, i = 1,2,...k-1 gives

$$-\theta(k-i) + \frac{\theta e^{-\theta p_i}}{1-e^{-\theta p_i}} - (n-k)\theta - \lambda = 0, \tag{4.6.9}$$

$\frac{\partial \log L}{\partial p_k} = 0$ gives

$$\frac{\theta e^{-\theta p_k}}{1-e^{-\theta p_k}} - (n-k)\theta - \lambda = 0 , \tag{4.6.10}$$

$\frac{\partial \log L}{\partial \lambda} = 0$ gives

$$\sum_{i=1}^{k} p_i + S - 1 = 0 , \tag{4.6.11}$$

$\frac{\partial \log L}{\partial S} = 0$ gives

$$\lambda = 0 , \tag{4.6.12}$$

Using (4.6.12) in (4.6.9), we obtain

$$-\theta(k-i) + \frac{\theta e^{-\theta p_i}}{1-e^{-\theta p_i}} - (n-k)\theta = 0$$

$$-(k-i) + \frac{e^{-\theta p_i}}{1-e^{-\theta p_i}} - (n-k) = 0$$

or, $\frac{e^{-\theta p_i}}{1-e^{-\theta p_i}} = (n-i)$

or, $-\theta p_i = \log\{(n-i)(1-e^{-\theta p_i})\}$

Therefore, $p_i = -\frac{1}{\theta}\log\{(n-i)+\log(1-e^{-\theta p_i})\}$,i =1,2,...k-1 , (4.6.13)

Comment: This is an estimating equation of p_i. This equation can not be solved analytically but it may be solved numerically. So the numerical solution of the above equation for known value of θ is the desired estimate of p_i.

Again using (4.6.12) in (4.6.10), we obtain

$$\frac{e^{-\theta p_k}}{1-e^{-\theta p_k}} - (n-k) = 0$$

Thus, $p_k = -\frac{1}{\theta}\log\{(n-k)+\log(1-e^{-\theta p_k})\}$, (4.6.14)

Which is same as equation (4.4.13) when $i = k$

The estimate of S may be obtained from (4.6.11).

Finally the estimate of θ may be obtained from the numerical solution of the following equation

$$-\sum_{i=1}^{k-1}(k-i)p_i + \sum_{i=1}^{k}\frac{p_i e^{-\theta p_i}}{1-e^{-\theta p_i}} - (n-k)\sum_{i=1}^{k}p_i = 0, \quad (4.6.15)$$

4.7 Estimation of parameter based on type 2 censoring for unknown distribution

The non-parametric likelihood function under type 2 censoring [using (4.3.1)] is given by

$$L(\theta, p_1, p_2..p_r) = \frac{n!}{(n-r)!}\{\prod_{i=1}^{r} e^{-\theta(p_1+p_2+.....+p_{i-1})}(1-e^{-\theta p_i})\}\{e^{-\theta F(T_r)}\}^{n-r}, \quad (4.7.1)$$

$$= \frac{n!}{(n-r)!} e^{-\theta\sum_{i=1}^{r}(p_1+p_2+...+p_{i-1})} \prod_{i=1}^{r}(1-e^{-\theta p_i}) e^{-\theta(p_1+p_2+....+p_r)(n-r)}$$

So the log-likelihood function can be written as

$$\log L(\theta, p_1, p_2, ..p_r) = \log(\frac{n!}{(n-r)!}) - \theta\sum_{i=1}^{r}(p_1 + p_2 + ...p_{i-1}) + \sum_{i=1}^{r}\log(1-e^{-\theta p_i}) - \theta(n-r)(p_1 + p_2 + ..p_r)$$

$$= \log(\frac{n!}{(n-r)!}) - \theta\sum_{i=1}^{r-1}(r-i)p_i + \sum_{i=1}^{r}\log(1-e^{-\theta p_i}) - \theta(n-r)(p_1 + p_2 + ... + p_r), \quad (4.7.2)$$

We want to maximize (4.7.2) subject to condition

$p_1 + p_2 + ... + p_r \leq 1$

i.e. $p_1 + p_2 + ... + p_r + S = 1,$ (4.7.3)

where, S is a non-negative slack variable.

Adding Lagrange multiplier λ in (4.7.2), we get the following equation

$$L(\theta, p_1, ..p_r) = \log(\frac{n!}{(n-r)!}) - \theta\sum_{i=1}^{r-1}(r-i)p_i + \sum_{i=1}^{r}\log(1-e^{-\theta p_i}) - \theta(n-r)(\sum_{i=1}^{r}p_i) - \lambda(\sum_{i=1}^{r}p_i + S - 1),$$

(4.7.4)

Therefore, the non-parametric maximum likelihood estimator of θ, p_i are obtained from the solution of the following equations

$$\frac{\partial \log L}{\partial \theta} = 0, \frac{\partial \log L}{\partial p_i} = 0, \frac{\partial \log L}{\partial \lambda} = 0, \frac{\partial \log L}{\partial S} = 0, \quad (4.7.5)$$

Now, $\frac{\partial \log L}{\partial \theta} = 0$ gives

$$-\sum_{i=1}^{r-1}(r-i)p_i + \sum_{i=1}^{r}\frac{p_i e^{-\theta p_i}}{1-e^{-\theta p_i}} - (n-r)(\sum_{i=1}^{r}p_i) = 0, \quad (4.7.6)$$

Similarly, $\frac{\partial \log L}{\partial p_i} = 0$,i = 1,2,.. r-1 gives

$$-(r-i)\theta + \frac{\theta e^{-\theta p_i}}{1-e^{-\theta p_i}} - \theta(n-r) - \lambda = 0, \quad (4.7.7)$$

$\frac{\partial \log L}{\partial p_r} = 0$ gives

$$\frac{\theta e^{-\theta p_r}}{1-e^{-\theta p_r}} - \theta(n-r) - \lambda, \tag{4.7.8}$$

$\frac{\partial \log L}{\partial \lambda} = 0$ gives

$$\sum_{i=1}^{r} p_i + S - 1 = 0, \tag{4.7.9}$$

$\frac{\partial \log L}{\partial S} = 0$ gives

$$\lambda = 0, \tag{4.7.10}$$

Using (4.7.10) in (4.7.7) we obtain

$$-(r-i)\theta + \frac{\theta e^{-\theta p_i}}{1-e^{-\theta p_i}} - \theta(n-r) = 0$$

$$\text{or}, -(r-i) + \frac{e^{-\theta p_i}}{1-e^{-\theta p_i}} - (n-r) = 0$$

$$\text{or}, \frac{e^{-\theta p_i}}{1-e^{-\theta p_i}} = n-i$$

$$\text{or}, -\theta p_i = \log(n-i) + \log(1-e^{-\theta p_i})$$

Therefore, $p_i = -\frac{1}{\theta}\{\log(n-i) + \log(1-e^{-\theta p_i})\}$, i = 1,2,..,r-1 , (4.7.11)

Comment: This is an estimating equation of p_i. This equation can not be solved analytically but it may be solved numerically. So the numerical solution of the above equation for known value of θ is the desired estimate of p_i

Again using (4.7.10) in (4.7.8), we obtain

$$\frac{\theta e^{-\theta p_r}}{1-e^{-\theta p_r}} - \theta(n-r) = 0$$

or, $\frac{e^{-\theta p_r}}{1-e^{-\theta p_r}} = (n-r)$

Therefore, $p_r = -\frac{1}{\theta}\{\log(n-r) + \log(1-e^{-\theta p_i})\}$, (4.7.12)

The estimate of S may be obtained from the equation (4.7.9).

Finally the estimate of θ can be obtain from the numerical solution of the following equation

$$\sum_{i=1}^{r-1}(r-i)p_i - \sum_{i=1}^{r}\frac{p_i e^{-\theta p_i}}{1-e^{-\theta p_i}} + (n-r)(\sum_{i=1}^{r} p_i) = 0 \ , \quad (4.7.13)$$

4.8 Conclusions and further research

In this thesis, at first we have developed cure rate models, after words we have tried to find an estimator to estimate the cure rate by considering uncensored and censored data. In uncensored data, eventually we have considered several cases. When we have assumed $f_0(.)$ and $F_0(.)$ are known and also assumed non-cured and cured group in uncensored data, we have found an analytic solution for the cure rate parameter θ. That is, we have found an estimator of θ which converges to the true value of the parameter. On the other hand, when we assume only non-cured group in uncensored data, we could not find an analytic solution of θ but we have found an estimating equation for θ which might be solved numerically and the numerical solution of the estimating equation would be the estimate of the parameter. Also in that situation we have found that our estimating equation converges to the true equation. Considering both non-cured and cured group, when we assume $f_0(.)$ and $F_0(.)$ are unknown, we found a non-parametric estimating equation of θ. Unfortunately we could not find an explicit solution for θ. But hopefully, this non-parametric estimating equation may be solved numerically by choosing any one

of appropriate method as described in appendix. Also we have found the same result when we consider non-cured group only. That is, in that case we have found a non-parametric estimating equation for θ. Our finding suggests that we have developed the estimator of cure rate in one case and for the remaining cases we have developed the estimating equation for θ in uncensored data. Similarly, we have considered several cases to estimate the parameter of cure rate in censored data. Basically, we have tried to estimate the parameter based on Type 1 and Type 2 censoring. For known $f_0(.)$ and $F_0(.)$ we have found an explicit solution for the parameter of cure rate model based on Type 1 and Type 2 censoring. On the other hand when $f_0(.)$ and $F_0(.)$ are unknown, then we have found non-parametric estimating equations for θ based on Type 1 and Type 2 censoring scheme. So our whole findings suggest that theoretically we have developed several estimating equation of cure rate model.

For further research one could perform the simulation study of the cure rate model based on uncensored and censored data. The fact that one could apply numerical method which has been described in appendix in order to get the estimation or to obtain numerical results of the parameter of cure rate model.

Bibliography:

1. Berkson J., Gage RP.,(1952). Survival curves for cancer patients following treatments. Journal of the American Statistical Association 47; 501-515.
2. Boag JW., (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of the Royal Statistical Society, series B, 11; 15-44.
3. Cantor AB., Shuster JJ.,(1992). Parametric versus non-parametric methods for estimating cure rates based on censored survival data. Statistics in Medicine 11; 931-937.
4. Cantor AB., Shuster J.,(1994). Parametric versus non-parametric methods for estimating cure rates based on censored survival data-reply. Statistics in Medicine 13; 986-986.
5. Casella George and Berger Roger L.,(1990). Statistical Inference. Wadsworth and Brooks, Pacifia Grove, California.
6. Chen, M-H., Ibrahim, J.G., and Sinha, D.,(1999). A new Bayesian model for surviving data with a surviving fraction. Journal of the American Statistical Association 94; 909-919.
7. Cox D.R., (1972). Regression model and life tables. Journal of Statistical Society B 34; 187-220.
8. De Angelis R, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A.,(1999). Mixture models for cancer survival analysis: application to population based data with covariates. Statistics in Medicine 18; 441-454.

9. Edward R. Dougherty (1990). Probability and Statistics for the Engineering, Computing, and Physical sciences. Prentice Hall, Englewood Cliffs, New Jersey, USA.

10. Farewell VT.,(1982). The use of mixture models for the analysis of survival data with long term survivors. Biometrics 38; 1041-1046.

11. Farewell VT., (1986). Mixture model in survival analysis: are they worth the risk? Canadian Journal of Statistics 14; 257-262.

12. Fox A.H. (1963). Fundamental of Numerical analysis. New York, the Ronald Press Company.

13. Garg, M.L., Rao, B.R. and Redmond, C.K.(1970). Maximum likelihood estimation of the parameter of the Gompertz Survival function 19; Journal of the American Statistical Association 19; 152-159.

14. Geiser PW, Chang MN, Rao PV, Shuster JJ, Pullen J.,(1998). Modeling cure rates using the Gompertz model with covariate information. Statistics in Medicine 17; 831-839.

15. Gerald Curtis F. and W.O. Patrick (1989). Applied Numerical Analysis. Reading, Mass: Addition- Wesley.

16. Ghitany ME., Maller RA., and Zhou S.,(1994). Exponential mixture models with long term survivors and covariates. Journal of Multivariate Analysis 49; 218-241.

17. Ghitany ME., Maller R., Zhou S.,(1995). Estimating the proportion of immunes in censored samples- a simulation study. Statistics in Medicine 14; 39-49.

18. Goldman AI.,(1984). Survivorship analysis when cure is a possibility: a Monte-Corlo study. Statistics in Medicine 3; 153-163.

19. Goldman AI.,(1991). The cure model and time confounded risk in the analysis of survival and other timed events. Journal of Clinical Epidemiology 44(12); 1327-1340.

20. Halpern J., Brown BW.,(1987). Cure rate models: power of the log rank and generalized wilcoxon tests. Statistics in Medicine 6; 483-489.

21. Hauck WW, Mckee LT, Turner BJ.,(1997). Two-part survival models applied to administrative data for determining rate and predictors for maternal-child transmission of HIV. Statistics in Medicine 16; 1683-1694.

22. Haybittle JL.,(1965). A two parameters model for the survival curve of treated cancer patients. Journal of the American Statistical Association 60; 16-26.

23. Haybittle JL.,(1959). The estimation of the proportion of patients cured after treatment for cancer of the breast. British Journal of Radiology 32; 725-733.

24. Kalbfleisch J., Prentice R.,(1980). The Statistical Analysis of Failure time Data. Wiley, New York.

25. Kuk AYC., Chen CH.,(1992). A mixture model combining logistic regression with proportional hazard regression. Biometrika 79; 531-541.

26. Kaplan, E.L. and Meier, P.,(1958). Non-parametric estimation from incomplete observations. Journal of the American Statistical Association 53; 457-481.

27. Klein J.P. and Moeschberger, M.L.,(1997). Survival Analysis, Techniques for Censored and Truncated data, Spinger-Verlag, New York.

28. Laska EM., Meisner MJ.,(1992). Non-parametric estimation and testing in a cure model. Biometrics 48; 1223-1234.

29. Lawless J F.(1982). Statistical Models and Methods for Lifetime Data. Wiley; New York.

30. Lawless J F.,(2003). Statistical Models and Methods for Lifetime Data. Wiley; New York.

31. Lee JW., Sather HN.,(1995). Group sequential methods for comparison of cure rates in clinical trials. Biometrics 51; 756-763.

32. Maller RA., Zhou S.,(1995). Testing for the presence of immune or cured individuals in censored survival data. Biometrics 51: 1197-1205.

33. Maller RA., Zhou S.,(1992). Estimating the proportion of immunes in a censored sample. Biometrika 79; 731-739.

34. Mould RF., Boag JW.,(1975). A test of several parametric models for estimating success rate in the treatment of Carcinoma cervix uteri. British Journal of Cancer 32; 529-550.

35. Murtha J. and Willard E.,(1973). Statistics and Calculus, a First course. Prentice-hall, Inc., Englewood Cliffs, New Jersey.

36. O' Neill TJ.,(1979). Distribution free of cure time. Biometrika 66; 184-187.

37. Peng Y.,(2003). Fitting semi-parametric cure models. Computational Statistics and Data Analysis 41; 481-490.

38. Peng Y., and Carriere K.C., (2002). An empirical comparison of parametric and semi parametric cure model. Biometrical Journal 44(8); 1002-1014.

39. Peng Y., Dear KB., and Denham JW.,(1998). A generalized F mixture model for cure rate estimation. Statistics in Medicine 17; 813-830.

40. Sposto R., Sather HN.,(1985). Determining the duration of comparative clinical trials while allowing for cure. Journal of Chronic Diseases 38; 683-690.

41. Sposto R., Sather HN.,(1992). Baker SA. A comparison of tests of the difference in the proportion of patients who are cured. Biometrics 48; 87-99.

42. Sposto R.,(2002). Cure model analysis in cancer: An application to data from the children's cancer group. Statistics in Medicine 21; 293-312.

43. Sy JP, Taylar JM.,(2000). Estimation in a Cox proportional hazard cure model. Biometrics 56; 227-236.

44. Tsodikov A.,(1998)., A proportional hazard model taking account of long-term survivors. Biometrics 54; 1508-1516.

45. Tsodikov A.,(2001). Estimation of Survival based on proportional hazards when cure is a possibility. Mathematical and Computer Modeling 33; 1227-1236.

46. Tsodikov A., Loeffler M., Yakovlev A.,(1998). A cure model with time-changing risk factor: an application to the analysis of secondary Leukemia. A report from the international Database on Hodskins Disease. Statistics in Medicine 17; 27-40

47. Yakovlev Ay.,(1994). Letter to the editor. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. Statistics in Medicine 13; 983-985.

48. Yakovlev A., and Tsodikov A.,(1996). Stochastic Models of tumor Latency and their Biometrical applications. World Scientific, Singapore.

49. Zhou S., Maller RA.,(1995). The likelihood ratio test for the presence of immunes in a censored sample. Statistics in Medicine 27; 181-201.

Appendix

A.1 Introduction to Numerical approach

In many applied problems, the equations involved can not be written in linear form, and the solution of non-linear equations plays a role in applied mathematics of importance equal to that of the solution of system of linear equations. Numerical methods (non linear method) are usually needed if the terms of the equations are in whole or in part transcendental functions.

A basic approach to the solution of numerical (non linear equation) equations involves a set of successive approximation to the precise solution. In some examples several different ways may be considered for setting up an approximation formula. The convergence of the particular scheme chosen must be established, and then the rate of convergence may be speeded up by an appropriate choice of method.

A.2 Iterative method

The Iterative method is more general and used method for any numerical solution. We start with the equation $g(\theta) = 0$ and rearrange it into an equivalent expression of the form $\theta = g(\theta)$ such that $f(\hat{\theta}) = 0$, $\hat{\theta} = g(\hat{\theta})$.

In general we will be concerned with only the real roots of these equations. The approximation scheme is given by

$$\theta_{n+1} = g(\theta_n), \qquad \text{(A.2.1)}$$

where, n= 1, 2, 3, ...

The equation (A.2.1) produces a sequence $\{\theta_n\}$ starting with an initial value θ_1 which may be guess, or the result of another approximation scheme or merely an

arbitrary value such as $\theta_1 = 0$. The convergence of a particular sequence may be tested in terms of the absolute error involved. Aitkinson(1978) shows that each error is about proportional to the previous error to the 1.85th power.

A.3 The Newton-Raphson method

The Newton-Raphson method is one of the most applied method of solving equations. This method is based on a linear approximation of the function. When the derivative of $g(\theta)$ is a simple expression which can be easily found, the real roots of the equation $g(\theta)=0$ can be found rapidly by a process called Newton-Raphson method after its discovers. This method consists in finding an approximate value of the desired root graphically or otherwise and then finding a correction term which must be applied to the approximate value to get the exact value of the root. We illustrate this technique by solving first for single parameter θ. Let us consider the following expression

$$g(\theta)=\frac{\delta \log L}{\delta \theta} \quad ,$$

(A.3.1)

Where, $L = L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$

The problem is to find the estimate of θ say $\hat{\theta}$ such that $g(\theta) = 0$. Therefore, $\hat{\theta}$ is the required maximum likelihood estimator of θ. If there is more than one value $\hat{\theta}$ such that $g(\hat{\theta}) = 0$, the choice of an initial value is very important. In most cases, the initial value obtained is in the neighborhood of the maximum likelihood estimator.

When in doubt, one should consider several different initial values. If $\hat{\theta}$ can not be obtained explicitly from solving equation $g(\hat{\theta}) = 0$, we may attempt a solution by means of the Newton-Raphson procedure. Suppose that $\hat{\theta}_0$ be an initial value of $\hat{\theta}$ and then finding a correction term h so that the equation $g(\hat{\theta}) = 0$ becomes $g(\hat{\theta}_0 + h) = 0$. Now we can expand $g(\hat{\theta}_0 + h)$ by Taylor's theorem as

$$g(\hat{\theta}_0 + h) = g(\hat{\theta}_0) + h\, g'(\hat{\theta}_0) + \frac{h^2}{2!} g''(\hat{\theta}_0) + \ldots\ldots = 0 , \qquad \text{(A.3.2)}$$

Suppose that h is very small and we may neglect the higher power of h. Then (A.3.2) becomes as

$$g(\hat{\theta}_0 + h) = g(\hat{\theta}_0) + h\, g'(\hat{\theta}_0) = 0 , \qquad \text{(A.3.3)}$$

Therefore, $g(\hat{\theta}_0) + h\, g'(\hat{\theta}_0) = 0$, (A.3.4)

which means that

$$h = - \frac{g(\hat{\theta}_o)}{g'(\hat{\theta}_0)} , \qquad \text{(A.3.5)}$$

Then the improved value of the root is given by

$$\theta_0^{(1)} = \hat{\theta}_0 - \frac{g(\hat{\theta}_o)}{g'(\hat{\theta}_0)} , \qquad \text{(A.3.6)}$$

Where, $\theta_0^{(1)}$ is called the first approximation of the desired root. Similarly in the same way, the p-th approximation of $\hat{\theta}$ is given by

$$\theta_0^{(p)} = \hat{\theta}_{p-1} - \frac{g(\hat{\theta}_{p-1})}{g'(\hat{\theta}_{p-1})}, \tag{A.3.7}$$

Where, $g'(\hat{\theta}_{p-1}) = \frac{dg(\theta)}{d\theta} \mid_{\theta=\hat{\theta}_{p-1}}$

Therefore, Newton –Raphson method consists of solving each iteration, the equation is $g(\hat{\theta}_p) + (\theta - \hat{\theta}_p) g'(\hat{\theta}_p) = 0$, (A.3.8)

It is evident from Newton-Raphson formula that the larger derivative $g'(\theta)$, the smaller is the correction term which must be applied to get the correct value of the root. This means that when the graph of $g(\theta)$ is nearly vertical where it crosses the x-axis, the correct value of the root can be found rapidly. But this method should not be used when the graph of $g(\theta)$ is nearly horizontal where it crosses the x-axis. The most important part of this method is to choose initial value because the Newton-Raphson method will converge to value that is not the maximum likelihood function $L(\theta)$. Newton did not publish an extensive discussion of this method but he solved a cubic polynomial in "Principia" [1687]. In 1975, Gross and Clark noted that although this method does not always guarantee a maximum, it is a safeguard in that more than one set of initial values is considered and it means that any particular values in convergence can be uncovered.

A.4 Lagrange Multiplier

In many problem of optimization i.e., of finding a maximum or minimum value of a function, the real world context imposes certain restrictions on the variables of the function. Lagrange has invented a method that can convert such a problem to one which yields to the standard technique of setting the partial derivatives equal to zero. The method is described as follows:

Given a function $f(x,y)$. Find (x_0,y_0) such that (i) $f(x_0,y_0)$ is optimal (either maximum or minimum) subject to the condition (ii) $g(x_0,y_0)=0$. It involves the following steps:

Step 1. Form a new function of three variables

$$F(x,y,\lambda)=f(x,y)-\lambda g(x,y), \tag{A.4.1}$$

Step 2. Take the partial derivatives of F with respect to its variables and set them equal to zero:

$$F'_x(x,y,\lambda)=\frac{d}{dx}F(x,y,\lambda)=0, \tag{A.4.2}$$

$$F'_y(x,y,\lambda)=\frac{d}{dy}F(x,y,\lambda)=0, \tag{A.4.3}$$

$$F'_\lambda(x,y,\lambda)=\frac{d}{d\lambda}F(x,y,\lambda)=0, \tag{A.4.4}$$

Step 3. Find the simultaneous solutions x_0, y_0, λ_0 of the equations (A.4.2), (A.4.3) and (A.4.4) and check the value $f(x_0, y_0)$ to determine whether it is the desired optimal solution.

Note that the value λ is called a Lagrange Multiplier. The auxiliary function $F(x, y, \lambda)$ is called the Lagrangean. The function $f(x, y)$ is called the objective function. The method can be more tersely described as introducing a Lagrangean and optimizing it instead of the original function $f(x, y, z)$. As with any general method, the practitioner or the researcher must be wary of using Lagrange Multiplier when a suitable direct approach is available. It should be pointed out that we offer no justification for the Lagrange Multiplier technique because doing so would require a fairly sophisticated excursion in to the properties of functions of several variables.