Journal of Data Analysis & Operations Research
ISSN 2384-4671
J. Data Ana. Oper. Res. 1, No. 1 (Jun. 2014)

Parametric Estimation for Partly Interval-Censored Failure Time Data Under Weibull Distribution

Azzah Mohammad Alharpy[1,2], Noor Akma Ibrahim[3]

[1]*Department of Mathematics, Faculty of Science, Taibah University, Almadinah Almunawwarah, Saudi Arabi*
[2]*Laboratory of Computational Statistics and Operations Research, Institute for Mathematical Research*.
[3]*Department of Mathematics, Faculty of Science, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

Received: 1 Jul. 2013, Revised: 27 Aug. 2013; Accepted: 7 Mar. 2014, Published: Jun. 2014

Abstract: This article discusses the parametric estimation in the existence of partly interval-censored failure time data. Partly interval-censored failure time data is composed of exact observations and interval-censored observations. This phenomenon often happens in clinical trials and health studies that are followed by periodic follow-ups. We have constructed a parametric estimation for this kind of failure time data under Weibull distributions by using direct approach (without imputation) and indirect approach (with multiple imputation technique). Simulation studies are established to assess the proposed estimate for survival function and to illustrate the difference between the two approaches. The simulation results indicate that the presented procedure works well for both approaches but the estimate under the indirect approach is better than the estimate under direct approach. A modified secondary data set from breast cancer study is used to illustrate the proposed estimation.

Keywords: Multiple imputation, Survival function, Partly interval-censored data.

1. Introduction

The Partly interval-censored (PIC) data often occurs in medical and health studies that are followed by periodic follow-ups. With PIC data, the failure times are exactly observed for some subjects but for the remaining subjects, the failure times are observed only to lie in an interval. An example of this kind of data is provided by the Framingham Heart Disease study. In this study, the time of the first emergence of angina pectoris in coronary heart disease patients was the event of interest. For some patients, the event times are recorded exactly, but for the remaining patients, times are recorded only to fall within the intervals between two clinical examinations, see Odell et al. (1992). Researchers that have addressed the PIC include Peto and Peto (1972) who handling PIC data, considering the exact data as an interval-censored with a very short interval. Huang (1999) derived the asymptotic properties of the nonparametric estimation for the distribution function with PIC data. Kim (2003) studied the maximum likelihood estimation in the presence of PIC data under the proportional hazards model. Zhao et al. (2008) developed a nonparametric test approach in the existence of the PIC data, which is based on the same idea of the generalized log-rank test for interval-censored data as in Sun et al. (2005). Alharpy and Ibrahim (2013) discussed the parametric test in the existence of the PIC data under Weibull distribution via multiple imputation. In this article we aspire to present parametric estimate for the survival functions in the presence of the PIC failure time data under Weibull distribution by using the direct and indirect approaches.

2. Parametric Estimation for Partly Interval-Censored Data

In this section we display parametric maximum likelihood estimate (PMLE) for PIC data under Wiebull distribution by using two methods. The first method is generalized parametric estimation without multiple imputation technique (Direct Approach) and the generalized parametric estimation via multiple imputation technique (Indirect Approach) is the second method. Let $T_i > 0$ be a random variable to signify the failure time of interest for i^{th} subject and n the

E-mail: azzahmath@hotmailcom, nakma@putra.upm.edu.my

number of subjects from sample with Wiebull survival function $S(t)=\exp[-(\frac{t}{b})^a]$ where a and b are the shape and scale parameters respectively. Assume that for this treatment the PIC data are available. This means that, we observe the exact failure time given by $O_1=\{\{t_i\}_{i=1}^{n_1};\}$ for n_1 subjects, and we observe interval-censored failure time for the rest of the subjects n_2 given by $O_2=\{(L_i,R_i],i=n_1+1,\cdots,n;\ n_2=n-n_1\}$ where (L_i, R_i] denote the interval in which T_i is observed, and L_i ,R_i are positive random variables independent of T_i , such that $L_i < R_i$ with probability one. When $L_i < R_i = \infty$, the failure time T_i is subject to right-censored observation. Also, when $0=L_i < R_i$, the failure time T_i is regarded as left-censored observation. Our aim is to estimate the survival function $S(t)$. As Alharpy and Ibrahim (2013, pp.622), pointed out the likelihood function for θ with PIC data can be shown as

$$L(\theta)=\prod_{i=1}^{n_1} f(t;\theta)\prod_{i=n_1+1}^{n}[S(L_i;\theta)-S(R_i;\theta)]$$

$$=\prod_{i=1}^{n_1} f(t;\theta)\prod_{i=n_1+1}^{n}[1-S(L_i;\theta)]^{\delta_i}\left[\frac{S(L_i;\theta)}{S(R_i;\theta)}-1\right]^{\gamma_i}\left[S(R_i;\theta)\right]^{1-\delta_i} \quad (1)$$

where $\delta_i=I(t_i\in(0,L_i])$ and $\gamma_i=I(t_i\in(L_i,R_i])$.

and under Weibull distribution the likelihood function as

$$L(a,b)=\prod_{i=1}^{n_1}(\frac{a}{b})(\frac{t_i}{b})^{a-1}\exp[-(\frac{t_i}{b})^a]$$
$$\prod_{i=n_1+1}^{n}[1-\exp[-(\frac{t_i}{b})^a]]^{\delta_i}$$
$$[\exp[(\frac{R_i}{b})^a-(\frac{L_i}{b})^a]-1]^{\gamma_i}$$
$$[\exp[-(\frac{R_i}{b})^a]]^{1-\delta_i} \quad (2)$$

In order to estimate the survival function we need to estimate the parameters. Under Weibull model the maximum likelihood estimates of a and b are the solutions of

$$U_a(a,b)=\frac{\partial\log L(a,b)}{\partial a}=0$$
and (3)
$$U_b(a,b)=\frac{\partial\log L(a,b)}{\partial b}=0$$

that can be solved by using Newton-Raphson method. This method is called, the direct approach and $\hat{S}(t)$ indicates the PMLE using direct approach.

3. Parametric Estimation with Indirect Approach

The imputation technique employed is to impute exact failure times for the interval-censored observations and then applying the parametric estimation. The imputation is for all interval-censored including finite and infinite. The imputation scheme that we use is Rubin's multiple imputation (Rubin, 1987) and the procedure of multiple imputation is as follows. Let Y be a pre-determined positive integer, and y be an integer satisfying $1< y < Y$. For $y=1,\cdots,Y$ do

Step 1. Let T_i^y be a random sample taken from the conditional probability function,

$$P_r(t_i^y=t_k\mid t_i^y\in(L_i,R_i])=\frac{\hat{S}(t_{k-1})-\hat{S}(t_k)}{\hat{S}(L_i)-\hat{S}(R_i)}$$

where

$t_k\in(L_i,R_i]$ for $i=1,\cdots,n_2;k=1,\cdots,M$

The $\{t_k\}_{k=1}^{M}$ denote the unique arrangement element of $\{0,t_j,L_i,R_i;\ j=1,\cdots,n_1,\ i=n_1+1,\cdots,n\}$

and $\hat{S}(t)$ is the estimate of survival function by using direct method. Then we have a set of exact data $\{t_i^y,i=1,\cdots,n_2\}$.

Step 2. Mix the exact data that we impute from the conditional probability function in step 1

with the exact data that we have in the original data. Then we have an exact data

$$\{t_i,i=1,\cdots n_1;t_i^y,i=n_1+1,\cdots,n\}$$

Step 3. Apply the parametric estimate for the exact data, meaning under Weibull distribution the likelihood function is

$$L(a,b)=\prod_{i=1}^{n}(\frac{a}{b})\,(\frac{t_i}{b})^{a-1}\exp\,[-(\frac{t_i}{b})^{a}]$$

and the maximum likelihood estimates of a and b are the solution of Eq(3) in which can be solved by using Newton-Raphson method see (Wong, R. K.W; 1977).

Step 4**.** Repeat step 1 to step 3 for each $y=1,\cdots,Y$.

Step 5. The estimate survival function given by $\tilde{S}(t)=\sum_{y=1}^{Y}\frac{\hat{S}(t)^{y}}{Y}$.

4. Simulation Study

To examine the accuracy of the parametric estimates which were proposed in previous section, we conducted a simulation study to calculate bias, variance (Var.) and the mean square error (MSE). To calculate the bias the following formula is considered:

$$bias(\hat{S}(t))=\frac{1}{N}\sum_{i=1}^{N}(\hat{S}(t)-S(t))$$
$$=E(\hat{S}(t))-S(t)\,;$$

where;

- N is the number of simulation.
- $\hat{S}(t)$ is the estimated value of survival function at time t.
- $S(t)$ is the true value of survival function at time t.

and the MSE for survival function estimate is calculated by :

$$MSE(\hat{S}(t))=Var\;(\hat{S}(t))+\left\{E(\hat{S}(t))-S(t)\right\}^{2}\;;$$

where

$$Var(\hat{S}(t))=\frac{1}{N}\sum_{i=1}^{N}\left[\;\hat{S}(t)-E(\hat{S}(t))\;\right]^{2}.$$

In order to compare between two approaches (direct and indirect) of the PMLE for the PIC data, the relative efficiency (RE) is considered as follow:

$$RE=\frac{MSE(\hat{S}_{W.O}(t))}{MSE(\hat{S}_{W.I}(t))}$$

where

- $\hat{S}_{W.O}(t)$ is the estimated value of survival function without imputation.
- $\hat{S}_{W.I}(t)$ is the estimated value of survival function with imputation.

To generate the data, it is assumed, the failure times from two Weibull distributions with shape parameter a and scale parameter b, are generated. The first Weibull distribution with *a=1.5* and *b=6* in which has increasing hazard function. The second Weibull distribution with *a=0.6* and *b=3* in which has decreasing hazard function. Also, it is supposed, there is only one sample containing $N=(n_1+n_2)$ observations, where n_1 represents the number of the exact observations $\{t_i, i=1,\cdots,n_1\}$ and n_2 represents the number of the interval-censored observations $\{(L_i,R_i], i=1,\cdots,n_2\}$.

To generate the interval-censored data, the set of follow-up studies that pre-specify examination times for studying subjects are constructed. Furthermore, each subject is assumed to be observed at 11 follow-up times: $\tau_r=1+1.5(r-1),\ r=1,\cdots,11$, but a subject may be absent from the follow-ups with probability q', where $0\le q'\le 1$ except at the first follow-up. Then, it is defined that the interval-censored failure time for the subject be the shortest distance between two success examination times which covers the generated true failure time. The left side of the interval for a subject will be defined as zero when the true failure time is less than the first success examination time, and the right side of the interval for a subject will be defined as infinity when the true failure time is greater than the terminal success examination time. A larger absence probability q' will yield the interval-censored with longer interval. The result recorded is based on 500 replications and the number of imputation is 10. Also, there are seven failure times of interest which were chosen from data, randomly.

Table 1. The Bias, Var. and MSE of the parametric estimate of the survival function for Weibull distribution with parameters (1.5,6) under direct approach with 50% exact observations and different sample size.

			N=50 $n_1=n_2=25$			N=100 $n_1=n_2=50$		
q'	*t*	S(t)	**Bias**	**Var.**	**MSE**	**Bias**	**Var.**	**MSE**
0.1	1	0.9342	0.0322	0.0002	0.0012	0.0315	0.0001	0.0011
	2.5	0.7642	0.0701	0.0015	0.0064	0.0677	0.0007	0.0053
	4	0.5802	0.0705	0.0031	0.0081	0.0678	0.0015	0.0061
	5.5	0.4158	0.0456	0.0039	0.0060	0.0442	0.0018	0.0038
	7	0.2836	0.0154	0.0035	0.0037	0.0155	0.0015	0.0017
	8.5	0.1852	-0.0068	0.0025	0.0025	-0.0062	0.0011	0.0011
	10	0.1163	-0.0175	0.0014	0.0017	-0.0171	0.0006	0.0009
0.3	1	0.9342	0.0328	0.0002	0.0013	0.0321	0.0001	0.0011
	2.5	0.7642	0.0721	0.0015	0.0067	0.0693	0.0007	0.0055
	4	0.5802	0.0728	0.0033	0.0086	0.0697	0.0014	0.0063
	5.5	0.4158	0.0473	0.0040	0.0062	0.0458	0.0017	0.0038
	7	0.2836	0.0163	0.0036	0.0039	0.0164	0.0015	0.0018
	8.5	0.1852	-0.0066	0.0026	0.0026	-0.0058	0.0011	0.0011
	10	0.1163	-0.0177	0.0014	0.0017	-0.0171	0.0006	0.0009
	1	0.9342	0.0337	0.0002	0.0013	0.0330	0.0001	0.0012
	2.5	0.7642	0.0745	0.0016	0.0072	0.0717	0.0007	0.0058
	4	0.5802	0.0755	0.0034	0.0091	0.728	0.0015	0.0068
0.5	5.5	0.4158	0.0492	0.0043	0.0067	0.0483	0.0018	0.0041
	7	0.2836	0.0170	0.0039	0.0042	0.0179	0.0016	0.0019
	8.5	0.1852	-0.0068	0.0027	0.0027	-0.0053	0.0011	0.0011
	10	0.1163	-0.0182	0.0015	0.0018	-0.0171	0.0006	0.0009

Table 2. The Bias, Var. and MSE of the parametric estimate of the survival function for Weibull distribution with parameters (0.6,3) under direct approach with 50% exact observations and different sample size.

		N=50 $n_1=n_2=25$			N=100 $n_1=n_2=50$			
q'	*t*	S(t)	**Bias**	**Var.**	**MSE**	**Bias**	**Var.**	**MSE**
0.1	1	0.5961	0.3103	0.0008	0.0971	0.3135	0.0002	0.0985
	2.5	0.4080	0.2974	0.0025	0.0909	0.3009	0.0009	0.0914
	4	0.3047	0.2059	0.0033	0.0457	0.2083	0.0014	0.0448
	5.5	0.2373	0.1132	0.0031	0.0159	0.1142	0.0015	0.0145
	7	0.1896	0.0410	0.0024	0.0041	0.0408	0.0012	0.0029
	8.5	0.1544	-0.0078	0.0017	0.0018	-0.0088	0.0009	0.0010
	10	0.1275	-0.0368	0.0010	0.0024	-0.0384	0.0005	0.0020
0.3	1	0.5961	0.3125	0.0008	0.0985	0.3157	0.0002	0.0999
	2.5	0.4080	0.3012	0.0024	0.0931	0.3048	0.0010	0.0939
	4	0.3047	0.0728	0.0033	0.0086	0.0697	0.0014	0.0465
	5.5	0.2373	0.0473	0.0040	0.0062	0.0458	0.0017	0.0152
	7	0.1896	0.0163	0.0036	0.0039	0.0164	0.0015	0.0031
	8.5	0.1544	-0.0066	0.0026	0.0026	-0.0058	0.0011	0.0010
	10	0.1275	-0.0177	0.0014	0.0017	-0.0171	0.0006	0.0019
0.5	1	0.5961	0.3156	0.0008	0.1004	0.3191	0.0002	0.1020
	2.5	0.4080	0.3065	0.0025	0.0964	0.3110	0.0010	0.0977
	4	0.3047	0.2145	0.0033	0.0493	0.2183	0.0015	0.0492
	5.5	0.2373	0.1192	0.0033	0.0175	0.1217	0.0016	0.0164
	7	0.1896	0.0444	0.0026	0.0046	0.0455	0.0013	0.0034
	8.5	0.1544	-0.0063	0.0018	0.0018	-0.0064	0.0009	0.0009
	10	0.1275	-0.0365	0.0011	0.0024	-0.0375	0.0006	0.0020

Table 3. The Bias, Var. and MSE of the parametric estimate of the survival function for Weibull distribution with parameters (1.5,6) under indirect approach with 50% exact observations and different sample size.

		N=50 $n_1=n_2=25$			N=100 $n_1=n_2=50$			
q'	t	S(t)	**Bias**	**Var.**	**MSE**	**Bias**	**Var.**	**MSE**
	1	0.9342	0.0246	0.0002	0.0008	0.0263	0.0001	0.0008
	2.5	0.7642	0.0496	0.0016	0.0041	0.0531	0.0008	0.0036
	4	0.5802	0.0451	0.0032	0.0052	0.0493	0.0014	0.0038
	5.5	0.4158	0.0237	0.0039	0.0045	0.0277	0.0016	0.0024
	7	0.2836	0.0016	0.0033	0.0033	0.0042	0.0014	0.0014
	8.5	0.1852	0.0131	0.0022	0.0024	0.0123	0.0010	0.0012
	10	0.1163	0.0190	0.0012	0.0016	0.0196	0.0006	0.0010
	1	0.9342	0.0220	0.0009	0.0014	0.0249	0.0001	0.0007
	2.5	0.7642	0.0446	0.0021	0.0041	0.0493	0.0007	0.0031
	4	0.5802	0.0394	0.0033	0.0049	0.0448	0.0014	0.0034
0.3	5.5	0.4158	0.0190	0.0037	0.0041	0.0239	0.0016	0.0022
	7	0.2836	0.0018	0.0032	0.0032	0.0016	0.0014	0.0014
	8.5	0.1852	0.0152	0.0022	0.0024	0.0135	0.0009	0.0011
	10	0.1163	0.0201	0.0013	0.0017	0.0199	0.0005	0.0009
	1	0.9342	0.0207	0.0002	0.0006	0.0221	0.0001	0.0006
	2.5	0.7642	0.0385	0.0016	0.0031	0.0423	0.0007	0.0025
	4	0.5802	0.0312	0.0031	0.0041	0.0368	0.0014	0.0028
0.5	5.5	0.4158	0.0113	0.0036	0.0037	0.0177	0.0015	0.0018
	7	0.2836	0.0076	0.0031	0.0032	0.0018-	0.0013	0.0013
	8.5	0.1852	0.0188	0.0021	0.0025	0.0146	0.0009	0.0011
	10	0.1163	0.0219	0.0012	0.0017	0.0197	0.0005	0.0009

Table 4. The Bias, Var. and MSE of the parametric estimate of the survival function for Weibull distribution with parameters (0.6,3) under indirect approach with 50% exact observations and different sample size.

		N=50 $n_1=n_2=25$			N=100 $n_1=n_2=50$			
q'	t	S(t)	**Bias**	**Var.**	**MSE**	**Bias**	**Var.**	**MSE**
	1	0.5961	0.3032	0.0005	0.0924	0.3070	0.0002	0.0944
	2.5	0.4080	0.2832	0.0019	0.0821	0.2882	0.0009	0.0840
	4	0.3047	0.1912	0.0028	0.0394	0.1946	0.0014	0.0393
	5.5	0.2373	0.1014	0.0027	0.0130	0.1025	0.0014	0.0119
	7	0.1896	0.0330	0.0022	0.0033	0.0322	0.0011	0.0021
	8.5	0.1544	-0.0127	0.0015	0.0017	-0.0146	0.0008	0.0010
	10	0.1275	-0.0395	0.0009	0.0025	-0.0420	0.0005	0.0023
	1	0.5961	0.3035	0.0005	0.0926	0.3075	0.0002	0.0948
	2.5	0.4080	0.2825	0.0019	0.0817	0.2886	0.0009	0.0842
	4	0.3047	0.1892	0.0027	0.0385	0.1946	0.0014	0.0393
0.3	5.5	0.2373	0.0986	0.0027	0.0124	0.1021	0.0014	0.0118
	7	0.1896	0.0299	0.0021	0.0030	0.0314	0.0012	0.0022
	8.5	0.1544	-0.0155	0.0014	0.0016	-0.0155	0.0008	0.0010
	10	0.1275	-0.0419	0.0009	0.0027	-0.0428	0.0005	0.0023
	1	0.5961	0.3038	0.0005	0.0928	0.3081	0.0002	0.0951
	2.5	0.4080	0.2817	0.0019	0.0813	0.2892	0.0009	0.0845
	4	0.3047	0.1872	0.0027	0.0377	0.1947	0.0014	0.0393
0.5	5.5	0.2373	0.0957	0.0027	0.0119	0.1018	0.0014	0.0118
	7	0.1896	0.0269	0.0022	0.0029	0.0308	0.0012	0.0021
	8.5	0.1544	-0.0182	0.0015	0.0018	-0.0161	0.0008	0.0011
	10	0.1275	-0.0441	0.0009	0.0028	-0.0434	0.0005	0.0024

Table 5. Relative efficiency to compare between two estimates (parametric with direct and parametric with indirect) of survival function under Weibull distribution with parameters (1.5,6) with different percent of exact observations and 100 sample size.

				RE		
q'	t	(20,80)	(40,60)	(50,50)	(60,40)	(80,20)
0.1	1	3	1.57	1.38	1.22	1.11
	2.5	2.71	1.69	1.47	1.28	1.09
	4	2.75	1.80	1.61	1.34	1.12
	5.5	2.16	1.77	1.50	1.30	1.17
	7	1.46	1.29	1.21	1.21	1.06
	8.5	1	1	0.92	1	1.09
	10	1.125	1	0.90	0.90	0.90
0.3	1	4	1.83	1.57	1.38	1.11
	2.5	3.75	2.07	1.77	1.47	1.16
	4	3.74	2.32	1.85	1.54	1.19
	5.5	2.69	2.11	1.73	1.5	1.21
	7	1.43	1.36	1.29	1.21	1.06
	8.5	1	1	1	0.92	1.09
	10	1.11	1	1	0.90	0.90
0.5	1	6.50	3	2	1.57	1.22
	2.5	6.60	3.21	2.32	1.83	1.27
	4	5.57	3.33	2.43	1.94	1.29
	5.5	3	2.53	1.46	1.86	1.33
	7	1.53	1.54	1	1.36	1.13
	8.5	1	1	1	1	1.09
	10	1.10	1.11	1	1	0.90

Table.6 Relative efficiency to compare between two estimates (parametric with direct and parametric with indirect) of survival function under Weibull distribution with parameters (0.6,3) with different percent of exact observations and 100 sample size.

				RE		
q'	t	(20,80)	(40,60)	(50,50)	(60,40)	(80,20)
0.1	1	1.13	1.06	1.04	1.03	1.01
	2.5	1.25	1.13	1.09	1.06	1.02
	4	1.35	1.19	1.14	1.10	1.04
	5.5	1.45	1.32	1.22	1.15	1.07
	7	1.60	1.41	1.38	1.22	1.13
	8.5	1	1	1	1	0.90
	10	0.90	0.87	0.87	0.91	0.95
0.3	1	1.16	1.08	1.05	1.03	1.01
	2.5	1.32	1.16	1.12	1.08	1.03
	4	1.49	1.26	1.18	1.12	1.05
	5.5	1.73	1.40	1.29	1.20	1.08
	7	2.06	1.52	1.41	1.30	1.13
	8.5	0.91	0.91	1	1	1
	10	0.83	0.83	0.83	0.87	0.91
0.5	1	1.19	1.10	1.07	1.05	1.02
	2.5	1.42	1.22	1.16	1.11	1.04
	4	1.68	1.35	1.25	1.17	1.07
	5.5	2.09	1.55	1.39	1.28	1.10
	7	2.56	1.80	1.62	1.45	1.21
	8.5	0.92	0.91	0.82	1	1
	10	0.77	0.80	0.83	0.87	0.91

5. Results and Discussions

Tables 1 and 2 show the results of the bias, Var. and MSE for the parametric estimation of the survival function from Weibull distributions with the parameters (1.5,6) and (0.6,3) respectively, under direct approach with 50% of exact data and different sample size at $q' = 0.1, 0.3, 0.5$.

The results in Table 1 illustrate that when the sample size increases the MSE and Var. decrease at all times of interest whereas the bias decreases at all times except at t=0.7. From the results of Table 2, it is seen that when the sample size increases the Var. and MSE decreases at all times except at t=1and t=2.5 the MSE increases. In addition, the bias increases as the sample size increases.

Then, based on these tables, it has been concluded that the parametric estimation of the survival function is good if the Weibull distribution has increasing hazard function such as in Table 1 whereas the parametric estimation of the survival function is quite good if the Weibull distribution has decreasing hazard function such as in Table 2. In addition, the bias has negative effect on the MSE in all cases.

Tables 3 and 4 show results of the bias, Var. and MSE of the parametric estimation of survival function from the Weibull distributions with the parameters (1.5,6) and (0.6,3) , respectively, on the based on the indirect approach with 50% of the exact data, and different sample size at $q' = 0.1, 0.3, 0.5$.

The results of Table 3 has illustrated that when the sample size increases the MSE and Var. decreases while the bias increases. From Table 4, it has been obtained that when the sample size increases the Var. decreases and the bias increases with the MSE unstable.

According to tables 1 to 4, it has been concluded that the parametric estimation is suitable if the Weibull distribution has the increasing hazard function as in Table 3 while the parametric estimation is quite reasonable if the Weibull distribution has decreasing hazard function as in Table 4. Also, the bias has negative effect on the MSE.

It should be noted that the MSE and the bias in Tables 3 and 4 are less than the MSE and the bias in Tables 1 and 2, that means the parametric estimation under the indirect approach is better than the parametric estimation under the direct approach.

To compare two parametric estimations using direct approach and indirect approach, the RE has been considered. For this purpose, two Tables 5 and 6 have been constructed for the Weibull distribution with parameters (1.5,6) and (0.6,3), respectively which depend on simulation data with 500 replications, 10 imputations and $q' = 0.1, 0.3, 0.5$ for 100 sample size, but for different percent of exact data. From the results of these tables we can conclude that the parametric estimation under the indirect approach is more efficient than the direct approach. Also, the efficiency increases when the percent of exact data decreases.

6. An Example

We applied the proposed estimations to the modified breast cancer data that was presented by Alharpy and Ibrahim (2013). The data consist of 61 patients treated by radiation therapy (R) and 63 patients treated by radiation therapy plus adjuvant chemotherapy (R+C). This study was implemented to compare the cosmetic effects of radiation therapy alone against radiotherapy and adjuvant chemotherapy on women with early breast cancer and the event of interest was the time to first occurrence of breast retraction and the patients were observed at clinic visits every 4 or 6 months, where the actual dates of event were recorded exactly if available. If not the interval of events were noted. The aim of this study is to compare two treatments with respect to their cosmetic effects. The modified data set is shown in Table 7. We used the Weibull distribution to obtain estimated survival function for the two treatments R and R+C by using direct (W.O) and indirect (W.I) approaches as shown in Figure 1. Both estimates approach in Figure 1 indicate that the patients in the R+C group develop breast retraction earlier than those in the R group, suggesting that the indirect approach provides an acceptable approximation to the estimate.

Table 7: Time to cosmetic deterioration (in months) in breast cancer patients with two treatments.

Radiotherapy	Radiotherapy plus chemotherapy
(0,7]; (0.8];(0,5];(4,11];(5,12];(5,11];(6,10]; (7,16];(7,14];(11,15];(11,18];≥15;≥17;(17,25]; (17,25];≥18;(19,35];(18,26];≥22;≥24;≥24; (25,37];(26,40];(27,34];≥32;≥33;≥34;(36,44]; (36,48];≥36;(37,44];≥37;≥37;≥37;≥38≥40≥;45; ≥46;≥46;≥46;≥46;≥46;≥46;≥46;≥37;12;15;17;4; 18;20;21;28;30;34;22;40;42;46;36	(0,22];(0,5];(4,9];(4,8];(5,8];(8,12];(8,21];(10,35] (10,17];(11,13];≥11;(11,17];≥11;(11,20];(12,20]; ≥13;(13,39];≥13;≥13;(14,17];(14,19];(15,22]; (16,24];(16,20];(16,24];(16,60];(17,27];(17,23]; (18,25];(18,24];(19,32];≥21;(22,32];(17,26];≥23; (24,31];(24,30];(30,34];(30,36];≥31;≥32(32,40]; ≥34;≥34;≥35;(35,39];(44,48];≥48;4;7; 10;13;15;19;22;14;18;25;29;30;32;34;40

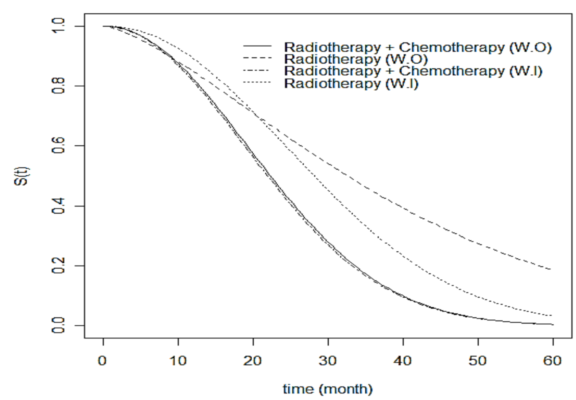

Figure 1: PMLE of survival functions of time to cosmetic deterioration.

7. Conclusion

In this article, the parametric estimate for PIC failure time data under Weibull distribution with direct (without imputation) and indirect (with imputation) approaches have been constructed. Simulation results show that the parametric estimation of the survival under the indirect approach is more efficient than the direct approach due to it's a small MSE. This is a realistic result, because in case of the interval-censored data the direct method depends on one value from the interval, while the indirect method depends on more than one values from the interval in which gives the opportunity to bring the correct value. Also, the efficiency increases when the percent of the exact data decreases, suggesting that the indirect approach is better to use when there is a large percent of the interval-censored data.

References

Alharpy, A. M., & Ibrahim N. A. (2013), Parametric tests for partly interval-censored failure time data under Weibull distribution via multiple imputation.

Huang, J. (1999), Asymptotic properties of nonparametric estimation based on partly interval-censored data. *Statist. Sinica.* 9, 501–519.

Kim, J. S. (2003), Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *J. R. Stat. Soc. Ser. B Stat. Methodol*, *65*, 489–502.

Odell, P. M., Anderson, K. M. and D'Agostino, R. B. (1992), Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics. 48*, 951–959.

Peto,R. and Peto, J.(1972). Asymptotically efficient rank invariant test procedures. *J. Roy. Statist. Soc. Ser. A 135*, pp.*185-207*

Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*.Wiley, New York.

Sun, J., Zhao, Q., & Zhao, X. (2005), Generalized log-rank tests for interval-censored failure time data. *Scand. J. Stat*. 32, pp. 49–57.

Wong, Raymond K. W. (1977), Weibull distribution, iterative likelihood techniques and hydrometerological data. *Journal of applied meteorology. 16*, 1360-1364.

Zhao, X., Zhao Q., Sun, J. and Kim, J. S. (2008), Generalized log-rank tests for partly interval-censored failure time data.*Biom. J., 50*, 375–385.

E-mail: azzahmath@hotmailcom, nakma@putra.upm.edu.my