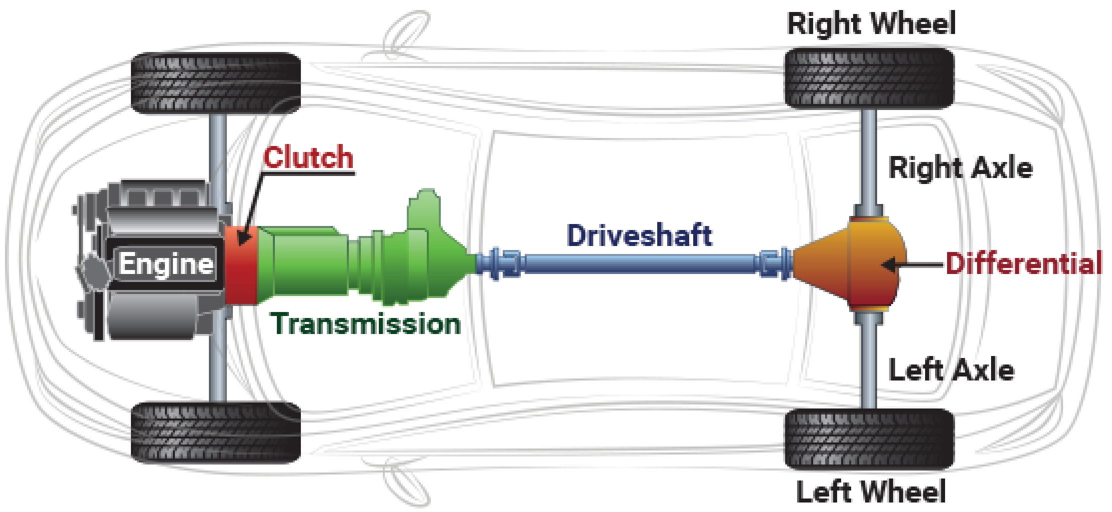
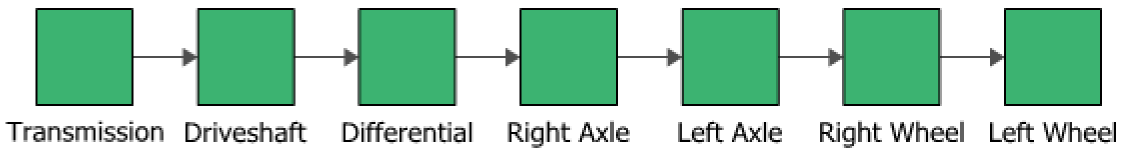
EM알고리즘을 이용한 기계부품의 고장 원인 및 신뢰성 분석

정나미, 손병관, 박찬석
부산대학교 산업공학과

Contents

A system in series has several components.





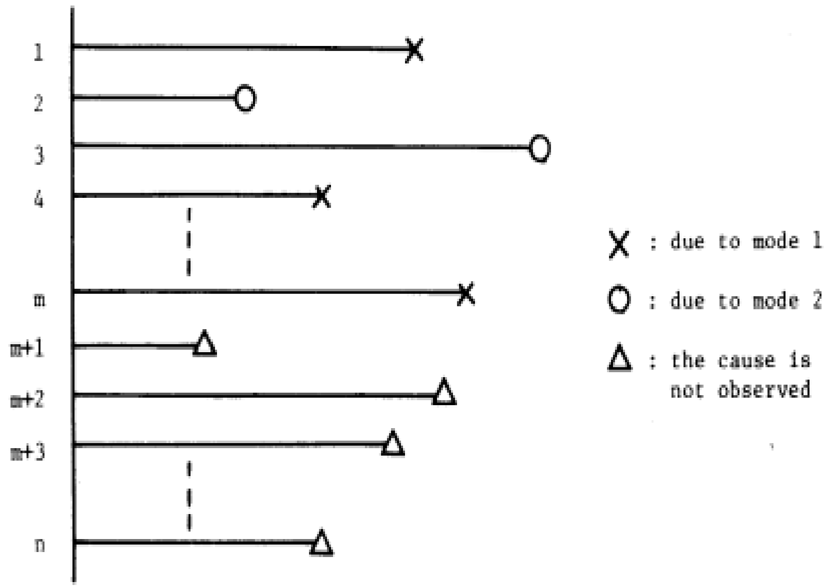
http://www.ssacstat.com/base/cs/cs_05.php?topmenu=5&left=5

Competing risks models

The failure of the whole system is caused by the earliest failure of any of the components, which is commonly referred to as competing risks.

Assumption - weakest link theory

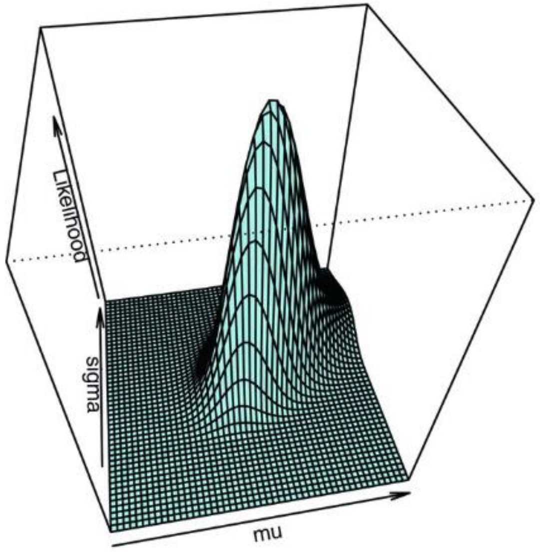
A1 The material contains inherently many strength-limiting defects, and its strength depends on the weakest defect of all of them.
A2 There are no interactions among the defects.



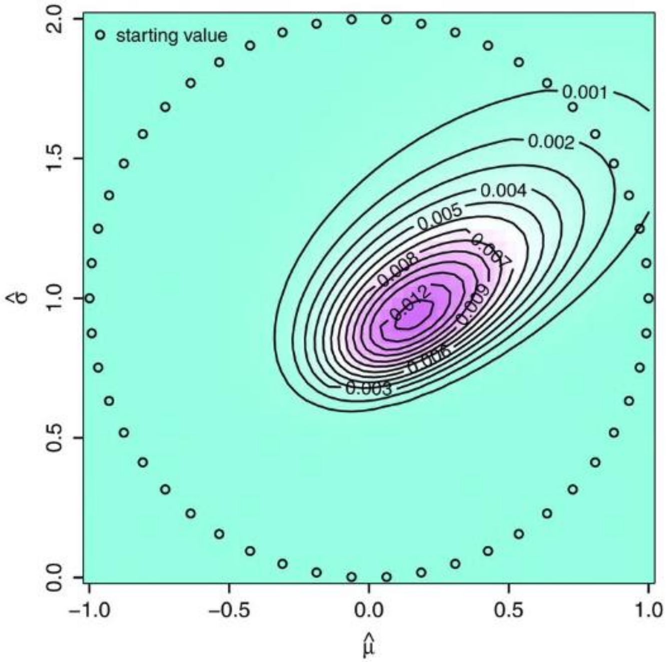
Park, C. S. and Padgett, W. J., (2006) Analysis of strength distributions of multi-modal failures using the EM algorithm, *Journal of Statistical Computation and Simulation*, 76(6), pp. 619~636.

M. Miyakawa, (1984) Analysis of Incomplete Data in Competing Risks Model, IEEE Transactions on Reliability, vol. R-33, no. 4, pp. 293-296.

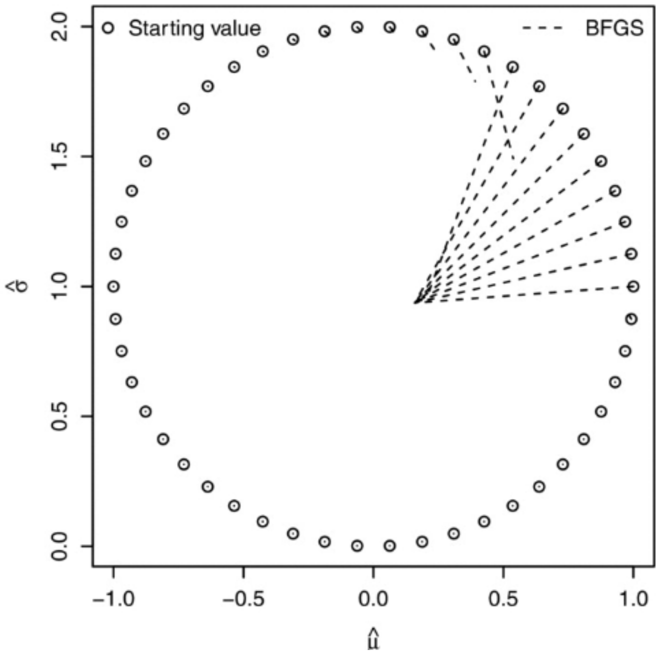
Illustration: Likelihood function with two parameters(BFGS에서 초기값의 영향)



(a) Contour plot



(b) Contour line

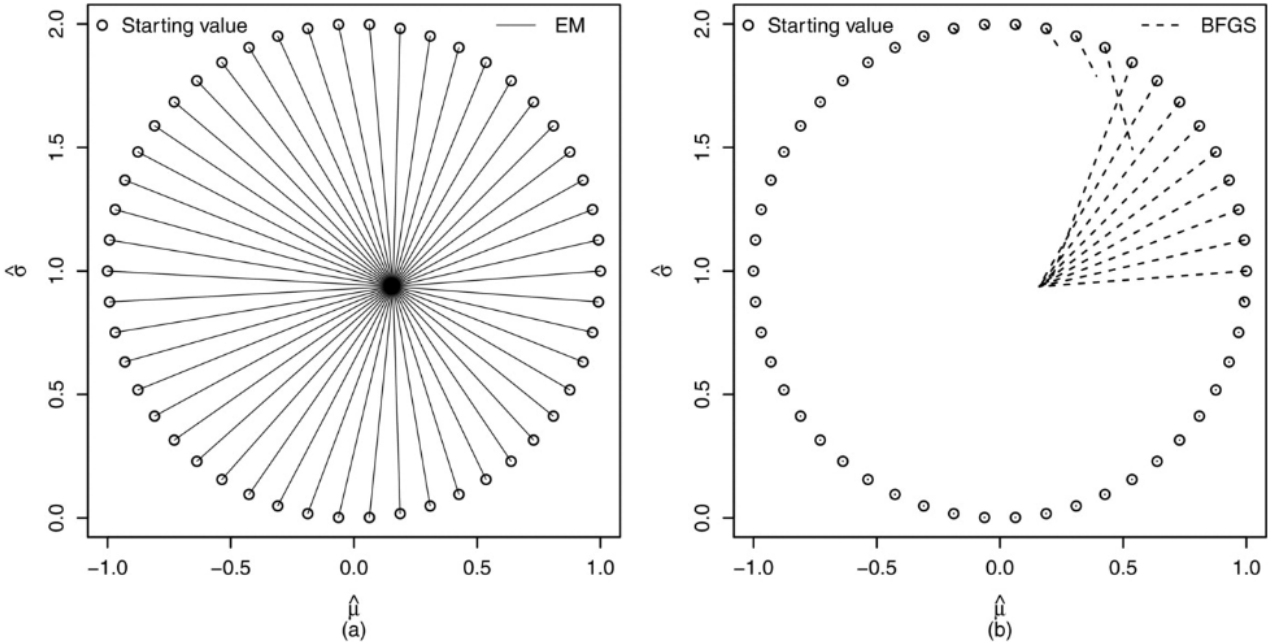


(c) Diagram of paths connecting the starting value a nd two estimates

Park C. (2013) Parameter estimation from load-sharing system data using the expectation–maximization algorithm. *IIE Transactions*.;45(2):147.

Illustration: Convergence of EM and BFGS(Newton-Raphson-type) estimation

- The EM algorithm was introduced by Dempster et al.(1977) to overcome the difficulty.



$^{50}/_{50} = 100\%$ $^{8}/_{50} = 16\%$

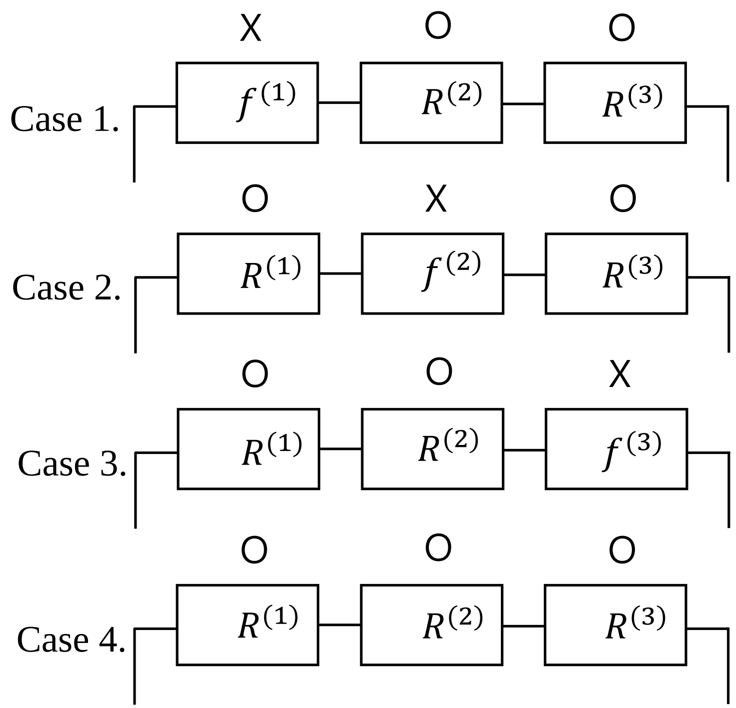
Park C. (2013) Parameter estimation from load-sharing system data using the expectation–maximization algorithm.*IIE Transactions*. 45(2), 147.
Dempster, A. P., N.M. Laird, and D. B. Rubin (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society B 39, 1-22.

Likelihood function with competing risk model

MLE는 Likelihood 함수를 최대가 되게 하는 값을 찾는 추정법

$$
\mathrm{L}(\Theta) \propto \prod_{i=1}^{n}\left[\left\{f^{(1)}(t_i)\prod_{\substack{j=1\\ j\neq 1}}^{J}R^{(j)}(t_i)\right\}^{\mathbb{I}_i(1)}\left\{R^{(j)}(t_i)\right\}^{\mathbb{I}_i(0)}\times\cdots\times\left\{f^{(J)}(t_i)\prod_{\substack{j=1\\ j\neq J}}^{J}R^{(j)}(t_i)\right\}^{\mathbb{I}_i(J)}\left\{R^{(j)}(t_i)\right\}^{\mathbb{I}_i(0)}\right]
$$

$$
=\prod_{i=1}^{n}\prod_{j=1}^{J}L_i\left(\theta^{(j)}\right)
$$



System states by four cases(System failure: Case:1, 2, 3, System Censoring case:4)

EM Algorithm

- A general iterative approach for computing the MLE of parametric models when there are no closed-form ML estimates, or the data are incomplete.
- The advantage is that it solves a difficult incomplete-data problem by constructing two easy steps.

s = s+1

No

E-step : $Q(\Theta;\Theta_s)$

M-step : $\Theta_{s+1} = \arg\max\, Q(\Theta;\Theta^{(s)})$

$\log L(\Theta^{(s+1)}) - \log L(\Theta^{(s)}) < \epsilon$
or
$|\lambda_{s+1}{}^{(j)} - \lambda_s{}^{(j)}| < \epsilon\lambda_{s+1}{}^{(j)}$
$|\alpha_{s+1}{}^{(j)} - \alpha_s{}^{(j)}| < \epsilon\alpha_{s+1}{}^{(j)}$

Yes

$\widehat{\Theta}_{mle}$

1

$Q(\Theta;\Theta_s)= \int \log L^C(\Theta|y,z)p(z|y,\Theta_s)dz$, Calculate the expected value of log likelihood using parameter estimates.

2

Derive parameter estimates that maximize the expected log likelihood $Q(\Theta;\Theta_s)$ at E-step.

3

The number of iterations increases and repeats until convergence when $\log L(\Theta^{(s+1)}) - \log L(\Theta^{(s)}) < \epsilon$ **(tolerance limit)**.

4

When the convergence criterion is reached, the EM sequence is done.

Park, C., & Lee, S. B. (2012) *Parameter estimation from censored samples using the expectation-maximization algorithm*

2 Failure modes data

- 30 observations of devices from a field-tracking study of a large system.
- Failure modes : **Censoring**(0), Surge(1), Wear(2)

Competing failure mode with 2 failure mode and censoring data

Interval Time (Cycle)	300,000	300,000	2,000	261,000	293,000	88,000	247,000	28,000	143,000	300,000
Failure mode	0	0	1	1	2	1	1	1	1	0
Interval Time (Cycle)	23,000	300,000	80,000	245,000	266,000	275,000	13,000	147,000	23,000	181,000
Failure mode	1	0	1	2	2	2	1	2	1	2
Interval Time (Cycle)	30,000	65,000	10,000	300,000	173,000	106,000	300,000	300,000	212,000	300,000
Failure mode	1	1	1	0	1	1	0	0	2	0

Meeker W.Q. and Escobar L.A., (1998) Statistical Methods for Reliability Data, John Wiley & Sons, Inc.

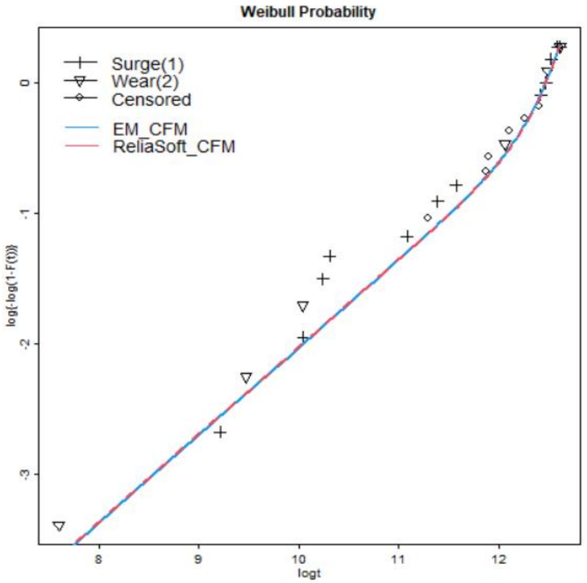
3 Failure modes data

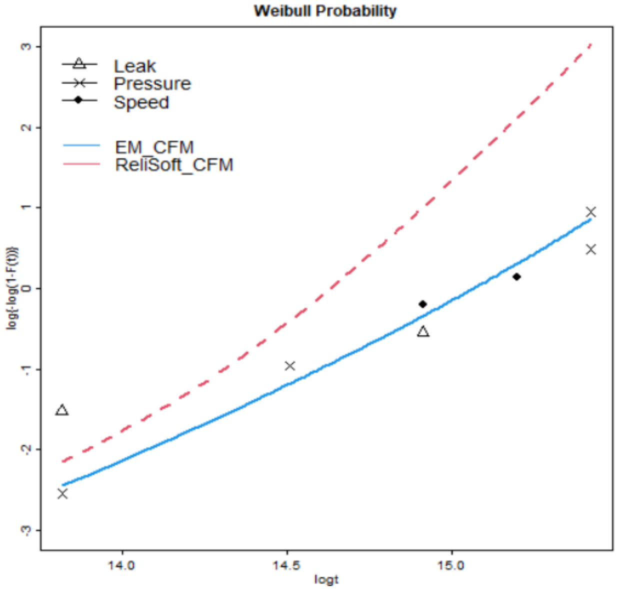
- 8 observations of complete data from a pneumatic cylinder(system)
- Failure modes : Leak(A), Pressure(B), Speed(C)

Competing failure mode with 3 failure mode

Sample Number	1	2	3	4	5	6	7	8
Interval Time(Cycle)	2,000,000	3,000,000	5,000,000	1,000,000	4,000,000	1,000,000	5,000,000	3,000,000
Failure mode	B	A	B	B	C	A	B	C

Chang, M. S., Choi, B. O., Kang, B. S., Park, J. W. and Lee, C. S., (2013) Reliability Analysis of a Mechanical Component with Multiple Failure modes, *Trans. Korean. Soc. Mech. Eng. A*, 37(9), pp. 1169~1174.





Estimation of Parameters by 2 Failure mode

Reliability Parameter	Failure mode 1 (Surge)		Failure mode 2 (Wear)	
	EM	ReliaSoft	**EM**	ReliaSoft
α(shape)	**0.671**	0.670	**4.337**	4.490
η(scale)	**449,442**	449,430	**340,379**	340,380

Estimation of Parameters by 3 Failure mode

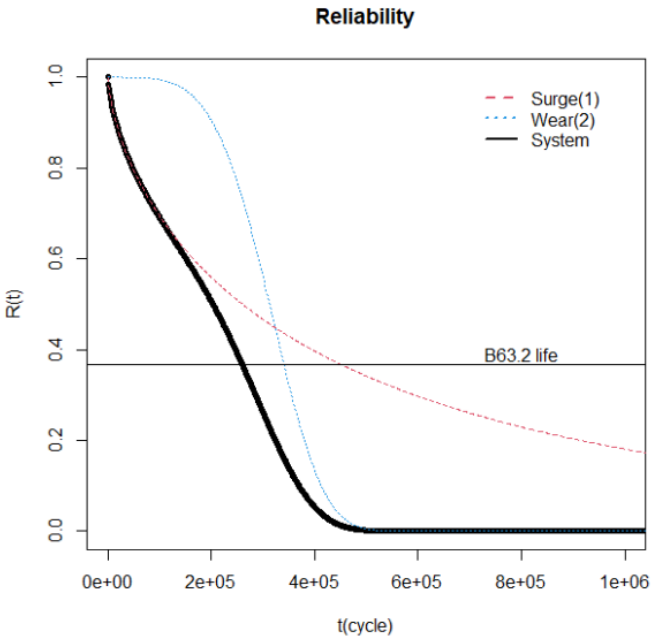
Reliability Parameter	Failure mode A (Leak)		Failure mode B (Pressure)		Failure mode C (Speed)	
	EM	ReliaSoft	**EM**	ReliaSoft	**EM**	ReliaSoft
α(shape)	**1.33**	1.35	**2.27**	4.25	**3.77**	3.97
η(scale)	**8,850,255**	5,806,500	**4,657,881**	2,660,800	**5,444,598**	3,316,600

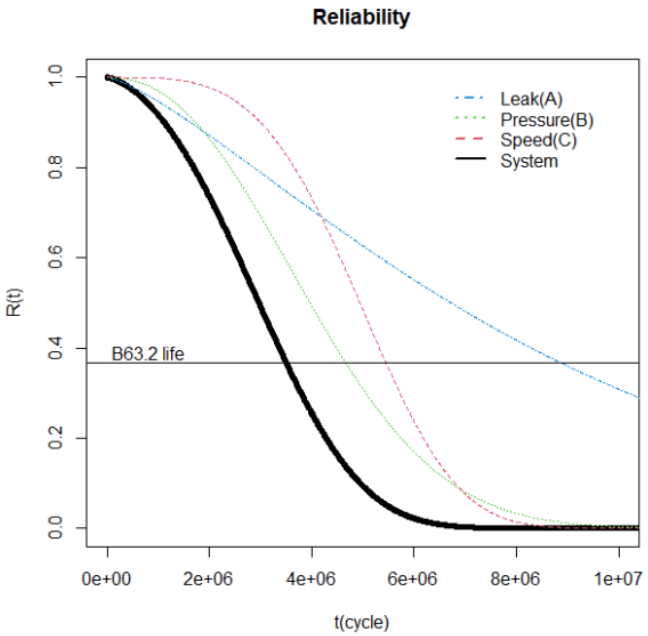
Comparison of EM and ReliaSoft Log likelihood and MSE by 2 Failure modes

	Log likelihood	MSE× 10^3
EM	**-317.293**	**0.831**
ReliaSoft	-317.519	0.855

Comparison of EM and ReliaSoft Log likelihood and MSE by 3 Failure modes

	Log likelihood	MSE× 10^3
EM	**-132.363**	**4.802**
ReliaSoft	-175.608	71.87





Bx life and MTBF in 2 failure modes

Cycle	Failure mode 1 (Surge)	Failure mode 2 (Wear)	System
B_1	473	117,855	473
B_5	5,374	171,616	5,373
B_{10}	15,709	202,597	15,709
$B_{63.2}$	449,442	340,379	259,496
MTTF	593,419	309,958	196,003

Bx life and MTTF in 3 failure modes

	Failure mode A (Leak)	Failure mode B (Pressure)	Failure mode C (Speed)	System
B_1	278,507	613,892	1,607,090	250,475
B_5	948,586	1,258,732	2,476,320	727,239
B_{10}	1,629,773	1,728,423	2,997,293	1,116,713
$B_{63.2}$	8,850,255	4,657,881	5,444,598	3,495,544
MTTF	8,137,733	4,125,929	4,918,588	3,013,835

1. EM Algorithm 과 기존 방법 간의 비교

- EM 알고리즘은 likelihood 함수가 unimodal의 경우 MLE는 global maximum이 되는 것을 보장
- 파라미터의 개수가 늘어날수록 MLE 수식의 복잡성이 증가하여 EM 알고리즘 형태가 적합
- 위의 한 사례 연구에서는 EM 알고리즘을 사용한 경우 MSE 값이 최대 10배 이상 차이 남

  → 기존의 분석 도구인 ReliaSoft는 MLE 가 최적화되어 있다고 보기 힘듦

2. Competing Failure Mode using EM Algorithm

- 직렬 시스템의 경우 한 부품의 고장 원인이 시스템 전체 수명에 큰 영향을 미침
- 시스템과 부품의 MTTF를 이용하여 신뢰도 향상을 위한 시스템-부품 관계 분석 가능

Masking Data

- Partial masking data
- Complete masking data

$$M_i = \{1,2\} \Longrightarrow 1 \text{ or } 2$$
$$M_i = \{2,3\} \Longrightarrow 2 \text{ or } 3$$
$$M_i = \{1,3\} \Longrightarrow 1 \text{ or } 3$$
$$M_i = \{1,2,3\} \Longrightarrow 1 \text{ or } 2 \text{ or } 3$$

Simulation example

Time	Cause	Masking	Time	Cause	Masking
0.712	1	{1}	2.149	2	{1, 2, 3}
1.629	1	{1, 2}	3.209	2	{2, 3}
0.063	1	{1}	1.606	2	{2, 3}
2.542	3	{1, 3}	5.276	1	{1, 3}
2.237	2	{2}	1.720	2	{1, 2}
2.054	1	{1}	0.598	1	{1}
4.028	3	{3}	0.033	1	{1, 2}

Simulation result

Causes	$\hat{\mu}^{(1)}$	$\hat{\sigma}^{(1)}$	$\hat{\mu}^{(2)}$	$\hat{\sigma}^{(2)}$	$\hat{\mu}^{(3)}$	$\hat{\sigma}^{(3)}$
Complete	1.549	2.000	1.469	0.839	1.691	0.602
Masked	1.575	2.026	1.508	0.886	1.622	0.531

```
> source("https://raw.githubusercontent.com/AppliedStat
+        /R-code/master/2006b/Rpp5.R")
>
> # lifetime observation
> X = c(1.9, 2.1, 3.2, 1.1, 2.1, 1.0, 2.0, 6.1, 3)
> # failure modes
> M = list(1, 1, 1, 2, 2, 3, 3, 0, c(1,2,3))
> weibull.cm.EM(X,M)   # Weibull Model
$alpha
[1] 2.085761 1.439960 1.354573

$lam
[1] 0.04159372 0.06104005 0.06704726

$iter
[1] 4

$conv
[1] TRUE
> #  We assume that there is partial masking in M.
> M = list(1, 1, 0, c(2,3), 2, 3, 3, c(1,2), c(1,2,3))
>
> weibull.cm.EM(X,M)   # Weibull Model
$alpha
[1] 2.324880 2.068188 1.277285

$lam
[1] 0.02657034 0.02271350 0.09832056

$iter
[1] 13

$conv
[1] TRUE
```

R language Usage

https://raw.githubusercontent.com/AppliedStat/R-code/master/2006b/Rpp5.R

C. Park, (2005) Parameter estimation of incomplete data in competing risks using the EM algorithm, IEEE Transactions on Reliability.