ランダム打ち切りを含む データによるハードディスクドライブの 信頼性の近似推定について

小　泉　大　城

概要

本稿では，ランダム打ち切りを含むデータを用いて，2パラメータワイブル分布の最尤推定問題を近似的に解くことで，デジタル機器の補助記憶装置として広く用いられてきたハードディスクドライブ（Hard Disk Drive, HDD）の稼働時間を主に信頼性工学の観点から考察を行った．パラメータ推定にあたっては，米国Backblaze社により公開されている2018年から2021年までの約18万件のHDDのデータを利用した．信頼性工学の観点からは，これらのデータはHDDの故障の有無に応じて，HDDに電源が初めて入ってから故障するまでの稼働時間としての完全データと，未故障状態における当座の稼働時間としての不完全データ，すなわち今回該当するランダム打ち切りデータの2種類に大別される．両種類のデータからワイブル分布の形状および尺度パラメータの最尤推定値を近似的に数値計算したうえで考察を行ったところ，以下のことが明らかになった．（1）故障したのべ約6,000件の寿命の完全データのみから推定されたワイブル分布の形状パラメータは1より大きくなり，これは摩耗故障期であることを強く示唆する．（2）完全データのみから推定された平均故障時間（MTTF）は，おおむね1-4年程度．（3）ランダム打ち切りを含むデータ約18万件から推定された形状パラメータも9種類で1より大きくなり，故障しなかったHDDのデータを考慮してもやはり摩耗故障期であることを強く示唆する．（4）ランダム打ち切りを含むデータ全体に占める完全データの割合が15-25%程度のデータから近似最尤推定されたHDDの平均故障時

間（MTTF）は，5-8年程度.（5）HDD業界で使われている寿命の指標としての年間故障率（AFR）は，本稿で推定された故障率関数による値とよく一致する.

1　はじめに

1．1　研究背景

製品やシステムなどの設計にあたり，確率モデルを用いた部品の信頼性の分析は，品質管理や信頼性工学などの視点からさかんに研究されてきた分野のひとつである．こうした分析に使われる確率分布としては，ポアソン分布，指数分布，二重指数分布，正規分布，対数正規分布，ガンマ分布，ワイブル分布などが挙げられる［1,2,3］．特にワイブル分布［4,5,6］は，この分野でよく使われる連続型確率密度関数のひとつである．Mann［7］によると，ワイブル分布は，1928年にFisherとTippettの極値分布に関する研究に関連して初めて取り扱われ，しばらくはFisher-Tippett Type IIIという名前で知られていた．1939年になってスウェーデンのWaloddi Weibullが同様の確率分布を導出［4］し，1951年にWeibull自身がさらにいくつかの応用例を発表［5］するなどの経緯から，今日までワイブル分布と呼ばれている．以後，形状および尺度パラメータにより定義される2パラメータワイブル分布を基礎として，統計学や信頼性工学などの分野にまたがり様々な研究成果が報告されており［8,9,10］，近年，執筆者もいくつか報告を行った［11,12,13,14,15］.

信頼性工学においては，統計学における確率密度関数（probability density function）の定義にもとづいて，累積分布関数（cumulative distribution function），信頼度関数（reliability function），故障率関数（hazard function），平均故障時間（mean time to failure），あるいは平均故障間隔（mean time between failure）などが導出され，分析対象の信頼性の指標として利用される．このうち，故障率関数について，縦軸に故障率関数，横軸に時間を取ったプロットは，バスタブ型曲線［3］としてよく知られている．バスタブ型曲線では，

部品の故障率が時間とともに３つのパタンに変化する様子が表現されており，第１のパタンとして，故障率が時間とともに単調減少するdecreasing failure rate（DFR），第２に故障率が時間に関して一定で推移するconstant failure rate（CFR），第３に故障率が時間とともに単調増加するincreasing failure rate（IFR）がある．これら３つのパタンで表現されるバスタブ型曲線により，部品の稼働開始直後の初期故障の発生期に始まり，やがて安定的に稼働する偶発故障の発生期を経て，いつしか故障して稼働できなくなる摩耗故障の発生期に至るまで，部品の典型的なライフサイクルの統計モデル化が可能である．特に確率密度関数として２パラメータワイブル分布を定義すると，その故障率関数は形状パラメータが１未満の場合にDFR,1に等しい場合にCFR,1より大きい場合にIFRに対応するということから，この分野でワイブル分布が頻繁に利用される根拠のひとつとなっている．

つぎに，観測データについて，同じく信頼性工学の観点から考えてみる．ある部品が稼働を始めてから故障するまでの時間が観測されたとき，このデータは寿命の完全データ［1,2］と呼ばれる．部品がまだ故障せず，稼働している状態で観測を止めて時間を記録した場合，そのデータは不完全データの一種として，特に打ち切りデータ［1,2］と呼ばれる．さらにこの打ち切りに特に統一的な規則性がない場合に得られたデータをランダム打ち切りデータ［1,2］と呼ぶ．本稿で分析対象とするハードディスクドライブ（Hard Disk Drive, HDD）の寿命データについても，稼働を始めてから故障した時間としての完全データと，まだ故障していないが稼働を始めてからの経過時間としてのランダム打ち切りデータの集合となっている．

最後に２パラメータワイブル分布のパラメータ推定法について，信頼性工学の分野において，ワイブル確率紙［2］は手軽にワイブル分布のパラメータ推定値を得られる手法のひとつとされてきた．これは，ワイブル分布の分布関数を対数変換すると直線になることを利用し，専用の図面に観測データをプロットしてから目盛を読むことで，形状および尺度パラメータの推定値を得る，というものである．この方法は，部品の寿命の完全データの数が少ない場合など

でも，現場で手軽にパラメータ推定を行うことができるという長所があるが，視覚的な手法であるため推定の基準が必ずしも明確でないという短所がある．一方で，ワイブル分布の２種のパラメータの最尤推定法については，McCool［9］やRockette［10］をはじめとした数多くの研究がある．また，打ち切りデータを含むワイブル分布の２種のパラメータの最尤推定法については，Cohen［8］により提案されている．その他，ワイブル分布の様々なパラメータ推定法の歴史や関係等については，Evansらによる解説［16］が詳しい．

１．２　研究目的

本稿では，米国Backblaze社が公開しているのべ約18万件のHDDの寿命データ［18］を対象に，HDDの型番ごとに２パラメータワイブル分布のパラメータを近似最尤推定し，寿命特性を分析することを目的とする．１．１節で述べたように，このデータはHDDの寿命の完全データと，ランダム打ち切りデータの両方を含む．不完全データとしての打ち切りデータは，本来観測されるはずの完全データの下界を意味するが，完全データに加えてこれらの打ち切りデータも活用することで，より精確な推定結果を得られる可能性がある．さらに，HDDの型番ごとのデータ数（完全データとランダム打ち切りデータの総和）は，多くの型番（計11種の型番のうち９種）で1,000件を超えていることから，最尤推定法の有効性も期待される．

HDDの寿命特性としては，形状パラメータの最尤推定値にもとづく，１．１節で述べたバスタブ曲線を利用した部品のライフサイクルの分析，および２つのパラメータの近似最尤推定量を代入することによる平均故障時間（MTTF）や故障率関数の推定を行う．これらをHDDの型番別に，完全データのみの場合と，ランダム打ち切りを含むデータの場合で行い，推定結果を考察する．一方でHDD業界では，年間故障率（Annual Failure Rate, AFR）がHDDの寿命の指標となっている．AFRは，型番ごとにHDDの故障発生回数を総稼働時間で割り，１年稼働した場合の故障率に換算するという単純な計算式によっているが，このAFRと本稿で得られる推定故障率とを比較することで，AFRの指

標としての妥当性についても考察する.

最後に本稿の構成を述べる.つぎの2章では,準備として,信頼性工学における2パラメータワイブル分布について基本的な定義を行う.3章では,2パラメータワイブル分布のパラメータの近似最尤推定法について整理する.4章では,HDDの寿命データの仕様を説明し,データ解析の手順を説明してから推定結果を述べる.5章で得られた結果にもとづく考察を行い, 6章でまとめを行う.

2 準備

2.1 信頼性工学におけるワイブル分布

いま,$T \geq 0$ を部品の寿命時間とし,これを連続型確率変数とする.T の実現値を $t \geq 0$ とし,パラメータ $m, \lambda > 0$ をそれぞれ形状,尺度パラメータとする2パラメータワイブル分布［3,6］の確率密度関数 $f(t)$ は,定義 2.1で与えられる.

定義 2.1（2パラメータワイブル分布の確率密度関数).

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

■

ワイブル累積分布関数 $F(t)$ は,$f(x)$ を x に関して0から t まで積分した累積確率として定式化され,定義 2.2で与えられる.

定義 2.2（ワイブル累積分布関数).

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

■

ワイブル信頼度関数 $R(t)$ は，$f(x)$ を x に関して t から $+\infty$ まで積分した累積確率として定式化され，定義 2.3で与えられる．

定義 2.3（ワイブル信頼度関数）.

$$\begin{aligned} R(t) &= \int_t^{+\infty} f(x)dx \\ &= 1 - F(t) \\ &= \exp\left[-\left(\frac{t}{\lambda}\right)^m\right]. \end{aligned} \tag{3}$$

■

ワイブル故障率関数 $h(t)$ は，時刻 t における瞬間故障率を意味し，定義2.4で与えられる．

定義2.4（ワイブル故障率関数）.

$$\begin{aligned} h(t) &= \frac{f(t)}{R(t)} \\ &= \left(\frac{m}{\lambda^m}\right) t^{m-1}. \end{aligned} \tag{4}$$

■

ここで，横軸に t，縦軸にワイブル故障率関数 $h(t)$ を取りグラフを描くと，1．1節で述べたバスタブ型曲線を表すことができる．さらに形状パラメータが m<1 の場合はDFR，m=1 の場合はCFR，m>1 の場合はIFRに対応するのも同じく1．1節で述べたとおりである．

平均故障時間（Mean Time to Failure, MTTF）は，定義 2.1で定義した確率密度関数 $f(t)$ の期待値として，定義 2.5で与えられる．

定義 2.5（ワイブル平均故障時間, MTTF）.

$$E(T) = \int_0^{+\infty} t\, f(x)dt \tag{5}$$

$$\quad = \lambda\,\Gamma\left(1 + \frac{1}{m}\right), \tag{6}$$

ただし，$\Gamma(\cdot)$ は式(7)で定義されるガンマ関数である．

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} \exp(-x)\, dx\,, \alpha > 0\,. \tag{7}$$

■

3　ワイブル分布のパラメータ推定

本章では, 3．1節, 3．2節でそれぞれ, 完全データ, ランダム打ち切りを含むデータの場合の2パラメータワイブル分布のパラメータの近似最尤推定法を整理する.

3．1　完全データによるパラメータ推定

定義 2.1の $f(t)$ のもとで長さ n の系列 $\{t_1, t_2, \ldots, t_n\}$ を観測したとき，パラメータ m, λ による尤度関数 $L_1(m, \lambda)$ は，定義 3.1で与えられる.

定義 3.1（完全データによる尤度関数）.

$$\begin{aligned} L_1(m,\lambda) &= \prod_{i=1}^{n} f(t_i) \\ &= \prod_{i=1}^{n} \frac{m}{\lambda}\left(\frac{t_i}{\lambda}\right)^{m-1} \exp\left[-\left(\frac{t_i}{\lambda}\right)^{m}\right]. \end{aligned} \tag{8}$$

■

式(8)を最大化する m, λ をそれぞれ $\hat{m}_1, \hat{\lambda}_1$ とすると，これらは最尤推定量として定義 3.2で与えられる.

定義 3.2（完全データによる最尤推定量）.

$$\begin{cases} \hat{m}_1 = \arg\max\limits_{m} L_1(m,\lambda)\;; \\ \hat{\lambda}_1 = \arg\max\limits_{\lambda} L_1(m,\lambda)\;. \end{cases} \tag{9}$$

■

最尤推定量が満たす極値条件より，$\hat{m}_1$, $\hat{\lambda}_1$ が満たす必要条件は，補題 3.1で与えられる［8］.

補題 3.1（完全データによる最尤推定量の必要条件［8］）.

$$\begin{cases} \left[\frac{\sum_{i=1}^{n}(t_i)^{\hat{m}_1}\ln t_i}{\sum_{i=1}^{n}(t_i)^{\hat{m}_1}} - \frac{1}{\hat{m}_1}\right] - \frac{1}{n}\sum_{i=1}^{n}\ln t_i = 0\,; \\ \\ \hat{\lambda}_1 = \left[\frac{\sum_{i=1}^{n}(t_i)^{\hat{m}_1}}{n}\right]^{\frac{1}{\hat{m}_1}}\ . \end{cases} \tag{10}$$

■

McCool［9］やRockette［10］によれば，式(10)を満たす$\hat{m}_1$, $\hat{\lambda}_1$の組は，ただひとつ存在することが示されている．ただし，本稿では数値計算の意味での近似最尤推定となる.

3．2 ランダム打ち切りデータを含むデータによるパラメータ推定

いま，長さnの系列 $\{t_1, t_2, \ldots, t_n\}$ を寿命時間に関する完全データ，長さrの系列 $\{u_{n+1}, u_{n+2}, \ldots, u_{n+r}\}$ をランダム打ち切りデータとする．ここで，ランダム打ち切りをしなかったとしたら得られるはずの完全データを $\{t_{n+1}, t_{n+2}, \ldots, t_{n+r}\}$ で表すと, ランダム打ち切りデータは観測されない完全データの下界を表すことから，j=1,2,...,rとしてu_{n+j}<t_{n+j}が成り立つ．以降ではこれらのu_{n+j}のことを単にu_jと表記する．この理由は，今回のデータ解析において，n個の完全データ，r個のランダム打ち切りデータともに順序関係が存在せず，これらのインデックスがハードディスクドライブの各個体に固有のシリアル番号に対応する名義尺度であることによる．なお，シリアル番号を含むデータの詳しい仕様については4．1節で述べる.

定義 2.1の$f(t)$ および, 定義 2.3の$R(t)$ もとでn個の完全データ $\{t_1, t_2, \ldots, t_n\}$，$r$個のランダム打ち切りデータ $\{u_1, u_2, \ldots, u_r\}$ を観測したとき，パラメータm,

λによる尤度関数 $L_2(m, \lambda)$ は，定義 3.3で与えられる.

定義 3.3（ランダム打ち切りを含むデータによる尤度関数）.

$$
\begin{aligned}
&L_2(m,\lambda) \\
&\quad = \frac{(n+r)!}{r!}\prod_{i=1}^{n} f(t_i)\prod_{j=1}^{r} R(u_j) \\
&\quad = \frac{(n+r)!}{r!}\prod_{i=1}^{n}\frac{m}{\lambda}\left(\frac{t_i}{\lambda}\right)^{m-1}\exp\left[-\left(\frac{t_i}{\lambda}\right)^{m}\right]\prod_{j=1}^{r}\exp\left[-\left(\frac{u_j}{\lambda}\right)^{m}\right]. \qquad (11)
\end{aligned}
$$

■

式(11)を最大化する m, λ をそれぞれ $\hat{m}_2, \hat{\lambda}_2$ とすると，これらは最尤推定量として定義 3.4で与えられる.

定義 3.4（ランダム打ち切りデータによる最尤推定量）.

$$
\begin{cases}
\hat{m}_2 = \arg\max\limits_{m} L_2(m,\lambda)\,; \\[2ex]
\hat{\lambda}_2 = \arg\max\limits_{\lambda} L_2(m,\lambda)\,.
\end{cases} \qquad (12)
$$

■

最尤推定量が満たす極値条件より，$\hat{m}_2, \hat{\lambda}_2$ が満たす必要条件は，補題 3.2で与えられる［8］.

補題 3.2（ランダム打ち切りデータによる最尤推定量の必要条件［8］）.

$$
\begin{cases}
\left[\dfrac{\sum_{i=1}^{n}(t_i)^{\hat{m}_2}\ln t_i+\sum_{j=1}^{r}(u_j)^{\hat{m}_2}\ln u_j}{\sum_{i=1}^{n}(t_i)^{\hat{m}_2}+\sum_{j=1}^{r}(u_j)^{\hat{m}_2}}-\dfrac{1}{\hat{m}_2}\right]-\dfrac{1}{n}\sum_{i=1}^{n}\ln t_i=0\,; \\[4ex]
\hat{\lambda}_2=\left[\dfrac{\sum_{i=1}^{n}(t_i)^{\hat{m}_2}+\sum_{j=1}^{r}(u_j)^{\hat{m}_2}}{n}\right]^{\frac{1}{\hat{m}_2}}.
\end{cases} \qquad (13)
$$

■

式(13)の数値計算に関して，Harter［17］は，定義 2.1の２パラメータワイブル分布にさらに未知の位置パラメータを加えた３パラメータワイブル分布の最尤推定を扱い，尤度関数のフィッシャー情報行列の正則性は，以下の３つのうちの少なくともひとつが満たされることと同値であることを明らかにした．これらはすなわち，（1）形状パラメータが２より大きい，（2）位置パラメータが既知，（3）下からの打ち切りデータが存在する，である．このうち（2）については，３．１節で述べた２パラメータワイブル分布における完全データの最尤推定量の最適性を保証している．また，（3）については，本稿で扱うランダム打ち切りデータが満たす条件であることから，式(13)の最尤推定量の最適性を保証しているものと考えられる．したがって，式(13)について数値計算により解を求めることは可能で，近似最尤推定も可能であるということになる．

4　ランダム打ち切りを含む HDD の寿命データの解析

本章では，４．１節でデータ仕様，４．２節でデータ解析の手順を述べたうえで，４．３節で完全データの解析結果を，４．４節でランダム打ち切りを含むデータの解析結果を示す．

４．１　データ仕様

本稿における解析対象のデータとして，米国Backblaze社が公開しているハードディスクドライブ（HDD）の寿命データ［18］を利用した．HDDとは，今日主流となっているノイマン型コンピュータにおける補助記憶装置の一種で，磁性体が塗布された円盤を高速回転させながら磁気ヘッドを移動させることで，コンピュータ内部の情報の読み書きを行う仕組みをもった精密機器のことである．近年はタブレット端末やスマートフォン端末などのモバイル機器の出現にともない，ソリッドステートドライブ（Solid State Drive, SSD）に取って代わられつつあるが，低価格でかつ記録可能な情報量が大容量であることか

ら，クラウドコンピューティング環境等におけるサーバ機器やパーソナルコンピュータ（PC）などの補助記憶装置として依然として欠かせないものとなっている．

今回対象としたのは，2018年1月1日から2021年12月31日までの4年間（1,461日）にクラウド系の情報システムで実際に稼働したHDDの状態をS.M.A.R.T（Self-Monitoring Analysis and Reporting Technology）により記録した，のべ約18万件のデータである[1)]．データ解析にあたり集計したのは，(1）日付［年月日]，(2）シリアル番号，(3）型番，(4）データ容量［bytes]，(5）稼働時間［hrs]，(6）状態［正常:0または異常:1］の6種類である．このうち，シリアル番号は個々のHDDに一意に設定されたいわゆるIDにあたる情報である．

表1に年ごとの日数，完全データの総数（n），ランダム打ち切りデータの総数（r）の内訳を示す．これらのnとrは，3．2節で定義したものであるが，これらは今回集計した項目のうち，前述の「(2）シリアル番号」の総数に相当する．そのうち，「(6）状態」が異常の場合の「(5）稼働時間」が，個々の完全データ$t_i, i=1,2,...,n$となり，「(6）状態」が正常の場合の「(5）稼働時間」が，個々のランダム打ち切りデータ$r_j, j=1,2,...,r$となる．

表1：HDDの寿命データの仕様［1/2］

年	自	至	日数	n	r
2021	01月01日	12月31日	365	1,573	126,888
2020	01月01日	12月31日	366	1,369	23,200
2019	01月01日	12月31日	365	2,154	15,071
2018	01月01日	12月31日	365	1,255	8,652
小計			1,461	6,351	173,811

つぎに表2にHDDの型番ごとのサイズ，データ容量，完全データ（n），ラ

1）執筆者は，2013-2016年ののべ約9万件のデータも過去に分析した［14,15]．

ンダム打ち切りデータ（r）の内訳を示す．表2では，2018年からの4年間の集計データから完全データの総数nが100以上となる計11種の型番を対象としている．

さらに，表3に完全データについて，HDDの型番ごとの各年のデータ数の内訳を示す．

最後に，表4にランダム打ち切りデータについて，HDDの型番ごとの各年のデータ数の内訳を示す．また，本稿末尾にある付録Aの図2から図12に，型番ごとの完全データの相対ヒストグラムを斜線で，ランダム打ち切りデータの相対ヒストグラムを灰色で示した．

表2：HDDの寿命データの仕様［2/2]

型番	サイズ	データ容量	n	r
ST12000NM0007	3.5″	12［TB]	1,937	38,313
ST4000DM000	3.5″	4［TB]	1,590	23,245
ST8000NM0055	3.5″	8［TB]	736	14,982
ST8000DM002	3.5″	8［TB]	446	10,065
ST12000NM0008	3.5″	12［TB]	373	20,528
TOSHIBA MG07ACA14TA	3.5″	14［TB]	364	38,389
TOSHIBA MQ01ABF050	2.5″	500［GB]	297	1,656
HGST HMS5C4040BLE640	3.5″	4［TB]	184	14,660
ST500LM012 HN	2.5″	500［GB]	150	609
TOSHIBA MQ01ABF050M	2.5″	500［GB]	140	442
HGST HUH721212ALN604	3.5″	12［TB]	134	10,922
小計			6,351	173,811

4．2 データ解析の手順

いま，k=1,2とし，k=1の場合を完全データによる推定，k=2の場合をランダム打ち切りを含むデータによる推定とする．表4に示したHDDの型番ごとに，補題 3.1または補題 3.2にしたがい，形状パラメータの近似最尤推定値$\hat{m}_k$を数値計算により求め，つぎに尺度パラメータの近似最尤推定値$\hat{\lambda}_k$を計算す

る．

さらに平均故障時間（MTTF）の推定値 $\hat{E}_k(T)$ を，$\hat{m}_k$, $\hat{\lambda}_k$ を代入することで計算する．この $\hat{E}_k(T)$ は，MTTFの定義 2.5を使い，定義 4.1で与えられる．

定義 4.1（MTTFの推定値）．

$$\hat{E}_k(T) \quad = \quad \hat{\lambda}_k \, \Gamma\left(1 + \frac{1}{\hat{m}_k}\right), \, k = 1, 2 \, . \tag{14}$$

■

なお，寿命データの単位が時間［hrs］であるため，式(14)で求まる平均故障時間 $\hat{E}_k(T)$ の単位も同じく時間［hrs］であることに注意されたい．また，数値計算にあたっては, 統計ソフトウェアR［19］のバージョン3.6.3を用いた．

表 3：HDD 寿命の完全データの内訳

型番	小計	2018 年	2019 年	2020 年	2021 年
ST12000NM0007	1,937	295	1,155	337	150
ST4000DM000	1,590	581	401	269	339
ST8000NM0055	736	126	222	177	211
ST8000DM002	446	92	124	91	139
ST12000NM0008	373	0	9	147	217
TOSHIBA MG07ACA14TA	364	9	10	102	243
TOSHIBA MQ01ABF050	297	50	80	96	71
HGST HMS5C4040BLE640	184	54	57	34	39
ST500LM012 HN	150	41	43	34	32
TOSHIBA MQ01ABF050M	140	6	22	32	80
HGST HUH721212ALN604	134	1	31	50	52
小計	6,351	1,255	2,153	1,364	1,574

表4：HDD寿命のランダム打ち切りデータの内訳

型番	小計	2018年	2019年	2020年	2021年
ST12000NM0007	38,313	1,067	14,120	21,802	1,324
ST4000DM000	23,245	4,022	272	341	18,610
ST8000NM0055	14,982	224	178	248	14,332
ST8000DM002	10,065	124	91	134	9,716
ST12000NM0008	20,528	0	109	218	20,201
TOSHIBA MG07ACA14TA	38,389	9	50	146	38,184
TOSHIBA MQ01ABF050	451	0	98	72	281
HGST HMS5C4040BLE640	14,660	1,854	34	69	12,703
ST500LM012 HN	609	121	34	32	422
TOSHIBA MQ01ABF050M	442	24	35	86	297
HGST HUH721212ALN604	10,922	2	50	52	10,818
小計	173,811	8,652	15,071	23,200	126,888

4．3　完全データの解析結果

表5に完全データを対象とした型番ごとの形状パラメータ $\hat{m}_1$，尺度パラメータ $\hat{\lambda}_1$，および平均故障時間MTTFの推定値 $\hat{E}_1(T)$ を示す.

また，表5で得られたパラメータ推定値をもとに，HDDの型番ごとのワイブル分布確率密度関数のグラフを，本稿末尾にある付録Aの図2から図12の破線として示した.

表5：完全データの推定結果

型番	$\hat{m}_1$	$\hat{\lambda}_1$	$\hat{E}_1(T)$
ST12000NM0007	1.752	13,590	12,102
ST4000DM000	3.882	39,523	35,763
ST8000NM0055	2.081	24,856	22,016
ST8000DM002	2.934	32,966	29,410
ST12000NM0008	1.796	10,602	9,429
TOSHIBA MG07ACA14TA	1.046	7,525	7,390
TOSHIBA MQ01ABF050	2.452	24,600	21,816

HGST HMS5C4040BLE640	2.400	29,413	26,074
ST500LM012 HN	3.262	34,410	30,848
TOSHIBA MQ01ABF050M	2.604	27,772	24,668
HGST HUH721212ALN604	1.306	12,510	11,544

4．4　ランダム打ち切りを含むデータの解析結果

表6にランダム打ち切りを含むデータを対象とした形状パラメータ $\hat{m}_2$，尺度パラメータ $\hat{\lambda}_2$，および平均故障時間MTTFの推定値 $\hat{E}_2(T)$ を示す．

また，表6で得られたパラメータ推定値をもとに，HDDの型番ごとのワイブル分布確率密度関数のグラフを，本稿末尾にある付録Aの図2から図12に実線で示した．

表6：ランダム打ち切りを含むデータの推定結果

型番	$\hat{m}_2$	$\hat{\lambda}_2$	$\hat{E}_2(T)$
ST12000NM0007	1.287	197,500	182,771
ST4000DM000	2.321	165,236	146,401
ST8000NM0055	1.329	373,954	343,888
ST8000DM002	1.842	250,001	222,093
ST12000NM0008	1.289	347,594	321,589
TOSHIBA MG07ACA14TA	0.973	1,252,238	1,267,318
TOSHIBA MQ01ABF050	1.674	53,649	47,920
HGST HMS5C4040BLE640	1.440	940,888	853,922
ST500LM012 HN	2.355	80,148	71,027
TOSHIBA MQ01ABF050M	2.491	51,824	45,978
HGST HUH721212ALN604	0.863	3,940,342	4,247,180

5　考察

本章では，5．1節において完全データの推定結果の考察を行い，つづく5．2節においてランダム打ち切りを含むデータの推定結果の考察を行う．

5．1　完全データの推定結果の考察

5．1．1　形状パラメータについて

表5の形状パラメータの推定値 $\hat{m}_1$ について，11種類のHDDの型番すべてで $\hat{m}_1>1$ となり，定義2.4の故障率関数が描くバスタブ曲線については，IFR，すなわち摩耗故障期であることを強く示唆する結果となった．完全データはシリアル番号のHDDが故障した時点における稼働時間であることから，今回分析したすべての型番のHDDが，本稿の分析が想定している部品のライフサイクルの中で，もっとも故障が発生しやすい時期であることを意味している．なお，$\hat{m}_1$ が最小であったのは，型番：TOSHIBA MG07ACA14TAで $\hat{m}_1$=1.046，最大となったのは，型番：ST4000DM000で $\hat{m}_1$=3.882となった．

5．1．2　尺度パラメータについて

表5の尺度パラメータの推定値 $\hat{\lambda}_1$ は，どれも $\hat{E}_1(T)$ の値の1割増程度の値となった．この理由についてはつぎの5．1．3節で考察するが，ほぼ似たような値であれば，平均故障時間を意味する $\hat{E}_1(T)$ のほうが理解しやすいので，詳細は5．1．3節で詳しく考察することとし，ここでは1点のみ，表5で $\hat{m}_1$=1.046≒1となったTOSHIBA MG07ACA14TAについてのみ考察する．

1．1節で述べたとおり，信頼性工学のバスタブ曲線において，$\hat{m}_1$=1はCFR，つまり部品が安定的に稼働する偶発故障期を意味する．この場合，時間を問わず故障率は一定であり，定義 2.4の右辺に m=1を代入して得られる故障率関数 $h(t)$ は，

$$h(t) = \frac{1}{\lambda},$$

となる．ここで，TOSHIBA MG07ACA14TAについては表5より，$\hat{\lambda}_1$=7,525であることから，このHDDは7,525［hrs］に1回故障すると推定される．これよりこの型番の平均故障間隔（Mean Time between Failure, MTBF）［1,2］の推定値は，7,525［hrs］になる．これは日数になおすと約313日で，連続稼働の条件下では，1年もたずに修理交換が必要，ということになる．ただし，

これは完全データのみによる推定結果からの考察である．ランダム打ち切りを含んだデータによる推定結果については５．２．２節で考察する．

５．１．３　平均故障時間について

今回扱ったHDDでは，表２で示したデータ容量が12-14［TB］のものは比較的大容量であるが，表５の型番で見ると，上から１番目のST12000NM0007が$\hat{E}_1(T)$=12,102［hrs］，５番目のST12000NM0008のそれが9,429［hrs］，６番目のTOSHIBA MG07ACA14TAのそれが7,390［hrs］，11番目のHGST HUH721212ALN604のそれが11,544［hrs］と，総じて早期の故障傾向を読み取ることができる．これらを年に換算するとそれぞれ，1.38，1.07，0.84，1.32年にあたる．

一方，ノート型PC等によく使われている2.5インチのHDDは，表５では上から７番目のTOSHIBA MQ01ABF050，９番目のST500LM012 HN，10番目のTOSHIBA MQ01ABF050Mの３種類であり，これらのデータ容量はいずれも500［GB］と小さめであるが，これらはいずれも平均故障時間が20,000［hrs］以上となっており，年換算すると約2.5年以上となって，3.5インチの12［TB］のものより長く稼働していた．残りの４-８［TB］のデータ容量のものは，22,000から35,000［hrs］程度，つまり2.5から4.1年程度稼働したあとに故障しており，これらの型番は，2.5インチのHDDと同等かそれ以上の信頼性があると考えてよさそうである．表５の最右列の平均故障時間の推定値 $\hat{E}_1(T)$ について，型番ごとの大小関係を見ると，おおむね $\hat{m}_1$ や $\hat{\lambda}_1$ のそれと一致していた．まず，$\hat{m}_1$ については，形状パラメータの推定値が小さい型番のものは比較的早く故障し，大きい型番のものはある程度の時間が経過してから故障しているという傾向があることがわかる．つぎに $\hat{\lambda}_1$ についても，k=1,$\hat{m}_1$>1 の条件下では，定義 4.1の $\hat{E}_k$ の右辺が，

$$\hat{E}_1(T) = \hat{\lambda}_1 \, \Gamma \left(1 + \frac{1}{\hat{m}_1} \right) ,$$

となっていることから，完全データについては型番ごとの $\hat{\lambda}_1$ と $\hat{E}_1(T)$ は正比

例の関係にあり，両者の大小関係は当然ながら一致する．

5．2 ランダム打ち切りを含むデータの推定結果の考察

5．2．1 形状パラメータについて

表6の形状パラメータの推定値$\hat{m}_2$について，すべての型番において表5にある$\hat{m}_1$の値よりも小さくなった．そこで，表7に$\hat{m}_1$，$\hat{m}_2$の値を転記し，$\hat{m}_2/\hat{m}_1$の値を計算して最右列に整理したところ，すべての型番でこの比は1未満の値となった．これは，ランダム打ち切りデータu_jが故障していない状態の稼働時間であり，真の寿命時間t_jの下界を表している以上，当然の結果であるが，$\hat{m}_2 \leq 1$となったのは，TOSHIBA MG07ACA14TA（$\hat{m}_2$=0.973）とHGST HUH721212ALN604（$\hat{m}_2$=0.863）の2種類のみであった．これらのうち前者については，完全データのみからの推定値$\hat{m}_1$=1.046に比べて93%の大きさにとどまったのに対し，後者については$\hat{m}_1$=1.306に比べて約3分の2の大きさとなった．

これらのつぎに小さな形状パラメータが得られたのは，ST12000NM0007（$\hat{m}_2$=1.287）とST12000NM0008（$\hat{m}_2$=1.289）で，$\hat{m}_1$に比べると$\hat{m}_2$の値は7割程度の大きさになり，これらはいずれもデータ容量が12［GB］の大容量HDDであった．また，データ容量が8［GB］であるST8000NM0055も$\hat{m}_2$=1.329となり，完全データのみの$\hat{m}_1$=2.081に比べると64%程度の大きさの推定値が得られた．表7の$\hat{m}_2/\hat{m}_1$の値の説明変数として，ランダム打ち切りデータの総数rと完全データの総数nの和に対するr比$r/(n+r)$が有効となりうるのか，両者の相関係数を計算してみたが，その値は0.405にとどまった．この要因としては，型番ごとにHDDの特性が違うと考えられることや，nとrの比も型番ごとにまちまちであること，さらに完全データとランダム打ち切りデータの間には単なるデータ数という量のみでは把握できない，質の相違が存在している可能性が考えられる．

表７：型番ごとの $\hat{m}_2/\hat{m}_1$ の値

型番	$\hat{m}_1$	$\hat{m}_2$	$\hat{m}_2/\hat{m}_1$
ST12000NM0007	1.752	1.287	0.735
ST4000DM000	3.882	2.321	0.598
ST8000NM0055	2.081	1.329	0.639
ST8000DM002	2.934	1.842	0.628
ST12000NM0008	1.796	1.289	0.718
TOSHIBA MG07ACA14TA	1.046	0.973	0.930
TOSHIBA MQ01ABF050	2.452	1.674	0.683
HGST HMS5C4040BLE640	2.400	1.440	0.600
ST500LM012 HN	3.262	2.355	0.722
TOSHIBA MQ01ABF050M	2.604	2.491	0.957
HGST HUH721212ALN604	1.306	0.863	0.661

５．２．２　尺度パラメータについて

５．１．２節と同様に，表６の中で形状パラメータが，$\hat{m}_2 \fallingdotseq 1$ と見なしうる型番を見てみると，上から，ST12000NM0007（$\hat{m}_2$=1.287），ST8000NM0055（$\hat{m}_2$=1.329），ST12000NM0008（$\hat{m}_2$=1.289），TOSHIBA MG07ACA14TA（$\hat{m}_2$=0.973），HGST HMS5C4040BLE640（$\hat{m}_2$=1.440），HGST HUH721212ALN604（$\hat{m}_2$=0.863）などが挙げられる．これらのHDDは，１．１節で述べた故障率関数のバスタブ曲線におけるCFR，つまり偶発故障期とみなせる．このように $\hat{m}_2 \fallingdotseq 1$ の場合は，５．１．２節で述べたように，尺度パラメータの推定値 $\hat{\lambda}_2$ の値がそのまま平均故障間隔（MTBF）の推定値となる．そこで，あらためてこれを整理すると，表８のようになり，これらのMTBFの推定値を年換算すると，上からそれぞれ，31，59，5,197，148，620［yrs］となり，かなり大きい値となった．これらの型番については，定義 2.4で述べた故障率関数 $h(t)$ を利用して，業界で利用されている年間故障率AFR（Annual Failure Rate）との比較が可能であるが，これについては５．２．４節で考察する．

表8：偶発故障期の型番ごとのMTBFの推定値

型番	$\hat{m}_2$	MTBF［hrs］
ST12000NM0007	1.287	197,500
ST8000NM0055	1.329	373,954
ST12000NM0008	1.289	347,594
TOSHIBA MG07ACA14TA	0.973	1,252,238
HGST HMS5C4040BLE640	1.440	940,888
HGST HUH721212ALN604	0.863	3,940,342

5．2．3　平均故障時間について

表6の最右列の平均故障時間 $\hat{E}_2(T)$ について，表5で得られた $\hat{E}_1(T)$ と比べると，これらはすべての型番について $\hat{E}_2(T)$ のほうが大きくなった．$\hat{E}_2(T)$ の値が特に大きくなった上位5つの型番は，HGST HUH721212ALN604とTOSHIBA MG07ACA14TA，HGST HMS5C4040BLE640，ST8000NM0055，ST12000NM0008などで，これらの平均故障時間を年に換算するとそれぞれ，484年，144年，97年，39年，37年という非現実的な値となった．これらに共通していえるのは，データ数全体に占める完全データ比 $r/(n+r)$ が5％未満ということである．表6で $\hat{E}_2(T)$ が小さくなったものとしては，TOSHIBA MQ01ABF050Mの45,978［hrs］，TOSHIBA MQ01ABF050の47,920［hrs］などで，これは年換算するとそれぞれ5.25，5.47年である．今回得られたデータの推定では少なくとも5年以上稼働する，ということになる．

表9の最右列に，表7と同様に，型番ごとの平均故障時間についての比 $\hat{E}_2(T)/\hat{E}_1(T)$ を計算して整理した．これらの比により，完全データのみによって推定された $\hat{E}_1(T)$ に比べて，ランダム打ち切りデータも併用して推定された $\hat{E}_2(T)$ がどれくらい大きくなったかがわかる．表9で特筆すべきは，最下行のHGST HUH721212ALN604の比率367.91と，上から5番目のTOSHIBA MG07ACA14TAの比率171.49であろう．これら2種の型番のHDDは，完全データのみによる平均故障時間の推定値は1万時間程度かそれ以下であったものが，ランダム打ち切りデータも併用して推定した結果，その推定値は100万

時間のオーダーになった．これらはそれぞれ，データ容量が12,14［TB］の大容量のHDDであり，ランダム打ち切りデータを考慮した結果，信頼性が特に高く推定されたことになる．

つぎに5．1．1節と同様に，型番ごとの $\hat{E}_2(T)/\hat{E}_1(T)$ の系列と，$n/(n+r)$ の系列との相関係数を計算したところ，-0.431となった．やはり前者を後者で線形的に説明するのは難しそうである．しかし，図1に示すように，縦軸に前者，横軸に後者を取った散布図を描いてみると，横軸の値 $n/(n+r)$，つまりデータ全体に占める完全データの割合が極めて小さい場合は縦軸の $\hat{E}_2(T)/\hat{E}_1(T)$ は極めて大きくなり（図1の左上の点），そこから横軸が増えるにつれて縦軸の値は急激に，かつほぼ単調に減少していることがわかる．表6または9において，$\hat{E}_2(T)$ の値が小さかったTOSHIBA MQ01ABF050M,TOSHIBA MQ01ABF050,ST500LM012 HNなどであるが，これらの平均故障時間を年換算するとおおむね5-8年程度で現実的な値と考えられる．これらの型番は，図1右下の横軸の範囲0.15 ～ 0.25に位置する3点と対応し，このときの縦軸の範囲は，1.86 ～ 2.30であった．

表9：型番ごとの $\hat{E}_2(T)/\hat{E}_1(T)$ の値

型番	$\hat{E}_1(T)$	$\hat{E}_2(T)$	$\hat{E}_2(T)/\hat{E}_1(T)$
ST12000NM0007	12,102	182,771	15.10
ST4000DM000	35,763	146,401	4.09
ST8000NM0055	22,016	343,888	15.62
ST8000DM002	29,410	222,093	7.55
ST12000NM0008	9,429	321,589	34.11
TOSHIBA MG07ACA14TA	7,390	1,267,318	171.49
TOSHIBA MQ01ABF050	21,816	47,920	2.20
HGST HMS5C4040BLE640	26,074	853,922	32.75
ST500LM012 HN	30,848	71,027	2.30
TOSHIBA MQ01ABF050M	24,668	45,978	1.86
HGST HUH721212ALN604	11,544	4,247,180	367.91

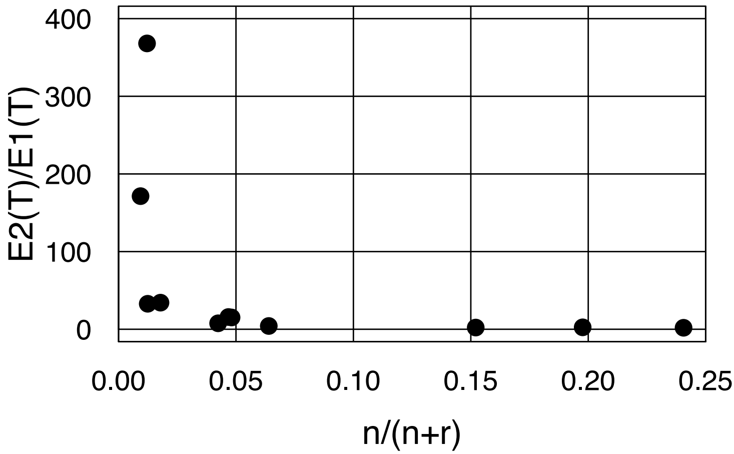

図1：$\hat{E}_2(T)/\hat{E}_1(T)$ と $n/(n+r)$ の散布図

5．2．4　年間故障率（AFR）と故障率関数との比較について

1．2節で述べたように，HDD業界では，製品の寿命の指標として年間故障率（Annual Failure Rate, AFR）が頻繁に使われている．これは，型番ごとにHDDの故障発生回数を総稼働時間で割り，1年稼働した場合の故障率に換算して求められている．この計算はノンパラメトリックな手法で手軽な四則演算ではあるものの，本稿で扱ったようなパラメトリックなワイブル分布とは異なる．そこで本章の最後に両者の比較考察を行ってみたい．

5.1.2節で述べたように，形状パラメータ $m=1$ の場合の故障率関数 $h(t)$ は，

$$h(t) = \frac{1}{\lambda},$$

で与えられる．本稿では時間データの単位はすべて［hrs］であることから，1年間を時間換算して $t=24\times365=8760$［hrs］とし，$h(8760)$ を計算すると，2パラメータワイブル分布による年間故障率を得ることができる．一方，HDD業界のいわゆるAFRについては，主なHDD型番についての故障回数と総稼働

時間等を米国Backblaze社が年ごとに公開していることから，本稿のデータと同じ2018-2021年の4年間について執筆者が計算した．これらをまとめて表10に示す．

表10：偶発故障期の型番ごとの故障率指標の比較［単位：%］

型番	$h(8760)$	AFR
ST12000NM0007	4.435	2.126
ST8000NM0055	2.342	1.366
ST12000NM0008	2.520	1.056
TOSHIBA MG07ACA14TA	0.700	0.800
HGST HMS5C4040BLE640	0.931	0.332
HGST HUH721212ALN604	0.222	0.451

表10の両者の系列の相関係数は0.942となり，よく一致しているといえそうである．ただし，ここで偶発故障期であるとした型番について，表8の$\hat{m}_2$の推定値を見ると，実際には完全に$m=1$となっているわけではない．そこで，両者の差分$\hat{m}_2-1$と，$h(8760)/\mathrm{AFR}$の値を表11に示す．

表11：形状母数の推定誤差と故障率指標の比

型番	$\hat{m}_2-1$	$h(8760)/\mathrm{AFR}$
ST12000NM0007	0.287	2.085
ST8000NM0055	0.329	1.714
ST12000NM0008	0.289	2.385
TOSHIBA MG07ACA14TA	-0.027	0.874
HGST HMS5C4040BLE640	0.440	2.800
HGST HUH721212ALN604	-0.137	0.493

表11の2種類の系列の相関係数は0.903となり，本稿で推定した$\hat{m}_2$の値が1からその差分が大きくなる（または小さくなる）ほど，$h(8760)/\mathrm{AFR}$の値も大きくなる(小さくなる)ことがわかる．さらに$\hat{m}_2-1$を説明変数x, $h(8760)/$

AFRを被説明変数yとした最小二乗法による単回帰式は，$y=3.749x+0.988$となり，この回帰式に$x=0$を代入したときには$y=0.988$となる．つまり，本稿の形状パラメータの推定値がちょうど$\hat{m}_2=1$となったとき，$h(8760)$とAFRの値はほぼ一致するものと思われる．以上の点から，HDD業界で使われている寿命の指標AFRは，本稿の2パラメータワイブル分布およびランダム打ち切りデータによる推定故障率関数から計算された年間故障率とよく一致することが確かめられた．

6 おわりに

本稿では，米国Backblaze社が公開しているハードディスクドライブ（HDD）の寿命データを使って，2パラメータワイブル分布の形状および尺度パラメータを数値計算し，HDDの信頼性の近似最尤推定を試みた．2018-2021年までの4ヵ年，のべ18万件のランダム打ち切りを含んだ計11種類の型番のHDD寿命を分析したところ，以下のことが明らかになった．（1）11種類すべての型番のHDDについて，故障したのべ約6,000件の寿命の完全データのみから推定されたワイブル分布の形状パラメータは1より大きくなり，これはIncreasing Failure Rate（IFR）の摩耗故障期であることを強く示唆する．（2）11種の型番について，完全データのみから推定された平均故障時間（MTTF）は，おおむね1年から4年の間である．（3）ランダム打ち切りを含むデータ約18万件から推定された形状パラメータも9種類で1より大きくなり，故障しなかったHDDのデータを考慮してもやはり摩耗故障期であることを強く示唆する．（4）ランダム打ち切りを含むデータ全体に占める完全データの割合が15-25%程度のデータから推定されたHDDの平均故障時間（MTTF）はおおむね5-8年程度．（5）HDD業界で使われている寿命の指標としての年間故障率（AFR）は，本稿で推定された故障率関数による値とよく一致する．

今後の課題としては，2022年以降のデータを用いた新たなデータ解析のほか，ワイブル分布の形状パラメータの推定値が1ではない場合の故障率の分析，

ワイブル分布のパラメータに回帰モデルを取り入れた分析，本稿で扱った点推定を一般化した区間推定によるデータ解析，あるいは本稿で仮定したランダム打ち切りデータによる統計モデル以外のアプローチによるパラメータ推定などが挙げられる．

参考文献

［1］Jerald F. Lawless, *Statistical Models and Methods for Lifetime Data,* John Wiley & Sons, 1982.

［2］Wayne B. Nelson著,柴田義貞 ,藤野和建 ,鎌倉稔成訳,「寿命データの解析」, 日科技連 , 1988年.

［3］Michael S. Hamada, Alyson G. Wilson, C. Shane Reese, and Harry F. Martz, *Bayesian Reliability,* Springer, 2008.

［4］Waloddi Weibull, “A Statistical Theory of the Strength of Materials,” Ingeniörsvetenskapsakademiens handlingar, nr. 151, 1939.

［5］Waloddi Weibull, “A Statistical Distribution Function of Wide Appli. cability,” Journal of Applied Mechanics, vol. 18, issue 3, pp. 293–297, Sep. 1951.

［6］Norman L. Johnson, Samuel Kotz, and N. Balakrishnan, *Continuous Univariate Distributions Volume 1*（*Second Edition*）*,* John Wiley & Sons, 1994.

［7］Nancy R. Mann, “Point and Interval Estimation Procedures for the Two-Parameter Weibull and Extreme-Value Distributions,” Technometrics, vol. 10, no. 2, May. 1968.

［8］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.

［9］John I. McCool, “Inference on Weibull Percentiles and Shape Parameter from Maximum Likelihood Estimates,” IEEE Transactions on Reliability, vol. R-19, no. 1, pp. 2–9, Feb. 1970.

[10] Howard Rockette, Charles Antle, and Lawrence A. Klimko, "Maximum Likelihood Estimation with the Weibull Model," Journal of the American Statistical Association, vol. 69, no. 345, pp. 246–249, Mar. 1974.

[11] 小泉大城,松嶋敏泰,"ある非定常なワイブル分布に従う時系列のベイズ最適な予測に関する一考察,"電子情報通信学会技術研究報告（信学技報）, vol. 115, no. 137, pp. 95–100, 2015年 7月.

[12] 小泉大城,"尺度パラメータが非定常なワイブル分布の応用について,"電子情報通信学会 2016年総合大会講演論文集, p. 50, 2016年 3月.

[13] Daiki Koizumi, "On the Weibull Distiribution with Nonstationary Scale Parameter," Proceeding of the International Society for Bayesian Analysis (ISBA) 2016 World Meeting, p. 323, Jun. 2016.

[14] 小泉大城,"ワイブル分布によるハードディスクドライブの寿命時間の最尤推定について,"情報処理学会研究報告（数理モデル化と問題解決）, vol. 2017-MPS-114, no. 3, pp. 1–4, 2017年 7月.

[15] Daiki Koizumi, "On the Maximum Likelihood Estimation of Weibull Distribution with Lifetime Data of Hard Disk Drives," Proceedings of the 2017 International Conference on Parallel and Distributed Process.ing Techniques and Applications (PDPTA'17), pp. 314–320, Jul. 2017.

[16] James W. Evans, David E. Kretschmann, and David W. Green, "Procedures for Estimation of Weibull Parameters," General Technocal Report FPL-GTR-264, Madison, WI: U.S. Department of Agriculture, Forest Service, Forest Products Laboratory, 2019.

[17] H. Leon Harter and Albert H. Moore, "Asymptotic Variances and Covariances of Maximum-Likelihood Estimators, from Censored Samples, of the Parameters of Weibull and Gamma Populations," the Annals of Mathematical Statistics, vol. 38, no. 2, pp. 557–570, Apl, 1967.

[18] Backblaze, Hard Drive Data and Stats [Online]. Available: https://www.backblaze.com/b2/hard-drive-test-data.html（参照：2022年7月5日）.

[19] The R Foundation, The R Project for Statistical Computing [Online]. Available: http://www.r-project.org/ (参照：2022年7月5日).

付録A　HDDの型番ごとの寿命データの相対ヒストグラムと推定確率密度関数のグラフ

以下，図2から図12までのグラフは，表2に示した11種類のHDDの型番ごとに，(1) ランダム打ち切りデータの相対ヒストグラム，(2) 完全データの相対ヒストグラム，(3) ランダム打ち切りを含むデータから推定されたワイブル分布の確率密度関数，(4) 完全データから推定されたワイブル分布の確率密度関数を可視化したものである.

いずれのグラフも横軸に時間 (Lifetime) [単位：hrs]，縦軸に相対度数または確率密度を取っており，凡例は，上記 (1) が灰色の棒 (Randomly Censored)，(2) が斜線の棒 (Complete)，(3) が実線 (p.d.f. (Randomly Censored))，(4) が破線 (p.d.f. (complete)) となっている.

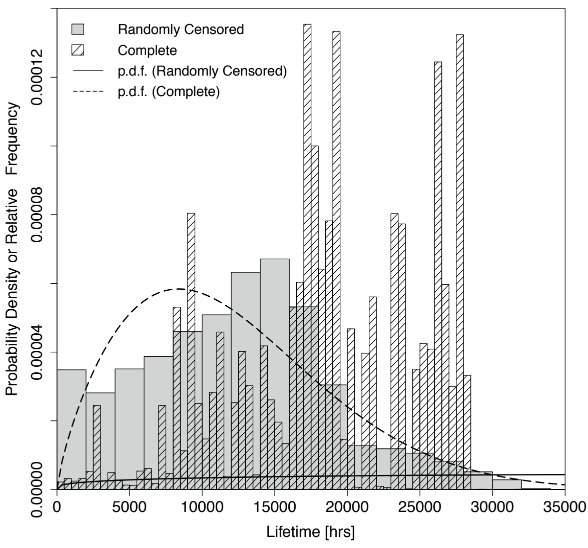

図2：ST12000NM0007の相対ヒストグラムと推定確率密度関数

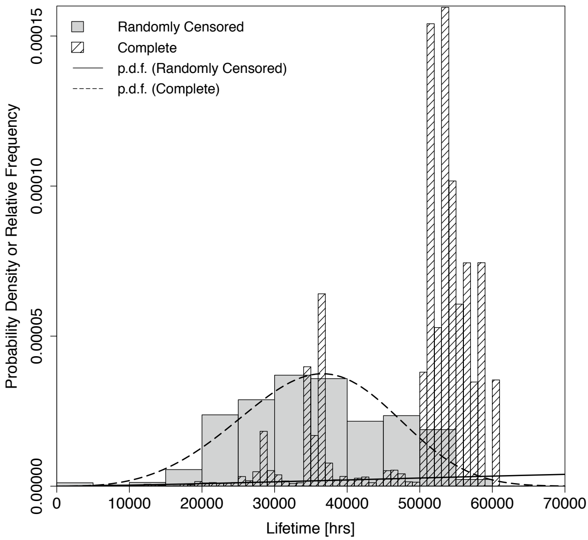

図３：ST4000DM000の相対ヒストグラムと推定確率密度関数

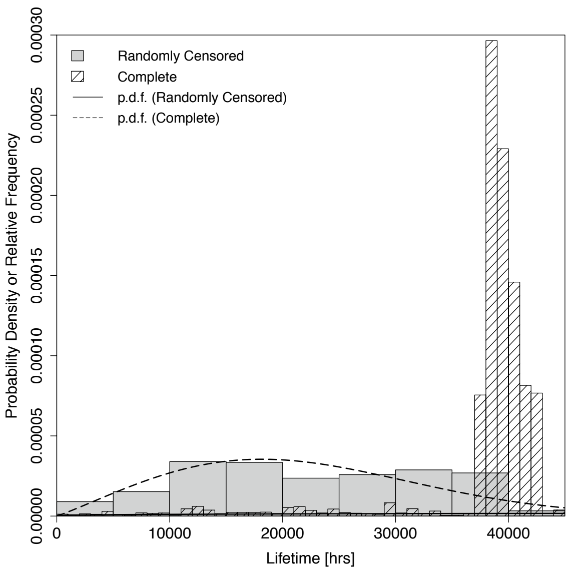

図４：ST8000NM0055の相対ヒストグラムと推定確率密度関数

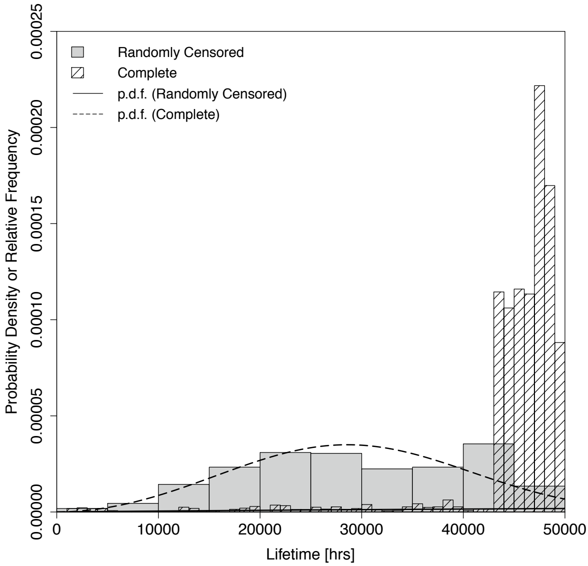

図５：ST8000DM002の相対ヒストグラムと推定確率密度関数

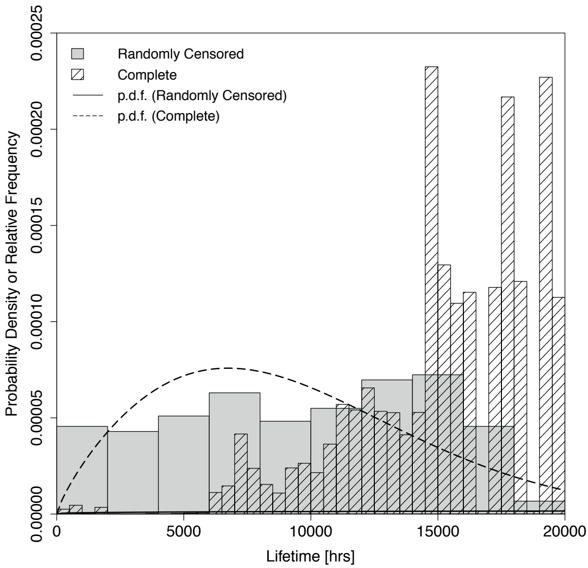

図６：ST12000NM0008の相対ヒストグラムと推定確率密度関数

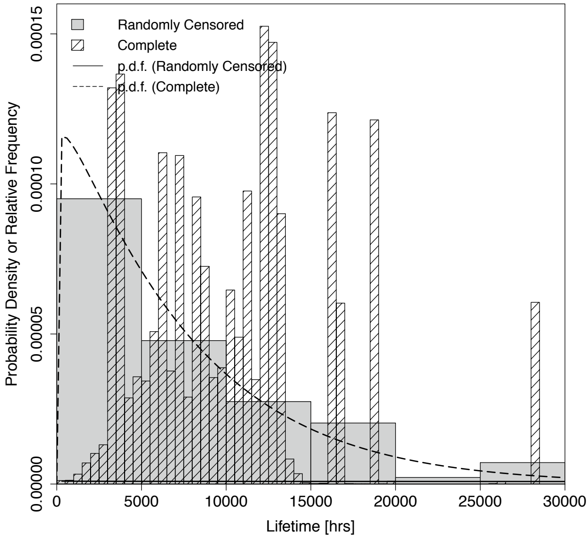

図７：TOSHIBA MG07ACA14TA の相対ヒストグラムと推定確率密度関数

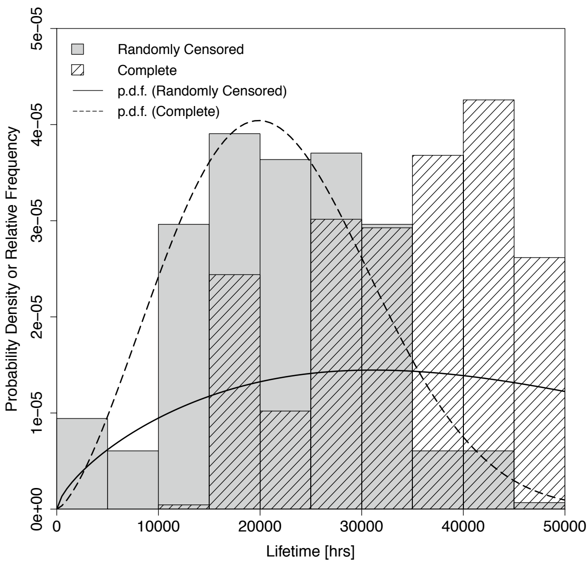

図８：TOSHIBA MQ01ABF050 の相対ヒストグラムと推定確率密度関数

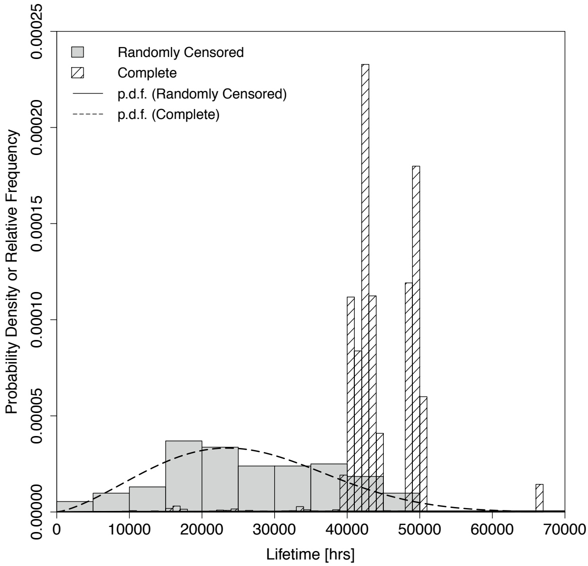

図 9：HGST HMS5C4040BLE640の相対ヒストグラムと推定確率密度関数

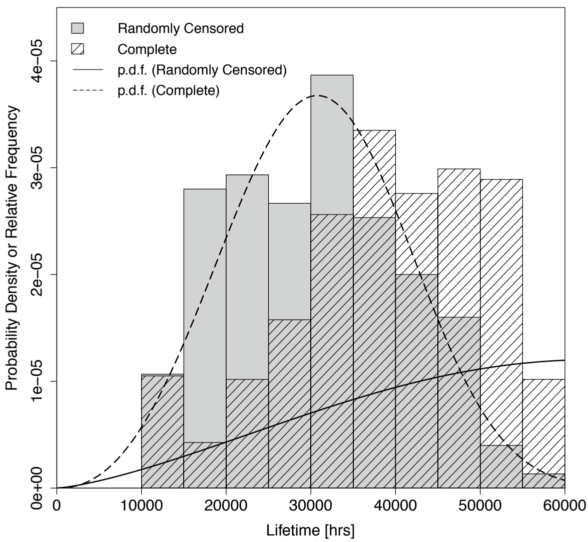

図10：ST500LM012 HNの相対ヒストグラムと推定確率密度関数

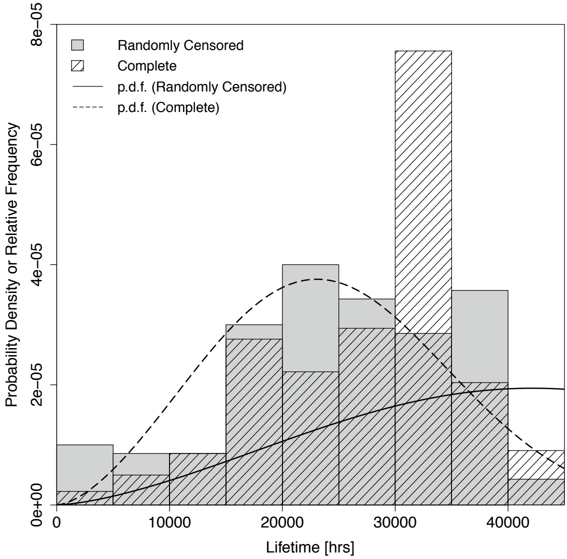

図11：TOSHIBA MQ01ABF050Mの相対ヒストグラムと推定確率密度関数

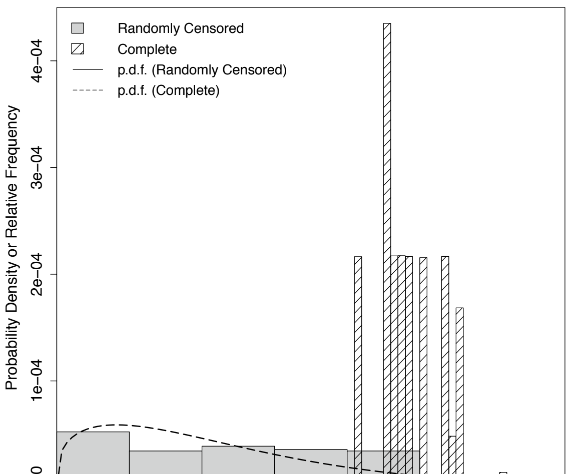

図12：HGST HUH721212ALN604の相対ヒストグラムと推定確率密度関数