1. 신뢰구간(Confidence Bound) 추정법

Confidence Bound를 추정하는 방법은 크게 Likelihood Ratio Confidence Bound, Fisher Matrix Confidence Bound, Beta Binomial Confidence Bound 등으로 알려져 있다. 이들은 각각 파라미터, 시간, 신뢰도(Reliability)에 대한 신뢰구간을 예측할 수 있으며, 다만 Beta Binomial Confidence Bound은 비파라미터 접근법이다. 여기서는 Likelihood Ratio Confidence Bound에 대해 기본개념 및 그 적용 예를 예제를 통해 설명한다.

2. Likelihood Ratio Confidence Bounds(Intervals)

가. 기본개념

고장율 데이터의 분포 파라미터를 추정하는 방법으로 가장 근접한 파라미터를 얻는 방법이다.

만약 x가 연속적인 랜덤 변수를 가지는 PDF(Probability Density Function)이라면 다음과 같이 표현된다.

$$f(x;\theta_1,\theta_2\theta_3,\ldots,\theta_k)$$

여기서, $\theta_1,\theta_2,\theta_3,\ldots,\theta_k$는 κ개 알려지지 않은 파라미터로서, R개의 독립적인 관찰로 얻어진 고장시간에 해당하는 $x_1,x_2,x_3,\ldots,x_R$을 가지고 추정되는데 필요하다.

그러면 Likelihood 함수는 다음과 같이 표현된다.

$$L(\theta_1,\theta_2\theta_3,\ldots,\theta_k|x_1,x_2,x_3,\ldots,x_R) = L = \prod_{i=0}^{\infty} f(x_i;\theta_1,\theta_2\theta_3,\ldots,\theta_k)$$

$$\Lambda = \ln L = \sum_{i=0}^{\infty} f(x_i;\theta_1,\theta_2\theta_3,\ldots,\theta_k)$$

따라서, Λ를 최대화 함으로써 Maximum Likelihood Parameter를 얻을 수 있으며, 아래의 식과 같다.

$$\frac{\partial \Lambda}{\partial \theta_j} = 0, \quad j = 1,2,\ldots,k$$

Likelihood Ratio 신뢰구간은 아래 식에 기초한다.

$$-2 \cdot \ln\left(\frac{L(\theta)}{L(\hat{\theta})}\right) \geq \chi^2{}_{\alpha;\kappa}$$

여기서, $L(\theta)$는 알려지지 않은 파라미터 벡터 θ에 대한 Likelihood 함수

$L(\hat{\theta})$는 추정된 벡터 $\hat{\theta}$ 로 계산된 Likelihood 함수

$x^2_{\alpha;\kappa}$는 자유도 α와 κ를 가지는 카이자승 분포함수이다. 여기서, κ는 추정되는 모집단의 수이고, α = δ (신뢰수준) 이다.

나. 파라미터에 대한 신뢰구간

2개의 파라미터를 가진 분포함수에 대해 신뢰수준(예 : 90%)이 주어지면 하나의 파라미터를 가정하면 위 식은 다음과 같이 표현된다.

$$-2 \cdot \ln\left(\frac{L(\theta_1,\theta_2)}{L(\widehat{\theta_1},\widehat{\theta_2})}\right) \geq x^2{}_{\alpha;1}$$

이를 다시 표현하면,

$$L(\theta_1,\theta_2) = L(\widehat{\theta_1},\widehat{\theta_2}) \cdot e^{-\frac{x^2{}_{\alpha;1}}{2}}$$

예제1)

5개의 고장율 데이터셋은 10, 20, 30, 40, 50시간이다. **와이블(Weibull) 분포를** 가정하여 분포파라미터를 추정하면, $\hat{\beta} = 2.2938$, $\hat{\eta} = 33.9428$로 계산되었다. 이 파라미터의 90%신뢰수준을 갖는 신뢰구간을 구하여라.

먼저 Likelihood 함수를 계하면,

$$L(\hat{\beta},\hat{\eta}) = \prod_{i=0}^{N} f(x_i;\hat{\beta},\hat{\eta}) = \prod_{i=0}^{N} \frac{\hat{\beta}}{\hat{\eta}} \cdot \left(\frac{x_i}{\hat{\eta}}\right)^{\hat{\beta}-1} \cdot e^{-\left(\frac{x_i}{\hat{\eta}}\right)^{\beta}}$$

$$L(\hat{\beta},\hat{\eta}) = \prod_{i=0}^{N} \frac{2.2938}{33.9428} \cdot \left(\frac{x_i}{33.9428}\right)^{1.2938} \cdot e^{-\left(\frac{x_i}{33.9428}\right)^{2.2938}} = 1.714714 \times 10^{-9}$$

$$L(\theta_1,\theta_2) - L(\widehat{\theta_1},\widehat{\theta_2}) \cdot e^{-\frac{x^2{}_{\alpha;1}}{2}} = 0$$

신뢰수준이 정해지면 카이자승분포의 값이 $x^2{}_{0.9;1} = 2.705543$ 으로 계산된다.

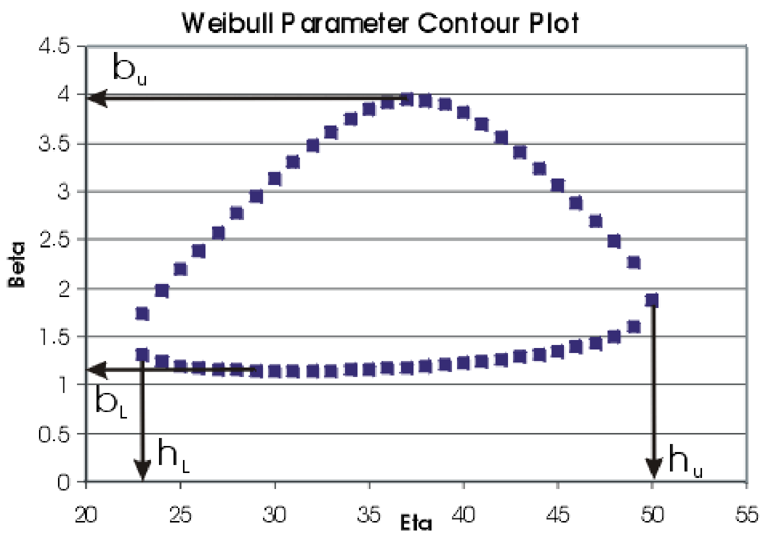
그러면 위 식은 다시,

$$L(\theta_1,\theta_2) - 1.714714 \times 10^{-9} \cdot e^{\frac{-2.705543}{2}} = 0$$

$$L(\theta_1,\theta_2) = 4.432926 \times 10^{-9}$$

그 다음 단계로, Iterative Process가 필요하다. 즉, 하나의 파라미터의 값을 변화시키면서 나머지 파라미터의 값을 구한다. 아래 표에 η에 따른 β의 변화를 나타내었다.

η	β_1	β_2	η	β_1	β_2
23	1.321	1.742	37	1.183	$\boxed{3.950}$
24	1.241	1.985	38	1.195	3.943
25	1.201	2.193	39	1.210	3.897
26	1.177	2.390	40	1.226	3.814
27	1.161	2.582	41	1.245	3.701
28	1.151	2.770	42	1.267	3.564
29	1.145	2.955	43	1.291	3.409
30	1.142	3.135	44	1.319	3.241
31	$\boxed{1.142}$	3.308	45	1.352	3.065
32	1.145	3.471	46	1.391	2.883
33	1.149	3.619	47	1.439	2.695
34	1.155	3.746	48	1.502	2.495
35	1.162	3.848	49	1.594	2.272
36	1.172	3.917	50	1.871	1.871



표와 그림에서 알 수 있듯이 90% 신뢰수준을 나타내는 파라미터의 신뢰구간은 η= 22.474, 49.967이고, β = 23, 50이다.

다. 시간에 대한 신뢰구간

파라미터에 대한 신뢰구간을 구하는 방식과 동일한 방법으로 구한다. 즉, 시간의 함수인 신뢰도(R)를 추정하는 방식이다.

예제2) 예제1에 주어진 데이터로 50% 신뢰도(R)를 가지는 시점에 대해 90% 신뢰수준을 가지는 시간에 대한 신뢰구간을 결정하라. 50% 신뢰도(R)를 가질 때의 Maximum Likelililihood 추정 시간은 28.930hr이다.

와이블 함수의 신뢰도식은 $R = e^{-(t/\eta)^\beta}$ 이다. 이것을 파라미터 η에 대해 다시 정리하면,

$$\eta = \frac{t}{(-\ln(R))^{\frac{1}{\beta}}}$$

이것을 Likelihood 식에 대입하면,

$$L(\beta, t) \ = \ \prod_{i=1}^{N} f(x_i; \beta, t, R),$$

$$= \prod_{i=1}^{5} \frac{\beta}{\left(\frac{t}{(-\ln(R))^{\frac{1}{\beta}}}\right)} \cdot \left(\frac{x_i}{\left(\frac{t}{(-\ln(R))^{\frac{1}{\beta}}}\right)}\right)^{\beta-1} \cdot \exp\left[-\left(\frac{x_i}{\left(\frac{t}{(-\ln(R))^{\frac{1}{\beta}}}\right)}\right)^{\beta}\right]$$

여기서 x_i 는 수집된 고장시간 데이터이다.

예제 1과 마찬가지로 신뢰구간에 대한 식은,

$$L(\theta_1, \theta_2) - L(\widehat{\theta_1}, \widehat{\theta_2}) \cdot e^{-\frac{\chi^2_{\alpha;1}}{2}} = 0$$

신뢰수준이 정해지면 카이자승분포의 값이 $\chi^2_{0.9;1} = 2.705543$ 으로 계산된다.

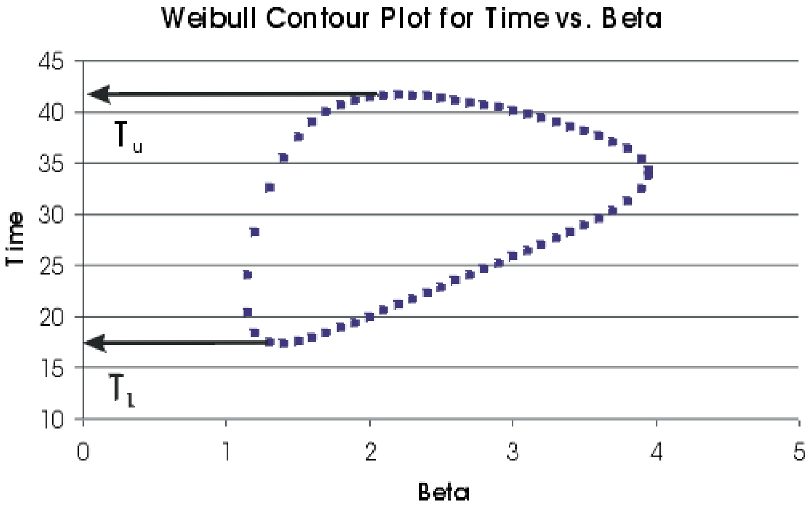
그러면 위 식은 다시,

$$L(\theta_1,\theta_2) - 1.714714 \times 10^{-9} \cdot e^{\frac{-2.705543}{2}} = 0$$

$$L(\theta_1,\theta_2) = 4.432926 \times 10^{-9}$$

그 다음 단계로, 예제1과 같이 Iterative Process를 사용하는 데 β의 값을 변화시키면서 t에 대한 값을 구하면 아래 표와 같다.

β	t_1	t_2	β	t_1	t_2
1.2	18.448	28.244	2.6	23.533	41.241
1.3	17.489	32.694	2.7	24.119	41.009
1.4	***17.389***	35.568	2.8	24.705	40.747
1.5	17.602	37.576	2.9	25.292	40.459
1.6	17.971	39.005	3.0	25.883	40.146
1.7	18.428	40.019	3.1	26.478	39.809
1.8	18.937	40.726	3.2	27.082	39.450
1.9	19.478	41.201	3.3	27.696	39.066
2.0	20.039	41.499	3.4	28.329	38.654
2.1	20.612	41.660	3.5	28.987	38.205
2.2	21.192	***41.714***	3.6	29.684	37.709
2.3	21.776	41.684	3.7	30.444	37.142
2.4	22.361	41.587	3.8	31.321	36.452
2.5	22.947	41.436	3.9	32.496	35.457



표와 그림에서 알 수 있듯이 시간에 대한 신뢰구간은 17.389, 41.714이다.

라. 신뢰도에 대한 신뢰구간

예제3) 예제1에 주어진 데이터로 시간이 45시간일 때 대해 90% 신뢰수준을 가지는 신뢰도(R)의 신뢰구간을 결정하라. t = 45일 때 Maximum Likelilihood 추정 신뢰도는 14.816%이다.

예제 2와 마찬가지로 η 에 대해 정리한 와이블 함수는,

$$\eta = \frac{t}{(-\ln(R))^{\frac{1}{\beta}}}$$

이것을 Likelihood 식에 대입하면,

$$L(\beta, R) \quad = \quad \prod_{i=1}^{N} f(x_i; \beta, t, R)$$

$$= \prod_{i=1}^{5} \frac{\beta}{\left(\frac{t}{(-\ln(R))^{\frac{1}{\beta}}}\right)} \cdot \left(\frac{x_i}{\left(\frac{t}{(-\ln(R))^{\frac{1}{\beta}}}\right)}\right)^{\beta-1} \cdot \exp\left[-\left(\frac{x_i}{\left(\frac{t}{(-\ln(R))^{\frac{1}{\beta}}}\right)}\right)^{\beta}\right]$$

예제 1, 2와 마찬가지로 신뢰구간에 대한 식은,

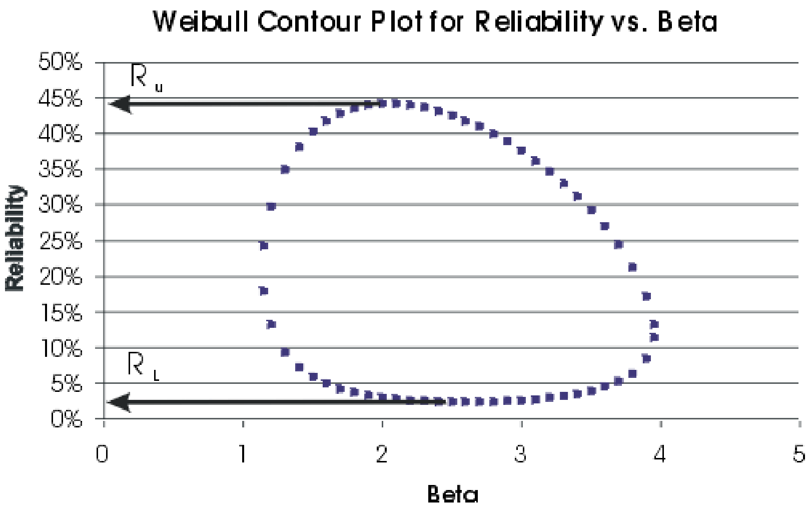
$$L(\theta_1, \theta_2) - L(\widehat{\theta_1}, \widehat{\theta_2}) \cdot e^{\frac{x^2{}_{\alpha;1}}{2}} = 0$$

이것에 Maximum Likelihood와 카이자승 값을 대입하면, 예제 2와 마찬가지로 Likelihood 함수는 아래와 같다.

$$L(\theta_1, \theta_2) = 4.432926 \times 10^{-9}$$

예제 2와 같이 β의 값을 변화시키면서 R에 대한 값을 구하면 아래 표와 같다.

β	R_1	R_2	β	R_1	R_2
1.2	0.1325	0.2975	2.6	$\boxed{0.0238}$	0.4191
1.3	0.0936	0.3499	2.7	0.0239	0.4104
1.4	0.0725	0.3816	2.8	0.0243	0.4004
1.5	0.0588	0.4032	2.9	0.0251	0.3892
1.6	0.0493	0.4184	3.0	0.0262	0.3767
1.7	0.0423	0.4291	3.1	0.0277	0.3629
1.8	0.0372	0.4362	3.2	0.0296	0.3478
1.9	0.0333	0.4406	3.3	0.0321	0.3311
2.0	0.0303	0.4426	3.4	0.0353	0.3128
2.1	0.0281	$\boxed{0.4426}$	3.5	0.0395	0.2925
2.2	0.0264	0.4409	3.6	0.0451	0.2699
2.3	0.0252	0.4375	3.7	0.0527	0.2442
2.4	0.0244	0.4327	3.8	0.0641	0.2136
2.5	0.0239	0.4266	3.9	0.0848	0.1727



따라서 90% 신뢰수준을 가지는 신뢰도(R)의 신뢰구간(확신구간)은 3.38%, 44.26%이다.

마. Simulation Based Bound

실제 적용에 있어서는 Monte-Carlo Simulation을 이용한 반복계산을 사용하여 신뢰구간을 구하는 것이 일반적이다.

2. Estimation Of Distribution Function Parameter

이 방법의 수행은 Y값의 변량의 제곱근의 합을 최소화하는 방법으로 구해지며, Least Square Method로 알려져 있다. 이 방법은 Linear 함수에 적용되는 방법으로서 실제 적용을 위해서는 분포함수를 Linear 함수로 변형해야 가능하다.
2-Parameter 와이블 함수를 예를 들어 보면, CDF(Cumulative Density Function)는 다음과 같다.

$$F(T) = 1 - e^{-\left(\frac{T}{\eta}\right)^{\beta}}$$

양변에 자연로그를 취하면,

$$\ln[1 - F(T)] = -\left(\frac{T}{\eta}\right)^{\beta}$$

$$\ln\{-\ln[1 - F(T)]\} = \beta \ln\left(\frac{T}{\eta}\right)$$

다시 정리하면,

$$\ln\{-\ln[1 - F(T)]\} = -\beta\ln(\eta) + \beta\ln(T)$$

$y = \ln\{-\ln[1 - F(T)]\}$, $a = -\beta\ln(\eta)$, $b = \beta$ 로 놓으면

$$y = a + bx$$

인 1차함수가 된다. 즉, Least Square법의 적용이 가능해 진다.

다음 단계로 Least Square법으로 통해 추정되는 a, b 값은 아래와 같은 식으로 표현된다.

$$\hat{a} = \frac{\sum_{i=1}^{N} y_i}{N} - \hat{b}\frac{\sum_{i=1}^{N} x_i}{N} = \bar{y} - \hat{b}\bar{x}$$

$$\hat{b} = \frac{\sum_{i=1}^{N} x_i y_i - \frac{\sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N}}{\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}}$$

여기서,

$$y_i = \ln\{-\ln[1 - F(T_i)]\} \quad , \quad x_i = \ln(T_i)$$

예제) 고장시간이 16, 34, 53, 75, 93, 120으로 측정되었다면, 와이블 분포파라미터를 구하라

Least Square법을 적용하기 위해 아래와 같이 도표화 할 수 있다.

N	T_i	$\ln(T_i)$	$F(T_i)$	y_i	$(\ln T_i)^2$	y_i^2	$(\ln T_i)y_i$
1	16	2.7726	0.1091	-2.1583	7.6873	4.6582	-5.9840
2	34	3.5264	0.2645	-1.1802	12.4352	1.3930	-4.1620
3	53	3.9703	0.4214	-0.6030	15.7632	0.3637	-2.3943
4	75	4.3175	0.5786	-0.1460	18.6407	0.0213	-0.6303
5	93	4.5326	0.7355	0.2851	20.5445	0.0813	1.2923
6	120	4.7875	0.8909	0.7955	22.9201	0.6328	3.8083
$\sum$		23.9068		-3.0070	97.9909	7.1502	-8.0699

그러면, b값을 구하는 식에 대입하면,

$$\widehat{b} = \frac{\sum_{i=1}^{6}(\ln T_i)y_i - (\sum_{i=1}^{6}\ln T_i)(\sum_{i=1}^{6}y_i)/6}{\sum_{i=1}^{6}(\ln T_i)^2 - (\sum_{i=1}^{6}\ln T_i)^2/6}$$

$$\widehat{b} = \frac{-8.0699 - (23.9068)(-3.0070)/6}{97.9909 - (23.9068)^2/6}$$

$$\widehat{b} = 1.4301$$

이결과를 가지고 a값을 구하는 식에 대입하면,

$$\widehat{a} = \overline{y} - \widehat{b}\overline{T} = \frac{\sum_{i=1}^{N} y_i}{N} - \widehat{b}\frac{\sum_{i=1}^{N}\ln T_i}{N}$$

$$\widehat{a} = \frac{(-3.0070)}{6} - (1.4301)\frac{23.9068}{6} = -6.19935$$

이로부터 β, η를 구할 수 있다.

$$\widehat{\beta} = \widehat{b} = 1.4301$$

$$\widehat{\eta} = e^{-\frac{\widehat{a}}{\widehat{b}}} = e^{-\frac{(-6.19935)}{1.4301}}, \quad \widehat{\eta} = 76.318\,\mathrm{hr}$$