Journal of the Society of Korea Industrial and Systems Engineering
Vol. 32, No. 1, pp.11－19, March 2009.

수리 가능 발전기 시스템의 고장추세 분석을 위한 변화점 접근방법

홍민표 · 배석주†

한양대학교 산업공학과

Change-point Approach for Analyzing Failure Trend in Repairable Generating Systems

Min Pyo Hong · Suk Joo Bae†

Dept. of Industrial Engineering, Hanyang University, Seoul, Korea

A number of trend test methods, i.e., Military Handbook test and Laplace test etc., have been applied to investigate recurrent failures trend in repairable systems. Existing methods provide information about only existence of trend in the system. In this paper, we propose a new change-point test based on the Schwarz Information Criterion(SIC). The change-point approach is more informative than other trend test methods in that it provides the number of change-points and the location of change-points if it exists, as well as the existence of change-point for recurrent failures. The change-point test is applied to nine 300MW generating units operated in East China. We confirm that the change-point test has a potential for establishing optimal preventive maintenance policy by detecting change-point of failure rate.

Keywords : Change-Point, Repairable System, Trend Test, Poisson Process

1. 서 론

수리가능 시스템은 고장이 발생하였을 때, 고장이 난 부분의 수리나 시스템 전체의 교체를 통하여 시스템을 운용 가능하게 하는 것을 말한다[16]. 오늘날 대부분의 많은 시스템들은 복잡한 수리가능 시스템으로 이루어져 있어 분석하기 어렵다. 하지만 체계적인 분석방법을 도입하여 시스템 신뢰도를 정확하게 예측할 수 있다면 효율적인 정비정책(Maintenance policy)을 수립하는 데 도움이 될 수 있을 것이다.

수리가능 시스템에서의 주관심사는 고장패턴을 감지하는 것이며, 고장간격시간(Inter-failures time)이 어떤 패턴에 따라 변할 때 우리는 추세(trend)를 가진다고 한다. 만약 단조추세로서 고장간격시간이 점점 길어진다면 이는 신뢰성이 개선됨을 의미하고 반대로 고장간격시간이 점점 짧아진다면 이는 신뢰성이 악화됨을 의미한다. 한편으로는 고장간격 시간이 짧아졌다 길어지는 주기성을 보이거나, 고장간격 시간이 짧아지다 일정하다 길어지는 욕조곡선(Bath-tub curve) 모양의 추세 보이는 경우도 있다. 고장의 패턴을 분석할 때 실질적으로 쓰이는 방법으로는 도시적(Graphical)인 방법과 통계적 검정(Statistical test) 방법이 있다.

논문접수일 : 2008년 04월 11일　논문수정일(1차 : 2008년 07월 03일, 2차 : 2009년 01월 21일)　게재확정일 : 2009년 02월 24일
† 교신저자 sjbae@hanyang.ac.kr
※ 이 논문은 2007년도 정부재원(교육인적자원부 학술 연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2007-331-D00539).

도시적 방법은 관찰된 고장데이터를 그림으로 보여줌으로써 데이터에 대한 전반적인 추세를 파악할 수 있도록 해준다. 예를 들면 고장시간과 누적 고장수를 도시하는 방법은 간단하면서도 매우 유용한 정보를 제공하여 주는데 관찰시간 동안의 도시된 모양이 선형을 띄면 고장률이 일정하다는 것을 의미하고, 고장간격시간 데이터가 각각 독립이고 지수분포의 형태를 띠면 재생과정(Renewal process) 또는 동질포아송과정(Homogeneous Poisson Process)이 적합함을 의미한다. 동질포아송과정은 수리가능 시스템에서의 고장 발생을 설명할 수 있는 가장 단순한 통계적 모형이라 할 수 있다. 동질포아송과정인지 아닌지를 구분하는 간단한 방법으로는 고장발생시간마다의 누적고장률을 도시하는 방법과 Duane plot(로그 변환 한 고장시간과 로그 변환 한 누적 고장비율에 대한 도시)을 하는 방법이 있다. 두 방법 모두 도시된 점들이 선형성을 띠면 동질포아송과정이 적합한 모델이라고 판단 할 수 있다. 그러나 동질포아송과정으로는 수리가능 시스템의 전체 수명주기 중 초기고장기간과 신뢰성이 악화되는 마모고장기간을 설명할 수 없으므로 시스템의 전체 수명주기는 비동질포아송과정을 이용하는 것이 일반적이다. 통계적 추세검정들은 고장추세의 변화에 대한 통계적 유의성을 검정하며, 고장추세가 동질포아송과정인지 단조추세인지를 알려준다. 라플라스 검정(Laplace test)과 밀리터리 핸드북 검정(Military handbook test)은 이러한 추세검정방법들 중 가장 일반적으로 사용되어지는 방법이다.

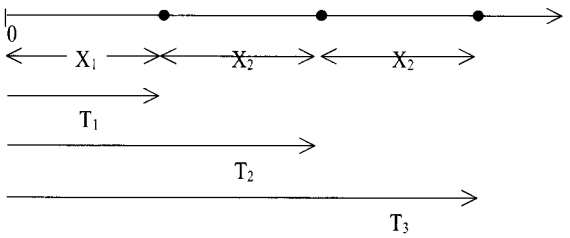
본 논문에서는 시스템의 고장패턴이 동질포아송과정(HPP)인지 비동질성포아송과정(NHPP)인지를 검정하는 새로운 방법으로 변화점검정(change-point test)을 제안 하고자 한다. 이 변화점검정은 고장간격시간들 사이에 분포의 이질성이 존재하는지에 대한 정보를 제공하여 준다. 이 방법이 기존의 다른 검정방법들과 다른 점은 고장간격시간들 사이에 이질성이 존재할 경우 시스템의 전체 데이터를 동일 분포를 가진 그룹끼리 나눌 수 있는 위치정보를 제공하여 준다는 점이다. 우리는 변화점이 존재하는 지를 판단하는 방법으로 SIC(Schwarz information criterion)를 사용한다. 이 이론은 SIC값의 최소화를 통해 변화점의 위치를 예측하는 것이다. 실례로서 변화점검정 방법을 중국 동부에서 가동 중인 증기터빈 발전설비 9곳의 고장간격시간 데이터를 적용하여, 라플라스 검정 및 밀리터리 핸드북 검정등과 그 결과를 비교하여 분석하였다.

2. 수리가능 시스템 모델

2.1 배경 이론 및 표기

재귀(recurrence) 데이터의 확률적 모형을 점과정(point-process)모델이라고 하며, 신뢰성에서의 시스템수리데이터는 재귀데이터로 볼 수 있다. 그 이유는 단일시스템의 수리과정에서 고장과 수리가 반복되는 시간 T_1, T_2, …, T_n을 재귀데이터로 볼 수 있기 때문이다. 일반적으로 수리시간은 정의된 어떤 특정 사건 이후 측정된 시스템의 나이 또는 시간으로 정의된다. 수리는 고정된 관찰시간구간 (t_0, t_1)에서 관찰되며 이때 $t_0 = 0$이다. 검정에 사용되는 데이터는 단일시스템 데이터를 사용하기도 하지만, 경우에 따라서는 여러 시스템 데이터를 결합하여 하나의 프로세스(process)형태로 사용하기도 하는데, 이러한 형태를 여러 점과정의 중첩(superposition of several point-processes) 또는 중첩된 점과정(superposed Point-process)이라고 한다.

본 논문에 사용되는 시스템의 고장시간정보를 <그림 1>에 나타내었다. 시스템의 수명시간 t는 0부터 시작하며, n번째 고장발생시간은 T_n으로 표기하며, $T_1, \cdots, T_n$은 서로 독립이며 시간구간$(0, t]$에서 발생한다. 고장간격시간은 $X_i = T_i - T_{i-1}$로 정의하며 시스템의 고장시간은 정수중단(미리 정한 시료수가 고장 날 때까지 시험) 또는 정시중단(미리 정한 시간까지 시험)방법을 통하여 얻는다.



<그림 1> 고장발생시간 T_i와 고장간격시간 X_i의 도시

3. 추세 분석

3.1 고장간격시간의 모수모형

점과정의 모수모형은 재귀데이터에 적합한 모델을 제공하며, 일반적으로 사용되는 모형으로는 동질포아송과정, 비동질포아송과정, 재생과정 등이 있다. 재귀데이터의 일반적인 예로는 단일 시스템의 고장발생시간을 들 수 있다.

3.1.1 동질포아송과정(HPP)

동질포아송과정은 일정한 고장률(intensity function)을 가지고 있는 포아송과정이다. 이 과정은 지수분포를 이용하여 다음과 같이 설명될 수 있다; 고장률이 λ인 동질포아송과정은 고장간격 시간들이 서로 독립이고 평균이 $1/\lambda$인 지수 분포를 따른다. 이 모델은 수리가능 시스템을 설명할 수 있는 가장 간단한 모델이지만 고장률이 일정하기 때문에 시스템의 신뢰성이 증가하는 부분이나 감소하는 부분을 설명할 수 없다는 제약이 있다. 따라서 이러한 경우 고장률이 일정하지 않은 비동질포아송과정을 사용한다.

3.1.2 비동질포아송과정(NHPP)

비동질포아송과정은 고장률 λ가 일정하지 않은 포아송과정을 말하며 다음과 같은 특징이 있다;

- $N(0)=0$
- 시간구간이 겹치지 않으면 사건의 수는 서로 독립이다.
- $P(N(t+h)-N(t)=1)=\lambda(t)+o(h),\ h\to 0$
- $P(N(t+h)-N(t)\geq 2)=o(h),\ h\to 0$

$N(t)$는 시간구간 $[0,\ t]$에서 발생된 사건의 수이며, $\lambda(t)$는 고장시스템에서의 고장발생비율(rate Of occurrence of failures : ROCOF)이며, 평균누적고장개수 $\Lambda(t)$는 다음과 같다.

$$\Lambda(t)=\int_0^t \lambda(t)dt \tag{1}$$

만약, 고장률이 일정하면($\lambda(t)\equiv\lambda$) 동질포아송과정으로서 어떠한 추세도 가지지 않음을 의미한다. 반면 고장률이 일정하지 않으면 비동질포아송과정이라하고, 일정한 추세를 가지고 있음을 나타낸다.

비동질포아송과정의 고장함수를 나타내는데 가장 자주 쓰이는 모형중 하나가 Power law process로서[2], 고장율 함수는 다음과 같다.

$$\lambda(t)=\frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}$$

Power law process는 시스템의 신뢰성이 개선되거나 악화되는 경우 모두를 나타낼 수 있으며, 흔히 와이블 과정(Weibull process)이라 하기도 한다. 위의 식에서 β와 θ는 각각 형상모수와 척도모수를 나타내며, 만약 $\beta>1$ 면 고장함수 $\lambda(t)$는 증가하고, $\beta<1$면 고장함수는 감소한다. 또한 $\beta=1$인 경우 Power law process는 간단히 고장률 $1/\theta$인 동질포아송과정으로 나타낼 수 있으며, 형상모수 β와 척도모수 θ는 각각 다음과 같이 추정 가능하다.

$$\hat{\beta}=\frac{n}{\sum_{i=1}^{n}\log(T/T_i)},\qquad \hat{\theta}=\frac{t}{n^{1/\hat{\beta}}}.$$

3.2 도시적 방법

3.2.1 TTT-Plot(Total Time on Test Plot)

수리가능 시스템의 TTT-Plot은 Barlow and Davis[5]에 의해 소개되었으며 비동질포아송과정을 기초로 한다. 고장률 $\lambda(t)$인 비동질포아송과정 존재시 전체 시간구간 $(a,b]$에서의 총 고장개수를 N이라 하고, 시스템의 k번째 고장시간을 T_k로 표시한다. $b=T$일 때, $0<T_1\leq\cdots\leq T_N\leq T$이 성립한다. $\lambda(u)$는 시간 u에서의 고장개수,

$$\Lambda(t)=\int_0^t \lambda(u)\,du \tag{2}$$

는 시간 0부터 t까지의 평균누적고장개수를 나타낸다. 비동질포아송과정의 TTT-plot은 비동질포아송과정의 누적고장개수 $k/N,\ k=1,\ \cdots,\ N$와 누적고장시간 $T_k/T,\ k=1,\ \cdots,\ N$을 도시한 그림이다. Nelson-Aalen plot은 TTT-plot과 좌표축이 바뀐 형태이다. Nelson-Aalen plot은 다음 장에 설명하기로 한다. TTT-plot의 결과 도시된 점들이 대각선에 위치하면 추세가 없는 동질포아송과정이라 판단할 수 있고, 대각선에 위치하지 않으면 추세가 있는 비동질포아송과정이라 판단할 수 있다. 다양한 TTT-plot 모양은 추세의 유형을 나타내며 <그림 2>에 나타나 있다. (a)는 고장률의 감소, (b)는 고장률의 증가, (c)는 욕조 모양의 고장률 변화를 나타낸다.

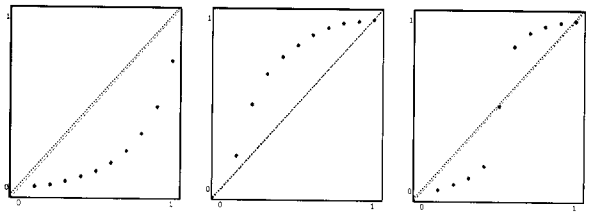
3.2.2 Nelson-Aalen Plot

누적 고장 함수 $\Lambda(t)=\int_0^t\lambda(u)du$에 대한 비모수 추정치는 다음과 같다.

$$\hat{\Lambda}(t)=\sum_{T_i\leq t}\frac{1}{N(T_i)} \tag{3}$$

여기서, $N(T_i)$는 시간 T_i이전까지 관찰된 고장의 수를 나타내고, $t<\min_i T_i$인 경우에 $\hat{\Lambda}(t)=0$이다. Nelson-Aalen plot은 시간 t와 $\hat{\Lambda}(t)$를 도시하여 도시된 점들이 직

선 가까이 위치하면 추세가 없다고 판단할 수 있고 그렇지 않으면 시스템이 추세를 가지고 있다고 판단 할 수 있다. 시스템이 하나일 경우에 Nelson-Aalen plot은 간단히 누적고장 개수와 관찰시간의 도시로 나타낼 수 있으며, 이는 하나의 수리가능시스템으로부터 관찰된 고장시간을 도시하는 일반적인 방법이다.



<그림 2> 비동질포아송과정의 대표적인 TTT plot 모양

3.3 통계적 추세검정

3.3.1 밀리터리 핸드북 검정(Military Handbook Test)

밀리터리 핸드북 검정은 시스템이 동질포아송과정이라는 귀무가설과 시스템이 단조 추세를 가지는 비동질포아송과정이다라는 대립가설을 검정한다. 시스템이 시간구간 $(a,b]$에서 단일 시스템일 경우의 검정통계량

$$M=2\sum_{j=1}^{\hat{n}}\ln\left(\frac{b-a}{T_j-a}\right) \tag{4}$$

- $T_j = j$번째 고장시간
- $(b-a)$ = 관찰된 시간구간
- $\hat{n}$ = 총 고장 개수(정시중단인 경우 n이며 정수중단인 경우 $n-1$)

은 귀무가설이 성립할 때 자유도 $2\hat{n}$인 χ^2분포를 따른다. 따라서 $M>\chi^2_\alpha(2\hat{n})$ 또는 $M<\chi^2_\alpha(2\hat{n})$면 귀무가설을 기각한다. 이때 $\chi^2_\alpha(2\hat{n})$은 자유도가 $2\hat{n}$인 χ^2분포의 α-임계값(critical point)을 의미한다. 시스템이 비동질포아송과정을 따른다는 귀무가설과 증가하는 추세의 시스템을 갖는 비동질포아송과정을 따른다는 대립가설을 검정하기 위해 밀리터리 핸드북 검정의 단측검정이 사용되는데, 이 검정은 대립가설이 power law를 따르는 증가하는 형태의 고장률함수를 가질 때 가장 적합하다. 이 검정방법은 U가 구간(0, 1)에서 일정하게 분포할 때, $-2\ln(U)$는 자유도 2인 χ^2분포를 따른다는 사실을 기초로 하고 있다. 따라서 귀무가설 하에서 $T_1, \cdots, T_{\hat{n}}$은 구간 (a,b)에서 일양분포를 따르는 순서통계량으로 식 (4)의 귀무가설하의 분포를 설명하고 있다. 만약, 단조증가추세를 가진다면, 검정통계량 M은 귀무가설과 비교하여 작아지게 될 것이다. 왜냐하면 이때의 고장시간 $T_1, \cdots, T_{\hat{n}}$은 구간 (a,b)에서의 일양분포를 따르는 순서통계량보다 커지기 때문이다. 반대로 감소하는 경향을 가진다면, M값은 귀무가설 하에서의 값보다 커지게 될 것이다. 프로세스가 하나 이상(m) 일 경우 식 (4)는 다음과 같이 일반화 된다.

$$M_C=2\sum_{i=1}^{m}\sum_{j=1}^{\hat{n}_i}\ln\left(\frac{b_i-a_i}{T_{ij}-a_i}\right) \tag{5}$$

이 값은 시스템이 동질포아송과정을 따른다는 귀무가설 하에서 자유도 $2q$인 χ^2분포를 따르며 $q=\sum_{i=1}^{m}\hat{n}_i$이다. 우리는 이것을 결합된 밀리터리 핸드북 검정(combined Military Handbook test)이라 부른다. 동일한 발생률을 가지는 독립인 동질포아송과정을 귀무가설로 하면 검정통계량

$$M_T=2\sum_{k=1}^{\hat{N}}\ln\left(\frac{T(S)}{T(S_k)}\right) \tag{6}$$

은 자유도 $2\hat{n}$인 χ^2분포를 따른다. 그리고 우리는 이것을 TTT를 기초로 한 밀리터리 핸드북 검정(TTT based Military Handbook Test)이라고 한다.

3.3.2 라플라스 검정(Laplace Test)

라플라스 검정은 동질포아송과정을 귀무가설, 단조 고장함수를 가지는 비동질포아송과정을 대립가설로하는 검정을 하기위한 방법이다. 검정통계량은 다음과 같다.

$$L=\frac{\sum_{j=1}^{\hat{n}}T_j-\frac{1}{2}\hat{n}(b+a)}{\sqrt{\frac{1}{12}\hat{n}(b-a)^2}} \tag{7}$$

만약, 귀무가설이 참이라면, 검정통계량은 점근적으로 표준정규분포가 된다. 따라서 시스템이 비동질포아송과정이라는 귀무가설은 $|L|>z_{\alpha/2}$이면 기각된다. 이때 $z_{\alpha/2}$는 표준정규분포에 대한 임계값(critical point)이다. 검정통계량 L값으로부터 추세형태를 알 수 있는데, 만약 $L<0$이면 고장률은 감소하는 추세이고, $L>0$이면 증가하는 추세이다. 하나이상의 프로세스를 검정하기 위한 라플라스 검정을 일반화하여, m개의 독립인 프로세스를 가지고 있다면, 식 (7)은 다음과 같이 일반화될 수 있다.

$$L_C=\frac{\sum_{i=1}^{m}\sum_{j=1}^{\hat{n}_i}T_{ij}-\sum_{i=1}^{m}\frac{1}{2}\hat{n}_i(b_i+a_i)}{\sqrt{\frac{1}{12}\sum_{i=1}^{m}\hat{n}_i(b_i-a_i)^2}} \quad (8)$$

이 통계량을 바탕으로 한 검정은 귀무가설이 동질포아송과정(프로세스 각각의 고장률이 다를 수 있음)이고 대립가설이 고장률 $\lambda(t)=e^{\alpha_i+\beta t}$($\beta$는 전체 프로세스의 공통 값, α_i는 각 프로세스의 특정 값을 나타냄)을 갖는 비동질포아송과정일 때 최적검정이 된다. 우리는 이 검정을 결합된 라플라스 검정(combined Laplace test)이라고 한다.

4. 변화점 검정(Change-Point test)

고장간 시간을 나타내는 x_1, x_2, …, x_n를 확률분포함수 F_1, F_2, … F_n을 가지는 독립확률변수의 순서라고 할때, 일반적으로 다중변화점(Multiple change-point)문제는 다음과 같은 귀무가설을 검정한다.

$$\begin{aligned}H_0 &: F_1=F_2=\cdots=F_n\\ H_1 &: F_1=\cdots=F_{k_1}\neq F_{k_1+1}=\cdots\\ &\quad =F_{k_2}\neq\cdots\neq F_{k_R+1}=\cdots=F_n\end{aligned} \quad (9)$$

이때 $1<k_1<k_2<\cdots<k_R<n$이고 R은 미지의 변화점(change-point)개수이고, k_1, k_2, …, k_R은 미지의 변화점위치를 각각 나타낸다. 이 값들은 모두 추정이 가능한데, 만약 분포 F_1, F_2, …, F_n이 일반적인 모수족(parameter family) $F(\theta)$에 속할 때($\Theta\in R^P$), 변화점문제는 모집단의 모수 θ_i, $i=1$, …, n에 대한 다음 귀무가설을 검정한다.

$$\begin{aligned}H_0 &: \theta_1=\theta_2=\cdots=\theta_n=\theta(\text{미지의 모수})\\ H_1 &: \theta_1=\cdots=\theta_{k_1}\neq\theta_{k_1+1}=\cdots\\ &\quad =\theta_{k_2}\neq\cdots\neq\theta_{k_R+1}=\theta_n\end{aligned} \quad (10)$$

이러한 가설은 프로세스에서의 변화점 존재여부, 변화점의 개수 및 변화점의 위치등의 변화점 추정 양상들을 함께 나타낼 수 있게 된다.

4.1 변화점을 이용한 접근

Maguire et al.[13]는 연구에서 고장의 재귀비율이 변하는 시간을 검정하기 위해서는 고장의 수를 세는 것보다 고장시간간격을 사용하는 것이 유리하다는 것을 설명하였다. 따라서 우리는 수리가능시스템의 고장과정을 고장간격시간을 이용하여 표현할 것이다. 고장간격시간 X_1, …, X_n을 모수가 각각 λ_1, …, λ_n인 독립 지수분포의 확률변수라 할 때 밀도함수는 $f(x_i;\lambda_i)=\lambda_i e^{-\lambda_i x_i}$, $i=1$, …, n와 같다. 만약 미지의 R개의 변화점 k_1, …, k_R이 존재한다면, $1\le k_1<\cdots<k_R<n$과 $\lambda_{k_r}\neq\lambda_{k_r+1}$, $r=1$, …, R을 만족한다. 만약 $R=0$이면 변화점이 없음을 뜻한다. 변화점의 개수는 모델의 차원으로 고려되며 그 수는 베이지안을 바탕으로 한 SIC(Schwarz Information Criteria)를 통하여 추정 가능하다. 여기서 주목할 점은 SIC는 지수족(exponential family)으로부터 독립이라는 것이다. 변화점의 개수를 R이라 가정하면, 추정하여야 하는 모수의 수는 $(2R+1)$개 λ_{k_1}, …, λ_{k_R+1}과 변화점 k_1, …, k_R이다. 수리시스템의 고장시간에 대한 추세는 변화점에 의해 결정된다. $\lambda_{k_1}>\lambda_{k_2}>\cdots>\lambda_{k_R+1}$은 시스템이 개선되는 경우이고, 반대로 $\lambda_{k_1}<\lambda_{k_2}<\cdots<\lambda_{k_R+1}$은 악화되는 경우이다.

만약, 변화점개수 R과 변화점위치 k_1, …, k_R이 주어진다면 귀무가설에서의 최대 로그우도는

$$\mathcal{L}_1(k_1,\cdots,k_R)=\sum_{j=0}^{R}\sum_{i=k_j+1}^{k_{j+1}}\left\{\log\hat{\lambda}_{k_{j+1}}-x_i\hat{\lambda}_{k_{j+1}}\right\} \quad (11)$$

와 같다. 그러나 변화점의 개수와 위치를 모르는 경우에는 변화점의 개수가 먼저 결정되어져야 한다. 변화점의 개수 R과 위치 k_1, …, k_R을 추정하기 위해서 우리는 SIC를 이용하며 이때 SIC는 다음과 같다.

$$SIC(R)=-2\log L(\hat{\Theta}_R)+R\log n \quad (12)$$

$L(\hat{\Theta}_R)$은 R차원의 모수공간 Θ_R에서 R개의 모수를 가진 모델의 최대우도함수이다. 명확히 SIC는 AIC(Akaike information criterion)의 수정된 형태로 나타난다.

4.2 변화점 검정과정

단일 수리가능 시스템의 귀무가설(동질포아송과정)에 대한 검정에서 귀무가설은 $\lambda_1=\cdots=\lambda_n\equiv\lambda$이고, 대립가설은 R개의 변화점 k_1, …, k_R이 존재한다는 것이다. 동질포아송과정에 대한 검증 문제는 다음과 같이 나타낼 수 있다.

$$H_0 : \lambda_1 = \cdots = \lambda_n \equiv \lambda$$
$$H_1 : \lambda_1 = \cdots = \lambda_{k_1} \neq \lambda_{k_1+1} = \cdots = \lambda_{k_R} \neq \lambda_{k_R+1} = \cdots = \lambda_n \quad (15)$$

대립가설 하에서의 SIC값과 귀무가설 하에서의 SIC 값을 비교하면 가설에 대한 검정을 할 수 있다(식 (15)). 귀무가설에 대한 SIC를 $SIC(n)$이라 하고, 그 때

$$\min SIC \equiv \min\{SIC(n),\ SIC(R)\} \quad (16)$$

값이 $SIC(n)$이라면 귀무가설을 채택하고, 아니라면 귀무가설을 기각하여 R개의 변화점 $\hat{k}_1, \cdots, \hat{k}_R$이 존재한다고 결론 내릴 수 있다. 감마확률변수가 서로 독립이고 동일한 분포인 경우의 SIC방법은 Hsu[10]가 연속한 주식가격과 항공교통흐름에서의 변화점을 연구하여 정립하였다.

변화점의 개수와 변화점의 위치는 보통 알려져 있지 않은 경우가 대부분이며, 따라서 검정절차를 수행하기 위해서는 R과 $k_1, \cdots, k_R$을 결정해야만 한다. 정보영역의 원칙에 따라 변화점 $k_1, \cdots, k_R$은 R을 고정시키고 SIC가 최소화되는 그 때의 값을 $\hat{k}_1, \cdots, \hat{k}_R$로 결정하게 된다. 모형의 차원으로서의 변화점의 개수 R은 변화점들을 추정하기 위하여 사전에 결정되어야만 한다. 변화점의 개수를 결정하는 방법은 다음과 같은 2단계로 표현될 수 있다.

1단계 :

고정된 R에 대하여 (11)을 최대화하는 $k_1, \cdots, k_R$ 값을 미지의 변화점에 대한 추정치 $\hat{k}_1, \cdots, \hat{k}_R$로 결정한다. 즉

$$M(R) = \max_{0<k_1<\cdots<k_R<n} \mathcal{L}_1(k_1, \cdots, k_R) \equiv \mathcal{L}_1(\hat{k}_1, \cdots, \hat{k}_R) \quad (17)$$

2단계 :

Schwarz criterion을 이용하여 $R \leq R_U$의 조건하에 다음 식을 최소화하는 그때의 R의 값을 $\hat{R}$으로 결정한다.

$$SIC(R)' = -2M(R) + (2R+1)\log n = 2\sum_{j=0}^{R}\sum_{i=\hat{k}_j+1}^{\hat{k}_{j+1}} \left\{x_i \hat{\lambda}_{k_{j+1}} - \log \hat{\lambda}_{k_{j+1}}\right\} + (2R+1)\log n \quad (18)$$

여기서 R_U는 R의 상계(upper bound)를 뜻한다. 외관상으로 SIC는 AIC(Akaike Information Criterion)에서 페널티 값만 변형한 형태이다. AIC의 경우 페널티 값을 $(2R+1)\log n$대신 $2(2R+1)$을 사용한다. 그러나 SIC는 참 모델에 대한 차수의 일관된 추정을 가능하게 해준다는 것이 알려져 있다[15]. 만약 변화점의 개수와 변화점의 위치가 알려져 있지 않다면 식 (12)의 $SIC(R)$ 대신 식 (18)의 $SIC(R)'$이 변화점의 존재를 평가하기 위하여 $SIC(n)$과 비교될 수 있다.

4.3 하나의 변화점이 있는 경우

오직 하나의 변화점이 k에 있는 경우를 고려해 보자. 식 (15)의 검정문제는 다음과 같이 표현될 수 있다 :

$$H_0 : \lambda_1 = \cdots = \lambda_n \equiv \lambda$$
$$H_1 : \lambda_1 = \cdots = \lambda_k \neq \lambda_{k_1+1} = \cdots = \lambda_n, \quad (19)$$

H_1하에 최대 로그-우도를 $\mathcal{L}_1(k)$라고 하면, 최대우도 추정치는

$$\hat{\lambda}_1 = \cdots = \hat{\lambda}_k = k/\sum_{i=1}^{k} x, \quad \widehat{\lambda_{k+1}} = \cdots = \hat{\lambda}_n = (n-k)/\sum_{i=k+1}^{n} x_i$$

이 되며,

$$\mathcal{L}_1(k) = \left(\frac{k}{\sum_{i=1}^{k} x_i}\right)^k \left(\frac{n-k}{\sum_{i=k+1}^{n} x_i}\right)^{n-k} e^n$$

이 된다. 또한 H_1하에 SIC는

$$SIC(k) = -2\mathcal{L}_1(k) + 2\log n = 2k\log\left(\sum_{i=1}^{k} x_i\right) + 2(n-k)\log\left(\sum_{i=k+1}^{n} x_i\right) - 2k\log k - 2(n-k)\log(n-k) + 2(n+\log n)$$

특정값 k에 대하여 만약 $SIC(n) > SIC(k)$이면 대립가설 H_1이 채택되며, 변화점 k의 위치는

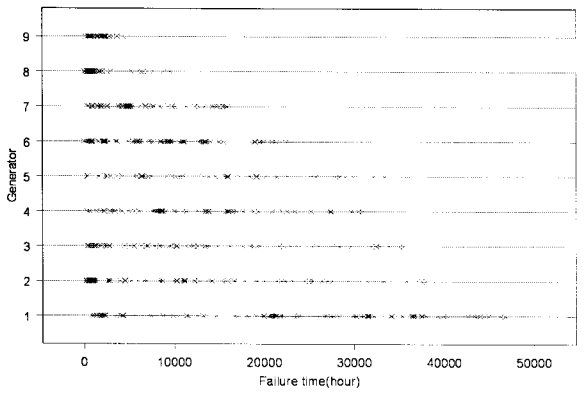
$$SIC(\hat{k}) = \min_{1 \leq k < n} SIC(k)$$

을 만족하는 그 때의 값 $\hat{k}$에 의해 추정된다.

5. 실제 데이터의 적용

이 장에서는 기존의 추세검정과 변화점검정의 비교를 위해 9개의 증기터빈 발전설비(Steam-turbine generating units)의 실제 고장시간데이터를 사용하였다. 증기터빈 발생설비는 대체로 복잡한 구조를 가지고 있으며 수리가능한 시스템이다. 각각의 증기터빈 발전장치는 발전기(Generator), 증기터빈(Steam turbine), 보일러(Boiler)등으로 구성되어 있다. 화석연료를 사용하는 증기터빈의 작동원리는 보일러에서 연료를 산화하여 증기가 발생시키면, 발생된 증기는 터빈의 날개(Blade)를 회전시키고,

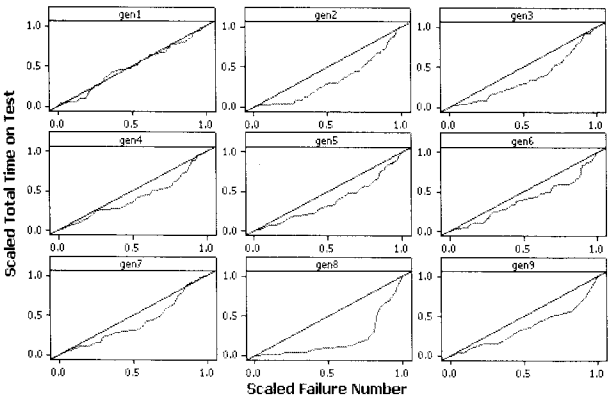
이 회전으로 인해 발전기의 축(Shaft)이 회전하면 전기가 발생한다. 화석연료를 사용하는 증기터빈 발전설비의 규모는 1메가와트(MW)부터 1000메가와트 이상이며, 본 논문에서 사용한 데이터는 중국 동부 여러 곳의 전기 발전소에서 사용되고 있는 $300MW$ 증기터빈 발전설비 9개로 측정한 데이터이다. 고장 데이터에 대한 그림은 <그림 3>에 나타나 있다. Interval plot(시스템의 고장시간들을 일직선상에 도시함)의 결과 고장간격시간들이 일정하지 않음을 알 수 있으며 이는 고장간격시간들이 서로 독립이 아닐 수도 있다는 분석을 가능하게 한다. 증기터빈 발생설비들에 사용되는 발전기, 증기터빈, 보일러는 각각 동일한 회사에서 생산되었기 때문에, 이 설비들의 시스템은 동일하다고 본다.



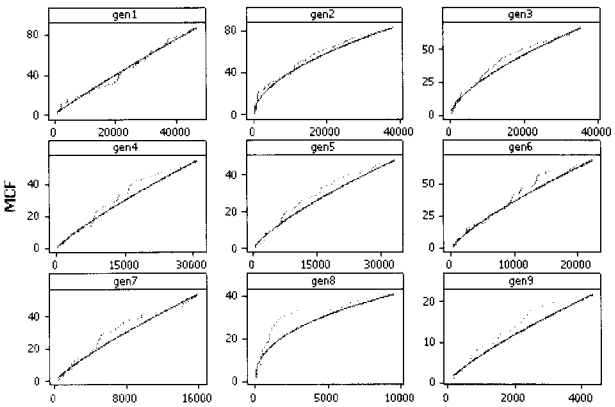
<그림 3> 발전설비 고장데이터의 Interval plot

5.1 도시적방법 적용

데이터를 도시화하는 방법으로 전체적인 추세를 파악할 수 있다. 본 논문에서는 TTT-plot, Nelson-Aalen plot을 실시하였으며 각각의 결과는 <그림 4>, <그림 5>이다 (gen$n$은 n번 발전설비를 의미함). <그림 4>와 <그림 5>를 살펴보면 1번 발전설비는 TTT-plot과 Nelson- Aalen plot에 대각선상에 위치한 조건으로 보아 동질포아송과정을 가정할 수 있다. 하지만 나머지 발전설비는 모두 도시된 점들이 대각으로부터 거리를 두고 있으므로 어떠한 추세를 가지고 있을 가능성이 높다고 판단 할 수 있다. 하지만 가능성이 높다고 판단할 수 있을 뿐이지 정확한 수치를 근거로 제시할 수는 없기 때문에 다음 장에 통계적 분석방법을 통해 추세를 가지고 있는지 검정하고자 한다.



<그림 4> 발전설비 고장데이터의 TTT plot



<그림 5> 발전설비 고장데이터의 Nelson-Aalen plot

5.2 통계적 추세분석방법 적용

앞 제 5.1절의 분석을 좀 더 설득력 있게 하기위해서는 통계적인 검정이 필요하다. 따라서 이 논문에서는 발전기의 고장데이터에 대한 밀리터리 핸드북 검정, 라플라스 검정을 실시하였으며, 이 두 검정을 통해 얻은 결과는 각각 <표 1>, <표 2>에 나타내었다. 밀리터리 핸드

<표 1> 발전설비 고장데이터의 Military Handbook 검정 결과

발전기	Test statistic	p-value	Result ($\alpha = 0.05$)
1	188.54	0.368	HPP
2	**305.70**	**0.0001**	**NHPP**
3	**205.64**	**0.0001**	**NHPP**
4	131.53	0.123	HPP
5	**127.16**	**0.026**	**NHPP**
6	**174.82**	**0.020**	**NHPP**
7	130.66	0.105	HPP
8	**190.11**	**0.0001**	**NHPP**
9	55.75	0.152	HPP

북 검정의 결과인 <표 1>을 살펴보면 $\alpha=0.05$로, 2번, 3번, 5번, 6번 및 8번 발전설비의 고장데이터는 비동질포아송과정을 따르며, 나머지 발전설비는 동질포아송과정을 따른다는 것을 알 수 있고, 라플라스 검정의 결과인 <표 2>를 살펴보면 1번 발전설비를 제외한 모든 발전설비의 고장데이터는 비동질포아송과정을 따름을 알 수 있다. 하지만, 동일한 발전설비에 대해 밀리터리 핸드북 검정과 라플라스 검정이 완전히 동일한 결과를 보이지는 않고 있다. 두 검정의 결과 6개의 발전설비(1번, 2번, 3번, 5번, 6번, 8번)에 대해서는 일치한 결과를 보이지만, 3개의 발전설비(4번, 7번, 9번)에 대해서는 불일치한 결과를 보이고 있다. 이는 두 검정방법이 일관된 검정결과를 제시하지 못하고 있음을 의미한다.

<표 2> 발전설비 고장데이터의 Laplace 검정 결과

발전기	Test statistic	p-value	Result ($\alpha=0.05$)
1	-0.57	0.572	HPP
2	**-5.72**	**0.0001**	**NHPP**
3	**-4.11**	**0.0001**	**NHPP**
4	**-2.35**	**0.019**	**NHPP**
5	**-2.92**	**0.003**	**NHPP**
6	**-2.90**	**0.004**	**NHPP**
7	**-2.32**	**0.020**	**NHPP**
8	**-6.44**	**0.0001**	**NHPP**
9	**-2.28**	**0.022**	**NHPP**

<표 3> 발전설비 고장데이터의 변화점 검정 결과

발전기	변화점(시간)		변화점으로 나뉜 구간의 λ값		
	첫 번째	두 번째			
1	·	·	·	·	·
2	**1279.48**	**17834.77**	**0.0172**	**0.0021**	**0.0006**
3	13897.7	·	0.0032	0.0011	·
4	16897.02	·	0.0024	0.0005	·
5	11135.45	·	0.0023	0.0010	·
6	13712.49	13493.65	0.0039	0.0183	0.0014
7	4586.69	·	0.0044	0.0030	·
8	1480.15	·	0.0182	0.0016	·
9	2426.9	·	0.0074	0.0021	·

5.3 변화점(Change-point) 검정

앞 제 4.2절에서 설명하였듯이 $SIC(n)$보다 작은 $SIC(R)$ 값이 존재하면, 변화점이 존재하는 것이고, 이때, $SIC(R)$을 최소화하는 점 k_i가 변화점이 존재하는 점이다. 검정결과를 <표 3>에 정리하였으며, 변화점위치(분포가 달라지는 고장시간의 위치)와 변화점을 기준으로 나뉜 구간의 고장률(λ)을 나타내었다. 실험 결과 1번 발전설비를 제외한 모든 발전설비에서 변화점이 존재하는 것을 알 수 있었다. 2번 발전설비(전체 고장시간은 37956.11시간) 검정결과를 예를 들어 설명하면 2번 발전설비의 고장데이터에 대한 변화점은 2개가 존재하며, 그 시간은 각각 1279.48와 17834.77에 해당한다. 이 변화점을 기준으로 구간을 (0, 1279.48), (1279.48, 1784.77), (17834.77, 37956.11)으로 나눌 수 있으며, 이 세구간은 각각 고장률(λ)이 0.0172, 0.0021, 0.0006인 지수분포를 따른다고 할 수 있다. 변화점으로 나뉜 구간의 고장률은 점점 작아지므로 발전시스템이 개선되고 있음을 알 수 있다. 이는 도시적방법 결과와 일치하는데, TTT-plot에서 고장데이터들이 대각선 라인 밑에 있음을 이전에 확인할 수 있었다. 6번 발전설비의 경우는 고장률이 증가하였다 감소하는 욕조곡선 형태임을 알 수 있다. 나머지 발전설비에 대해서도 같은 방법으로 해석가능하다.

5.4 Power Law Process 적용

Power law process를 통한 모수추정결과와 신뢰구간은 <표 4>과 같다. 모든 발전설비의 형상모수추정치($\hat{\beta}$)가 1보다 작은 값을 가지므로 시간에 따라 고장률이 점점 감소하는 개선시스템임을 알 수 있다. 하지만 1번 발전설비의

<표 4> 발전설비 고장데이터의 모수추정치와 신뢰구간

발전기	$\hat{\beta}$	95% 신뢰구간		$\hat{\theta}$	95% 신뢰구간	
1	0.9229	0.7297	1.1149	370.19	103.12	847.12
2	0.5430	0.4267	0.6587	11.096	2.9909	25.627
3	0.6516	0.4965	0.8058	55.897	13.567	135.02
4	0.8363	0.6169	1.0542	256.48	54.571	649.32
5	0.7549	0.5431	0.9650	197.83	38.34	58.357
6	0.6871	0.5258	0.8473	47.219	11.688	112.96
7	0.8265	0.6077	1.0438	127.93	26.875	324.69
8	0.4208	0.2918	0.5487	1.4851	0.2517	3.808
9	0.7893	0.4664	1.1082	86.056	8.7813	239.05

경우 추정된 값이 1에 가까운 0.92287이고 95% 신뢰구간이 1을 포함하므로 동질포아송과정이라 판단된다. 4번, 7번, 9번 발전설비의 경우 각각의 신뢰구간이 1을 포함하지만 추정된 값이 1과는 많은 차이를 보이며, 95% 신뢰구간의 상한(upper bound)이 1과 근사한 값을 가지므로 비동질포아송과정이라 결론 내릴 수 있다. 따라서 1번 발전설비를 제외한 모든 발전설비는 시간에 따라 고장률이 감소하는 비동질포아송과정이라 결론 내릴 수 있으며, 이는 앞 제 5.3절의 변화점 검정의 결과가 타당함을 보여준다.

6. 결 론

본 논문에서 우리는 실제 사용되고 있는 발전설비의 고장 데이터를 이용하여 시스템의 추세검정을 실시해 보았다. 기존의 검정방법 중 도시적인 방법으로는 Nelson-Aalen plot, TTT-plot을 실시하였으며, 통계적 검정방법으로는 라플라스 검정, 밀리터리 핸드북 검정을 실시하였다. 도시적 방법은 도시를 통한 비교적 손쉬운 방법으로 시스템의 추세를 분석 할 수는 있지만, 수치화된 정확한 정보를 제공하지는 못하며, 기존의 통계적 추세방법인 라플라스 검정 및 밀리터리 핸드북 검정은 수치화된 정보를 제공하지만 서로 간에 불일치한 결과가 존재하여 분석자로 하여금 혼란을 야기할 수 있다. 또한 추세에 대한 단편적인 정보(동질포아송과정인지 비동질포아송과정인지)만을 제공하는 단점을 가지고 있다. 하지만 새로운 추세검정 방법으로 제시하는 변화점검청 결과는 기본적으로 기존의 검정방법을 통한 정보(변화점이 존재한다는 것은 시스템이 비동질 포아송 과정임을 뜻한다) 외에 변화점의 유무, 변화점의 위치정보, 시스템의 추세정보를 제공한다. 변화점의 위치정보가 주어지면 데이터 분할을 통한 정확한 모형화를 가능하게 하며, 정확한 모형정보는 예방보전(preventive maintenance)을 위한 기초자료로 충분히 활용할 수 있을 것으로 기대한다.

참고문헌

[1] Akaike, H.; Information Theory and an Extension of the Maximum Likelihood Principle, *2nd International Symposium of Information Theory*(Petrov, B. N. and Csaki, E., eds.), Akademiai Kiado, Budapest, 267-281, 1973.

[2] Ascher, H. and Feingold, H.; *Repairable systems Reliability : Modeling, Inference, Misconceptions and Their Causes*, Marcel Dekker, New York, 75, 1984.

[3] Bain, L. J., Engelhardt, M., and Wright, F. T.; Tests for an Increasing Trend in the Intensity of a Poisson Process : A Power Study, *Journal of the American Statistical Association*, 80 : 419-422, 1985.

[4] Barlow, R. E., Bartholomew, D. J., Bremner, J. M., and Brunk, H. D.; *Statistical Inferences Under Order Restrictions*, Wiley, New York, 1972.

[5] Barlow, R. E. and Davis, B.; Analysis of time between failures for repairable components. *Nuclear Systems Reliability Engineering and Risk Assessment*(eds J. B. Fussell and G. R. Burdick), SIAM, Philadelphia, 543-561, 1977.

[6] Boswell, M. T.; Estimating and Testing Trend in a Stochastic Process of Poisson Type, *Annals of Mathematical Statistics*, 37 : 1564-1573, 1966.

[7] Bo Henry Lindqvist; on the statistical modeling and analysis of repairable systems, *Statistical Science*, 21 : 532-551, 2006.

[8] Cox, D. R.; Some Statistical Methods Connected With Series of Events, *Journal of the Royal Statistical Society*, Ser. B, 17 : 129-164, 1955.

[9] Duane, J. T.; Learning Curve Approach to Reliability Monitoring, *IEEE Transactions on Aerospace*, 2 : 563-566, 1964.

[10] Hsu, D. A.; Detecting Shifts of Parameter in Gamma Sequences with Applications to Stock Price and Air Traffic Flow Analysis, *Journal of the American Statistical Association*, 74 : 31-40, 1979.

[11] Jie Chen, A. K. Gupta; *Parametric Statistical Change Point Analysis*, Birkhäuser, Boston, 2000.

[12] Kvaløy, J. T. and Lindqvist, B. H.; TTT-Based Tests for Trend in Repairable Systems Data, *Reliability Engineering and System Safety*, 60 : 13-28, 1998.

[13] Maguire, B. A., Pearson, E. S., and Wynn, A. H. A.; The Time Intervals Between Industrial Accidents, *Biometrika*, 39 : 168-180, 1952.

[14] MIL-HDBK-18; *Reliability Growth Management*, Headquaters, U.S Army Communications Research and Development Command, ATTN : DRDCOPT, Fort Monmouth, NJ 07702, 1981.

[15] Schwarz, G.; Estimating the Dimension of a Model, *The Annals of Statistics*, 6 : 461-464, 1978.

[16] Steven E. Rigdon, Asit P. Basu; *Statistical Methods for the Reliability of Repairable Systems*, John Wiley and Sons, New York, 2000.

[17] William Q. Meeker, Luis A. Escobar; *Statistical Methods for Reliability Data*, John Wiley and Sons, New York, 1998.