ANALYSIS OF PARTIALLY COMPLETE TIME AND TYPE OF FAILURE DATA

DEBASHIS KUNDU
DEPARTMENT OF MATHEMATICS AND STATISTICS
INDIAN INSTITUTE OF TECHNOLOGY, KANPUR
INDIA
ORGANIZATIONS

• DATA DESCRIPTION

• ASSUMPTIONS

• NON-PARAMETRIC METHODS

• PARAMETRIC METHODS

• BAYESIAN TECHNIQUES

• CONCLUSIONS
WE HAVE THE FOLLOWING OBSERVATIONS

\[ T = \text{Failure Time} \in (0, \infty) \]

\[ \delta = \text{Cause of Failure} \in \{1, \ldots, K\}. \]

- **Type: 1** \( \{T = t, \delta = j\} \)
- **Type: 2** \( \{T > t\} \)
- **Type: 3** \( \{T > t, \delta = j\} \)
- **Type: 4** \( \{T = t\} \).

If the data consist of these types of observations, how to analyze the data?

It involves by specifying a distribution of \((T, \delta)\).
Different Data Sets

Data Set 1: Brain Cancer Data

172 patients were given the radiation therapy having a particular type of brain cancer. Data were recorded from the day of the therapy till death/withdrawal. Almost 50% have a censored value of $T$. Over 65% have an unknown value of $\delta$. No Type-4 data.

Data Set 1: Prostate Cancer Data

79 male patients were given the therapy having the stage 4 prostate cancer. Data were recorded from the day of the therapy till death/withdrawal. Almost 35% have a censored value of $T$. Over 50% have an unknown value of $\delta$. 16% belong to Type-4.
NON-PARAMETRIC APPROACH

Let us define:

\[ S(t) = P(T > t) \quad \delta_j(t) = P(\delta = j | T = t). \]

Likelihood Contribution:

- **Type 1:** \( \delta_j(t)dS(t) \)
- **Type 2:** \( S(t) \)
- **Type 3:** \( \int_t^\infty \delta_j(u)dS(u) \)
- **Type 4:** \( dS(t) \)

\[
dS(t) = \begin{cases} 
  -S'(t)dt & \text{if } T \text{ is absolutely cont.} \\
  S(t-) - S(t) & \text{if } T \text{ is discrete}
\end{cases}
\]

Assumption: Missing \( \delta \) happens at random
Suppose $t_1, \ldots, t_K$ are the observed time points where some events have taken place. At each time point $t_k$ divide the number of observations into four groups.

$$S(t_m) = \prod_{l=1}^{m} p_l, \quad dS(t_m) = (1 - p_m) \prod_{l=1}^{m-1} p_l$$

$$p_l = \frac{S(t_l)}{S(t_{l-1})}, \quad \delta_{jm} = \delta_j(t_m)$$

$$\delta_{j,K+1} = P(\delta = j|T > t_K)$$

It is possible to obtain the MLEs using the EM algorithm.
PARAMETRIC METHOD
(*T is assumed to be continuous*)

LATENT FAILURE TYPE MODEL (Cox)

$T_j =$ The random variable is the time of failure if the individual fails due to cause $j$.

We do not observe $T_1, \ldots, T_K$.

We observe:
$T = \min\{T_1, \ldots, T_K\}$
$\delta = \{j; T_j < T_l, l = 1, \ldots, K, l \neq j\}$

It is assumed that $T_1, \ldots, T_K$ are independent. It is an assumption and it cannot be verified from the observed data.

For parametric studies, $T_i$’s are assumed to have different parametric forms.
CAUSE SPECIFIC HAZARD FUNCTION APPROACH

OVERALL HAZARD FUNCTION

\[ \lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t \mid T \geq t)}{\Delta t} \]

CAUSE SPECIFIC HAZARD FUNCTION

\[ \lambda_j(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t, \delta = j \mid T \geq t)}{\Delta t} \]

\[ \lambda(t) = \sum_{j=1}^{K} \lambda_j(t) \quad S(t) = e^{-\int_0^t \lambda(u)\,du} \]

THE PROBABILITY DENSITY FUNCTION DUE TO CAUSE \( j \) CAN BE WRITTEN AS

\[ f_j(t) = \lambda_j(t)S(t) \]

DIFFERENT PARAMETRIC FORMS OF \( \lambda_j \)'S ARE ASSUMED. IT IS OBSERVED THAT IN CASE OF EXPONENTIAL AND WEIBULL MODELS THE TWO FORMULATIONS ARE EQUIVALENT.
Latent Failure Time Model

Exponential Model:

Suppose $T_j$’s are independent exponential random variables with mean $\frac{1}{\lambda_j} = \theta_j$.

It is possible to obtain the MLEs and UMVUEs of $\lambda_j$’s.

The MLE of $\theta_j$ does not exist. We propose to use the conditional MLEs. The conditional probability density functions of $\hat{\theta}_j$ is a mixture of gamma density functions. It can be used to construct the confidence interval also.
**WEIBULL FAILURE DISTRIBUTION**

The distribution function of $T_j$’s have the following form:

$$F_j(t) = 1 - e^{-\lambda_j t^\alpha}$$

The MLEs of $\alpha$ and $\lambda_1, \ldots, \lambda_K$ do not have explicit forms. EM algorithm can be used to compute the MLEs but it takes a long time to converge. A fixed point type algorithm has been proposed and the convergence is quite fast.
PARAMETRIC AND NON-PARAMETRIC ESTIMATIONS

DATA SET 1

Figure 1: Empirical survival function and the estimated survival functions using exponential and Weibull models.
DATA SET 2

Figure 2: Empirical survival function and estimated survival functions using MLE’s and UMVUE’s. Here lifetime distributions are exponential.
DATA SET 2

Figure 3: Empirical survival function and the estimated survival functions using MLE’s and modified MLE’s. Here the lifetime distributions are Weibull.
Bayesian Analysis

Exponential Model

$\lambda_j$s have gamma independent priors

Two Cases depending on ($\text{# Type 4} - \text{# Type 3}$) observations

- If ($\text{# Type 4} - \text{# Type 3}$) $\geq 0$ then explicit solutions
- If ($\text{# Type 4} - \text{# Type 3}$) $< 0$ then no explicit solution
- If ($\text{# Type 4} - \text{# Type 3}$) $= 0$ then very convenient form and explicit credible intervals
- If ($\text{# Type 4} - \text{# Type 3}$) $> 0$ then it can be obtained as mixtures and conservative credible intervals can be constructed
The posterior density functions of $\lambda_j$ given $\alpha$ can be obtained explicitly.

The posterior density function of $\alpha$ is log-concave.