

A Parameter Estimation Method using Nonlinear Least Squares

Suna Oh^a · Jongwoo Song^{a,1}

^aDepartment of Statistics, Ewha Woman University

(Received March 5, 2013; Revised April 23, 2013; Accepted May 14, 2013)

Abstract

We consider the problem of estimating the parameters of heavy tailed distributions. In general, maximum likelihood estimation(MLE) is the most preferred method of parameter estimation because it has good properties such as asymptotic consistency, normality and efficiency. However, MLE is not always the best solution because MLE is unstable or does not exist in some cases. This paper proposes another parameter estimation method, non-linear least squares(NLS) and compares its performance to MLE. The NLS estimator is achieved by minimizing sum of squared difference between empirical cumulative distribution function(CDF) and a theoretical distribution function. In this article, we compare the NLS method to MLE using simulated data from heavy tailed distributions. The NLS method is shown to perform better than MLE in Burr distribution when the sample size is small; in addition, it performs well in a Frechet distribution.

Keywords: Parameter estimation, MLE, NLS.

1. 서론

통계학에서 확률분포의 모수(parameter)를 추정하는 것은 가장 기본적이면서도 중요한 일이다. 관측된 자료로부터 적절한 확률모형과 그에 해당하는 적절한 모수를 추정함으로써, 다양한 통계적 방법론들을 더 정확하게 사용할 수 있으며, 더 나아가 앞으로 발생할 수 있는 사건에 대한 예측도 가능하게 된다. 일반적으로 사용되는 모수 추정 방법으로는 최대우도추정량(Maximum likelihood estimators; MLE), 베이지 추정량(Bayes estimators), 직률추정량(Method of moments estimators; MME), 최소제곱추정량(Least squares estimators; LSE), 최소분산불편추정량(Minimum variance unbiased estimator; MVUE), 최량선형불편추정량(Best linear unbiased estimator; BLUE) 등이 있다. 그 중에서도 가장 보편적으로 사용되고 있는 추정량은 MLE이다.

MLE는 우도함수(likelihood function)를 최대화하는 모수에 대한 추정치이다. MLE는 적절한 조건이 만족되었을 때, 일관성(consistency), 정규성(normality), 효율성(efficiency), 충분성(sufficiency), 불변성(invariance) 등 여러 방면에서 좋은성질을 가지고 있기 때문에 가장 선호되는 모수 추정 방법이다

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(No. 2012-0003739).

¹Corresponding author: Associate professor, Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 120-750, Korea. E-mail: josong@ewha.ac.kr

(Hogg 등, 2012). 하지만 이러한 성질들이 모든 경우에 적용되는 것은 아니다. 우도함수가 잘 정의되지 않거나, 자료의 손실이 있을 때에는 MLE가 존재하지 않는 경우도 있다. 또 MLE가 존재하더라도 위에서 언급한 좋은 성질들을 모두 만족하지 않을 수도 있다. MLE는 수리적으로 계산 가능한 경우도 있지만 수치적(numerically)으로 계산해야 하는 경우도 많다. 이러한 경우에는 다양한 최적화 방법론(optimization method)을 사용해서 MLE를 계산 할 수 있다.

또 다른 모수 추정 방법 중 하나인 비선형 최소제곱법(nonlinear least squares; NLS)은 경험분포함수(empirical CDF)와 이론적 분포함수 차이의 제곱합을 최소화 하는 방법이다. NLS 추정량은 어떠한 분포에도 사용 가능하다는 장점이 있지만, MLE와 마찬가지로 최적화 문제가 존재한다. 또한 초기치에 크게 의존할 수 있다는 단점을 가지고 있다. Song과 Song (2011) 논문에서는 NLS의 이러한 단점을 보완한 새로운 모수 추정 방법론인 NLS-2를 제안하였고, 형상 모수(shape parameter)가 0보다 큰 일반화파레토분포(generalized pareto distribution; GPD)에 적용시켜 본 결과 MLE 보다 훨씬 좋은 성능을 나타냈다. 본 논문에서는 NLS-2 방법론을 일반화파레토분포가 아닌 다른 분포에 적용시켜 보고, 그 성능을 MLE와 비교해 보았다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 이 논문에서 사용되는 MLE와 NLS방법론을 간략히 소개한다. 제 3장에서는 모의 실험을 통하여 여러 가지 분포에서 NLS-2 방법론과 MLE의 성능을 비교하여 본다. 마지막으로 제 4장에서는 결론이 제시된다.

2. 모수 추정

2.1. MLE

최대우도추정량(MLE)은 우도함수 $L(x_1, \dots, x_n; \theta)$ 를 최대화 하는 모수, θ 이다.

$$\hat{\theta} = \arg \max_{\theta} L(\theta; X).$$

MLE는 일정한 조건이 만족된 경우에 일관성, 정규성, 효율성, 충분성, 불변성 등의 다양한 좋은 성질들을 가진다. 특히 일관성은 관측치의 수가 증가함에 따라, MLE가 확률적으로 참값 θ_0 에 수렴한다는 것으로 추정량의 가장 좋은 성질 중에 하나이다.

$$\lim_{n \rightarrow \infty} p\left(\left|\hat{\theta} - \theta_0\right| > \epsilon\right) = 0, \quad \forall \epsilon > 0.$$

MLE가 일관성을 가지기 위해서는 다음과 같은 조건들이 만족되어야 한다 (Hogg 등, 2012).

1. pdf는 서로 구분된다. 즉, $\theta \neq \theta_0 \Leftrightarrow f(\bullet|\theta) \neq f(\bullet|\theta_0)$ 이다.
2. pdf는 모든 θ 에 대해 공동 받침(support)을 갖는다.
3. 점 θ_0 는 Ω 안에 있는 점이다.

X_1, \dots, X_n 이 다음 세 가지 조건을 만족한다고 하자. 여기서 θ_0 는 참 모수이고, $f(x; \theta)$ 는 Ω 안에 있는 θ 에 대해 미분 가능하다. 우도식이 유일해 $\hat{\theta}$ 를 가지면, $\hat{\theta}$ 는 θ_0 의 일치추정량이다.

MLE가 일관성을 가지게 하는 3가지 정칙 조건을 만족하고 다음의 5가지 조건을 만족한다면 근사적으로 정규분포를 따른다.

1. $\hat{\theta} \in \Omega$.
2. $f(X; \theta) > 0$ 이고 θ 에서 연속적으로 두 번 미분 가능하다.

3. $\int \sup_{\theta \in N} \|\nabla_{\theta} f(X; \theta)\| dx < \infty$ 이고, $\int \sup_{\theta \in N} \|\nabla_{\theta\theta} f(X; \theta)\| dx < \infty$ 이다.
4. $I = E(\nabla_{\theta} \ln f(X; \theta_0) f(X; \theta_0)')$: 피셔정보(Fisher information)가 존재하고, 정칙(nonsingular)이다.
5. $E(\sup_{\theta \in N} \|\nabla_{\theta} f(X; \theta)\|) < \infty$ 이다.

$$\sqrt{n}(\hat{\theta} - \theta_0) \rightarrow N\left(0, \frac{1}{I(\theta_0)}\right).$$

우도함수가 미분 불가능해서 $\partial \rho(\theta)/\partial \theta = 0$ 을 계산 할 수 없는 경우, 또는 자료의 손실이 존재하는 경우 MLE를 직접적으로 계산 할 수 없다. 이러한 경우 반복적인(iterative) 계산을 통하여 구할 수 있다. 수치적 방법으로 MLE를 구하는 방법론은 Newton-Raphson 방법과 Fisher scoring 방법 (Jennrich와 Sampson, 1976; Givens와 Heiting, 2005), EM 알고리즘 (Dempster 등, 1977) 등이 있다.

2.2. NLS-2

비선형최소제곱법(NLS)은 경험적 누적분포함수(Empirical cumulative distribution function; ECDF)를 반응변수로, 이론적 누적분포함수(Theoretical cumulative distribution function)를 독립변수로 놓고, 잔차제곱합(residual sum of squares; RSS)를 최소화하는 모수를 추정하는 방법이다. NLS-2 방법은 2단계에 걸쳐 NLS 방법을 사용한 것으로, 데이터의 분포가 넓게 퍼져있는, 두꺼운 꼬리분포(heavy tailed distribution)의 모수를 추정하는 경우에 적합하다. 먼저, 경험적 누적분포함수와 이론적 누적분포함수에 각각 로그를 취해서 NLS 회귀분석을 실시한다. 그 결과 얻어진 추정치를 원래의 경험적 누적분포함수와 이론적 누적 분포함수의 적합 절차에 초기치로 사용하여 최종 추정치를 얻는다.

1. $F_n(x)$ 를 경험적 누적분포함수, $F(x)$ 를 이론적 누적분포함수라 하자. 각각 로그를 취해서 경험적 누적분포함수를 반응변수로, 이론적 누적분포함수를 독립변수로 하여 회귀식을 세운다.

$$\log(F_n(x_i)) = \log(F(x_i)) + \epsilon_i, \quad i = 1, \dots, n.$$

잔차제곱합을 최소화시키는 모수를 추정한다.

$$\hat{\theta}^{NLS1} = \arg \min_{\theta} \sum_{i=1}^n (\log(F_n(x_i)) - \log(F(x_i)))^2.$$

2. $\hat{\theta}^{NLS1}$ 를 초기치로 두었을 때, $F_n(x_i)$ 와 $F(x_i)$ 을 각각 반응변수와 독립변수로 한 회귀분석 식의 잔차제곱합을 최소화하는 모수를 추정한다.

$$\hat{\theta}^{NLS2} = \arg \min_{\theta} \sum_{i=1}^n (F_n(x_i) - F(x_i))^2.$$

이 때 얻어진 추정치 $\hat{\theta}^{NLS2}$ 를 NLS-2 추정치라 한다.

잔차제곱합이 최소가 되는 추정량을 얻기 위해 최적화 방법으로 R의 “optim” 함수를 사용한다. 이 함수는 최적화 문제에서 가장 많이 사용되고 있다. 본 논문에서는 Song과 Song (2011)에서 사용한 바와 같이 디폴트 최적화 알고리즘인 “Nelder-Mead” (Nelder와 Mead, 1965)를 사용하였다. 이 NLS-2 모수추정방법론은 일반화 파레토 분포를 이용한 POT 방법 (McNeil과 Saladin, 1997)에서 분포가 두꺼운 꼬리를 가지는 경우에 가장 좋은 성능을 보여주었다.

Table 3.1. RMSE for Burr distribution ($c = 1, k = 0.05, 0.5, 0.1, 1$)

$c = 1$		\hat{k}				\hat{c}			
$n \setminus k$		0.05	0.1	0.5	1	0.05	0.1	0.5	1
10	MLE	0.046970	0.084835	0.253475	0.465075	6.127164	10.58759	12.15393	4.281369
	NLS2	0.143761	0.146727	0.263571	0.500559	1.077350	1.856663	2.701746	1.542783
25	MLE	0.037046	0.059599	0.144267	0.240938	4.754534	6.382732	0.749314	0.226920
	NLS2	0.087224	0.091468	0.161672	0.246598	1.187711	1.852230	0.880990	0.243912
50	MLE	0.028235	0.040504	0.099771	0.164187	3.142021	2.681809	0.201382	0.145829
	NLS2	0.058922	0.064813	0.113209	0.172311	1.200085	1.394026	0.237558	0.156750
100	MLE	0.019582	0.027432	0.069308	0.111647	1.436834	0.600239	0.130667	0.096042
	NLS2	0.040596	0.046809	0.079493	0.118953	1.021705	0.873140	0.152710	0.106858
200	MLE	0.013355	0.018822	0.049018	0.079074	0.490935	0.219291	0.086973	0.066205
	NLS2	0.028751	0.033442	0.056335	0.084968	0.750708	0.462103	0.102148	0.073088
500	MLE	0.008716	0.011994	0.030596	0.049779	0.193185	0.121580	0.053621	0.041957
	NLS2	0.017889	0.021569	0.035480	0.053311	0.415212	0.234882	0.063502	0.046643

3. 모의 실험

모의실험에서는 여러 가지 분포에 대한 모의 자료를 생성하고, 각 경우에 대해서 NLS-2 추정치와 MLE의 성능을 비교해 보았다. 모의실험을 진행한 분포는 Burr 분포, 코시분포(Cauchy distribution), 프레셰 분포(Frechet distribution)이다. 모의실험의 과정은 다음과 같다.

1. 독립이고 동등한 분포를 따르는 표본을 각각 10, 25, 50, 100, 200, 500개 추출한다.
2. 선택된 표본들에서 NLS-2 방법과 ML 방법을 사용하였을 때의 추정치를 각각 얻는다.
3. 단계 1, 2를 10,000번 반복하여 평균제곱근오차(RMSE)를 구한다.

3.1. Burr 분포

Burr 분포는 2개의 형상모수 k, c 를 가지고 있다.

$$f(x; k, c) = kc \frac{x^{c-1}}{(1+x^c)^{k+1}}, \quad k, c > 0.$$

burr 분포는 어떠한 수명 단봉 자료에서도 적용 가능하다는 장점을 가지고 있어, 공학, 생존 분석, 산업 과학 그리고 보험료 측정 시 주로 사용되고 있다 (Dasgupta, 2011; Todd 등, 2010; Wingo, 1983).

Table 3.1에서는 c 가 1로 고정 되어 있을 때 k 값을 0.05, 0.1, 0.5, 1로 변화 시켜가면서 모의실험을 진행하였다. 표 안에 있는 숫자는 평균 제곱근 오차를 나타내며, 각각의 경우에 대해 첫 번째 줄에 있는 숫자는 mle에 관한 결과이고, 두 번째 줄에 있는 숫자는 nls-2에 관한 결과이다. 예를 들어 Table 3.1에서 가장 왼쪽 위에 있는 0.046970는 k 가 0.05, c 는 1인 Burr 분포에서 10개의 표본을 추출하여 k 에 대한 MLE를 계산하는 것을 10000번 반복해서 나온 RMSE 값이다. 6.127164는 같은 방법으로 c 에 대한 MLE를 계산해서 나온 RMSE 값이다. Table 3.2는 반대로 모수 k 가 1일 때 모수 c 를 0.1, 3, 5, 10으로 변화시켜가면서 모의 실험을 진행한 결과이다. 나머지는 Table 3.1과 동일하다.

c 가 1로 고정되어 있는 경우, 표본의 수가 10개일 때에는 \hat{c} 의 RMSE가 모든 경우에서 NLS-2 방법이 MLE보다 월등하게 작았다. 표본의 수가 25, 50개 일 때에도, c 가 1로 고정되어 있을 때, k 값이 0.05, 0.1인 경우, \hat{c} 의 추정을 NLS-2 방법론이 훨씬 잘하였다. 표본의 수가 증가할 수록, 두 방법 모두 실

Table 3.2. RMSE for Burr distribution ($c = 0.1, 3, 5, 10, k = 1$)

$k = 1$	$n \setminus c$	\hat{k}				\hat{c}			
		0.1	3	5	10	0.1	3	5	10
10	MLE	0.456600	0.468554	0.451072	0.415376	0.636282	11.16739	2.502470	33.15333
	NLS2	0.432449	0.454085	0.417414	0.615075	0.145621	5.041943	6.660558	12.26294
25	MLE	0.236900	0.231670	0.239659	0.235361	0.022214	0.682381	1.059827	2.059414
	NLS2	0.243868	0.242304	0.243207	0.237998	0.022254	0.731130	1.115363	2.184877
50	MLE	0.157307	0.161673	0.158402	0.159556	0.013793	0.432154	0.711455	1.468682
	NLS2	0.164464	0.166789	0.168400	0.164158	0.015185	0.454781	0.758268	1.56061
100	MLE	0.108763	0.109403	0.106856	0.116280	0.009446	0.284582	0.480871	0.998302
	NLS2	0.116535	0.116463	0.115078	0.122001	0.010114	0.308573	0.530691	1.044763
200	MLE	0.079066	0.078969	0.078329	0.079121	0.006703	0.199055	0.332406	0.663842
	NLS2	0.084290	0.083943	0.083813	0.084029	0.007435	0.219570	0.368382	0.735240
500	MLE	0.048538	0.049925	0.050304	0.050352	0.004125	0.126062	0.215452	0.415856
	NLS2	0.051060	0.053628	0.054005	0.053927	0.004644	0.139255	0.237506	0.472037

계 값에 근사한 추정을 하여 RMSE가 작아졌으며, 그 속도가 MLE의 경우가 더 빨라서 표본의 수가 100일 때에는 $c = 1, k = 0.05$ 인 경우 이외에는 MLE의 성능이 더 좋았다. 또한, 표본이 200개 이상일 때에는 MLE가 모든 경우에 대해 NLS-2보다 좋았다. 하지만, c 추정시 표본의 수가 클 때, 두 추정량의 RMSE값의 차이는 작은 반면, 표본의 수가 작을 때에는 NLS-2가 MLE 보다 월등하게 잘했다. 모의실험 결과, MLE의 경우, 표본의 수가 작은 경우에는 참 값에서 굉장히 많이 떨어져 있는 값으로 추정하기도 하였다. 예를 들어, 표본의 수가 10개 일 때 $c = 1, k = 0.05$ 인 경우 NLS-2의 RMSE가 MLE 보다 6배가량 작게 나타났다. 반면에 k 에 대한 추정량은 거의 대부분의 경우에 MLE의 성능이 NLS-2 보다 좋게 나타났다.

k 가 고정되었을 경우 또한, 표본의 개수가 10개일 때에는 k 와 c 추정 시 모두, NLS-2의 성능이 MLE 보다 좋았다. 표본의 수가 클 때에는 NLS-2와 MLE 간에 큰 차이가 나타나지 않는 것에 비해, 표본의 수가 작을 때에는 NLS-2 방법론이 MLE 보다 훨씬 근사하게 참값을 추정하였다.

3.2. 프레셰 분포(Frechet distribution)

프레셰 분포는 일반화 극단치 분포(Generalized extreme value distribution; GEV)의 하나로 형상모수 ξ (shape parameter)와 위치 모수 μ (location parameter), 척도모수 σ (scale parameter)를 가지고 있다 (Abd-Elfattah와 Omima, 2009). ξ 값이 클수록 두꺼운 꼬리분포를 갖는다. Song과 Song (2011) 논문에서와 같이 위치모수는 우리의 관심영역이 아니므로 0으로 두고 실험을 진행하였다.

$$F(x; \theta, \xi, \mu) = e^{-\left(1 - \xi \left(\frac{x - \mu}{\sigma}\right)\right)^{-\frac{1}{\xi}}}, \quad 1 - \xi \left(\frac{x - \mu}{\sigma}\right) > 0.$$

Table 3.3은 형상모수가 2이고, 척도모수는 각각 0.1, 1, 5, 10일 때의 결과이다. 각각의 경우에 대해 첫 번째 줄은 mle에 대한 결과이고, 두 번째 줄은 nls-2에 대한 결과이다. Table 3.4는 반대로 척도모수가 1일 때, 형상모수를 0.2, 0.5, 1, 3으로 바꿔가며 실험한 결과이다. 나머지는 Table 3.3과 같다.

Table 3.3의 결과 모든 경우에 대해 NLS-2 방법론을 사용하였을 때의 평균제곱근오차 값이 최대 우도 추정 방법을 사용하였을 때보다 작게 나타났다. 또한 Table 3.4에서도 MLE보다 NLS-2의 성능이 더 좋은 경우가 많았다. 특히, 척도모수가 1일 때 형상모수가 1 이상인 경우에 대해서는 대부분 NLS-2가 더 참값에 가까웠다. 척도모수가 고정되어 있을 때, ξ 값이 작을 때는 ML 방법이 NLS-2 방법보다 더

Table 3.3. RMSE for Frechet distribution ($\xi = 2, \sigma = 0.1, 1, 5, 10$)

$\xi = 2$		$\hat{\sigma}$				$\hat{\xi}$			
$n \setminus \sigma$		0.1	1	5	10	0.1	1	5	10
10	MLE	1209839	11635072	281773.2	114993	3.023063	4.174457	4.252675	4.41018
	NLS2	0.327653	1.825353	16.65919	19.31991	1.743241	1.270816	1.51692	1.51593
25	MLE	499.7174	832.4911	26621.91	10548.19	1.178201	1.271968	1.406737	1.544023
	NLS2	0.089315	0.358411	1.206336	2.865569	0.776794	0.746086	0.735103	0.718412
50	MLE	7.545466	939.0829	309.4099	119.2557	0.569418	0.609804	0.581187	0.616264
	NLS2	0.018295	0.168471	0.807894	1.63844	0.521240	0.509060	0.490990	0.487384
100	MLE	2.339807	20.16039	14.55087	47.84729	0.384974	0.439101	0.384436	0.394686
	NLS2	0.012635	0.113552	0.571048	1.149248	0.351836	0.324456	0.328351	0.329407
200	MLE	0.114386	3.763678	36.03459	102.9583	0.409124	0.393661	0.396735	0.43787
	NLS2	0.008785	0.077691	0.395415	0.767146	0.237803	0.226949	0.224578	0.221492
500	MLE	0.350933	17.56801	4.796533	14.71905	0.402904	0.418483	0.403948	0.434897
	NLS2	0.005115	0.047876	0.236549	0.491466	0.145135	0.137858	0.138392	0.142923

Table 3.4. RMSE for Frechet distribution ($\xi = 0.2, 0.5, 1, 3, \sigma = 1$)

$\sigma = 1$		$\hat{\sigma}$				$\hat{\xi}$			
$n \setminus \xi$		0.2	0.5	1	3	0.2	0.5	1	3
10	MLE	0.352663	0.540314	20.7705	801236812	0.705035	1.24583	2.210809	6.137367
	NLS2	0.88976	0.919504	1.31615	2.68222	0.656349	0.732367	0.891679	1.927381
25	MLE	0.18303	0.23176	0.341115	12577.79	0.217622	0.242652	0.372135	1.786802
	NLS2	0.26905	0.30408	0.337078	0.26645	0.317141	0.396664	0.521734	0.988366
50	MLE	0.134418	0.160606	0.236293	237199	0.136363	0.160769	0.208546	1.402819
	NLS2	0.161397	0.146900	0.174178	0.197617	0.232795	0.280618	0.361083	0.704654
100	MLE	0.091243	0.112824	0.153576	28415.41	0.085396	0.107174	0.134678	1.461909
	NLS2	0.105405	0.102037	0.100646	0.140260	0.162789	0.202548	0.242952	0.477081
200	MLE	0.063919	0.079067	0.106366	123.4869	0.058954	0.073104	0.094573	1.510718
	NLS2	0.072773	0.068104	0.069162	0.091743	0.119093	0.141160	0.167914	0.312797
500	MLE	0.040614	0.049950	0.091759	95.98339	0.036055	0.042759	0.073769	1.609991
	NLS2	0.045094	0.045157	0.044215	0.057810	0.074641	0.087341	0.113629	0.196084

참값에 유사한 추정을 한다. 하지만 ξ 값이 커질수록, 즉 상대적으로 두꺼운 꼬리 분포를 가질 때에는 MLE의 일관성이 성립하지 않는다. 특히 척도모수를 추정할 때에는 MLE의 RMSE가 매우 크게 나타나며, 표본의 수가 500으로 클 때에도 $\xi = 3, \sigma = 1$ 일 때 척도모수에 대한 RMSE 값이 95.98339로 굉장히 크게 나타났다. 반면 그에 비해, NLS-2 방법론은 두꺼운 꼬리 분포일 때에도 유사한 추정치를 제공하며, MLE에 비해 월등하게 낮은 RMSE 값을 나타냈다.

3.3. 코시 분포(Cauchy distribution)

코시분포는 꼬리가 두꺼운 분포 중 하나로, 위치모수 θ (location parameter)와 척도모수 λ (scale parameter)를 가지고 있다.

$$f(x; \theta, \lambda) = \frac{1}{\lambda \left(1 + \left(\frac{x - \theta}{\lambda} \right)^2 \right)}$$

Table 3.5. RMSE for Cauchy distribution ($\theta = 0, \lambda = 0.1, 1, 10, 100$)

$\theta = 0$		$\hat{\theta}$				$\hat{\lambda}$			
$n \setminus \lambda$		0.1	1	10	100	0.1	1	10	100
10	MLE	0.053559	0.537291	5.270496	52.5554	0.052209	0.526825	5.340631	54.4628
	NLS2	4.772902	27.30024	111.6494	1882.199	33.20395	189.7635	780.8036	13091.01
25	MLE	0.030582	0.302836	3.035478	29.93974	0.029979	0.301680	3.028429	30.22494
	NLS2	0.035032	0.351691	3.52708	34.79186	0.035317	0.357360	3.587513	35.83997
50	MLE	0.020549	0.206786	2.044716	20.86388	0.020530	0.204849	2.042532	20.5358
	NLS2	0.023508	0.235477	2.353727	23.61669	0.022196	0.222406	2.226494	22.35854
100	MLE	0.014360	0.144111	1.445856	14.66646	0.014188	0.143243	1.436995	14.33894
	NLS2	0.016229	0.165619	1.644603	16.50566	0.014754	0.149334	1.501317	14.90891
200	MLE	0.009998	0.09962	0.999069	10.04669	0.010082	0.10002	1.007531	10.08553
	NLS2	0.011421	0.113266	1.141059	11.34303	0.010291	0.102089	1.028919	10.26559
500	MLE	0.006313	0.062514	0.637424	6.536671	0.006363	0.063044	0.631187	6.410645
	NLS2	0.007140	0.071184	0.724757	7.155302	0.006417	0.06366	0.636753	6.423471

Table 3.6. RMSE for Cauchy distribution ($\theta = 0.1, 1, 10, 100, \lambda = 1$)

$\lambda = 1$		$\hat{\theta}$				$\hat{\lambda}$			
$n \setminus \theta$		0.1	1	10	100	0.1	1	10	100
10	MLE	0.535001	0.53860	0.537774	0.523542	0.55168	0.544060	0.541278	0.551450
	NLS2	2.132007	39.98436	15.73881	162.3274	14.06759	278.0601	109.3361	1129.599
25	MLE	0.30339	0.299760	0.304287	0.303743	0.30447	0.302052	0.305073	0.304109
	NLS2	0.350729	0.350378	0.355245	0.351949	0.359718	0.356067	0.361005	0.36015
50	MLE	0.204012	0.204916	0.204197	0.205952	0.203653	0.207832	0.205621	0.205686
	NLS2	0.233851	0.235789	0.232129	0.236130	0.222105	0.226149	0.223423	0.223143
100	MLE	0.143576	0.143085	0.143847	0.14132	0.142642	0.143370	0.145179	0.144097
	NLS2	0.163057	0.161650	0.164966	0.162414	0.148335	0.149406	0.151981	0.149925
200	MLE	0.100298	0.100748	0.100977	0.100878	0.100481	0.101033	0.099660	0.100586
	NLS2	0.113536	0.114625	0.115974	0.115268	0.102674	0.103067	0.101722	0.102798
500	MLE	0.062567	0.063930	0.063446	0.062993	0.062100	0.062645	0.063253	0.062838
	NLS2	0.070816	0.072261	0.072129	0.071644	0.062694	0.063166	0.063793	0.063495

Table 3.5는 위치모수가 0일 때에 대해 척도모수가 각각 0.1, 1, 10, 100인 경우이다. 각각의 경우에 대해 위의 줄은 mle에 대한 결과를, 아래 줄은 nls-2에 대한 결과를 나타낸다.

Table 3.6은 척도모수가 1일 때에 대해 위치모수가 0.1, 1, 10, 100으로 변하는 경우이다. 나머지는 Table 3.5와 동일하다.

대부분의 경우, NLS-2 추정량이 MLE보다 좋지 못하였다. 특히 표본의 수가 10개로 매우 작은 경우에는 NLS-2를 사용하였을 때의 RMSE가 매우 크게 나타나, NLS-2가 참값을 제대로 추정하지 못한다는 것을 알 수 있다. 또한 척도모수의 값이 증가함에 따라 두 가지 방법론을 사용하였을 때 모두 오차가 증가하였다. 하지만 표본의 수가 증가하면 두 방법론이 아주 유사한 성능을 보여주는 것을 볼 수 있다.

4. 결론

본 논문에서는 Song과 Song (2011) 논문에서 제시한 새로운 모수 추정 방법인 NLS-2를 일반화파레토 분포가 아닌 다른 다양한 분포에 적용시켜 보았다. 많은 경우 MLE가 NLS-2보다 성능이 좋았으나, 표

본의 수가 작을 때의 Burr 분포에서, 그리고 ξ 값이 클 때, 프레세 분포에서는 NLS-2의 성능이 더 좋게 나타났다. 또한 MLE가 성능이 더 좋은 경우라고 하더라도 NLS-2의 결과도 상당히 근접해 있는 것을 볼 수 있다.

코시 분포의 경우 일반적으로 척도모수가 알려졌을 때, 위치모수에 대한 우도함수가 다봉분포(multimodal distribution)이므로 위치모수의 MLE를 찾는 데 어려움이 있다고 알려져 있다 (Johnson 등, 1994). 하지만 모의실험 결과, 수치적 방법으로 MLE를 구할 경우 척도모수의 값이 커질수록 오차 값이 커지기는 하나, 일정한 추정 값을 제시하였다. 또한 NLS-2 방법을 사용하였을 때보다 항상 오차가 작게 나타났으며, 수치적인 방법론을 이용한 MLE가 꽤 정확한 추정량을 제공한다는 것을 알 수 있었다. 그 이외에도 와이블 분포, 로그 로지스틱 분포, 로그 감마분포에 대해 적용해본 결과 NLS-2의 성능이 MLE 보다 좋지 못하였다.

우리는 표본의 수가 작은 경우에 NLS-2 방법론의 성능을 높이기 위하여 경험적 누적분포 함수에 여러 가지 평활법(smoothing method)을 시도해 보았다 (Fernholz, 1991). 먼저 주어진 관측치를 이용해서 KDE(kernel density estimation) 방법을 이용하여 분포함수를 추정하고, 수치적 방법론을 이용하여 경험적 누적분포 함수를 다시 만들었다. 그리고 NLS-2 방법론을 적용해 보았다. 두 번째는 경험적 누적분포 함수를 스플라인 함수(spline function)를 이용하여 평활한(smooth) 경험적 누적분포함수로 변환한 후에 NLS-2 방법론을 적용시켰다. 두 가지 방법 모두 표본의 수가 작은 경우에 우리가 임의로 표본의 수를 늘릴 수 있다는 점에서 기존의 경험적 누적분포 함수를 이용한 NLS-2 보다 좋은 결과를 가져 올 수 있을 것이라 예상했다. 하지만 모의실험 결과 모두 기존의 NLS-2 방법론보다 성능이 좋지 못하였다. 그 이유는 역시, 주어진 관측치보다 더 많은 정보를 만들 수는 없기 때문이라고 생각한다. 평활법을 사용하여 관측치를 늘려도 결국 정보량이 늘어나는 것이 아니고 오히려 노이즈만 더 늘어나서 결과가 좋지 않게 나오는 것이라고 판단된다.

우리는 본 논문을 통해서 우리가 제안한 방법이 두터운 꼬리를 가지는 분포의 경우에 표본의 수가 작으면 MLE보다 나은 성능을 보일 수 있다는 것을 보였다. 하지만 많은 경우에 이론적인 결과가 말해주듯이 MLE의 성능이 가장 우수한 것을 볼 수 있었다.

References

- Abd-Elfattah, A. M. and Omima, A. M. (2009). Estimation of the Unknown parameters of the generalized frechet distribution, *Journal of Applied Sciences Research*, **5**, 1398–1408.
- Dasgupta, R. (2011). On the distribution of Burr with applications, *Indian Statistical Institute 2011*.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, **39**, 1–39.
- Fernholz, L. (1991). Almost sure convergence of smoothed empirical distribution functions, *Scandinavian Journal of Statistics*, **18**, 255–262.
- Givens, G. H. and Hoeting, J. A. (2005). *Computational Statistics*, John Wiley & Sons, New Jersey.
- Hogg, R. V., Craig, A. and McKean, J. W. (2012). *Introduction to Mathematical Statistics*, Pearson.
- Jennrich, R. I. and Sampson, P. F. (1976). A Newton-Raphson and related algorithms for maximum likelihood variance component estimation, *Technometrics*, **18**, 11–17.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, John Wiley & Sons, New York.
- McNeil, A. J. and Saladin, T. (1997). The Peaks over thresholds method for estimating high quantiles of loss distribution, *Proceedings of 28th international ASTIN Colloquium*.
- Nelder, J. A. and Mead, R. (1965). A simplex algorithm for function minimization, *Computer Journal*, **7**, 308–313.

- Song, J. and Song, S. (2011). A quantile estimation for massive data with generalized Pareto distribution, *Computational Statistics and Data Analysis*, **56**, 143–150.
- Todd, C., Headrick, M. D. P. and Yanyan, S. (2010). On Simulating Univariate and Multivariate Burr Type III and Type XII Distributions, *Applied Mathematical Sciences*, **4**, 2207–2240.
- Wingo, D. R. (1983). Maximum likelihood methods for the burr type XII distribution to life test data, *Biometrical Journal*, **25**, 77–84.

비선형 최소제곱법을 이용한 모수추정 방법론

오선아^a · 송종우^{a,1}

^a이화여자대학교 통계학과

(2013년 3월 5일 접수, 2013년 4월 23일 수정, 2013년 5월 14일 채택)

요약

우리는 두꺼운 꼬리를 갖는 분포의 모수를 추정하는 방법론을 연구하였다. 일반적으로 MLE(최대우도 추정량)가 모수추정 방법론중에 가장 많이 사용되는데, 이는 MLE가 점근적 일치성과 정규성 그리고 효율성을 가지고 있기 때문이다. 하지만 MLE가 늘 가장 좋은 추정법은 아니다. 어떤 경우에는 MLE가 존재하지 않을 수도 있고 계산이 안정적이지 않을 수도 있다. 본 논문에서는 비선형 최소제곱추정법을 이용한 모수추정 방법론을 제시하고 그 성능을 MLE와 비교하였다. NLS 추정량은 empirical CDF와 이론적 CDF의 차이의 제곱을 최소화 하는 방법론이다. 본 논문에서는 두꺼운 꼬리를 가지는 다양한 분포하에서 우리가 제안하는 NLS방법론과 MLE와의 성능을 비교하였다. 그 결과, Burr 분포에서 표본의 수가 적을 때 우리의 방법론이 MLE보다 좋은 성능을 보여주었고, Frechet 분포에서도 좋은 결과를 얻을 수 있었다.

주요용어: 모수 추정, 최대우도 추정량, 최소제곱법.

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업이다 (No. 2012-0003739)(송종우).

¹교신저자: (120-750) 서울시 서대문구 이화여대길 52, 이화여자대학교 통계학과, 부교수.

E-mail: josong@ewha.ac.kr