On the Maximum Likelihood Estimation of Weibull Distribution with Lifetime Data of Hard Disk Drives

Daiki Koizumi†
†Department of Information and Management Science, Otaru University of Commerce, Hokkaido, Japan

Abstract— *The maximum likelihood estimations (MLEs) of the shape and scale parameters under the two-parameter Weibull distribution are considered with both complete and randomly censored data. These estimation methods are applied to real lifetime data of hard disk drives (HDDs) where the number of them is more than 90,000 for almost 4 years (from 2013 to 2016). Then the mean time to failure (MTTF) of each HDD is estimated. It turned out that almost all estimated shape parameters were more than one that indicates the Increasing Failure Rate (IFR). Furthermore, the estimated MTTF from complete data can be interrupted as the empirical lower bound of MTTF. But some lower bounds were much smaller than the official MTTF value guaranteed by a manufacturer. Finally, it also turned out that the ratio of the estimated MTTF from randomly censored data over that from complete data can measure the efficiency of this empirical lower bound.*

Keywords: Reliability Engineering, Probability Model, Weibull Distribution, Maximum Likelihood Estimation (MLE), Mean Time to Failure (MTTF), and Hard Disk Drive (HDD)

1. Introduction

1.1 Background

The lifetime data analysis is one of important research fields not only in the quality management of products but also in the reliability engineering [1], [2], [3]. In order to guarantee the certain lifetime for products, some probabilistic approaches are often chosen. Typically, a probability distributions have been assumed for the lifetime data and its parameters are estimated by observed data from reliability testing. For the probability distributions of lifetime data, the Weibull, exponential, normal, log normal, Poisson, and Gamma distributions etc. have been assumed [1], [2], [3].

Among those distributions, the Weibull distribution is one of famous probability density functions in the reliability engineering [1], [2], [3]. Especially the two-parameter Weibull distribution [4], [5] has the shape and scale parameters. If the shape parameter is less than one, the target device has the decreasing failure rate (DFR) [3]. If it equals to one, the target device has the constant failure rate (CFR) and the Weibull distribution corresponds to the exponential distribution [3]. If it is more than one, the target device has the increasing failure rate (IFR) [3]. These characteristics have been often referred as a part of *Bathtub curve* [2], [3].

On the other hand, there also exist some types of observed data in the reliability engineering. If the exact lifetime data are observed, those are called *complete data* [1], [2]. Otherwise the device is still alive and the complete lifetime data cannot be observed during the reliability testing. In this case, the reliability testing would be then forcibly stopped. This action has been called *censoring* [1], [2] and the rest of data except for the complete data would be recorded as *censored data* [1], [2]. There are several types of censoring such as Type I, Type II, and *Random censoring* [1], [2]. The random censoring is the most general type among those. The censored data are not the exact lifetime data but the lower bounds for the exact lifetime data. Therefore there is some possibility of their information to compensate the information of small number of complete data. The maximum likelihood estimation (MLE) methods of the Weibull distribution for the complete and randomly censored data have been proposed [4]. Recently, its application for the rocket engine tests has been also reported [5].

1.2 Previous Studies

In lifetime data analysis, it often happens that numbers of the complete or censored data are small. In this case, some graphical analysis techniques such as probability papers, probability plotting, and hazard plotting etc. have been developed and widely applied to estimate the parameters of the probability density functions for the lifetime data [2]. One of problems for those methods is that the objective functions for the estimations are undefined. However, the above situation has not been the case in some engineering fields. For example, the cloud technology has been widely spread in the network server communities. Cloud technology requires a huge number of hard disk drives (HDDs) as storage devices. The set of the lifetime data is now big data and those have been analyzed [6], [7] in some reliability engineering approaches. Especially, Schroeder et al. examined the lifetime data of HDDs [7]. They have been reported that the Weibull and Gamma distributions are accepted for the lifetime of HDDs but the exponential and log normal distributions are rejected according to their Chi-Square tests. They also tried to plot probability distribution curves including Weibull plot, however, they have not define any detail estimation methods [7].

1.3 Purpose of This Study

This paper assumes the two-parameter Weibull distribution and estimates those parameters by the maximum likelihood estimation (MLE) for the large numbers of lifetime data of HDDs. Fortunately, the lifetime data of HDDs in the cloud systems have been widely monitored and those data have been reported [8] from 2013 to 2016 where the total number of lifetime data is close to hundred thousands. This paper regards those data as both complete and randomly censored lifetime data where no outliers as well as no other operating factors such as temperatures, vibrations, and disk access error rates etc. are considered. Then, both shape and scale parameters of Weibull distribution as well as the mean time to failure (MTTF) of the HDDs would be estimated by MLE. Even if the number of complete data is small, it may be possible for randomly censored data to provide the sufficient information for the estimations.

As a result, the following three points would be at least observed. The first is that almost all estimated shape parameters are more then one and it indicates that each lifetime of objective HDDs has the Increasing Failure Rate (IFR). The second is that the estimated MTTF from complete data can be interrupted as the empirical lower bounds for MTTF. But some lower bounds are much smaller than the official MTTF value guaranteed by a manufacturer. The last is that the ratio of the MTTF from randomly censored data over that of complete data can measure the efficiency of the empirical lower bound.

The rest of this paper is organized as follows: the next section 2 gives definitions in terms of the reliability engineering under the Weibull distribution. Section 3 derives the maximum likelihood estimators under the Weibull distribution for both complete and randomly censored data. Section 4 considers to estimate the shape parameters, scale parameters, and the MTTF from the real lifetime data of hard disk drives. Section 5 discusses those results. The final section 6 gives concluding remarks.

2. Preliminaries

Let $T \geq 0$ denote the lifetime of the target device. From the probabilistic viewpoint, let $T = t$, i.e. $t \geq 0$ denotes the observed value of a continuous random variable T.

Suppose that the probability density function of t is the following two parameter Weibull distribution with the shape parameter $m > 0$ and the scale parameter $\lambda > 0$.

Definition 2.1 (Probability density function [1], [2]*):*

$$f(t) = \frac{m}{\lambda}\left(\frac{t}{\lambda}\right)^{m-1}\exp\left[-\left(\frac{t}{\lambda}\right)^{m}\right], \quad t \geq 0,\ m > 0,\ \lambda > 0\,. \tag{1}$$

■

Based on Definition 2.1, the following cumulative distribution function of $F(t)$, the reliability function of $R(t)$, and the mean time to failure (MTTF) or the expectation of $E(t)$ can be defined.

Definition 2.2 (Cumulative distribution function [1], [2]*):*

$$\begin{aligned} F(t) &= \int_0^t f(x)dx \\ &= 1 - \exp\left[-\left(\frac{t}{\lambda}\right)^{m}\right]. \end{aligned} \tag{2}$$

■

Definition 2.3 (Reliability function [1], [2]*):*

$$\begin{aligned} R(t) &= \int_t^\infty f(x)dx \\ &= \exp\left[-\left(\frac{t}{\lambda}\right)^{m}\right]. \end{aligned} \tag{3}$$

■

Definition 2.4 (Mean Time to Failure (MTTF) [1], [2]*):*

$$\begin{aligned} E(t) &= \int_0^\infty x\, f(x)dx \\ &= \lambda\,\Gamma\left(1 + \frac{1}{m}\right), \end{aligned} \tag{4}$$

where $\Gamma(\cdot)$ is the following Gamma function,

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\exp(-x)\,dx\,, \alpha > 0\,. \tag{5}$$

■

3. Maximum Likelihood Estimation (MLE) of Weibull Distribution

Suppose that a reliability test about a device is executed. A sufficient number of products of the device have been prepared and a sufficient number of the lifetime data can be observed. The main purpose of this test is to estimate and evaluate the mean time to failure (MTTF) of the device. In order to estimate MTTF, the shape and scale parameters of the Weibull distribution should be estimated. This paper considers the maximum likelihood estimation (MLE) and assumes two types for the observed data, i.e. the complete and the randomly censored data. The following subsection derives the maximum likelihood estimators for those two types.

3.1 MLE for Complete Data

Suppose that the complete data sequence $t_1, t_2, \ldots, t_n$ is observed where t_i, $i = 1, 2, \ldots, n$ denotes the ith observed lifetime data of the device. Let $L_1(m, \lambda)$ denotes the likelihood function where m, λ denote the shape and scale parameters of the Weibull distribution, respectively. Then, the maximum likelihood estimators $\hat{m}_1$ and $\hat{\lambda}_1$ can be formulated as follows:

Definition 3.1 (Likelihood function [4]):

$$L_1(m,\lambda) = \prod_{i=1}^{n} f(t_i) = \prod_{i=1}^{n} \frac{m}{\lambda}\left(\frac{t_i}{\lambda}\right)^{m-1} \exp\left[-\left(\frac{t_i}{\lambda}\right)^{m}\right]. \quad (6)$$

■

Definition 3.2 (Maximum likelihood estimators):

$$\begin{cases} \hat{m}_1 &= \arg\max_{m} L_1(m,\lambda)\,; \\ \hat{\lambda}_1 &= \arg\max_{\lambda} L_1(m,\lambda)\,. \end{cases} \quad (7)$$

■

Then, the log likelihood equations with respect to Definition 3.1 and 3.2 become,

$$\begin{cases} \left.\frac{\partial \ln L_1(m,\lambda)}{\partial m}\right|_{m=\hat{m}_1} &= 0\,; \\ \left.\frac{\partial \ln L_1(m,\lambda)}{\partial \lambda}\right|_{\lambda=\hat{\lambda}_1} &= 0\,. \end{cases} \quad (8)$$

The above equations (8) give the following conditions for the maximum likelihood estimators of both $\hat{m}_1$ and $\hat{\lambda}_1$:

Lemma 3.1 (Conditions for estimators [4]):

$$\begin{cases} \left[\frac{\sum_{i=1}^{n}(t_i)^{\hat{m}_1}\ln t_i}{\sum_{i=1}^{n}(t_i)^{\hat{m}_1}} - \frac{1}{\hat{m}_1}\right] - \frac{1}{n}\sum_{i=1}^{n}\ln t_i = 0\,; \\ \hat{\lambda}_1 = \left[\frac{\sum_{i=1}^{n}(t_i)^{\hat{m}_1}}{n}\right]^{\frac{1}{\hat{m}_1}}. \end{cases} \quad (9)$$

■

3.2 MLE for Randomly Censored Data

Suppose that the observation consists of both complete and randomly censored data. Assume that the former sequence is $t_1, t_2, \ldots, t_n$ and the latter sequence is $u_{n+1}, u_{n+2}, \ldots, u_{n+r}$. Note that the exact lifetime data $t_1, t_2, \ldots, t_n$ is observed and $t_{n+1}, t_{n+2}, \ldots, t_{n+r}$ is not observed but $u_{n+j} < t_{n+j}$, $j = 1, 2, \ldots, r$ is observed in this case. This means that the censored data express the lower bounds for unknown complete data. In order to simplify the notation, u_j would be used instead of u_{n+j} for the rest of this paper.

Then, the ML estimation problem for the shape and scale parameters of m and λ can be formulated as follows:

Definition 3.3 (Likelihood function [4]):

$$L_2(m,\lambda) = \frac{(n+r)!}{r!}\prod_{i=1}^{n} f(t_i)\prod_{j=1}^{r} R(u_j) = \frac{(n+r)!}{r!}\prod_{i=1}^{n}\frac{m}{\lambda}\left(\frac{t_i}{\lambda}\right)^{m-1}\exp\left[-\left(\frac{t_i}{\lambda}\right)^{m}\right] \prod_{j=1}^{r}\exp\left[-\left(\frac{u_j}{\lambda}\right)^{m}\right]. \quad (10)$$

■

Definition 3.4 (Maximum likelihood estimators):

$$\begin{cases} \hat{m}_2 &= \arg\max_{m} L_2(m,\lambda)\,; \\ \hat{\lambda}_2 &= \arg\max_{\lambda} L_2(m,\lambda)\,. \end{cases} \quad (11)$$

■

The above equations (11) give the following conditions for the maximum likelihood estimators of both $\hat{m}_2$ and $\hat{\lambda}_2$:

Lemma 3.2 (Conditions for estimators [4]):

$$\begin{cases} \left[\frac{\sum_{i=1}^{n}(t_i)^{\hat{m}_2}\ln t_i + \sum_{j=1}^{r}(u_j)^{\hat{m}_2}\ln u_j}{\sum_{i=1}^{n}(t_i)^{\hat{m}_2} + \sum_{j=1}^{r}(u_j)^{\hat{m}_2}} - \frac{1}{\hat{m}_2}\right] - \frac{1}{n}\sum_{i=1}^{n}\ln t_i = 0\,; \\ \hat{\lambda}_2 = \left[\frac{\sum_{i=1}^{n}(t_i)^{\hat{m}_2} + \sum_{j=1}^{r}(u_j)^{\hat{m}_2}}{n}\right]^{\frac{1}{\hat{m}_2}}. \end{cases} \quad (12)$$

■

4. Lifetime Data Analysis for Hard Disk Drives (HDDs)

4.1 Data Specifications

For data analysis, the lifetime data of hard disk drives (HDDs) in data center of Backblaze, Inc. [8] is used. The data contain daily status of HDDs based on the Self-monitoring, Analysis and Reporting Technology (S.M.A.R.T.). As of Apr. 2017, the total number of both complete and randomly censored data is 92,369 for almost 4 years (1362 days). Note that each unit of lifetime data is `[hrs]` according to the specification of S.M.A.R.T. Each number of both n and r corresponds to single physical device of HDD with single serial number. Furthermore, a single HDD model consists of plural serial numbers. Additionally no outliers as well as no other operating factors such as temperatures, vibrations, and disk access error rates etc. are considered. The basic data specifications are summarized in Table 1 and 2.

Table 1: HDDs Lifetime Data Specifications [1/2]

Year	From	To	Days	n	r
2016	Jan. 1st.	Dec. 31st.	366	1,422	78,162
2015	Jan. 1st.	Dec. 31st.	365	1,427	4,151
2014	Jan. 1st.	Dec. 31st.	365	2,200	3,405
2013	Apr. 10th	Dec. 31st.	266	740	3,062
	Total		1,362	5,789	85,044

4.2 Estimation Results for Complete Data

The complete lifetime data of t_i, $i = 1, 2, \ldots, n$ are from the serial numbers of single model of HDD. For each model of HDDs, the shape and scale parameters of Weibull distribution have been estimated as $\hat{m}_1$, $\hat{\lambda}_1$ by Eq. (9). From

Table 2: HDDs Lifetime Data Specifications [2/2]

	Complete Data t_i	Randomly Censored Data u_j
Total Manufacturers	6	6
Total Models	67	87
Total Serial Numbers	$n = 5,789$	$r = 85,044$
Max Data Capacity	8 [TB]	8 [TB]
Min Data Capacity	80 [GB]	80 [GB]

Eq. (4), the estimated MTTF [hrs] of $\hat{E}_1(x)$ can be calculated as the following:

$$\hat{E}_1(t) = \hat{\lambda}_1 \Gamma\left(1 + \frac{1}{\hat{m}_1}\right) \text{ [hrs]}. \quad (13)$$

In order to obtain those estimators, numerical calculations have been executed by the statistical computing software R version 3.3.3 [9]. The results for the main ten models have been summarized in Table 3.

Table 3: Estimation Results for Complete Data

Models	n	$\hat{m}_1$	$\hat{\lambda}_1$	$\hat{E}_1(t)$
ST4000DM000	1,807	1.256	12,188	11,339
ST3000DM001	1,720	5.376	18,891	17,417
ST31500541AS	397	4.683	39,261	35,912
Hitachi HDS722020ALA330	229	4.698	38,359	35,093
ST31500341AS	216	3.405	38,604	34,684
WDC WD30EFRX	162	1.019	8,292	8,227
Hitachi HDS5C3030ALA630	134	2.120	28,689	25,408
HGST HMS5C4040ALE640	103	1.104	8,198	7,902
ST1500DL003	90	1.558	10,236	9,202
ST320LT007	88	4.417	26,384	24,051

4.3 Estimation Results for Randomly Censored Data

For randomly censored data, a sequence of both t_i, $i = 1, 2, \ldots, n$ and u_j, $j = 1, 2, \ldots, r$ has been used to estimate the shape and scale parameters of Weibull distribution as $\hat{m}_2$, $\hat{\lambda}_2$ by Eq. (12). The estimated MTTF [hrs] of $\hat{E}_2(x)$ has been also calculated by Eq. (13). The results for the main ten models have been summarized in Table 4.

5. Discussions

5.1 Estimated Shape Parameters

As mentioned in Section 1.1, the value of the shape parameter in the Weibull distribution classifies at least three types of failure rates: DFR for $\hat{m} < 1$, CFR for $\hat{m} = 1$, and IFR for $\hat{m} > 1$.

For complete data, Table 3 shows that all of the estimated shape parameters are $\hat{m}_1 > 1.00$ and have classified into IFR. The only exception can be WDC WD30EFRX. This model has $\hat{m}_1 = 1.019$ and it is almost CFR. Note that a manufacturer has been estimated $\hat{m} = 0.55$ [10] for the first year MTTF and Schroeder et al. have been estimated $0.71 \leq \hat{m} \leq 0.76$ [7]. Thus there exist remarkable differences between our result and previous results.

For randomly censored data, Table 4 shows that two models i.e. WDC WD30EFRX and HGST HMS5C4040ALE640 have the shape parameters of $\hat{m}_2 = 0.638, 0.666$, respectively and those mean DFRs. The rest of models have the estimated shape parameters $\hat{m}_2 > 1.00$ that means IFRs.

Furthermore, Table 5 shows that the ratios of estimated shape parameters for both complete and randomly censored data i.e. $\hat{m}_2/\hat{m}_1$. Those ratios indicated that $\hat{m}_2/\hat{m}_1 < 1.00$ except for ST3000DM001. Thus it can be concluded that there is a strong tendency to be $\hat{m}_2 \leq \hat{m}_1$.

5.2 Estimated Scale Parameters

For complete data, the estimated scale parameters of $\hat{\lambda}_1$ range from 8,198 to 39,261 according to Table 3. A manufacturer obtained the $\hat{\lambda} = 3{,}787{,}073$ for the first year [10]. The corresponding HDD models from this manufacturer are ST4000DM000, ST3000DM001, ST31500541AS, ST31500341AS, ST1500DL003, and ST320LT007. Those estimated $\hat{\lambda}_1$ range from 10,236 to 39,261 and each of them is much smaller than the above $\hat{\lambda}$.

For randomly censored data, the estimated scale parameters of $\hat{\lambda}_2$ range from 11,873 to 12,288,903 according to Table 4. Furthermore, Table 5 shows the ratios of $\hat{\lambda}_2/\hat{\lambda}_1$. All ratios in Table 5 become $\hat{\lambda}_2/\hat{\lambda}_1 > 1.00$. It means that all estimated scale parameters from randomly censored data is larger than those from complete data.

5.3 Estimated MTTFs

For complete data, the estimated MTTFs of $\hat{E}_1(t)$ range from 7,902 [hrs] to 35,912 [hrs] according to Table 3. A manufacturer obtained the $\hat{m} = 0.55, \hat{\lambda} = 3{,}787{,}073$ [10]. From Eq. (4), these estimators give the following estimated MTTF for this manufacturer:

$$\begin{aligned} \hat{E}(t) &= \hat{\lambda} \cdot \Gamma\left(1 + \frac{1}{\hat{m}}\right) \\ &= 3{,}787{,}073 \cdot \Gamma\left(1 + \frac{1}{0.55}\right) \\ &= 6{,}447{,}223 \text{ [hrs]}. \end{aligned}$$

On the other hand, Schroeder et al. reports that "The MTTF for today's highest quality disks range from 1,000,000 hours to 1,500,000 hours" [7]. Our result from the complete lifetime data is extremely smaller than the above estimations.

For randomly censored data, the estimated MTTFs of $\hat{E}_2(t)$ range from 10,709 [hrs] to 16,366,251 [hrs] according to Table 4. Table 7 shows the ratios of $\hat{E}_2(t)/\hat{E}_1(t)$. According to Table 7, all ratios satisfy $\hat{E}_2(t)/\hat{E}_1(t) > 1.00$. It means that all MTTFs of $\hat{E}_2(t)$ for randomly censored data are larger than those of $\hat{E}_1(t)$ for complete data. But each degree varies depending on the HDD model. For example, the ratios of ST3000DM001 and ST1500DL003 are 1.14

Table 4: Estimation Results for Randomly Censored Data

Models	$n+r$	n	r	$\hat{m}_2$	$\hat{\lambda}_2$	$\hat{E}_2(t)$
ST4000DM000	36,532	1,807	34,725	1.122	227,663	218,260
ST3000DM001	4,537	1,720	2,817	5.833	21,362	19,787
ST31500541AS	2,087	397	1,690	3.583	68,530	61,737
Hitachi HDS722020ALA330	4,765	229	4,536	2.954	129,479	115,546
ST31500341AS	662	216	446	3.173	56,735	50,795
WDC WD30EFRX	1,289	162	1,127	0.638	490,229	683,660
Hitachi HDS5C3030ALA630	4,661	134	4,527	1.460	458,109	414,993
HGST HMS5C4040ALE640	7,162	103	7,059	0.666	12,288,903	16,366,251
ST1500DL003	106	90	16	1.511	11,873	10,709
ST320LT007	98	88	10	4.210	27,296	24,814

Table 5: The Ratios of the Estimated Shape Parameters

Models	$\hat{m}_2/\hat{m}_1$
ST4000DM000	0.89
ST3000DM001	1.09
ST31500541AS	0.77
Hitachi HDS722020ALA330	0.63
ST31500341AS	0.93
WDC WD30EFRX	0.63
Hitachi HDS5C3030ALA630	0.69
HGST HMS5C4040ALE640	0.60
ST1500DL003	0.97
ST320LT007	0.95

Table 6: The Ratios of the Estimated Scale Parameters

Models	$\hat{\lambda}_2/\hat{\lambda}_1$
ST4000DM000	18.68
ST3000DM001	1.13
ST31500541AS	1.75
Hitachi HDS722020ALA330	3.38
ST31500341AS	1.47
WDC WD30EFRX	59.12
Hitachi HDS5C3030ALA630	15.97
HGST HMS5C4040ALE640	1498.97
ST1500DL003	1.16
ST320LT007	1.03

and 1.16, respectively on Table 7. For these models, $\hat{E}_1(t)$ is slightly smaller than $\hat{E}_2(t)$. Therefore $\hat{E}_1(t)$ is a superior empirical lower bound of the MTTF of each HDD model. On the other hand, the ratio of HGST HMS5C4040ALE640 is 2071.28 on Table 5. It means that $\hat{E}_1(t) << \hat{E}_2(t)$. In this case, $\hat{E}_1(t)$ is an inferior empirical lower bound of the MTTF for this model.

5.4 Empirical Lower Bounds for MTTFs and their Efficiencies

From the observations so far, it turns out that the each $\hat{E}_1(t)$ from complete data gives the empirical lower bound of the MTTF for the respective HDD model. But those efficiencies must vary depending on those HDD models. The typical examples for two HDD models would be considered in the following three figures.

Fig. 1 depicts the relative frequency histograms and estimated Weibull probability density functions (p.d.f.) for ST3000DM001. In Fig. 1, the grey and dashed bars represent the relative frequencies of randomly censored and complete data, respectively. Moreover, solid and dashed lines represent the estimated Weibull p.d.f. from randomly censored and complete data, respectively. In Fig. 1, two p.d.f. lines are located near each other since they are estimated from two histograms of completed and randomly censored data and those histograms are close together. In this HDD model, the MTTF from dashed line is slightly smaller than that from solid line. In fact, the ratio of $\hat{E}_2(t)/\hat{E}_1(t)$ from Table 3 and 4 becomes $19{,}787/17{,}417 = 1.14$ which means that $\hat{E}_1(t)$ is superior empirical lower bound for MTTF of this HDD model.

Fig. 2 also depicts the relative frequency histograms and estimated Weibull p.d.f. for HGST HMS5C4040ALE640 and Fig. 3 is enlarged version of Fig. 2 with the magnified vertical axis. Additionally the legends in Fig. 2 and Fig. 3 are exactly same as Fig. 1. In this HDD model, two estimated p.d.f. lines are located far each other since two histograms are far each other. In this HDD model, the MTTF from dashed line is much smaller than that from solid line. The ratio becomes $\hat{E}_2(t)/\hat{E}_1(t) = 12{,}288{,}903/8{,}198 = 2071.28$ which means that $\hat{E}_1(t)$ is inferior empirical lower bound for MTTF of this HDD model.

Thus the efficiency of $\hat{E}_1(t)$ as the lower bound of the MTTF can be measured by the ratio of $\hat{E}_2(t)/\hat{E}_1(t)$. The closer this ratio gets to 1.00 (from larger side), the superior the efficiency of $\hat{E}_1$ as the lower bound of MTTF is. Table 7 shows all ratios and those efficiencies take various values depending on HDD models.

6. Concluding Remarks

This paper assumes the two-parameter Weibull distribution and considers the maximum likelihood estimation

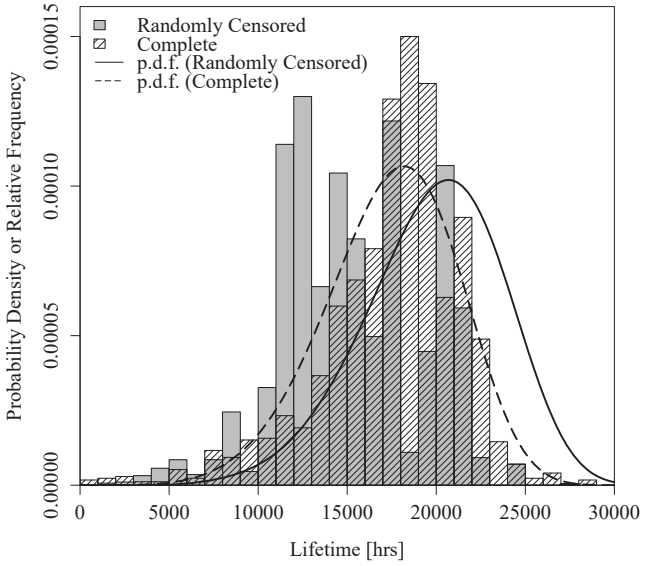

Fig. 1: Relative Frequency Histogram and Estimated Weibull p.d.f for ST3000DM001 ($n = 1{,}720, r = 2{,}817$).

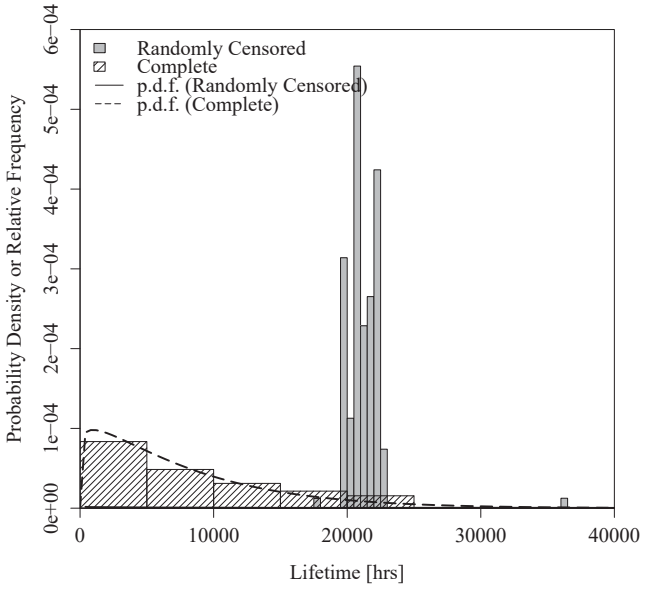

Fig. 2: Relative Frequency Histogram and Estimated Weibull p.d.f for HGST HMS5C4040ALE640 ($n = 103, r = 7{,}059$).

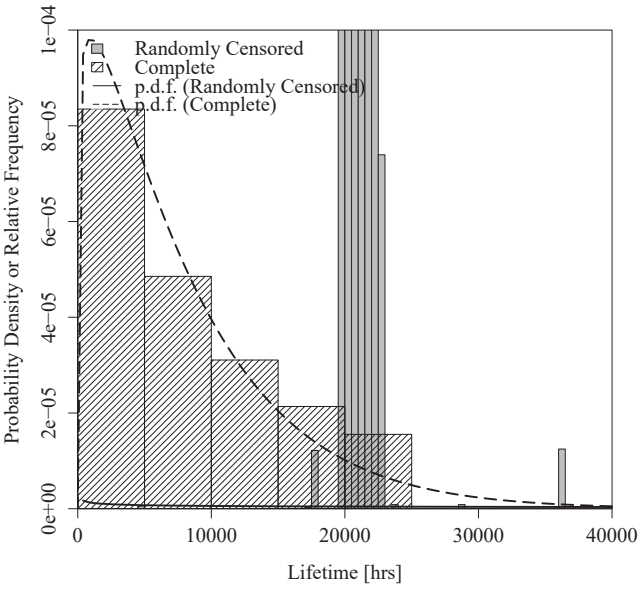

Fig. 3: Enlarged Figure of Fig. 2 with Magnified Vertical Axis.

Table 7: The Ratios of the Estimated MTTFs

Models	$\hat{E}_2(t)/\hat{E}_1(t)$
ST4000DM000	19.25
ST3000DM001	1.14
ST31500541AS	1.72
Hitachi HDS722020ALA330	3.29
ST31500341AS	1.46
WDC WD30EFRX	83.10
Hitachi HDS5C3030ALA630	16.33
HGST HMS5C4040ALE640	2071.28
ST1500DL003	1.16
ST320LT007	1.03

(MLE) for the shape and scale parameters to estimate the mean time to failure (MTTF) of the hard disk drive (HDDs). For large number of real lifetime data of HDDs, the complete and randomly censored data have been considered where the number of them is more than 90,000 for almost 4 years. The estimation results showed that almost all estimated shape parameters were more than one that means the Increasing Failure Rate (IFR). Furthermore, each estimated MTTF from the complete data can be interrupted as the empirical lower bound of the MTTF. But some of those bounds were much smaller than the official MTTF that was guaranteed by a manufacturer. Finally, the ratio of the estimated MTTF from the randomly censored data over that from complete data can measure the efficiency of this empirical lower bound.

One of open problems would be to evaluate the validity of the Weibull distribution to express the lifetime data of HDD. It would be able to be evaluated by measuring and

comparing the estimation errors among several probability density functions.

References

[1] Jerald F. Lawless, *Statistical Models and Methods for Lifetime Data,* John Wiley & Sons, 1982.
[2] Wayne B. Nelson, *Applied Life Data Analysis,* John Wiley & Sons, 1982.
[3] Michael S. Hamada, Alyson G. Wilson, C. Shane Reese, and Harry F. Martz, *Bayesian Reliability,* Springer, 2008.
[4] A. Clifford Cohen, "Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples," Technometrics, vol.7, no.4, pp. 579–588, Nov. 1965.
[5] Haibo Li, Zhengping Zhang, Yanping Hu, and Deqiang Zheng, "Maximum likelihood estimation of Weibull distribution based on random censored data and its application," Proceeding of the 8th International Conference on Reliability, Maintainability and Safety, pp.302–304, Jul. 2009.
[6] Eduardo Pinheiro, Wolf-Dietrich Weber and Luiz Andre Barroso, "Failure Trends in a Large Disk Drive Population," Proceeding of the 5th USENIX Conference on File and Storage Technologies (FAST'07), pp. 17–29, Feb. 2007.
[7] Bianca Schroeder and Garth A. Gibson, "Disk failures in the real world: What does an MTTF of 1,000,000 hours mean to you?," Proceeding of the 5th USENIX Conference on File and Storage Technologies (FAST'07), Article no. 1, Feb. 2007.
[8] Backblaze, Hard Drive Data and Stats [Online]. Available: https://www.backblaze.com/b2/hard-drive-test-data.html
[9] The R Foundation, The R Project for Statistical Computing [Online]. Available: http://www.r-project.org/
[10] Gerry Cole, "Estimating Drive Reliability in Desktop Computers and Consumer Electoronics Systems," Technology Paper from Seagate, TP–338.1, Nov. 2000.