

## MDPDE의 조율모수 선택에 관한 연구<sup>†</sup>

유동현<sup>1</sup> · 김병수<sup>2</sup>

<sup>1</sup>계명대학교 통계학과 · <sup>2</sup>영남대학교 통계학과

접수 2015년 4월 20일, 수정 2015년 5월 19일, 게재확정 2015년 5월 21일

### 요약

MDPDE는 이상치에 강건한 성질을 가진 추정량으로써 최대우도추정량의 대안으로 많은 연구자들에 의해 연구되어 왔다. MDPDE는 조율모수에 따라 성질이 변하게 되는데, 로버스트성과 점근 효율성이 서로 상충하는 현상으로 인해 최적의 조율모수를 선택하는 것은 쉽지 않다. 본 연구에서는 MDPDE의 최적의 조율모수를 선택하는 방법으로 Fujisawa와 Eguchi (2006)가 제시한 방법과 Warwick (2006)이 제시한 방법을 소개하고, 모의실험을 통해 비교분석하였다. 연구 결과 Warwick (2006)의 방법은 특정한 경우 매우 작은 조율모수를 선택하게 될 수도 있다는 사실을 알 수 있었는데, 같은 경우에 Fujisawa와 Eguchi (2006)의 방법은 이러한 현상을 보이지 않았다. 따라서, Fujisawa와 Eguchi (2006)의 방법이 범용적으로 사용하기에 적절하다고 판단된다.

주요용어: 로버스트성, 점근효율성, 조율모수, 최소거리추정량.

### 1. 서론

최대우도추정량 (maximum likelihood estimator)은 다양한 분야의 통계적 추론에서 가장 널리 사용되고 있는 추정량이다. 최대우도추정량은 점근적으로 효율적 (asymptotically efficient)이라는 장점이 있지만, 이상치에 매우 취약한 단점 또한 갖고 있다. 이러한 최대우도추정량의 단점을 보완하기 위하여 많은 연구자들이 최소거리추정량 (minimum distance estimator)을 고안하고, 그 성질을 규명하였다 (Beran, 1977; Tamura와 Boos, 1986; Simpson, 1987; Basu와 Lindsay, 1994; Cao 등, 1995). 최소거리추정량은 이상점에 강건한 성질인 로버스트성 (robustness)을 가지기 때문에, 최대우도추정량의 좋은 대안이 될 수 있다. 최소거리추정량 중 특히 Basu 등 (1998)이 제안한 MDPDE (minimum density power divergence estimator)는 밀도함수를 기반으로 한 다이버전스 (divergence)를 이용하여 도출되는 추정량으로써 강한 로버스트성을 가질 뿐만 아니라, 앞서 언급된 연구들에서의 추정량과는 달리 자료의 참 밀도함수 (true density function)를 추정할 필요가 없어 띠너비 선택 (bandwidth selection)과 같은 참 밀도함수를 추정하는 과정에서 만날 수 있는 어려움들을 피해갈 수 있는 장점이 있다. 이러한 장점들로 인해 MDPDE는 많은 연구자들의 관심을 받고 연구되어 왔다. 예를 들면, Mattheou 등 (2009)은 모형 선택 문제에, Mollah 등 (2006, 2007)은 독립성분분석 (independent component analysis)에 MDPDE를 이용하였다. 또한, 시계열자료 분석에도 응용되어 Lee와 Na (2005)는 시계열 자료의 변화점 검정법을 개발하는데 이 추정량을 이용하였고, Kim과 Lee (2011, 2013)는 다변량 시계열자료의 공분산 행렬을 추정하는데 MDPDE를 적용하였다. MDPDE는 조율모수 (tuning parameter)에 의해 그 성질이 다르게 나타나는데, 일반적으로 조율모수 값이 커지면 로버스트성은 강해지는 반면, 최대우도추정량

<sup>†</sup> 이 연구는 2014년도 영남대학교 학술연구조성비에 의한 것임 (214A054005).

<sup>1</sup> (704-701) 대구광역시 달서구 달구벌대로 1095, 계명대학교 통계학과, 조교수.

<sup>2</sup> 교신저자: (712-749) 경상북도 경산시 대학로 280, 영남대학교 통계학과, 조교수. E-mail: bkim@yu.ac.kr

에 비해 점근효율성 (asymptotic efficiency)은 떨어지게 된다. 이와 같이, 조율모수 값에 따라 로버스트성과 점근효율성이 서로 상충 (trade-off)하는 현상이 나타나기 때문에, 자료 분석시에 만족할만한 로버스트성과 점근효율성을 동시에 갖게 하는 최적의 조율모수 (optimal tuning parameter)를 선택하는 것은 쉬운 일이 아니라고 할 수 있다. MDPDE는 특정한 모수적 분포족 (parametric distribution family)을 차용하여 모수를 추정하게 되는데 (자세한 내용은 2절에 소개), Fujisawa와 Eguchi (2006)는 자료의 참 밀도함수와 고려된 모수적 분포족의 차이를 Cramer-von Mises 다이버전스로 측정하고, 여기에 교차검증 (cross-validation) 아이디어를 도입하여 최적의 조율모수를 추정하는 방법을 제안하였다. 한편, Warwick (2005)은 점근평균제곱오차 (asymptotic mean squared error)를 이용하여 로버스트성과 점근효율성을 동시에 측정하는 측도를 제시하고, 이를 최소화 하는 값을 최적의 조율모수로 제안하였다. 본 연구에서는 MDPDE의 조율모수 선택에 대한 방법을 제시한 두 연구를 모의실험을 통해 비교분석하여 연구자들이 MDPDE를 사용할 때 최적의 조율모수를 선택하는데 도움이 되고자 한다.

본 논문의 구성은 다음과 같다. 2절에서는 MDPDE와 조율모수를 선택하는 두 연구에 대해 소개한다. 3절에서는 추정할 모수가 한 개인 경우와 두 개인 경우에 대한 모의실험을 통해 조율모수를 선택하는 두 방법을 비교분석한다. 마지막으로 4절에서는 본 연구의 결론에 대해 언급한다.

## 2. MDPDE와 조율모수의 선택 방법

### 2.1. MDPDE와 점근분포

미지의 모수  $\theta \in \Theta \subset \mathbb{R}^m$ 와 르벡 측도 (Lebesgue measure)에 대해 밀도함수  $\{f_\theta\}$ 를 갖는 모수적 분포족  $\{F_\theta\}$ 를 고려하자. 또한,  $\mathcal{G}$ 를 르벡 측도에 대해 밀도함수를 갖는 모든 분포들의 모임이라 하자. Basu 등 (1998)은 미지의 모수  $\theta$ 를 추정하기 위해 먼저 다음과 같은 DPD (density power divergence)  $d_\alpha$ 를 제안하였다.

$$d_\alpha(g, f) := \begin{cases} \int \{f^{1+\alpha}(z) - (1 + \frac{1}{\alpha})g(z)f^\alpha(z) + \frac{1}{\alpha}g^{1+\alpha}(z)\} dz, & \alpha > 0, \\ \int g(z)(\log g(z) - \log f(z)) dz, & \alpha = 0, \end{cases}$$

단,  $f$ 와  $g$ 는 밀도함수이고,  $\alpha$ 는 조율모수이다. 이를 이용하면 모든  $\mathcal{G}$ 의 원소  $G$ 에 대하여 DPD를 최소화 하는 범함수 (functional)  $T_\alpha(G)$ 를 다음과 같이 정의할 수 있다.

$$d_\alpha(g, f_{T_\alpha(G)}) = \min_{\theta \in \Theta} d_\alpha(g, f_\theta),$$

단,  $g$ 는  $G$ 의 밀도함수이다. 여기서 주목할 점은, 만약  $G$ 가  $\{F_\theta\}$ 에 속하면  $\Theta$ 의 특정 원소  $\theta$ 에 대하여  $T_\alpha(G) := \theta_\alpha = \theta$ 가 만족된다는 것이다. 이들을 바탕으로 Basu 등 (1998)은 분포  $G$ 를 따르는 임의표본  $X_1, \dots, X_n$ 이 주어졌을 때 MDPDE  $\hat{\theta}_{\alpha, n}$ 을 다음과 같이 정의하였다.

$$\hat{\theta}_{\alpha, n} = \operatorname{argmin}_{\theta \in \Theta} H_{\alpha, n}(\theta),$$

단,  $H_{\alpha, n}(\theta) = \frac{1}{n} \sum_{i=1}^n V_\alpha(\theta; X_i)$  이고,

$$V_\alpha(\theta; x) = \begin{cases} \int f_\theta^{1+\alpha}(z) dz - (1 + \frac{1}{\alpha})f_\theta^\alpha(x), & \alpha > 0, \\ -\log f_\theta(x), & \alpha = 0. \end{cases} \quad (2.1)$$

식 (2.1)로부터  $\alpha = 0$ 이면 MDPDE는 최대우도추정량이 되고,  $\alpha = 1$ 이면  $L^2$ -거리추정량 ( $L^2$ -distance estimator)이 됨을 알 수 있다. 이 사실로 미루어보면, 조율모수  $\alpha$ 가 증가할수록 로버스트성은 강해지지만 점근효율성은 떨어짐을 짐작할 수 있고, 실제로 MDPDE의 로버스트성과 점근효율성

은 서로 상충하는 현상을 보인다. 너무 높은  $\alpha$  값은 이상치에 강건한 성질은 가지겠지만 점근효율성이 떨어지게 되고, 너무 작은  $\alpha$  값은 그 반대의 현상을 보이게 되므로 로버스트성과 점근효율성을 동시에 고려한 최적의 조율모수 선택 문제는 매우 의미있는 연구주제라고 할 수 있다.

Basu 등 (1998)의 연구에 따르면 MDPDE  $\hat{\theta}_{\alpha,n}$ 은  $\theta_\alpha$ 의 일치추정량이고, 다음과 같은 점근분포를 갖는다.

$$\sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_\alpha) \xrightarrow{d} N(0, J^{-1}KJ^{-1}),$$

$$\begin{aligned} \text{단, } J = J(\theta_\alpha) &= \int u_{\theta_\alpha}(z)u_{\theta_\alpha}(z)'f_{\theta_\alpha}^{1+\alpha}(z)dz \\ &\quad + \int \{i_{\theta_\alpha}(z) - \alpha u_{\theta_\alpha}(z)u_{\theta_\alpha}(z)'\} \{g(z) - f_{\theta_\alpha}(z)\} f_{\theta_\alpha}^\alpha(z)dz, \end{aligned}$$

$$K = K(\theta_\alpha) = \int u_{\theta_\alpha}(z)u_{\theta_\alpha}(z)'f_{\theta_\alpha}^{2\alpha}(z)g(z)dz - \xi\xi',$$

$$\xi = \xi(\theta_\alpha) = \int u_{\theta_\alpha}(z)f_{\theta_\alpha}^\alpha(z)g(z)dz,$$

$$u_\theta(z) = \frac{\partial}{\partial \theta} \log f_\theta(z),$$

$$i_\theta(z) = -\frac{\partial}{\partial \theta} u_\theta(z).$$

일반적으로 MDPDE는 조율모수 뿐만 아니라 분포족의 선택에도 영향을 받지만, 본 연구의 목적인 조율모수에 따른 영향에 집중하기 위해 여기서는  $g \in \{f_\theta\}$ 인 경우만을 고려한다. 이 때, 참모수 (true parameter)를  $\theta_0$ 라 하면 MDPDE의 점근분포는 다음과 같이 나타낼 수 있다.

$$\sqrt{n}(\hat{\theta}_{\alpha,n} - \theta_0) \xrightarrow{d} N(0, J^{-1}KJ^{-1}),$$

$$\text{단, } J = J(\theta_0) = \int u_{\theta_0}(z)u_{\theta_0}(z)'f_{\theta_0}^{1+\alpha}(z)dz,$$

$$K = K(\theta_0) = \int u_{\theta_0}(z)u_{\theta_0}(z)'f_{\theta_0}^{1+2\alpha}(z)dz - \xi\xi',$$

$$\xi = \xi(\theta_0) = \int u_{\theta_0}(z)f_{\theta_0}^{1+\alpha}(z)dz.$$

## 2.2. 조율모수의 선택: Fujisawa와 Eguchi (2006)의 방법

Fujisawa와 Eguchi (2006)는 다이버전스와 교차검증을 이용한 최적의 조율모수 선택 방법을 제안하였다. 만일 자료의 참 밀도함수  $g$ 를 알고 있다면, 조율모수를 다음과 같은 방법으로 선택할 수도 있을 것이다.

$$\hat{\alpha} = \operatorname{argmin}_\alpha D(g(x), f_{\hat{\theta}_{\alpha,n}}(x)),$$

단,  $D$ 는 다이버전스이다. 편의상  $x_1, \dots, x_n$ 를 크기 순서대로 나열한 관측치라 하자. Fujisawa와 Eguchi (2006)는 다이버전스  $D$ 로 다음과 같은 경험적 형태 (empirical version)를 갖는 Cramer-von Mises 다이버전스를 고려하였다.

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{i-0.5}{n} - F_\theta(x_i) \right\}^2$$

여기에 교차검증의 아이디어를 적용하여 다음과 같이 조율모수를 선택하도록 제안하였다.

$$\begin{aligned}\hat{\alpha} &= \operatorname{argmin}_{\alpha} \left( \frac{1}{n} \sum_{i=1}^n \left\{ \frac{i-0.5}{n} - F_{\hat{\theta}_{\alpha,n}^{(-i)}}(x_i) \right\}^2 \right) \\ &:= \operatorname{argmin}_{\alpha} FE(\alpha),\end{aligned}$$

단,  $\hat{\theta}_{\alpha,n}^{(-i)}$ 는  $i$ 번째 관측치를 제외하고 구한 MDPDE이다. 즉, 최적의 조율모수는 위에서 정의된 함수  $FE(\alpha)$ 를 최소로 하는  $\alpha$ 로 정의된다. 이 방법은  $i = 1, \dots, n$ 에 대해  $\hat{\theta}_{\alpha,n}^{(-i)}$ 를 각각 구해야하므로 계산이 번거로울 수 있는데,  $\hat{\theta}_{\alpha,n}^{(-i)}$ 는 모든 관측치를 이용한 MDPDE와 영향력 함수 (influence function)를 이용한 식으로 근사하여 나타낼 수 있다. 자세한 내용은 Fujisawa와 Eguchi (2006)의 5절 내용을 참고하길 바란다.

### 2.3. 조율모수의 선택: Warwick (2005)의 방법

Warwick (2005)은 다음과 같은 점근평균제곱오차를 이용한 최적의 조율모수 선택 방법을 제시하였다.

$$\text{AMSE} = (\theta_{\alpha} - \theta_0)(\theta_{\alpha} - \theta_0)' + \text{As.var}(\hat{\theta}_{\alpha,n}),$$

단,  $\text{As.var}(\hat{\theta}_{\alpha,n})$ 는  $\hat{\theta}_{\alpha,n}$ 의 점근분산-공분산행렬로써, 2.1절에서 언급된 바와 같이  $J^{-1}KJ^{-1}$  형태를 갖는다. 또한,  $\theta_{\alpha}$ 의 추정량으로 MDPDE  $\hat{\theta}_{\alpha,n}$ 을 사용하고,  $\theta_0$ 의 추정량으로는  $L^2$ -거리추정량  $\hat{\theta}_{1,n}$ 을 사용하여 다음과 같은 점근평균제곱오차 추정량을 제시하고,

$$\text{AM}\hat{\text{SE}} = (\hat{\theta}_{\alpha,n} - \hat{\theta}_{1,n})(\hat{\theta}_{\alpha,n} - \hat{\theta}_{1,n})' + \text{As.var}(\hat{\theta}_{\alpha,n}), \quad (2.2)$$

$\text{AM}\hat{\text{SE}}$ 의 대각합 (trace)인  $\text{tr}(\text{AM}\hat{\text{SE}})$ 를 최소로 하는  $\alpha$ 를 최적의 조율모수로 제안하였다. 즉,  $\text{tr}(\text{AM}\hat{\text{SE}}) := \text{War}(\alpha)$ 로 정의하면 최적의 조율모수는 다음과 같이 선택된다.

$$\begin{aligned}\hat{\alpha} &= \operatorname{argmin}_{\alpha} \text{tr}(\text{AM}\hat{\text{SE}}) \\ &:= \operatorname{argmin}_{\alpha} \text{War}(\alpha)\end{aligned}$$

예를 들어, 모수가  $\mu$ 와  $\sigma$  두 개인 경우, 최적의 조율모수는 다음과 같이 선택된다.

$$\hat{\alpha} = \operatorname{argmin}_{\alpha} \{ (\hat{\mu}_{\alpha,n} - \hat{\mu}_{1,n})^2 + (\hat{\sigma}_{\alpha,n} - \hat{\sigma}_{1,n})^2 + \text{As.var}(\hat{\mu}_{\alpha,n}) + \text{As.var}(\hat{\sigma}_{\alpha,n}) \}$$

Warwick (2005)은 식 (2.2) 오른쪽 변의 첫 번째 항은 MDPDE  $\hat{\theta}_{\alpha,n}$ 의 로버스트성을 측정하는 측도로 쓰일 수 있고, 두 번째 항으로는 점근효율성을 평가해 볼 수 있어 점근평균제곱오차가 로버스트성과 효율성을 동시에 측정할 수 있는 측도로 사용될 수 있다고 주장하였다. 점근평균제곱오차의 추정량을 정의하는 과정에서  $\theta_0$ 의 추정량으로는 어떠한 로버스트성을 가진 추정량을 사용해도 되지만, Warwick (2005)은  $L^2$ -거리추정량인  $\hat{\theta}_{1,n}$ 를 사용하는 것이 가장 효과적이라고 주장하였다.  $\hat{\theta}_{1,n}$ 이 포함된 항이 MDPDE  $\hat{\theta}_{\alpha,n}$ 의 로버스트성을 측정하는 측도라는 점을 감안하면, 강한 로버스트성을 가진  $L^2$ -거리추정량을 참 모수의 추정량으로 사용하는 것은 합당하다고 생각된다.

### 3. 모의실험을 통한 두 방법의 비교

#### 3.1. 모수가 한 개인 경우: 지수분포

본 절에서는 모의실험을 통해 2절에서 언급한 두 방법을 비교분석하고자 한다. 먼저, 한 개의 모수를 추정하는 경우를 고려하는데, 이를 위해 참모수  $\theta_0$ 가 1인 지수분포 (Exp(1))로부터 100개의 자료를 생성하였다. 자료의 크기를 다르게 한 경우에도 연구 결과에 차이가 없었으므로 본 연구에서는 자료의 크기가 100인 경우만을 고려한다. 참고로 이 자료의 최댓값은 4.2003이고, 최솟값은 0.0071이다. 오염된 자료 (contaminated data)를 생성하기 위해 100개의 자료 중 일부를 임의로 선택하여 이상치 10으로 대체하였는데, 본 연구에서는 자료 중 1개, 5개, 10개의 자료를 이상치로 대체하는 3가지 경우를 고려하였다. 이상치로 선택된 10은 제3사분위수+1.5\*사분위수범위를 초과하는 값이다. 오염된 자료를 이용하여 MDPDE  $\hat{\theta}_{\alpha,n}$ 을 구하기 위해 식 (2.1)에 지수분포의 밀도함수  $f_{\theta}(x) = \theta \exp(-\theta x)$ 를 대입하면,  $\hat{\theta}_{\alpha,n}$ 를 구하기 위해서는 다음과 같은 추정방정식 (estimating equation)의 해를 구해야 함을 알 수 있다.

$$\theta = \frac{\frac{1}{n} \sum_{i=1}^n \exp(-\alpha \theta X_i) - \frac{\alpha}{(1+\alpha)^2}}{\frac{1}{n} \sum_{i=1}^n \exp(-\alpha \theta X_i) X_i}$$

그러나 주어진 추정방정식의 해석해 (analytic solution)를 구하는 것은 매우 어렵기 때문에 다음과 같은 방법으로 수치해 (numerical solution)를 구하였다. 먼저, 초기치  $\hat{\theta}_{\alpha,n,1}$ 을 최대로 추정량  $n / \sum_{i=1}^n X_i$ 으로 하고,  $\hat{\theta}_{\alpha,n,k+1}$ ,  $k = 1, 2, \dots$ 을 다음과 같이 구한다.

$$\hat{\theta}_{\alpha,n,k+1} = \frac{\frac{1}{n} \sum_{i=1}^n \exp(-\alpha \hat{\theta}_{\alpha,n,k} X_i) - \frac{\alpha}{(1+\alpha)^2}}{\frac{1}{n} \sum_{i=1}^n \exp(-\alpha \hat{\theta}_{\alpha,n,k} X_i) X_i}, \quad k = 1, 2, \dots$$

이 과정을 반복하여  $\hat{\theta}_{\alpha,n,k}$ 와  $\hat{\theta}_{\alpha,n,k+1}$ 이 매우 가까워지면 그 값을  $\hat{\theta}_{\alpha,n}$ 으로 한다. Warwick (2005)의 방법으로 최적의 조율모수를 구할 때 필요한 MDPDE의 점근분산은

$$\text{As.var}(\hat{\theta}_{\alpha,n}) = \frac{(1+\alpha)^2}{(1+\alpha^2)^2} \left\{ \frac{(1+\alpha)^4(1+4\alpha^2)}{(1+2\alpha)^3} - \alpha^2 \right\} \theta_0^2$$

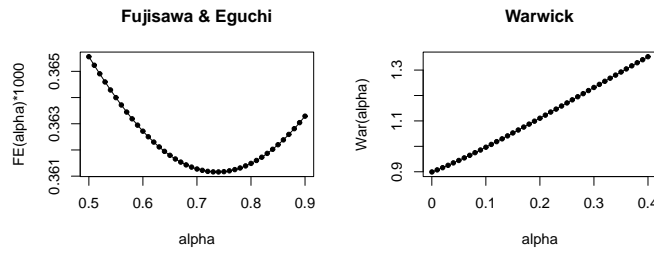
이고,  $\theta_0$ 의 추정량으로  $\hat{\theta}_{\alpha,n}$ 을 사용하였다. 또한, 최적의 조율모수를 구할시 대역적 최솟값 (global minimum)을 얻기 위해  $\alpha$ 의 초기치로 0부터 1까지 0.01씩 증가시킨 값들을 사용하였다.

Table 3.1은 Fujisawa와 Eguchi (2006)와 Warwick (2005)의 방법을 이용하여 선택된 최적의 조율모수와 그 조율모수에 해당하는 MDPDE를 보여준다. Figures 3.1~3.3은 각각 이상치가 1개, 5개, 10개인 경우에 2.2절과 2.3절에서 정의한 조율모수에 대한 함수  $FE(\alpha)$ ,  $War(\alpha)$ 를 다양한  $\alpha$ 에 대해 나타낸 그림이다. Table 3.1과 Figures 3.1~3.3을 보면, Warwick (2005)의 방법의 경우 이상치의 개수와 무관하게 최적의 조율모수가 0으로 결정됨을 알 수 있다. 이러한 현상의 원인은 Table 3.2를 보면 알 수 있다. Table 3.2는 이상치의 개수가 1개와 5개인 경우, 식 (2.2)에서 로버스트성을 측정하는 측도와 효율성을 측정하는 측도 (즉, 추정량의 점근분산)를 다양한  $\alpha$  값에 따라 나타낸 것인데, 효율성의 측도가 로버스트성의 측도에 비해 매우 큰 값을 나타냄을 알 수 있다. 따라서, 최적의 조율모수가 효율성의 측도인 추정량의 점근분산에 크게 기여하여 선택된 것이다. 이 결과로부터 Warwick (2005)의 방법은 고려된 분포족의 선택, 참모수, 이상치의 크기와 개수 등에 따라 어떤 경우에는 지나치게 작은 조율모수를 선택하게 될 수도 있다는 사실을 알 수 있었다. Table 3.1을 보면 최적의 조율모수에 해당하는 MDPDE의 값 또한 Fujisawa와 Eguchi (2006)의 방법이 참모수 1에 더욱 가까운 것을 알 수 있다. 이상치가 1개인 경우, Fujisawa와 Eguchi (2006)의 방법을 통해 얻은 최적의 조율모수는 0.74인데,

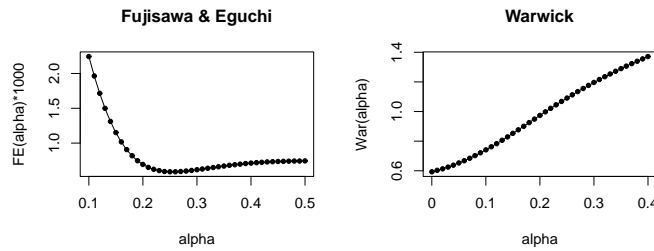
이 경우  $\alpha$ 에 따른  $FE(\alpha)$ 의 차이가 매우 미미하였다. 이전 연구들을 참조해 보면, Fujisawa와 Eguchi (2006)의 방법은 이상치의 개수가 많아지면 로버스트성이 강조되어 전반적으로 최적의 조율모수가 커지는 경향이 있다. 이는 3.2절에서 설명할 두 개의 모수를 추정하는 경우에서도 확인할 수 있다. Fujisawa와 Eguchi (2006)의 방법에 의해 구한 최적의 조율모수에 대응하는 MDPDE는 이상치의 개수와 관계 없이 모든 경우에서 최대우도추정량보다 참모수에 가깝게 추정되나, 이상치의 개수가 많아질수록 정확도는 떨어진다. 마지막으로, Table 3.3은 이상치를 15로 증가시켜 더 강한 이상치를 원자료에 포함시킨 결과이다. Warwick (2005)의 방법은 이상치가 10인 경우와 마찬가지로 매우 작은  $\alpha$  값이 도출되었고, Fujisawa와 Eguchi (2006)의 방법과 비교하였을 때 참모수를 잘 추정하지 못했다. Fujisawa와 Eguchi (2006)의 방법을 이용했을 때는 이상치가 10인 경우와 비교하여 최적의 조율모수가 오히려 작게 선택되었는데, 이는 더 작은 조율모수로도 충분한 로버스트성을 확보할 수 있었음을 의미한다. 또한, 이 결과는 최적의 조율모수가 로버스트성만이 아닌 집근효율성도 고려하여 결정됨을 의미한다. 실제로 참모수 1과의 차이를 살펴보면, 이상치가 10인 경우와 15인 경우가 매우 비슷하다는 사실을 알 수 있다.

**Table 3.1** Optimal tuning parameter and corresponding MDPDE (Exp(1), outlier=10)

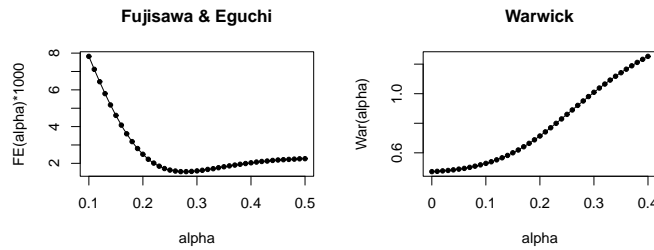
Number of outliers	MLE	Fujisawa and Eguchi		Warwick	
	$\hat{\theta}_{0,n}$	$\hat{\alpha}$	$\hat{\theta}_{\hat{\alpha},n}$	$\hat{\alpha}$	$\hat{\theta}_{\hat{\alpha},n}$
1	0.9475	0.7372	0.9947	0	0.9475
5	0.7188	0.2549	0.9672	0	0.7188
10	0.5414	0.2772	0.8882	0	0.5414



**Figure 3.1**  $FE(\alpha)$  (left) and  $War(\alpha)$  (right) for various tuning parameter  $\alpha$  (Exp(1), 1 outlier)



**Figure 3.2**  $FE(\alpha)$  (left) and  $War(\alpha)$  (right) for various tuning parameter  $\alpha$  (Exp(1), 5 outliers)



**Figure 3.3**  $FE(\alpha)$  (left) and  $War(\alpha)$  (right) for various tuning parameter  $\alpha$  (Exp(1), 10 outliers)

**Table 3.2** Comparison between robustness and efficiency

	Number of outlier=1					
	$\alpha = 0$	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$
Robustness measure	0.0015	0.0001	0.0003	0.0002	0	0
Efficiency measure	0.8977	1.1107	1.3524	1.5782	1.7646	1.9103
	Number of outlier=5					
	$\alpha = 0$	$\alpha = 0.2$	$\alpha = 0.4$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 1$
Robustness measure	0.0763	0.0038	0.0003	0.0004	0.0001	0
Efficiency measure	0.5167	0.9697	1.3712	1.6259	1.8078	1.9435

**Table 3.3** Optimal tuning parameter and corresponding MDPDE (Exp(1), outlier=15)

Number of outliers	MLE	Fujisawa and Eguchi		Warwick	
	$\hat{\theta}_{0,n}$	$\hat{\alpha}$	$\hat{\theta}_{\hat{\alpha},n}$	$\hat{\alpha}$	$\hat{\theta}_{\hat{\alpha},n}$
1	0.9046	0.7386	0.9949	0	0.9046
5	0.6093	0.1869	0.9734	0	0.6093
10	0.4261	0.2053	0.8959	0.038	0.4739

**3.2. 모수가 두 개인 경우: 정규분포**

이제 두 개의 모수를 추정하는 경우를 고려하기 위해 참모수가  $\mu_0 = 0, \sigma_0 = 1$ 인 정규분포 ( $N(0, 1)$ )로 부터 100개의 자료를 생성하였다. 자료의 최댓값은 2.7585이고, 최소값은 -2.162이다. 3.1절과 같은 방법으로, 생성된 자료 중 랜덤하게 선택된 1개, 5개, 10개의 자료를 이상치 10으로 대체하였다. 이상치 10은 제3사분위수+1.5\*사분위수범위를 초과하는 값이다. 식 (2.1)에 정규분포의 밀도 함수를 대입하여 MDPDE를 구하려면 다음과 같은 추정방정식의 해를 구해야 하는데,

$$\mu = \frac{\sum_{i=1}^n \exp\left(-\frac{\alpha(X_i - \mu)^2}{2\sigma^2}\right) X_i}{\sum_{i=1}^n \exp\left(-\frac{\alpha(X_i - \mu)^2}{2\sigma^2}\right)},$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n \exp\left(-\frac{\alpha(X_i - \mu)^2}{2\sigma^2}\right) (X_i - \mu)^2}{\sum_{i=1}^n \left\{ \exp\left(-\frac{\alpha(X_i - \mu)^2}{2\sigma^2}\right) - \frac{1}{\sqrt{1+\alpha}} \frac{\alpha}{1+\alpha} \right\}}},$$

이 역시 해석해를 구하기 힘든 형태이므로, 초기치로 최대우도추정량  $\hat{\mu}_{\alpha,n,1} = \sum_{i=1}^n X_i/n, \hat{\sigma}_{\alpha,n,1} = \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2/n}$ 을 사용하여  $k = 1, 2, \dots$ 에 대하여 다음과 같은 방법으로 수치해를 구한다.

$$\hat{\mu}_{\alpha,n,k+1} = \frac{\sum_{i=1}^n \exp\left(-\frac{\alpha(X_i - \hat{\mu}_{\alpha,n,k})^2}{2\hat{\sigma}_{\alpha,n,k}^2}\right) X_i}{\sum_{i=1}^n \exp\left(-\frac{\alpha(X_i - \hat{\mu}_{\alpha,n,k})^2}{2\hat{\sigma}_{\alpha,n,k}^2}\right)},$$

$$\hat{\sigma}_{\alpha,n,k+1} = \sqrt{\frac{\sum_{i=1}^n \exp\left(-\frac{\alpha(X_i - \hat{\mu}_{\alpha,n,k})^2}{2\hat{\sigma}_{\alpha,n,k}^2}\right) (X_i - \hat{\mu}_{\alpha,n,k})^2}{\sum_{i=1}^n \left\{ \exp\left(-\frac{\alpha(X_i - \hat{\mu}_{\alpha,n,k})^2}{2\hat{\sigma}_{\alpha,n,k}^2}\right) - \frac{1}{\sqrt{1+\alpha}} \frac{\alpha}{1+\alpha} \right\}}}$$

MDPDE의 접근 분산은 다음과 같고,

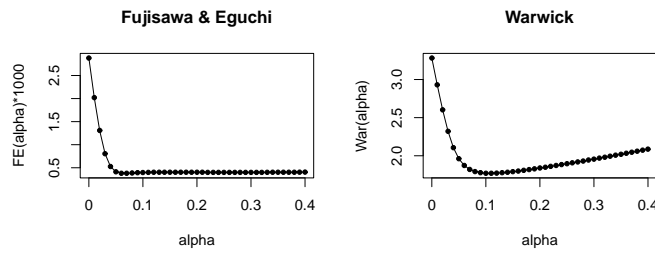
$$\begin{aligned} \text{As.var}(\hat{\mu}_{\alpha,n}) &= \left(1 + \frac{\alpha^2}{1 + 2\alpha}\right)^{3/2} \sigma_0^2, \\ \text{As.var}(\hat{\sigma}_{\alpha,n}) &= \frac{(1 + \alpha)^2}{(2 + \alpha^2)^2} \left\{ \frac{2(1 + \alpha)^3(1 + 2\alpha^2)}{(1 + 2\alpha)^{5/2}} - \alpha^2 \right\} \sigma_0^2, \end{aligned}$$

$\sigma_0$ 의 추정량으로는  $\hat{\sigma}_{\alpha,n}$ 을 사용하였다.

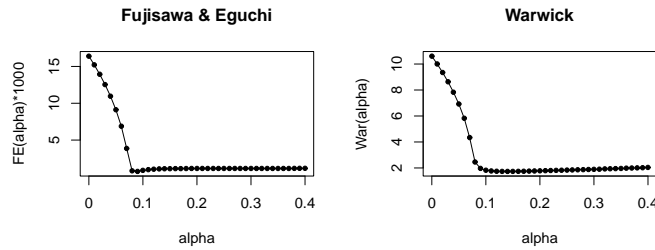
Table 3.4는 두 방법을 이용하여 선택된 최적의 조율모수와 그에 해당하는 MDPDE를 보여준다. 또한, Figures 3.4-3.6은 각각 이상치의 개수가 1개, 5개, 10개인 경우  $FE(\alpha)$ 와  $War(\alpha)$  함수를  $\alpha$ 에 따라 그린 그림이다. Figures 3.4-3.6를 보면  $FE(\alpha)$ 와  $War(\alpha)$ 가 매우 유사한 형태를 보이고 있음을 알 수 있는데, Table 3.4를 보면 실제 선택된 최적의 조율모수도 매우 유사하다. 대체적으로 두 방법 모두 특정한  $\alpha$  값 이후로는  $FE(\alpha)$ 나  $War(\alpha)$ 가 비슷한 값을 갖고, 이상치의 개수가 증가하면 최적의 조율모수 또한 증가하는 경향을 보였다. Table 3.5는 이상치를 15로 크게 한 경우인데 대부분의 경우 3.1절의 결과와 같이 조금 더 작은 조율모수가 선택되었고, 이 경우에도 두 방법은 큰 차이를 보이지 않았다.

**Table 3.4** Optimal tuning parameter and corresponding MDPDE (N(0,1), outlier=10)

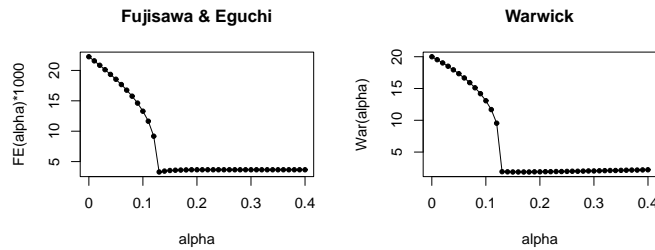
Number of outliers	MLE		Fujisawa and Eguchi			Warwick		
	$\hat{\mu}_{0,n}$	$\hat{\sigma}_{0,n}$	$\hat{\alpha}$	$\hat{\mu}_{\hat{\alpha},n}$	$\hat{\sigma}_{\hat{\alpha},n}$	$\hat{\alpha}$	$\hat{\mu}_{\hat{\alpha},n}$	$\hat{\sigma}_{\hat{\alpha},n}$
1	0.1696	1.4449	0.0653	0.0737	1.0964	0.1074	0.0667	1.0750
5	0.5210	2.3980	0.0846	0.0380	1.1360	0.1396	0.0189	1.0580
10	1.010	3.1601	0.1253	0.0131	1.1140	0.1591	0.0062	1.0919



**Figure 3.4**  $FE(\alpha)$  (left) and  $War(\alpha)$  (right) for various tuning parameter  $\alpha$  (N(0,1), 1 outlier)



**Figure 3.5**  $FE(\alpha)$  (left) and  $War(\alpha)$  (right) for various tuning parameter  $\alpha$  (N(0,1), 5 outliers)



**Figure 3.6**  $FE(\alpha)$  (left) and  $War(\alpha)$  (right) for various tuning parameter  $\alpha$  (N(0,1), 10 outliers)



**Table 3.5** Optimal tuning parameter and corresponding MDPDE (N(0,1), outlier=15)

Number of outliers	MLE		Fujisawa and Eguchi			Warwick		
	$\hat{\mu}_{0,n}$	$\hat{\sigma}_{0,n}$	$\hat{\alpha}$	$\hat{\mu}_{\hat{\alpha},n}$	$\hat{\sigma}_{\hat{\alpha},n}$	$\hat{\alpha}$	$\hat{\mu}_{\hat{\alpha},n}$	$\hat{\sigma}_{\hat{\alpha},n}$
1	0.2196	1.8216	0.0360	0.0724	1.0903	0.0651	0.0675	1.0676
5	0.7710	3.4172	0.0534	0.0223	1.0593	0.0779	0.0197	1.0473
10	1.5102	4.6072	0.3336	-0.0024	1.1081	0.1032	0.0066	1.0757

#### 4. 결론

이상치에 강건한 통계적 추론방법에 대한 연구는 많은 연구자들의 관심의 대상이 되어왔다. 부분선형 모형에서 이상치에 강건한 반응 변수 변환과정을 제안한 Seo와 Yoon (2013)의 연구 등이 좋은 예이다. MDPDE는 이상치에 강건한 성질을 갖는 추정량으로써 최대우도추정량과  $L^2$ -거리추정량을 연결하는 매우 유용한 추정량이다. MDPDE는 조율모수에 따라 로버스트성과 점근효율성이 서로 상충하게 되므로, 이 두 가지를 모두 고려한 최적의 조율모수를 찾는 연구는 매우 중요한 의미를 갖는다. 본 연구에서는 MDPDE의 최적의 조율모수를 선택하는 방법으로써 Fujisawa와 Eguchi (2006)의 연구와 Warwick (2005)의 연구를 소개하고, 모의실험을 통해 두 방법을 비교하였다. 두 방법은 유사한 결과를 보이기도 하지만, Warwick (2005)의 방법은 경우에 따라 점근효율성에 강하게 기인하여 매우 작은 조율모수를 선택하게 될 수도 있다는 사실을 알 수 있었다. 따라서, Fujisawa와 Eguchi (2006)의 방법이 범용적으로 사용하기에 적절하다고 판단된다. 두 방법이 유사한 결과를 보인 3.2절의 결과를 보면, 최적의 조율모수에 대응하는 MDPDE가 참모수를 매우 가깝게 추정하고 있어 이상점에 강건한 성질을 확인할 수 있었다.

#### References

- Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, **85**, 549-559.
- Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: Efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics*, **46**, 683-705.
- Beran, R. (1977). Minimum hellinger distance estimates for parametric models. *Annals of Statistics*, **5**, 445-463.
- Cao, R., Cuevas, A. and Fraiman, R. (1995). Minimum distance density-based estimation. *Computational Statistics & Data Analysis*, **20**, 611-631.
- Fujisawa, H. and Eguchi, S. (2006). Robust estimation in the normal mixture model. *Journal of Statistical Planning and Inference*, **136**, 3989-4011.
- Kim, B. and Lee, S. (2011). Robust estimation for the covariance matrix of multivariate time series. *Journal of Time Series Analysis*, **32**, 469-481.
- Kim, B. and Lee, S. (2013). Robust estimation for the covariance matrix of multivariate time series based on normal mixtures. *Computational Statistics & Data Analysis*, **57**, 125-140.
- Lee, S. and Na, O. (2005). Test for parameter change based on the estimator minimizing density-based divergence measures. *Annals of the Institute of Statistical Mathematics*, **57**, 553-573.
- Mattheou, K., Lee, S. and Karagrigoriou, A. (2009). A model selection criterion based on the BHHJ measure of divergence. *Journal of Statistical Planning and Inference*, **139**, 228-235.
- Mollah, M. N. H., Minami, M. and Eguchi, S. (2006). Exploring latent structure of mixture ICA models by the minimum beta-divergence method. *Neural Computation*, **18**, 166-190.
- Mollah, M. N. H., Minami, M. and Eguchi, S. (2007). Robust prewhitening for ICA by minimizing beta-divergence and its application to fast ICA. *Neural Processing Letters*, **25**, 91-110.
- Seo, H. S. and Yoon, M. (2013). Regression diagnostics for response transformations in a partial linear model. *Journal of the Korean Data & Information Science Society*, **24**, 33-39.
- Simpson, D. G. (1987). Minimum Hellinger distance estimation for the analysis of count data. *Journal of the American Statistical Association*, **82**, 802-807.

- Tamura, R. N. and Boos, D. D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance. *Journal of the American Statistical Association*, **81**, 223-229.
- Warwick, J. (2005). A data-based method for selecting tuning parameters in minimum distance estimators. *Computational Statistics & Data Analysis*, **48**, 571-585.

## A study on tuning parameter selection for MDPDE<sup>†</sup>

Donghyeon Yu<sup>1</sup> · Byungsoo Kim<sup>2</sup>

<sup>1</sup>Department of Statistics, Keimyung University

<sup>2</sup>Department of Statistics, Yeungnam University

Received 20 April 2015, revised 19 May 2015, accepted 21 May 2015

### Abstract

The MDPDE is an attractive alternative to maximum likelihood estimator because of the strong robustness properties that it inherently possess. The characteristics of MDPDE can be varied with the tuning parameter, in general, there is a trade-off between robustness and asymptotic efficiency. Hence, selection of optimal tuning parameter is important but complicated task. In this study, we introduce two optimal tuning parameter selection methods proposed by Fujisawa and Eguchi (2005) and Warwick (2006). Through simulation study, we found out that Warwick's method yields excessively small optimal tuning parameter in certain cases while Fujisawa and Eguchi's method performs well. Therefore, we think Fujisawa and Eguchi's method can be used commonly for finding optimal tuning parameter of MDPDE.

*Keywords:* Asymptotic efficiency, minimum distance estimator, robustness, tuning parameter.

---

<sup>†</sup> This work was supported by the 2014 Yeungnam University Research Grant (214A054005).

<sup>1</sup> Assistant professor, Department of Statistics, Keimyung University, Daegu 704-701, Korea.

<sup>2</sup> Corresponding author: Assistant professor, Department of Statistics, Yeungnam University, Kyongsan 712-749, Korea. E-mail: bkim@yu.ac.kr