The Korean Journal of Applied Statistics (2017) DOI: https://doi.org/10.5351/KJAS.2017.30.1.095
30(1), 95–102

Comparing the efficiency of dispersion parameter estimators in gamma generalized linear models

Seongil Jo[a] · Woojoo Lee[b,1]

[a]Department of Statistics and Applied Probability, National University of Singapore;
[b]Department of Statistics, Inha University

(Received October 4, 2016; Revised October 27, 2016; Accepted October 28, 2016)

Abstract

Gamma generalized linear models have received less attention than Poisson and binomial generalized linear models. Therefore, many old-established statistical techniques are still used in gamma generalized linear models. In particular, existing literature and textbooks still use approximate estimates for the dispersion parameter. In this paper we study the efficiency of various dispersion parameter estimators in gamma generalized linear models and perform numerical simulations. Numerical studies show that the maximum likelihood estimator and Cox-Reid adjusted maximum likelihood estimator are recommended and that approximate estimates should be avoided in practice.

Keywords: gamma generalized linear model, dispersion parameter, maximum likelihood estimate, efficiency

1. 서론

감마 일반화 선형 모형(gamma generalized linear model)은 포아송 분포(Poisson distribution) 또는 이항 분포(binomial distribution)에 기반한 일반화 선형 모형에 비해 그 동안 상대적으로 적은 관심을 받아왔다. 따라서 포아송 또는 이항 일반화 선형 모형과는 달리 감마 일반화 선형 모형에서는 오래전에 제안된 통계기법이 아직도 사용되는 경우가 많다. 오래전에 만들어진 기법이라고 해서 모두 다 교체되어야 하는 것은 아니지만, 훨씬 더 우수한 성능을 가지고 있는 대안이 있다면 교체를 심각하게 검토해 볼 필요성이 있다. 본 논문에서는 감마 일반화 선형 모형의 산포 모수(dispersion parameter)와 관련된 추론 문제에서 이러한 문제점을 지적하고, 최대 가능도 추정량(maximum likelihood estimator)과 기존의 근사 추정량(approximate estimator)의 성능을 수치 연구를 통해 비교해 보고자 한다.

현재 가장 최신 버젼인 R(3.3.1 버전)에서도 감마 일반화 선형 모형을 구하는 경우 산포 모수에 대한 적률추정치(method of moment estimate)만을 제공하고 있다. 또한, 많은 연구자들이 실제 분석에서 산포 모수에 대한 근사 추정량를 사용하고 있다. 만약 산포 모수에 대한 최대 가능도 추정량을 구하기가 매우 어렵다면, 근사 추정량의 사용은 정당화 될 수 있으나 실제 최대 가능도 추정량을 구하기는 어렵지

This research was supported by a grant [MPSS-NH-2015-79] through the Disaster and Safety Management Institute funded by Ministry of Public Safety and Security of Korean government.
[1]Corresponding author: Department of Statistics, Inha University, 100, Inha-ro, Nam-gu, Incheon 22212, Korea. E-mail: lwj221@gmail.com

가 않다. 3절에서 자세히 설명되겠지만 1차원 변수에 대한 뉴턴-랩슨(Newton-Raphson)식의 몇 차례 반복 계산이면 충분하다. 따라서, 최대 가능도 추정량의 성능이 유한 표본에서 기존의 여러 추정량에 비해 우수하다면 근사 추정량을 사용하는 것은 지양되어야 할 것으로 보인다.

본 논문의 구성은 다음과 같다. 2절에서는 감마 일반화 선형 모형의 기초를 제공하고, 3절에서는 산포모수와 관련한 기존의 근사 추정량과 McCullagh와 Nelder (1989)에서 이야기되는 적률추정법에 기반한 추정량, 최대 가능도 추정량을 소개한다. 4절에서 이 들의 성능비교를 한후, 5절에서 결론을 맺는다.

2. 감마 분포

i번째 관측치 Y_i가 감마분포를 따르면서 평균 μ_i와 산포모수 ϕ 를 갖는다고 할 때, 해당하는 확률 밀도 함수의 표현은 다음과 같이 주어진다.

$$\frac{\nu^\nu y_i^{\nu-1}\exp(-\nu y_i/\mu_i)}{\mu_i^\nu \Gamma(\nu)}dy_i.$$

이 분포는 $Y_i \sim \text{Gamma}(\mu_i,\phi)$으로 나타내고, $\phi = 1/\nu$의 관계를 가진다. 여기서, $E(Y_i) = \mu_i$이고 $\text{Var}(Y_i) = \phi\mu_i^2 = \mu_i^2/\nu$이 된다. 적률 생성 함수(moment generating function)는 $K_i(t) = E(\exp(tY_i)) = \nu\log(1-\mu_i t/\nu)$으로 주어진다. Y_i에 알려진 가중치 w_i가 있는 경우 $Y_i \sim \text{Gamma}\,(\mu_i,\phi/w_i)$으로 표현한다.

실제 자료분석에서는 감마분포(gamma distribution)의 정준 연결함수(canonical link function)인 역함수(reciprocal function)가 아닌 로그 연결함수(log link function)를 훨씬 더 자주 사용하므로, 본 논문에서는 $\mu_i = \exp(x_i^T\beta)$을 고려할 것이다. 여기서 x_i는 i번째 설명변수(explanatory variable) 벡터이고 β는 대응되는 회귀계수(regression coefficient)이다. β의 추론을 위해서 최대 가능도 추정량이 사용되는데, 이를 구하기 위한 반복 가중최소 제곱법(iterative weighted least squared method)은

$$\widehat{\beta}_{(r+1)} = \left(X^TW_{(r)}X\right)^{-1}\left(X^TW_{(r)}z_{(r)}\right)$$

인데, 여기서 X는 i번째 행벡터가 x_i가 되는 모형 행렬이고, $z_{(r)} = X\widehat{\beta}_{(r)} + (Y-\mu_{(r)})/\mu_{(r)}$이 된다. Y와 μ는 각각 i번째 성분이 Y_i와 μ_i가 되는 벡터를 나타낸다. 아랫첨자 (r)은 r번째 반복에서 계산된 값을 사용하고 있다는 뜻이고, 실제 $W_{(r)}$은 $n\times n$ 항등행렬(identity matrix)로 주어지기 때문에 r에 따라 달라지지 않는다. 위의 식에서 주목할만한 점은 $\widehat{\beta}$를 구할 때, ϕ의 값이 사용되지 않는다는 점이다. 또한 ℓ_i를 Y_i의 로그 가능도 함수(log-likelihood function)라 할 때,

$$E\left(\frac{\partial^2\ell_i}{\partial\mu_i\partial\phi}\right) = 0$$

이 되기 때문에 $\widehat{\beta}$와 $\widehat{\phi}$의 점근 공분산 행렬(asymptotic covariance matrix)의 비대각 성분이 0이 되므로, ϕ를 추론할 때 β는 $\widehat{\beta}$으로 알려져 있다고 생각하고 ϕ를 추론하는 것으로 충분하다. 한편, $\widehat{\beta}$만의 점근 공분산 행렬은

$$\text{Cov}\left(\widehat{\beta}\right) = \phi\left(X^TWX\right)^{-1}$$

으로 주어지기 때문에, β에 대한 가설 검정을 위해서는 ϕ에 대한 올바른 추론이 필수적임을 확인할 수 있다.

3. 산포모수에 대한 최대 가능도 추정량

$\widehat{\beta}$가 주어졌을 때, ν에 대한 최대 가능도 추정량을 얻기 위해서는 우리는 감마 로그 가능도 함수를 ν에 대해 미분해서 얻는 다음의 식을 풀어야 한다.

$$\sum_{i=1}^{n} w_i \{\log(w_i\nu) - \psi(w_i\nu)\} = \frac{D_p(y, \widehat{\mu})}{2}, \tag{3.1}$$

여기서 $\psi(\cdot)$은 digamma 함수를 의미하고, $D_p(y, \widehat{\mu}) = -2\sum_{i=1}^{n} w_i[\log(y_i/\widehat{\mu}_i) - (y_i - \widehat{\mu}_i)/\widehat{\mu}_i]$으로 정의되며, $\widehat{\mu}_i$는 μ_i의 최대 가능도 추정치이다. 평균 모형에 절편이 설명변수로 포함되어 있으면, $\sum_i[w_i(y_i - \widehat{\mu}_i)/\widehat{\mu}_i] = 0$이기 때문에 $D_p(y, \widehat{\mu})$는 $-2\sum_i w_i \log(y_i/\widehat{\mu}_i)$으로 더 단순화되어 표현이 가능하다.

위의 식 (3.1)에서 특별한 경우로 모든 w_i가 1인 경우를 고려해보자. 점수 방정식(score equation)은

$$\log(\nu) - \psi(\nu) = \frac{D_p(y, \widehat{\mu})}{2n}$$

로 주어진다. 기존의 여러 문헌에서는 최대 가능도 추정량 $\widehat{\phi}(= 1/\widehat{\nu})$을 얻는 것 대신에, ϕ가 작다는 가정하에

$$\log \nu - \psi(\nu) \approx \frac{1}{2\nu} + O\left(\nu^{-2}\right)$$

의 근사식에서 $O(\nu^{-2})$ 항을 무시하고 ν에 대해 식을 풀면

$$\widehat{\phi} \approx \frac{D_p(y, \widehat{\mu})}{n} \tag{3.2}$$

이 얻어진다.

위에서 사용된 테일러 근사식(Talyor expansion)보다 한 차수 높은 근사식은,

$$\log \nu - \psi(\nu) \approx \frac{1}{2\nu} + \frac{1}{12\nu^2} + O\left(\nu^{-4}\right)$$

으로 주어지는데, 이 때

$$\widehat{\phi} \approx \frac{2D_p(y, \widehat{\mu})}{n\left(1 + \left(1 + \frac{2D_p(y,\widehat{\mu})}{(3n)}\right)^{\frac{1}{2}}\right)} \tag{3.3}$$

이 된다. 이 식은 Cordeiro와 McCullagh (1991)의 식 (2.3)과 일치한다.

그렇지만, 컴퓨터의 발전을 통해 반복 계산을 매우 빠르게 실행할 수 있으므로, 현재 우리가 위와 같은 근사식을 고집할 필요가 없다. 먼저

$$f_w(\nu) = \left[\sum_{i=1}^{n} w_i \left(\log(w_i\nu) - \psi(w_i\nu)\right)\right] - \frac{D_p(y, \widehat{\mu})}{2}$$

라 두면,

$$f'_w(\nu) = \sum_{i=1}^{n} w_i \left(\frac{1}{\nu} - \psi'(w_i\nu)w_i\right)$$

이 된다. 여기서 $\psi'(x)$은 trigamma 함수를 뜻한다. 여기서 우리가 원하는 것은 $f_w(\nu^*) = 0$ 만족시키는 값을 찾는 것이며, 이를 위해 뉴턴-랩슨 알고리즘(Newton-Raphson algorithm)은 아래와 같이 주어진다.

$$\nu^{(r+1)} \leftarrow \nu^{(r)} - \frac{f_w(\nu^{(r)})}{f'_w(\nu^{(r)})},$$

여기서 $\nu^{(r)}$은 r번째 반복에서의 값을 의미한다.

가중치가 모두 1로 같은 경우, 즉 $w_i = 1,\ i = 1, \ldots, n$에서는 $f(\nu) = \log(\nu) - \psi(\nu) - D_p(y, \widehat{\mu})/(2n)$이 되고 $f'(\nu) = 1/\nu - \psi'(\nu)$으로 단순화된다. 뉴턴-랩슨 알고리즘은

$$\nu^{(r+1)} \leftarrow \nu^{(r)} - \frac{f(\nu^{(r)})}{f'(\nu^{(r)})}$$

이 된다. 이 알고리즘은 실제 모든 $\mu_i = \mu$일 때의 Choi와 Wette (1969)에 의해서 제안된 알고리즘과 일치한다. Choi와 Wette (1969)는 임의의 $\nu > 0$에 대해 $f'(\nu) < 0$이 됨을 보였고, 따라서 이 알고리즘은 수렴됨이 보장된다. 현재 R의 MASS 패키지 안에서 "gamma.dispersion"이라는 함수가 이를 구현하여 준다.

평균 모형에 있는 회귀계수를 추정하기 되면서 생기는 편의(bias)는 ϕ의 추정에 영향을 줄 수 있다. 다행스럽게도 감마 모형에서는 β와 ϕ는 서로 직교(orthogonal)하기 때문에 편의를 보정하는 식이 간단하게 주어진다. 여기서 직교란 Cox와 Reid (1987)에서 주어진 개념으로, 로그 가능도 함수를 β와 ϕ로 편미분한 함수의 기대값이 0인 경우를 말한다. 이 때, Cox-Reid의 수정된 점수 방정식(adjusted score equation)은

$$\log\nu - \psi(\nu) - \frac{p}{2n\nu} = \frac{D_p(y, \widehat{\mu})}{2n} \tag{3.4}$$

으로 주어진다.

먼저 $g(\nu) = \log(\nu) - \psi(\nu) - p/(2n\nu) - D_p(y, \widehat{\mu})/(2n)$이고 $g'(\nu) = 1/\nu - \psi'(\nu) + p/(2n\nu^2)$이라고 하자. 여기서 우리는 뉴턴-랩슨 식인

$$\nu^{(r+1)} \leftarrow \nu^{(r)} - \frac{g(\nu^{(r)})}{g'(\nu^{(r)})}$$

으로, $g(\nu^*) = 0$을 만족하는 ν^*를 찾을 수 있다. 이 때 최대 가능도 추정량을 초기치로 활용하는 것이 계산상 편리하다.

한편, McCullagh와 Nelder (1989)에서는 작은 ϕ일 때의 근사추정량에 대한 대안으로 적률기반 추정량인

$$\widehat{\phi}^{\text{MME}} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{(y_i - \mu_i)^2}{\mu_i^2} \tag{3.5}$$

을 제시하였다. 이 추정치가 일치성(consistency)을 가진다는 것은 잘 알려져있다.

4. 모의실험

본 절에서는 3절에서 설명한 산포모수 ϕ의 추정치들에 대한 성능을 비교해보기 위해 모의실험을 실시하였다. 모의실험은 반복실험을 위해 총 100개의 독립적인 데이터를 생성하였고, 각 데이터는 0과 1사이

Table 4.1. Summary of the root mean squared errors

True dispersion	Sample size	Approx.	CM	MLE	Adj. MLE	MME
$\phi = 0.01$	$n = 50$	0.002	0.002	0.002	0.002	0.002
	$n = 100$	0.001	0.001	0.001	0.001	0.001
	$n = 200$	0.001	0.001	0.001	0.001	0.001
	$n = 500$	0.001	0.001	0.001	0.001	0.001
$\phi = 0.1$	$n = 50$	0.021	0.022	0.022	0.019	0.020
	$n = 100$	0.014	0.014	0.014	0.014	0.014
	$n = 200$	0.010	0.010	0.010	0.010	0.011
	$n = 500$	0.005	0.005	0.005	0.005	0.006
$\phi = 1.0$	$n = 50$	0.206	0.207	0.206	0.182	0.245
	$n = 100$	0.194	0.126	0.127	0.127	0.171
	$n = 200$	0.153	0.079	0.078	0.075	0.123
	$n = 500$	0.162	0.059	0.059	0.058	0.107
$\phi = 1.5$	$n = 50$	0.324	0.269	0.263	0.236	0.383
	$n = 100$	0.354	0.203	0.201	0.196	0.326
	$n = 200$	0.321	0.138	0.132	0.128	0.220
	$n = 500$	0.318	0.097	0.089	0.087	0.158

Approx. = approximate estimation; CM = Cordeiro and McCullagh's (1991) estimation; MLE = maximum likelihood estimation; Adj. MLE = adjusted maximum likelihood estimation; MME = method of moments estimation.

의 균등분포(uniform distribution)에서 생성한 5개의 설명변수와 감마 분포에서 생성된 반응변수로 이루어져있다. 구체적인 모형은 다음과 같다.

$$\begin{aligned} Y_i|x_{i1},\ldots,x_{i5} &\sim \text{Gamma}\left(\mu_i,\phi\right), \quad i=1,\ldots,n, \\ \log(\mu_i) &= 0.8 + x_{i1} - 0.1x_{i2} + 0.2x_{i3} + 0.5x_{i4} - 0.3x_{i5}, \quad i=1,\ldots,n, \end{aligned} \tag{4.1}$$

여기서 $x_{ij} \sim \text{Unif}(0,1)$, $j = 1,\ldots,5$이다. n은 50, 100, 200, 500을 고려하였다.

모의실험을 통해 얻은 결과는 Table 4.1과 4.2에 나타나 있다. Table 4.1은 5가지 방법에 의해 추정된 $\widehat{\phi}$값과 실제 자료를 생성할 때 사용된 ϕ값 사이의 제곱근 평균제곱오차(root mean squared error; RMSE)를 제공한다. 즉, k번째 모의실험 데이터셋에서 계산된 ϕ의 추정치를 $\widehat{\phi}_k$라고 할 때,

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{100}\left(\widehat{\phi}_k - \phi\right)^2}{100}}$$

으로 계산된다. Table 4.2는 각 방법에 의해 추정된 산포모수값의 평균값과 그것의 표준오차(standard error)를 보여준다. 괄호 안에 있는 값들이 표준오차이다. 참고로 Table 4.1과 4.2는 서로 대응된다. 예를 들어 $\phi = 1.5$이고 $n = 50$일 때, Table 4.1에서의 RMSE는 0.324이고 Table 4.2에서 대응되는 세팅에서의 편의는 0.129 $(= 1.629 - 1.5)$이고, 분산은 $(0.030 \times \sqrt{100})^2$이므로 Table 4.2에서 구해진 RMSE는 $\sqrt{0.129^2 + (0.030 \times 10^2)} = 0.326$이다. 이 둘의 차이는 수치 오차(nuemrical error)로 볼 수 있다.

Table 4.1과 4.2에서 Approx는 식 (3.2), CM은 식 (3.3), MLE는 최대가능도 추정치를, Adj.MLE는 Cox-Reid 수정된 점수 방정식에서 얻어진 추정치를, MME는 적률기반 추정치를 각각 의미한다. CM,

Table 4.2. Average estimates and standard errors for dispersion parameter ϕ

True dispersion	Sample size	Approx.	CM	MLE	Adj. MLE	MME
$\phi = 0.01$	$n = 50$	0.009 (0.000)	0.009 (0.000)	0.009 (0.000)	0.010 (0.000)	0.010 (0.000)
	$n = 100$	0.009 (0.000)	0.009 (0.000)	0.009 (0.000)	0.010 (0.000)	0.010 (0.000)
	$n = 200$	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)
	$n = 500$	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)	0.010 (0.000)
$\phi = 0.1$	$n = 50$	0.088 (0.002)	0.086 (0.002)	0.086 (0.002)	0.096 (0.002)	0.096 (0.002)
	$n = 100$	0.095 (0.001)	0.093 (0.001)	0.093 (0.001)	0.098 (0.001)	0.097 (0.001)
	$n = 200$	0.099 (0.001)	0.098 (0.001)	0.098 (0.001)	0.100 (0.001)	0.100 (0.001)
	$n = 500$	0.100 (0.001)	0.098 (0.001)	0.098 (0.001)	0.099 (0.001)	0.099 (0.001)
$\phi = 1.0$	$n = 50$	0.998 (0.021)	0.868 (0.016)	0.875 (0.016)	0.950 (0.018)	0.863 (0.020)
	$n = 100$	1.113 (0.016)	0.958 (0.012)	0.966 (0.012)	1.005 (0.013)	0.929 (0.016)
	$n = 200$	1.121 (0.009)	0.965 (0.007)	0.973 (0.007)	0.993 (0.007)	0.963 (0.012)
	$n = 500$	1.144 (0.008)	0.982 (0.006)	0.991 (0.006)	0.999 (0.006)	0.997 (0.011)
$\phi = 1.5$	$n = 50$	1.629 (0.030)	1.328 (0.021)	1.354 (0.022)	1.460 (0.023)	1.253 (0.030)
	$n = 100$	1.738 (0.026)	1.404 (0.018)	1.434 (0.019)	1.488 (0.020)	1.380 (0.030)
	$n = 200$	1.769 (0.018)	1.427 (0.012)	1.458 (0.013)	1.485 (0.013)	1.435 (0.021)
	$n = 500$	1.795 (0.012)	1.445 (0.008)	1.477 (0.009)	1.488 (0.009)	1.478 (0.016)

Approx. = approximate estimation; CM = Cordeiro and McCullagh's (1991) estimation; MLE = maximum likelihood estimation; Adj. MLE = adjusted maximum likelihood estimation; MME = method of moments estimation.

MLE, Adj. MLE와 MME는 모든 경우에서 표본 수가 증가하면 RMSE가 감소하는 것으로 나타났다. 하지만 Approx는 표본 수가 증가하더라도 오히려 RMSE가 증가하는 경우가 있는데, 이는 Approx가 갖는 추정량의 편의가 표본 수가 커지더라도 지속적으로 남아있기 때문임을 Table 4.2에서 쉽게 확인할 수 있다. CM은 ϕ가 작을 때 MLE와 성능이 거의 유사함이 확인되었다. 그러나 ϕ가 상대적으로 클 때, 즉 본 시뮬레이션에서 ϕ가 1.5인 경우에는 MLE와의 차이가 나타났고, 이 차이는 CM이 가지고 있는 지속적인 편의가 만들어내고 있음을 Table 4.2에서 확인할 수 있다. 또한 CM의 주목할만 한 점은 표본 수가 작은 경우에서는 CM의 표준오차가 MLE의 표준오차보다 작은 경우가 있어서 편의가 있더라도 RMSE값이 더 작을 때도 생긴다는 것이다. Adj.MLE와 MLE는 ϕ의 값이 작을 때에는 RMSE의 차이가 나지 않다가 ϕ가 커지면서 차이가 나타나기 시작한다. Cox-Reid의 수정항 덕분에 Adj.MLE의 편의가 MLE의 편의 보다는 일관되게 작게 나타나지만, 표준오차가 커지는 경우가 있어서 RMSE 측면에서 일관되게 Adj.MLE가 MLE보다 좋은 것은 아니다. 하지만 전체적으로 Adj.MLE의 우수한 성능이 Table 4.1과 4.2에서 나타나고 있다. MME 역시 ϕ가 작을 때에는 MLE와 차이가 거의 나지 않거나, 일부의 경우 더 우수한 성능이 나타나기도 하였는데 ϕ가 커지면서 표준오차가 MLE에 비해 상당히 크기 때문에 RMSE 측면에서 MLE나 Adj.MLE에 비해 열등한 것으로 나타나고 있다.

5. 결론

본 논문에서는 감마 일반화 선형모형의 산포 모수에 대한 여러가지 추정량들을 설명하고 모의실험을 통해 추정량들의 성능을 알아보았다. 모의실험 결과 일반적으로 작은 산포 모수 ϕ값을 가정한 근사 추정량은 사용되지 않는 것이 바람직하며, 현재 R에서 감마 일반화 선형 모형 적합 시 기본적으로 제공하는 적률기반 추정량은 주의해서 사용해야 할 필요성이 있음을 확인하였다. 특히 산포 모수값이 크다고 기대되어 질 때에는 최대 가능도 추정량이나 Cox-Reid의 수정된 최대 가능도 추정량의 성능이 더 우수할

것으로 기대되므로, 이를 같이 고려하는 것이 자료 분석에 있어서 좋은 실천방안이 될 것이다.

그러나 최대 가능도 추정량의 사용에도 주의해야할 점이 있음을 주지해야 한다. 특히 감마 분포와 관련하여 최대 가능도 추정량의 사용이 어떠한 상황에서도 항상 선호되어야 하는 것은 아니라는 점은 McCullagh와 Nelder (1989)에 의해 지적된 적이 있는데, 감마 분포의 최대 가능도 추정량은 0에 매우 가까운 값을 갖는 관측치가 있는 경우 반올림 오차(rounding errors) 등에 대해 민감할 수 있기 때문이다. 이러한 특성을 확인하기 위해 4절의 모의실험과 같은 모형에서 $\phi = 0.1$이고 $n = 100$인 경우를 고려해보았고 한개의 반응변수 값만 10^{-6}으로 바꾸어준 경우, MME의 RMSE는 0.019, MLE의 RMSE는 0.260으로 나타난 것으로 부터 이러한 지적사항을 쉽게 확인할 수 있었다. 또한 일차, 이차 적률까지는 모형을 가정할 수 있으나 그 이후의 적률에 대한 가정이 상당히 의심스러운 경우에는 최대 가능도 추정량보다 적률기반 추정량이 더 좋은 성능을 가질 수 있으므로 분포 가정에 대한 불확실성의 여부에 따라 추정량의 선택이 달라질 수 있음을 기억해 두는 것이 좋다.

References

Choi, S. C. and Wette, R. (1969). Maximum likelihood estimation of the parameters of the gamma distribution and their bias, *Technometrics*, **11**, 683–690.

Cordeiro, G. M. and McCullagh, P. (1991). Bias correction in generalized linear models, *Journal of the Royal Statistical Society Series B (Methodological)*, **53**, 629–643.

Cox, D. R. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference, *Journal of the Royal Statistical Society Series B (Methodological)*, **49**, 1–39.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models* (2nd), Chapman and Hall, London.

감마 일반화 선형 모형에서의 산포 모수 추정량에 대한 효율성 연구

조성일[a] · 이우주[b,1]

[a]싱가포르 국립대학교 통계 및 응용확률 학과, [b]인하대학교 통계학과

(2016년 10월 4일 접수, 2016년 10월 27일 수정, 2016년 10월 28일 채택)

Abstract

감마 일반화 선형모형은 포아송 분포 또는 이항 분포에 기반한 일반화 선형모형에 비해 적은 관심을 받아왔다. 따라서 감마 일반화 선형모형에서는 오래전에 개발된 통계적인 기법이 아직도 사용되고 있으며, 특히 산포 모수에 대해서는 근사 추정치가 여전히 사용되고 있다. 본 논문에서는 감마 일반화 선형 모형의 산포 모수에 대해 다양한 추정량들을 알아보고 수치 연구를 통해 그들의 효율성을 비교한다. 수치 실험의 결과 최대 가능도 추정량과 Cox-Reid의 수정된 최대 가능도 추정량이 기존의 근사 추정량에 비해 좋은 성능을 보임을 확인하였다.

주요용어: 감마 일반화 선형모형, 산포 모수, 최대 가능도 추정, 효율성

본 연구는 정부(국민안전처)의 재원으로 재난안전기술개발사업단의 지원을 받아 수행된 연구임[MPSS- 자연-2015-79].

[1]교신저자: (22212) 인천광역시 남구 인하로 100, 인하대학교 통계학과. E-mail: lwj221@gmail.com