

고장시간 자료의 통계적 분석을 위한
PWF-GPH 방법
(PWF-GPH method for the statistical analysis of
failure time data)

김선영, 윤복식*

Abstract

In this paper, a life distribution fitting method based on generalized phase-type distributions(GPH) is presented. By fitting the life distribution to a GPH, we can utilize various useful properties of the GPH. Two different approaches are used according to the properties of the given failure time data. One is an approximation to a GPH through the piecewise Weibull failure rate(PWF) model and the other is a direct approximation to a GPH using the empirical distribution function. Two numerical examples are also presented. In the first example, both of the two approaches are utilized and compared for an incomplete data set. And in the second example, the direct approximation method from an empirical distribution is utilized for the analysis of a complete data set. In both cases, we could confirm the validity of the proposed method.

* 홍익대학교

1. 서론

어떠한 부품의 고장시간 자료를 분석하여 수명 분포(고장시간 분포)를 추정하는 것은 신뢰성 및 보수정책의 분석의 출발점이다. 그러나 고장시간 자료의 본래적인 불완전성이나 고장률의 육조곡선 형태의 변화성 때문에 일반적인 상황하에서의 수명 분포의 추정은 간단한 문제가 아니다. 본 논문에서는 양의 구간에서 정의된 임의의 분포를 근사화 할 수 있는 GPH 분포를 이용하여 수명분포의 추정을 자동화할 수 있는 방법을 제안하고자 한다. GPH 분포는 그 형태의 일반성과 더불어 확률적으로 유용한 성질들을 가지고 있어[6] 응용가능성이 높는데 특히 부품의 고장시간 분포들을 얻은 후 이들을 이용하여 전체 시스템의 고장시간 분포를 분석하는 데 편리할 수 있다[2-2절 참조].

본 논문에서는 GPH분포로 근사화시키는 과정에서 자료의 특성에 따라 두 가지 다른 접근 방법을 사용한다. 불완전 자료(incomplete data)의 경우에는 PWF모형[11]을 적용하여 구간별 Weibull분포의 형태로 고장시간 분포를 우선 설정하고 이것으로부터 GPH 근사화를 시도하고 완전 자료(complete data)의 경우에는 직접 경험적 분포(empirical distribution)로부터 GPH 근사화를 시도한다.

본 서론에 이어서 2장에서는 GPH분포의 정의 및 성질을 소개하고 3장에서는 GPH분포로의 근사화 방법을 분포함수가 알려진 경우와 경험적 자료만 주어진 경우, 두 가지로 구분하여 설명하며, 경험적 자료의 경우, 다시 완전자료(주어진 자료가 모두 고장시간 자료)인 경우와 불완전자료(고장시간 자료와 검열에 의한 제거의 시간자료로 구성된 자료)인 경우로 나누어 근사화 방법을 설명하고 3-3절에서는

GPH분포로의 근사화에 있어 효과적인 계산 방법을 설명한다. 4장에서는 PWF모형의 정의와 의미를 알아보고, 5장에서는 누적고장률의 추정방법 및 PWF모형에서의 구간변화점 및 규모모수와 형태모수의 추정방법을 설명한다. 6장에서는 분포함수를 구한 뒤 GPH분포로 근사화하는 적용예제를 불완전 자료인 경우와 완전자료인 경우로 구분하여 알아본다. 7장에서는 결론적으로 GPH분포의 근사화의 유용성 및 본 논문의 미비점을 알아보고 앞으로의 연구 과제를 기술한다.

2. GPH분포의 특성

2.1 GPH분포의 정의

$(g(n))_0^\infty$ 을 이산확률변수 L 의 분포함수라 하고 $(E_n)_0^\infty$ 을 평균이 $\frac{1}{\lambda}$ 인 i.i.d.(independent & identically distributed) 지수확률 변수열이라 할 때 (단, $E_0 = 0$ w.p.1)

$$X = \sum_{n=0}^L E_n$$

를 GPH(a generalized phase-type)확률변수라 한다.[6] 이 때 X 의 분포함수는 다음과 같다.

$$F_X(x) = \sum_{n=0}^{\infty} G(n) \frac{e^{-\lambda x} (\lambda x)^n}{n!}$$
$$G(n) = \sum_{k=0}^n g(k)$$
$$x \geq 0, n = 0, 1, 2, \dots$$

(L 의 누적 확률분포)

위에서 주어진 확률분포함수 F_X 를 generalized phase-type distribution (GPH)라 하며, GPH

(λ, g) 로 나타낸다.

2.2 GPH분포의 성질

GPH분포는 다음과 같은 성질들을 가지고 있다 [6].

1. GPH류는 유한 개의 convolution들에 대해 닫혀 있다.
2. X_1 과 X_2 를 각각 두 개의 독립적인 $GPH(\lambda_1, g_1)$, $GPH(\lambda_2, g_2)$ 분포를 따르는 확률변수들이라 하면 $X = X_1 - X_2$ 는 (λ, g) 의 BPH(Bilateral phase-type) 확률변수가 된다. 여기서

$$g(n) = \begin{cases} \sum_{k=0}^{\infty} g_1(k+n)g_2^*(k) & n = 1, 2, 3, \dots \\ g_1(0)g_2(0) & n = 0 \\ \sum_{k=0}^{\infty} g_2(k-n)g_1^*(k) & \end{cases}$$

이고 $i = 1, 2$ 에 대해

$$\begin{aligned} g_i^*(n) &= \sum_{k=0}^{\infty} \binom{n+k}{k} \left(\frac{1}{2}\right)^{n+k+1} g_i(k+1) \\ g_i^*(0) &= \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{k+1} g_i(k+1) + g_i(0) \end{aligned} \quad (2-1)$$

이 된다.

3. X_1 과 X_2 를 각각

$GPH(\lambda_1, g_1)$, $GPH(\lambda_2, g_2)$ 분포를 따르는

두 개의 독립적인 확률변수들이라 하면

$X = \max\{X_1 - X_2, 0\}$ 는 $GPH(\lambda, g)$ 분

포를 따른다. 여기서

$$g(n) = \begin{cases} \sum_{k=0}^{\infty} g_1(k) \sum_{r=k}^{\infty} g_2^*(r) & n = 0 \\ \sum_{k=0}^{\infty} g_1(k+n)g_2^*(k) & n = 1, 2, \dots \end{cases}$$

이고 g_2^* 는 (2-1)과 같이 정의된다.

4. GPH류는 mixture에 대해 닫혀 있다.
5. GPH류는 coherent system의 형성에 대해 닫혀 있다. 다시 말해서 component들이 GPH이면 coherent system도 GPH이다.

3. GPH분포의 추정방법

3.1 분포함수가 알려진 경우의 추정방법

분포함수 F 가 $[0, \infty]$ 에서 정의된 유계함수이면

$$F_\lambda(t) = \sum_{n=0}^{\infty} F\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad (3-1)$$

$t \geq 0, \lambda > 0$

는 $\lambda \rightarrow \infty$ 함에 따라 $F(t)$ 로 수렴한다[12].

위에서 주어진 식 (3-1)은 GPH분포를 따르는데 λ 가 커짐에 따라 $F_\lambda(t)$ 가 $F(t)$ 로 수렴하므로 분포함수 $F(t)$ 가 GPH분포로 근사화되는 것을 알 수 있다.

3.2 경험적 자료로부터의 추정방법

샘플규모가 m 인 자료 U_i ($i = 1, 2, \dots, m$)에 대해 분포함수를 $F(x)$, $x \geq 0$ 라고 하면 F 의 경험

적인 분포함수 $F_m(x)$ 는 다음과 같이 구해진다.

$$F_m(x) = \sum_{i=1}^m \frac{I(U_i \leq x)}{m}$$

$$I(A) = \begin{cases} 1, & A \text{가 참이면} \\ 0, & \text{그 외의 경우} \end{cases}$$

이때 $m \rightarrow \infty$ 함에 따라 $F_m(x) \rightarrow F(x)$

(w.p. 1)이다 따라서 (3-1)식의 $F(x)$ 대신에 $F_m(x)$ 를 대입하면

$$F_{m,\lambda}(x) = \sum_{n=0}^{\infty} F_m\left(\frac{x}{\lambda}\right) \frac{e^{-\lambda x} (\lambda x)^n}{n!} \dots (3-2)$$

는 모든 고정된 m 에 대해, $\lambda \rightarrow \infty$ 하면

$$F_{m,\lambda}(x) \rightarrow F(x) \text{이다. [12]}$$

따라서 경험적 분포함수 $F_m(x)$ 가 GPH분포의 형태로 근사화됨을 알 수 있다.

3.2.1 완전자료의 경우

이제 완전자료와 불완전자료로 나누어 이들의 경험적 분포함수를 구하는 방법을 알아보면 다음과 같다.

예로 $t_1 \leq t_2 \leq \dots \leq t_m$ (t_i 는 모두 고장시간)인 데이터 t_i 에 대해

$$F(t_j) = \sum_{i=1}^m \frac{I(t_i \leq t_j)}{m} = \frac{j}{m}$$

이 된다. 이로부터 확률함수를 구하면 다음과 같다.

$$f(t_j) = \frac{F(t_j) - F(t_{j-1})}{F(t_0) = 0, 1 \leq j \leq m}$$

$$f(t_j) = \frac{j}{m} - \frac{j-1}{m} = \frac{1}{m} \dots (3-3)$$

3.2.2 불완전자료의 경우

고장이 계속 일어난다면 (3-3)식을 적용하는데 무리가 없으나 불완전자료의 경우처럼 검열에 의한 제거의 시간이 포함된 경우는 달리 고려를 해야 한다. 따라서 그런 경우는 샘플 규모 m 이 줄어들었다고 간주할 수 있으므로 m 의 값이 하나씩 줄어들게 되어 $f(t_j)$ 의 값에 영향을 미친다.

다시 말해서

$t_1 \leq t_2 \leq \dots \leq t_m$ 에 대해

t_1 : 고장시간

t_2 : 검열에 의한 제거 시간

t_3 : 고장시간

⋮

이라면

$(0, t_1]$ 구간에서는 “1개 고장, 나머지는 고장 아님”을 의미하고,

$(t_1, t_2]$ 구간에서는 “고장 아닌 부품 1개 제거, 나머지는 고장 아님”을 의미하며,

$(t_2, t_3]$ 구간에서는 “1개 고장, 나머지는 고장 아님”을 의미한다.

⋮

따라서 $f(t_1)$ 은 샘플 규모 m 을 가지게 되어 경험적 분포함수로부터 $f(t_1) = \frac{1}{m}$ 을 가지게 되고,

$f(t_2)$ 는 검열에 의한 제거로 샘플 $m-1$ 개가 하나 줄어들게 되어 샘플규모가 이 되므로

$f(t_2) = \frac{1}{m-1}$ 이 된다. $f(t_3)$ 는 계속 샘플규모

$m-1$ 을 가지고 "1개 고장, 나머지는 고장 아님"을 의미하므로 경험적 분포함수에서 구한 (3-3)식을 이용하면 $f(t_3) = \frac{1}{m-1}$ 이 된다.

이를 일반화하면 다음과 같다.

첫째, $(t_{j-1}, t_j]$ 에서 검열에 의한 제거가 일어난 경우

$$f(t_j) = \frac{1}{(f(t_{j-1})\text{의 샘플규모}) - 1} = \frac{1}{\frac{1}{f(t_{j-1})} - 1}$$

이 되며,

둘째, $(t_{j-1}, t_j]$ 에서 고장이 일어난 경우

$$f(t_j) = f(t_{j-1})$$

이 된다. 따라서

$$F(t_j) = F(t_{j-1}) + f(t_j) \\ f(t_1) = f(t_1), 1 \leq j \leq m$$

이 된다. 단, 분포함수 값들을 구할 때, 고장시간 자료에 대해서만 구하고 검열에 의한 제거 시간의 경우에는 구하지 않는다.

3.3 GPH분포의 효과적인 계산방법

식(3-1)을 계산하기 위해

$$\sum_{n=0}^K \frac{e^{-\lambda t} (\lambda t)^n}{n!} > 1 - \epsilon$$

인 최소의 K 를 구하면

$$\sum_{n=K+1}^{\infty} F\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda t} (\lambda t)^n}{n!} < \sum_{n=K+1}^{\infty} \frac{e^{-\lambda t} (\lambda t)^n}{n!} < \epsilon$$

이 된다. 따라서 식(3-1)을 계산하는 대신

$$\sum_{n=0}^K F\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad (3-4)$$

로 계산하면 된다.[12] 이때 K 의 값은 $\epsilon = 10^{-7}$ 인 경우 대략적으로

$$K \cong \lambda t + 5\sqrt{\lambda t} \quad (3-5)$$

이면 충분하다. 따라서 식(3-4)는

$$\sum_{n=\lambda t - 5\sqrt{\lambda t}}^{\lambda t + 5\sqrt{\lambda t}} F\left(\frac{n}{\lambda}\right) \frac{e^{-\lambda t} (\lambda t)^n}{n!} \quad (3-6)$$

로 구하면 충분할 것이다.

식(3-6)을 계산함에 있어서 $n = [\lambda t]$ 일 때

$$P_n = \frac{e^{-\lambda t} (\lambda t)^n}{n!} = \frac{e^{-\lambda t} (\lambda t)^{\lambda t}}{(\lambda t)!} \quad (3-7)$$

를 먼저 계산하고 관계식

$$P_{n+1} = \frac{\lambda t}{n+1} P_n \quad (3-8)$$

$$P_{n-1} = \frac{n}{\lambda t} P_n \quad (3-9)$$

를 이용하여 식(3-7)과 식(3-8)로 부터

$$P_{n+1}, P_{n+2}, \dots, P_{n+5\sqrt{\lambda t}}$$

를 차례로 구하고, 식(3-7)과 식(3-9)로 부터

$$P_{n-1}, P_{n-2}, \dots, P_{n-5\sqrt{\lambda t}}$$

를 차례로 구하여 식(3-6)의 계산을 편리하게 수행할 수 있다.

4. PWF모형의 정의 및 의미

Weibull분포의 분포함수와 확률밀도함수(probability density function)는 각각 다음과 같

다.[11]

$$\begin{aligned} F(t) &= 1 - e^{-\lambda t^m} & , & & t > 0 \\ f(t) &= \lambda m t^{m-1} e^{-\lambda t^m} & , & & t > 0 \end{aligned} \quad (4-1)$$

단조고장률 분포 중에서 Weibull분포는 형태모수 (shape parameter) m 의 변화에 따라 고장률 함수가 단조 감소(decreasing failure rate), 일정, 또는 단조증가 함수(increasing failure rate)로 표현된다.

$R(t)=1-F(t)$ 라 하고 고장률 함수(순간 고장률)를 $Z(t)$ 라 하면

$$Z(t) = \frac{f(t)}{R(t)}, \quad (4-2)$$

이 되어

$$\begin{aligned} \int_0^t Z(x) dx &= \int_0^t \frac{f(x)}{R(x)} dx \\ &= -\ln R(t) \\ &= -\ln(1 - F(t)) \end{aligned}$$

이 된다. 따라서

$$F(t) = 1 - \exp\left(-\int_0^t Z(x) dx\right)$$

이 되며

$$\begin{aligned} H(t) &= \int_0^t Z(x) dx \\ &\quad (H(t) : \text{누적 고장률 함수}) \end{aligned}$$

로 정의하면

$$\begin{aligned} F(t) &= 1 - e^{-H(t)} \\ R(t) &= e^{-H(t)} \end{aligned} \quad (4-3)$$

이 된다.

(4-1)과 (4-3)를 이용하여

$$Z(t) = \frac{dH}{dt} = \lambda m t^{m-1} \quad (4-4)$$

를 구하면 고장률함수가 $0 < m < 1$ 일 때는 감소함수가 되고, $m = 1$ 일 때는 일정한 상수가 되며, $m > 1$ 일 때는 증가함수가 됨을 알 수가 있다.

고장률함수의 변화구간의 개수를 P 라고 할 때

$$Z(t) = \sum_{i=1}^P \lambda_i m_i (t - \tau_{i-1})^{m_i-1} \cdot I(\tau_{i-1}, \tau_i](t)$$

$$I(a, b](x) = \begin{cases} 1, & a < x \leq b \\ 0, & \text{otherwise} \end{cases} \quad (4-5)$$

λ_i : i 번째 구간의 Weibull분포 규모모수

m_i : i 번째 구간의 Weibull분포 형태모수

τ_i : i 번째 구간의 끝나는 시간

이 되며 위와 같이 정의되는 PWF모형은 각 구간별로 대응되는 고장 형태에 대해 그 고장 형태의 고장률 함수를 그 구간의 자료만으로 추정한다는 데 의미가 있다[11].

5. PWF모형의 모수 추정방법

5.1 누적고장률의 추정방법

경험적 고장률 추정치를 $\hat{Z}(t)$ 라 하면 식(4-2)에

의해

$$\hat{Z}(t) = \frac{n(t_i) - n(t_{i+1})}{n(t_i)\Delta t_i} \quad t_i < t \leq t_{i+1}$$

$$\begin{aligned} \Delta t_i &= t_{i+1} - t_i \\ n(t_i) &: \text{시간 } t_i \text{에 남아있는 부품의 수} \end{aligned}$$

이므로, i 번째 고장과 $i+1$ 번째 고장사이의 고장률

을 Z_i , $n(t_i) = n_i$ 라 하면

$$Z_i = \frac{n_i - n_{i+1}}{n_i \Delta t_i} = \frac{1}{n_i \Delta t_i}$$

이 된다. 따라서

$$\begin{aligned}
 H(t_i) &= \int_0^{t_i} Z(x) dx \\
 &= \sum_{k=1}^i \int_{t_{k-1}}^{t_k} Z_k dx \\
 &= \sum_{k=1}^i Z_k \Delta t_k = \sum_{k=1}^i \frac{1}{n_k}
 \end{aligned}$$

이 되며 위험치 h_i 에 대해

$$h_i = \frac{i}{n_i}$$

이 성립한다. 따라서 고장시간의 경우, 그 이전의 누적고장률에 그 시간의 위험치를 더하여 경험적 누적고장률을 구한다[11].

5.2 최소 자승법에 의한 누적 고장률 함수의 모수추정 방법

4장의 4.1절에서 Weibull분포의 누적 고장률 함수가

$$\begin{aligned}
 H(t) &= \int_0^t Z(x) dx \\
 &= \int_0^t \lambda m x^{m-1} dx \\
 &= \lambda t^m
 \end{aligned} \tag{5-1}$$

가 됨을 보였으며 여기서 m 은 0보다 큰 실수가 된다.

식(5-1)을 최소 자승법에 의해 직접 추정하면 λ 와 m 을 동시에 추정하여야 하므로 비선형 회귀 (regression)모형이 되어 버린다. 따라서 (5-1)의 양변에 \log 를 씌워

$$\log H(t) = \log \lambda + m \log t \tag{5-2}$$

로 만든다. 그러면 $\log H(t)$ 가 $\log t$ 의 직선 함수로 표현되어 직선의 기울기가 형태모수 m 이 되고

y 절편이 $\log \lambda$ 가 된다.

따라서 다음과 같이 추정한다.

고장시간이 t_1, t_2, \dots, t_n 라면 5-1절에서의

누적 고장률 추정방법에 따라 t_1, t_2, \dots, t_n

에서의 누적고장률을 $\widehat{H}_1, \dots, \widehat{H}_n$ 과 같이

추정한다. \widehat{H}_i 를 추정된 후에

$$x_1 = \log t_1, x_2 = \log t_2, \dots, x_n = \log t_n,$$

$$Y_1 = \log \widehat{H}_1, \dots, Y_n = \log \widehat{H}_n \text{ 라 하면 1차}$$

회귀모형

$$Y = \beta_0 + \beta_1 x + \epsilon \tag{5-3}$$

에 의해 최소 자승 추정치 β_0, β_1 을 구할 수 있다.[11] 즉,

$$\begin{aligned}
 \widehat{\beta}_1 &= \frac{C_{xY}}{S_x^2} \\
 &= \frac{\sum_{i=1}^n x_i Y_i - \frac{(\sum_{i=1}^n x_i)(\sum_{i=1}^n Y_i)}{n}}{\sum_{i=1}^n (x_i)^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}
 \end{aligned} \tag{5-4}$$

$$\widehat{\beta}_0 = \bar{Y} - \widehat{\beta}_1 \bar{x} \tag{5-5}$$

이 되어 (5-2)와 (5-3)으로부터 m 과 λ 의 추정치를 다음과 같이 구할 수 있다.

$$\begin{aligned}
 \widehat{m} &= \widehat{\beta}_1 \\
 &= \frac{\sum (\log t_i)(\log \widehat{H}_i) - \frac{(\sum \log t_i)(\sum \log \widehat{H}_i)}{n}}{\sum (\log t_i)^2 - \frac{(\sum \log t_i)^2}{n}}
 \end{aligned}$$

$$\hat{\lambda} = e^{\beta_0} \\ = \exp\left(\frac{\sum \log \hat{H}_i}{n} - \hat{\beta}_1 \frac{\sum \log t_i}{n}\right)$$

[11]에 의해 검정할 수 있다.

5.3 구간변화점의 추정방법

주어진 고장시간 자료가 Weibull분포를 따를 때 $\log \hat{H}_i$ 를 $\log t$ 축에 대해 그려보면 $\log t_i$ 와 $\log \hat{H}_i$ 는 직선형태를 보인다. 이때 관측된 모든 고장시간이 직선 형태로 적합될 수 있다면 전체 고장 자료에 대해 하나의 Weibull누적 고장률 함수가 추정될 수 있다. 그러나 만약 어느 시간부터 직선 형태에서 벗어나기 시작했다면 바로 그 시간이 Weibull분포에서 벗어나기 시작하는 최초의 시간, 다시 말해서 DFR에서 IFR로 변하기 시작하는 시간이 될 것이다. 이 시간은 그래프를 그려보고 직관적으로 추정할 수가 있으나 이는 주관적이고 번거롭기 때문에 객관적이고 자동적인 방법이 필요하다.

직선형태를 벗어나기 시작하는 시점은 단순 회귀모형보다 2차 이상의 항을 포함하는 회귀 모형이 주어진 관측자료를 더욱 잘 적합시키기 시작하는 시점이 될 것이다.

이 시점은 곧 다음의 2차 회귀모형

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon \quad (5-6)$$

에서 x^2 항이 필요해지기 시작하는 시점으로 볼 수 있다. 따라서 구간 변화점을 구하는 문제는 결국 가설

$$H_0 : \beta_2 = 0 \\ H_1 : \beta_2 \neq 0$$

을 기각하는 시점을 구하는 문제가 된다. 이것은 (5-6)에서 x^2 항이 필요한가 알아보는 부분 F검정법

6. 적용예제

6.1 불완전자료의 경우

6.1.1 PWF모형을 이용한 GPH근사화

전체 샘플규모가 64개이고 그 중 고장시간이 40개인 불완전자료(Nelson의 Group 6 자료)의 경우, 고장시간과 검열시간 모두에 대해 위험치(5-1절 이용)를 구하고 이를 이용하여 경험적 누적고장률을 구하면 위험치는 시간자료에 대해 그 시간까지 사용 혹은 실험 중이던 Unit의 개수의 역수가 되고, 경험적 누적고장률은 그 이전까지의 누적 고장률에 그 시간의 위험치를 더한 값이 된다.

다시 말해서 시간 t_i 에서의 위험치 h_i 는

$$h_i = \frac{1}{64 - (i - 1)} = \frac{1}{65 - i}$$

로 구하면 되고, 이를 이용한 누적고장률 H_i 는

$$H_i = H_{i-1} + h_i \quad (\text{여기서 } H_1 = h_1)$$

로 계산하여 고장시간에 대해서만 구하면 되고 검열에 의한 제거의 경우는 계산하지 않는다. 여기서 구한 누적고장률과 식(4-3) ($F(t) = 1 - e^{-H(t)}$)

을 이용한 고장시간의 분포함수 값 $\hat{F}(t)$ 들을 구하여 표에 나타내 보면 <표6-1>과 같다. <표6-1>에서는 고장을 하나의 형태로 간주하여 고장인 경우는 1, 검열에 의한 제거의 경우는 0으로 나타내었다.

<표 6-1>에서 주어진 자료를 가지고 PWF모형의 모수추정방법을 이용하여 λ 와 m 을 추정하면

$$\hat{\lambda} \cong 2.6 \times 10^{-4}$$

$$\hat{m} \cong 1$$

이 된다. 따라서 수명분포(고장시간 분포) $F(t)$ 는 다음과 같이 추정된다.

$$F(t) = 1 - e^{-\lambda t^m}$$

$$= 1 - e^{(-2.6 \times 10^{-4})t} \quad (6-1)$$

이는

$$\lambda = 2.6 \times 10^{-4}$$

$$g(n) = \begin{cases} 1 & , \quad n = 1 \\ 0 & , \quad otherwise \end{cases}$$

인 $GPH(\lambda, g)$ 에 해당된다.

따라서 PWF모형을 이용하여 분포함수 $F(t)$ 를 구하면 $F(t)$ 자체가 GPH 분포가 되므로 따로 GPH 분포로 근사화시킬 필요가 없다.

6.1.2 경험적 자료로부터의 분포함수를 이용한 GPH 근사화

6.1.1절의 “Nelson의 Group 6 자료”를 3.2절의 3.2.2를 이용하여 직접 분포함수 값을 구하면 다음과 같다.

$$t_1 \leq t_2 \leq \dots \leq t_{64} \text{ 에 대해}$$

$$t_1: \text{고장시간}$$

$$t_2: \text{고장시간}$$

$$t_3: \text{검열에 의한 제거 시간}$$

$$t_4: \text{고장시간}$$

$$\vdots$$

이므로

$$f(t_1) = \frac{1}{64} ,$$

$$f(t_2) = \frac{1}{64} ,$$

$$f(t_3) = \frac{1}{64-1} = \frac{1}{63} ,$$

$$f(t_4) = \frac{1}{63} , \dots$$

이 되고

$$F_m(t_1) = f(t_1) = \frac{1}{64}$$

$$F_m(t_2) = f(t_1) + f(t_2) = \frac{1}{64} + \frac{1}{64} ,$$

$F_m(t_3)$ 는 계산하지 않고,

$$F_m(t_4) = F(t_2) + f(t_4)$$

$$= \frac{1}{64} + \frac{1}{64} + \frac{1}{63} ,$$

...

과 같이 된다. 이 값 $F_m(t)$ 들을 계산하여 앞의 6.1.1의 $\hat{F}(t)$ 값들과 함께 표로 나타내면 <표 6-1>과 같다. <표 6-1>에서는 6.1.1에서 구한 위험치 h_i 와 누적고장률 \hat{H} , 이를 이용한 $\hat{F}(t)$ 그리고 6.1.2에서 구한 확률밀도함수 $f(t)$ 와 이로부터의 경험적 분포함수 $F_m(t)$ 를 동시에 나타내어 비교할 수 있게 하였다.

<표 6-1> Nelson의 Group 6 자료의 $\hat{F}(t)$ 값들과 $F_m(t)$ 값들의 비교

순서	시간(t)	고장형태	h	\hat{H}	$\hat{F}(t)$	f(t)	$F_m(t)$
1	45	1	0.0156	0.0156	0.0155	0.0156	0.0156
2	148	1	0.0159	0.0315	0.0310	"	0.0312
3	378	0	0.0161	.	.	0.0159	.
4	412	1	0.0164	0.0479	0.0468	"	0.0471
5	437	1	0.0167	0.0646	0.0626	"	0.0630
6	470	0	0.0169	.	.	0.0161	.
7	487	1	0.0172	0.0818	0.0785	"	0.0791
8	496	1	0.0175	0.0993	0.0945	"	0.0952
9	535	1	0.0178	0.1172	0.1106	"	0.1113
10	546	1	0.0182	0.1354	0.1266	"	0.1274
11	649	1	0.0185	0.1539	0.1426	"	0.1435
12	687	1	0.0189	0.1728	0.1587	"	0.1596
13	703	1	0.0192	0.1920	0.1747	"	0.1757
14	713	1	0.0196	0.2116	0.1907	"	0.1918
15	816	1	0.0200	0.2316	0.2067	"	0.2079
16	881	1	0.0204	0.2520	0.2228	"	0.2240
17	1103	0	0.0208	.	.	0.0164	.
18	1117	1	0.0213	0.2733	0.2391	"	0.2404
19	1238	1	0.0217	0.2950	0.2555	"	0.2568
20	1312	1	0.0222	0.3400	0.2882	"	0.2732
21	1312	1	0.0227	0.3400	0.2882	"	0.2896
22	1321	1	0.0233	0.3632	0.3046	"	0.3060
23	1342	1	0.0238	0.3870	0.3209	"	0.3224
24	1489	1	0.0244	0.4114	0.3373	"	0.3388
25	1585	1	0.0250	0.4364	0.3536	"	0.3552
26	1588	1	0.0256	0.4621	0.3700	"	0.3716
27	1738	0	0.0263	.	.	0.0167	.
28	1789	1	0.0270	0.4891	0.3868	"	0.3883
29	2070	0	0.0278	.	.	0.0169	.
30	2113	1	0.0286	0.5177	0.4041	0.0169	0.4052
31	2157	1	0.0294	0.5471	0.4214	"	0.4221
32	2203	1	0.0303	0.5774	0.4386	"	0.4390

순서	시간(t)	고장형태	h	\hat{H}	$\hat{F}(t)$	f(t)	$F_m(t)$
33	2432	1	0.0313	0.6086	0.4559	"	0.4559
34	2433	0	0.0323	.	.	0.0172	.
35	2757	1	0.0333	0.6420	0.4738	"	0.4731
36	2762	1	0.0345	0.6765	0.4916	"	0.4903
37	2899	1	0.0357	0.7122	0.5094	"	0.5075
38	3008	1	0.0370	0.7492	0.5273	"	0.5247
39	3164	1	0.0385	0.7877	0.5451	"	0.5419
40	3319	1	0.0400	0.8277	0.5629	"	0.5591
41	3380	1	0.0417	0.8693	0.5808	"	0.5763
42	3412	1	0.0435	0.9128	0.5986	"	0.5935
43	3484	0	0.0455	.	.	0.0175	.
44	3552	0	0.0476	.	.	0.0179	.
45	3688	0	0.0500	.	.	0.0182	.
46	3830	1	0.0526	0.9654	0.6192	"	0.6117
47	3855	0	0.0556	.	.	0.0185	.
48	3892	0	0.0588	.	.	0.0189	.
49	3901	0	0.0625	.	.	0.0192	.
50	3937	0	0.0667	.	.	0.0196	.
51	3985	0	0.0714	.	.	0.0200	.
52	4004	1	0.0770	1.0424	0.6474	"	0.6317
53	4006	0	0.0833	.	.	0.0204	.
54	4017	1	0.0909	1.1333	0.6780	"	0.6521
55	4056	0	0.1000	.	.	0.0208	.
56	4129	1	0.1111	1.2444	0.7119	"	0.6729
57	4137	0	0.1250	.	.	0.0213	.
58	4319	0	0.1429	.	.	0.0217	.
59	4368	0	0.1667	.	.	0.0222	.
60	4395	0	0.2000	.	.	0.0227	.
61	4413	0	0.2500	.	.	0.0233	.
62	4693	0	0.3333	.	.	0.0238	.
63	4816	0	0.5000	.	.	0.0244	.
64	5448	0	1.0000	.	.	0.0250	.

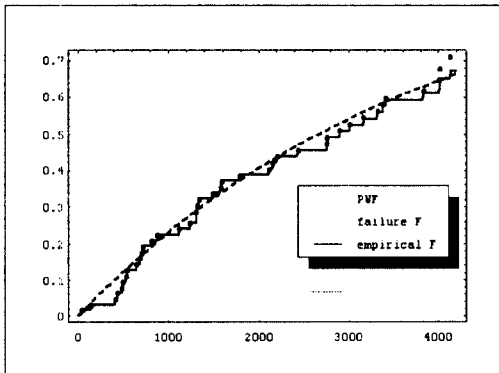
<표6-1>로부터 분포함수 $F_m(t)$ 를

$$F_m(t) = \begin{cases} 0, & t < t_1 \\ 0.0156, & t_1 \leq t < t_2 \\ \vdots & \vdots \\ 0.6521, & t_{39} \leq t < t_{40} \\ 0.6729, & t_{40} \leq t \end{cases}$$

(6-2)

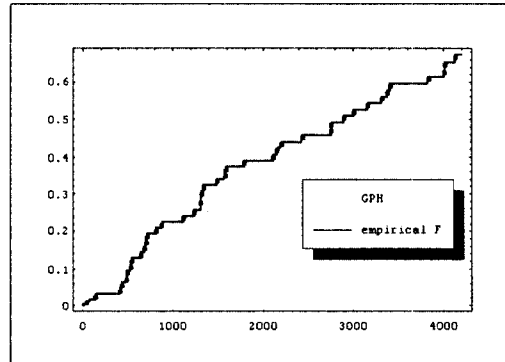
와 같이 구한 뒤 $F_m(t)$ 의 그래프와 6.1.1에서 구한

$\hat{F}(t)$ 값들, 그리고 PWF모형을 이용한 수명분포함수 식 (6-1)의 그래프를 동시에 나타내어 비교하면 <그림 6-1-1>과 같고, 이 <그림 6-1-1>에서 $F_m(t)$ 의 그래프는 선으로 $\hat{F}(t)$ 값들은 점으로 PWF모형에 의한 분포함수는 점선으로 표시되는데 여기서 $\hat{F}(t)$ 값들과 $F_m(t)$ 값들이 거의 일치함을 알 수 있다.



<그림 6-1-1> $\hat{F}(t)$ 와 $F_m(t)$ 및 PWF에 의한 분포함수의 비교

식(6-2)를 식(3-6)을 이용하여 GPH분포로 근사화시켜 GPH분포함수 값들을 매스매티카 프로그램을 이용하여 구한 뒤 여기서 구한 GPH분포함수 값들과 $F_m(t)$ 의 그래프를 동시에 나타내어 비교하면 다음의 <그림 6-1-2>가 되고 여기서 $F_m(t)$ 의 그래프는 선으로 GPH분포함수 값들은 점으로 표시되는데 그림에서 알 수 있듯이 $F_m(t)$ 값들과 GPH분포함수 값들이 거의 일치하여 $F_m(t)$ 가 GPH분포로 근사화가 잘 이루어짐을 알 수 있다. 따라서 불완전 자료의 경우, 경험적 분포함수가 GPH분포로 근사화가 잘 이루어짐을 알 수 있다.



<그림 6-1-2> 경험적 자료로부터의 분포함수와 GPH분포함수의 비교

6.2 완전자료의 경우

완전 자료의 예로 “Calvin 의 시뮬레이션 자료”를 이용하는 데 이는 49개의 고장시간 자료만으로 이루어진 것으로 3장의 3.2.1을 이용하여 경험적 분포함수 값을 구하면 다음과 같다.

$$F_m(t_i) = \frac{i}{49}$$

(t_i : i 번째 고장시간)

이 된다. 이를 표로 나타내면 <표6-2>가 되고, <표6-2>에서도 <표6-1>에서와 마찬가지로 고장형태의 표시에서 고장인 경우는 1, 고장이 아닌 경우는 0으로 나타내었다.

<표 6-2> Calvin 의 시뮬레이션 자료

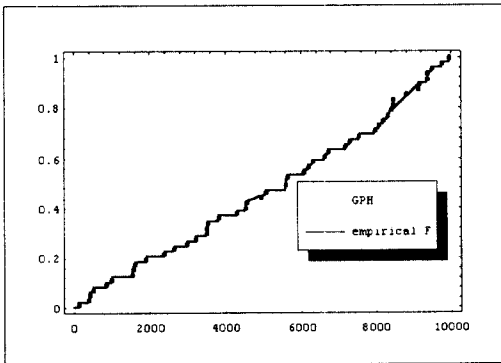
순서	시간(t)	고장형태	$F_m(t)$	순서	시간(t)	고장형태	$F_m(t)$
1	118	1	0.0204	26	5640	1	0.5306
2	390	1	0.0408	27	6070	1	0.5510
3	413	1	0.0612	28	6190	1	0.5714
4	504	1	0.0816	29	6310	1	0.5918
5	840	1	0.1020	30	6630	1	0.6122
6	990	1	0.1224	31	6730	1	0.6327
7	1540	1	0.1429	32	7190	1	0.6531
8	1561	1	0.1633	33	7310	1	0.6735
9	1609	1	0.1837	34	7540	1	0.6939
10	1908	1	0.2041	35	7960	1	0.7143
11	2385	1	0.2245	36	8060	1	0.7347
12	2640	1	0.2449	37	8180	1	0.7551
13	2979	1	0.2653	38	8310	1	0.7755
14	3220	1	0.2857	39	8400	1	0.7959
15	3500	1	0.3061	40	8440	1	0.8163
16	3510	1	0.3265	41	8750	1	0.8367
17	3540	1	0.3469	42	8800	1	0.8571
18	3820	1	0.3673	43	9125	1	0.8776
19	4310	1	0.3878	44	9140	1	0.8980
20	4530	1	0.4081	45	9360	1	0.9184
21	4570	1	0.4286	46	9375	1	0.9388
22	4940	1	0.4490	47	9480	1	0.9592
23	5060	1	0.4694	48	9740	1	0.9796
24	5590	1	0.4898	49	9950	1	1.0000
25	5590	1	0.5102				

<표6-2>로부터 분포함수 $F_m(t)$ 를

$$F_m(t) = \begin{cases} 0, & 0 < t < t_1 \\ \frac{1}{49}, & t_1 \leq t < t_2 \\ \vdots & \vdots \\ \frac{48}{49}, & t_{48} \leq t < t_{49} \\ 1 & t_{49} \leq t \end{cases}$$

(t_i : i 번째 고장시간)

와 같이 구한 뒤 식 (3-6)을 이용하여 GPH분포로 근사화시켜 GPH분포함수의 값들을 계산한 뒤 이 값들과 $F_m(t)$ 의 그래프를 동시에 나타내면 다음의 <그림 6-2>와 같다. <그림6-2>에서 $F_m(t)$ 의 그래프는 선으로, GPH분포함수 값들은 점으로 표시되었으며 이 그림에서 $F_m(t)$ 의 값들과 GPH분포함수 값들이 거의 일치함을 알 수 있다. 따라서 $F_m(t)$ 가 GPH분포로 근사화가 잘 이루어짐을 알 수 있으며 더 나아가 완전자료의 경우도 경험적 분포함수가 GPH분포로 근사화가 잘 이루어짐을 알 수 있다.



<그림 6-2> GPH분포와 경험적 자료로부터의 분포함수의 비교

7. 결론

지금까지 고장시간 분포를 추정하는 데 있어서 PWF모형을 이용하여 추정하는 방법과 경험적 자료로부터 직접 분포함수를 추정하는 방법 두 가지에 대하여 알아본 후 이들을 GPH분포로 근사화시키는 방법에 대해 알아보았다. 특히 6.1절의 경우 똑같은 불완전 자료에 대해 PWF모형을 이용한 경우와 경험적 자료로부터 직접 분포함수를 구한 후 GPH분포로 근사화시키는 두 가지의 방법을 비교하였는데 PWF모형을 이용하여 분포함수를 구한 6.1.1절의 경우 분포함수가 지수분포의 형태가 되므로 분포함수 자체가 GPH분포가 되어 따로 GPH분포로 근사화할 필요가 없었다. 6.1.2에서는 자료로부터 직접 분포함수를 구한 후 GPH분포로 근사화하는 방법을 예시하였고 6.2절에서는 완전자료의 경우 6.1.2와 같은 방법을 이용하여 GPH분포로 근사화가 잘 이루어짐을 보였다.

위와 같이 고장시간 분포가 GPH분포로 근사화가 되면 고장시간분포에 대해 GPH분포의 여러 가지 유용한 성질들을 이용할 수 있다. 특히 부품의 고장시간 분포를 GPH로 구하면 GPH분포의 성질 (2.2절의 GPH분포의 성질 5)을 이용하여 전체 시스템의 고장시간 분포가 다시 GPH분포가 되므로 GPH분포의 여러 가지 유용한 성질들을 전체 시스템에 적용할 수 있는 이점이 있다.

참 고 문 헌

- [1] Blachman, Nancy, "*Mathematica: A Practical Approach*", Prentice Hall, 1992
- [2] Neuts, M. F., "*Matrix-Geometric Solution in Stochastic Models: An Algorithmic Approach*", The Johns Hopkins University Press, Baltimore, 1981
- [3] Neuts, M. F., "*Structured Stochastic Matrices of M/G/1 Type and Their Applications*", Marcel Dekker, New York, 1989
- [4] Ross, S. M., "*Introduction To Probability Models*", University of California, Berkeley, 1972
- [5] Ross, S. M., "*Stochastic Process*", John and sons, Inc. 1983
- [6] Shanthikumar, J. G., "*Bilateral Phase-Type Distributions*", Naval Research Logistics Quarterly, Vol.32. , 1985
- [7] Wolfram, S., "*Mathematica : A System for Doing Mathematics by computer* ", Addison Wesley, 1991
- [8] William, T. Shaw, Jason Tigg, "*Applied Mathematica*", Addison wesley, 1994
- [9] 강 상권, 남 기원, 전 규창, "메스매티카 입문에서 활용까지", 성안당, 1993
- [10] 박 을용, 김 치영, 박 한식, "수학 대사전", 한국사전 연구원, 1990
- [11] 윤 복식, "*Piecewise Weibull 고장률 함수 모형에 의한 수명분포 추정방법*", 서울 대학교 석사 학위논문, 1982
- [12] 윤 복식, 박 광우, 이 창훈, "GPH분포에 의한 확률적 근사화", 한국 경영과학회지 제19권, 1994