Analysis of Partially Complete Time and Type of Failure Data

Debasis 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 = Failure Time $\in (0, \infty)$

δ = Cause of Failure $\in \{1, \ldots, \mathrm{K}\}$.

- Type: 1 $\{T = t, \delta = \mathrm{j}\}$
- Type: 2 $\{T > t\}$
- Type: 3 $\{T > t, \delta = \mathrm{j}\}$
- 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, δ).

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 δ. 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 δ. 16% belong to Type-4.

Non-Parametric Approach

Let us define:

$S(t) = P(T > t) \qquad \delta_j(t) = P(\delta = j | T = t).$

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: $dS(t)$

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

Assumption: Missing δ happens at random

Non-Parametric MLEs

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.

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

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

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

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 can not 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

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t}$$

Cause Specific Hazard Function

$$\lambda_j(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t, \delta = j | T \ge t)}{\Delta t}$$

$$\lambda(t) = \sum_{j=1}^{K} \lambda_j(t) \qquad S(t) = e^{-\int_0^t \lambda(u) du}$$

The probability density function due to cause j can be written as

$$f_j(t) = \lambda_j(t) S(t)$$

Different parametric forms of λ_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 ransom variables with mean $\frac{1}{\lambda_j} = \theta_j$

It is possible to obtain the MLEs and UMVUEs of λ_j's

The MLE of θ_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 α 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

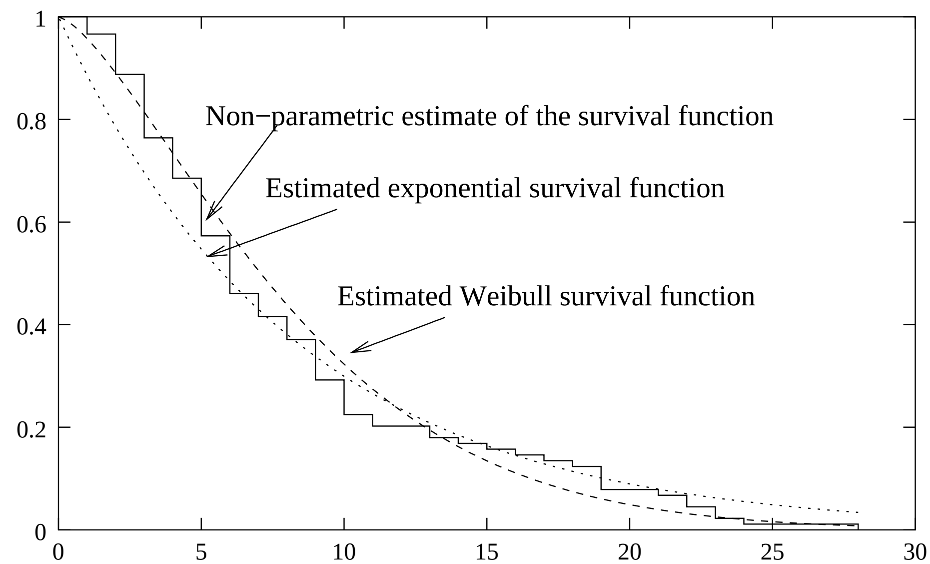

Figure 1: Empirical survival function and the estimated survival functions using exponential and Weibull models.

Data Set 2

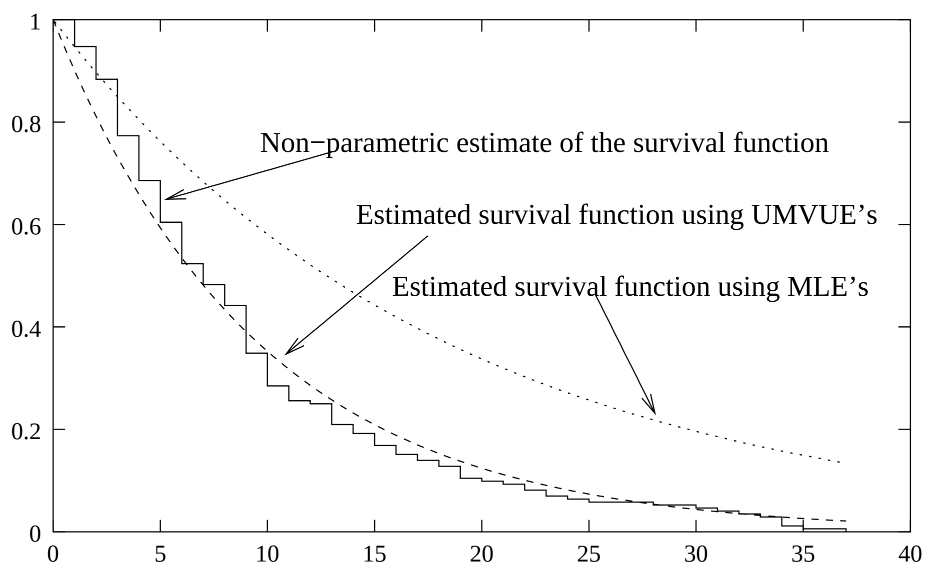

Figure 2: Empirical survival function and estimated survival functions using MLE's and UMVUE's. Here lifetime distributions are exponential.

Data Set 2

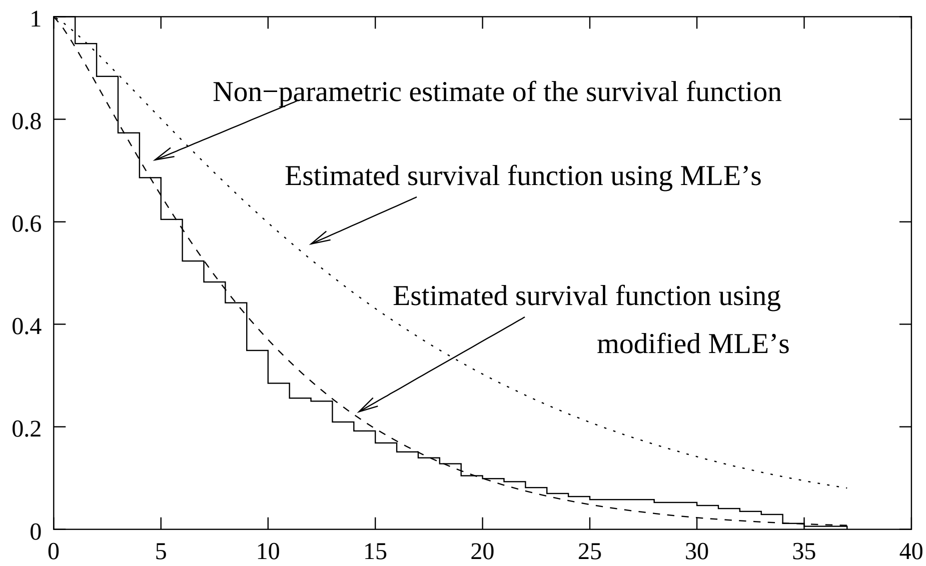

Figure 3: Empirical survival function and the estimated survival functions using MLE's and modified MLE's. Here the lifetime distributions are Weibull.

Bayesian Analysis

Exponential Model

λ_js have gamma independent priors

Two Cases depending on (# Type 4 - # Type 3) observations

- If (# Type 4 - # Type 3) ≥ 0 then explicit solutions
- If (# Type 4 - # Type 3) < 0 then no explicit solution
- If (# Type 4 - # Type 3) $= 0$ then very convenient form and explicit credible intervals
- If (# Type 4 - # Type 3) > 0 then it can be obtained as mixtures and conservative credible intervals can be constructed

Weibull Model

λ_js have independent gamma priors and we do not need to assume any specific form of the prior on α. It is assumed that it has the support on $(0, \infty)$ and it has a log-concave density function.

The posterior density functions of λ_j given α can be obtained explicitly.

The posterior density function of α is log-concave