

Information Matrix에 따른 Generalized Logistic 분포의 최우도 추정량 정확도에 관한 연구

A Study on the Accuracy of the Maximum Likelihood Estimator of the Generalized Logistic Distribution According to Information Matrix

신 흥 준* / 정 영 훈** / 허 준 행***

Shin, Hongjoon / Jung, Younghun / Heo, Jun-Haeng

Abstract

In this study, we compared the observed information matrix with the Fisher information matrix to estimate the uncertainty of maximum likelihood estimators of the generalized logistic (GL) distribution. The previous literatures recommended the use of the observed information matrix because this is convenient since this matrix is determined as the part of the parameter estimation procedure and there is little difference in accuracy between the observed information matrix and the Fisher information matrix for large sample size. The observed information matrix has been applied for the generalized logistic distribution based on the previous study without verification. For this purpose, a simulation experiment was performed to verify which matrix gave the better accuracy for the GL model. The simulation results showed that the variance-covariance of the ML parameters for the GL distribution came up with similar results to those of previous literature, but it is preferable to use of the Fisher information matrix to estimate the uncertainty of quantile of ML estimators.

keywords : generalized logistic distribution, maximum likelihood estimator, observed information matrix, Fisher information matrix

요 지

본 연구에서는 generalized logistic(GL) 분포의 최우도 추정량(maximum likelihood estimate)에 대한 불확실성 추정을 위하여 사용되는 관측정보행렬(observed information matrix)과 Fisher 정보행렬(Fisher information matrix)의 정확도를 비교해 보고자 하였다. 타 분포형에 대한 기존의 연구결과에서 표본의 크기가 클 경우 매개변수 추정시 관측정보행렬이 동시에 추정되어 계산시간도 단축되고 Fisher 정보행렬의 정확도와도 차이도 거의 없어 관측정보행렬의 사용이 추천된 바 있으나, 최근 사용이 증가되고 있는 GL 분포에 대한 연구결과는 아직 전무한 실정이며 기존 연구문헌의 결과를 토대로 구체적인 연구 없이 관측정보행렬을 사용하고 있는 상황이다. 따라서 본 연구에서는 이를 위해 모의실험을 수행하였으며, 모의 결과 최우도법에 의한 매개변수의 분산 및 공분산은 기존의 연구 결과와 비슷한 결과를 보이나, quantile에 대한 불확실성 추정에는 관측정보행렬보다 Fisher 정보행렬의 사용이 더 적절할 것으로

* 교신저자, 연세대학교 사회환경시스템공학부 박사과정

Corresponding Author, Ph.D student, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 120-749, Korea (e-mail: sinong@yonsei.ac.kr)

** 연세대학교 사회환경시스템공학부 박사과정

Ph.D student, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 120-749, Korea

*** 연세대학교 사회환경시스템공학부 교수

Professor, School of Civil and Environmental Engineering, Yonsei Univ., Seoul 120-749, Korea

로 판단되었다.

핵심용어 : generalized logistic 분포, 최우도 추정량, 관측정보행렬, Fisher 정보행렬

1. 서론

수문통계학에서 최우도 추정(maximum likelihood estimation)은 유전학자이자 통계학자인 Ronald A. Fisher가 처음으로 제시한 것으로서, 점 추정(point estimation)의 한 방법이다. 즉, 최우도법은 미지의 모집단에 대한 매개변수(unknown population parameter)를 추정하는 과정으로 우도함수(likelihood function)를 최대로 하는 특정값을 결정함으로써 매개변수를 찾는 과정이다. 여기서 매개변수 θ 를 k 개 가지고 있는 확률밀도함수 $f(x; \theta_1, \dots, \theta_k)$ 를 가정해보면, 우도함수(likelihood function)와 대수우도함수(log-likelihood function)는 다음 Eqs. (1)~(2)와 같다.

$$L(\theta; x) = \prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_k) \quad (1)$$

$$LL(\theta; x) = \log L(\theta; x) = \sum_{i=1}^n \log f(x_i; \theta_1, \dots, \theta_k) \quad (2)$$

한편 최우도법을 이용하여 얻어진 추정량의 불확실성을 판단하기 위해서는 근사적 분산-공분산 행렬을 이용하게 되는데, 매개변수 θ 의 최우도 추정값 $\hat{\theta}$ 에 대한 근사적 분산-공분산 행렬은 Fisher 정보행렬의 역행렬, 즉 대수우도함수의 매개변수에 대한 이계 편미분 도함수(second order partial derivatives)항의 기대값에 -를 곱한 값을 이용하여 구하게 된다. 이는 CRLB(Cramer-Rao Lower Bound)를 이용한 것으로, CRLB는 Eq. (3)과 같이 매개변수의 불편이(unbiased) 추정치의 분산이 Fisher 정보행렬(Fisher information matrix)의 역행렬 $I(\theta)^{-1}$ 에 의해 하한값을 갖는 것을 말한다.

$$\text{Var}(\hat{\theta}) \geq I^{-1}(\theta) \quad (3)$$

한편, Efron and Hinkley (1978)와 Hinkley (1978)는 위치매개변수와 크기매개변수로 이루어진 분포형 집단의 경우 관측정보행렬(observed information matrix)의 역행렬에 의해 추정된 매개변수의 분산-공분산 행렬이 Fisher 정보행렬의 역행렬에 의해 추정된 것보다 잘 추정되며, 이를 통해 quantile의 분산 역시 관측정보행렬을 사용하는 것이 더 우수하다고 하였다. 또한 Cohen (1965)은 완전표본(complete sample)과 절단표본

(censored sample)을 이용하여 2변수 Weibull 분포에 대하여, Lemon (1975)은 절단표본을 이용하여 3변수 Weibull 분포에 대하여 매개변수에 대한 분산-공분산 최우도 추정에 관측정보행렬의 사용을 주장한 바 있다. 또한 Prescott and Walden (1983)는 절단표본을 이용하여 generalized extreme value(GEV) 분포의 매개변수에 대한 분산-공분산 행렬을 유도하였으며, 모의를 통하여 이 경우에도 관측정보행렬의 사용이 우수한지 확인한 바 있다. 한편 Rao and Hamed (2000)는 generalized logistic(GL) 분포의 매개변수 및 quantile에 대한 분산-공분산 추정 시 객관적인 증명 없이 위와 같은 이유로 관측정보행렬의 사용을 주장한 바 있다. 하지만 다른 분포형의 경우 관측정보행렬의 사용이 Fisher 정보행렬에 의한 결과보다 우수하고 매개변수 추정과정의 일부로서 쉽게 구해진다고 해서, GL분포형의 경우에도 매개변수의 분산-공분산 추정 시 관측정보행렬을 사용하는 것은 성급한 결정이다. 아직 GL 분포에 대한 분산-공분산 추정 시 관측정보행렬의 사용에 관한 구체적인 연구결과가 발표된 적이 없으며, 앞서 언급된 결과는 Cohen (1965)이 밝힌 바와 같이 엄밀하게 말해 표본 크기가 매우 큰 경우나 편이(bias)가 작은 중간크기의 표본에서만 성립한다고 하였기 때문에, Rao and Hamed (2000)에 의한 GL 분포의 매개변수 및 quantile의 분산-공분산 추정 시 관측정보행렬의 사용은 잘못된 방법이라고 할 수 있다. 따라서 본 연구에서는 관측정보행렬과 Fisher 정보행렬을 사용하여 GL 분포의 최우도법을 이용한 매개변수 및 quantile의 정확도를 비교하여 정보행렬의 사용에 기준을 제시해 보고자 하였다.

2. 최우도(maximum likelihood)와 정보행렬(information matrix)

2.1 Generalized Logistic 분포와 최우도법

GL 분포형은 2변수 logistic 분포형의 일반화한 모형으로 kappa 분포형의 특수한 경우의 모형이다. 만약 형상매개변수가 0이라면 이 분포형은 logistic 분포형이 되며, 이 분포형의 일반화 과정은 Ahmad *et al.* (1988)의 log-logistic 분포형을 재매개변수화 한 것으로, generalized Pareto 분포형이나 GEV 분포형 같이 분포

형의 유사성을 반영하기 위해 이와 같은 이름이 붙여진 것이다 (Hosking and Wallis, 1997). GL 분포형의 누가 분포함수와 확률밀도함수는 각각 다음과 같다.

$$F(x) = \left[1 + \left\{ 1 - \frac{\beta}{\alpha}(x - \epsilon) \right\}^{1/\beta} \right]^{-1} \quad (4)$$

$$f(x) = \frac{1}{\alpha} \left\{ 1 - \frac{\beta}{\alpha}(x - \epsilon) \right\}^{1/\beta - 1} \left[1 + \left\{ 1 - \frac{\beta}{\alpha}(x - \epsilon) \right\}^{1/\beta} \right]^{-2} \quad (5)$$

$$\begin{aligned} \epsilon + \alpha/\beta \leq x < \infty \text{ for } \beta < 0 \\ -\infty < x \leq \epsilon + \alpha/\beta \text{ for } \beta > 0 \end{aligned}$$

여기서, ϵ 는 위치매개변수, α 는 크기매개변수, β 는 형상매개변수이다.

따라서 Eq. (5)를 이용하여 GL 분포형의 대수우도함수를 구해보면 다음과 같다.

$$\begin{aligned} \log L &= -N \ln(\alpha) + \left(\frac{1}{\beta} - 1 \right) \sum_{i=1}^N \ln \left\{ 1 - \frac{\beta}{\alpha}(x_i - \epsilon) \right\} \\ &\quad - 2 \sum_{i=1}^N \ln \left[1 + \left\{ 1 - \frac{\beta}{\alpha}(x_i - \epsilon) \right\}^{1/\beta} \right] \\ &= -N \ln(\alpha) - (1 - \beta) \sum_{i=1}^N y_i - 2 \sum_{i=1}^N \ln(1 + e^{-y_i}) \end{aligned} \quad (6)$$

여기서, $y_i = -\frac{1}{\beta} \ln \left\{ 1 - \frac{\beta}{\alpha}(x_i - \epsilon) \right\}$.

Eq. (6)의 대수우도함수를 ϵ , α , β 에 대해서 편미분하고 0으로 놓으면 다음과 같다.

$$-\frac{\partial \log L}{\partial \epsilon} = \frac{Q}{\alpha} = 0 \quad (7)$$

$$-\frac{\partial \log L}{\partial \alpha} = \frac{P+Q}{\alpha\beta} = 0 \quad (8)$$

$$-\frac{\partial \log L}{\partial \beta} = \frac{R}{\beta} - \frac{(P+Q)}{\beta^2} = 0 \quad (9)$$

여기서, $P = N - 2 \sum_{i=1}^N \frac{e^{-y_i}}{1 + e^{-y_i}}$,

$$Q = (\beta - 1) \sum_{i=1}^N e^{\beta y_i} + 2 \sum_{i=1}^N \frac{e^{(\beta-1)y_i}}{1 + e^{-y_i}},$$

$$R = N - \sum_{i=1}^N y_i + 2 \sum_{i=1}^N \frac{y_i e^{-y_i}}{1 + e^{-y_i}}.$$

GL 분포형의 최우도법에 의한 추정치 $\hat{\theta} = (\epsilon, \alpha, \beta)^T$ 는 Eqs. (7)~(9)를 이용하여 구할 수 있으며, 다음의 Eq. (10)을 통해 반복 계산하게 된다.

$$\theta^{(k+1)} = \theta^{(k)} + J(\hat{\theta})^{-1} \text{grad} L(\theta^{(k)}) \quad (10)$$

여기서, $\text{grad} L(\theta)$ 은 Eq. (11)과 같이 나타낼 수 있고, $J(\hat{\theta})$ 는 관측정보행렬로 Eq. (12)와 같이 나타낼 수 있으며, 관측정보행렬의 각 항인 이계 편미분 도함수는 신흥준 등 (2007)의 부록 A에 나타나 있다.

$$\text{grad} L(\theta) = (\partial \log L / \partial \epsilon, \partial \log L / \partial \alpha, \partial \log L / \partial \beta)^T \quad (11)$$

$$J(\hat{\theta}) = \begin{bmatrix} -\frac{\partial^2 \log L}{\partial \epsilon^2} & -\frac{\partial^2 \log L}{\partial \epsilon \partial \alpha} & -\frac{\partial^2 \log L}{\partial \epsilon \partial \beta} \\ -\frac{\partial^2 \log L}{\partial \alpha \partial \epsilon} & -\frac{\partial^2 \log L}{\partial \alpha^2} & -\frac{\partial^2 \log L}{\partial \alpha \partial \beta} \\ -\frac{\partial^2 \log L}{\partial \beta \partial \epsilon} & -\frac{\partial^2 \log L}{\partial \beta \partial \alpha} & -\frac{\partial^2 \log L}{\partial \beta^2} \end{bmatrix} \quad (12)$$

2.2 Fisher 정보행렬과 관측정보행렬

Fisher 정보행렬은 θ 의 우도함수 $L(\theta) = f(x; \theta)$ 에 의존하는 미지의 매개변수 θ 를 가지는 무작위 변수 x 에 대한 정보의 양을 측정하는 방법이다. Fisher 정보행렬은 다음 Eq. (13)과 같이 나타낸다.

$$I(\theta) = E \left[\left\{ \frac{\partial \log f(x; \theta)}{\partial \theta} \right\}^2 \right] \quad (13)$$

여기서 $0 \leq I(\theta) < \infty$ 이다.

만약 위 식이 다음 정칙조건(regularity condition)을 만족한다면,

$$\int \frac{\partial^2}{\partial \theta^2} f(x; \theta) dx = 0 \quad (14)$$

Fisher 정보는 다음과 같이 나타낼 수 있다.

$$I(\theta) = -E \left[\frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \right] \quad (15)$$

그러므로 Fisher 정보는 θ 에 대한 $\log f(x; \theta)$ 의 이계 편미분 도함수에 기대값을 취하고 -를 곱한 것으로, θ 의 최우도 추정값 근처에서 대수우도함수의 변화 정도를 파악할 수 있다. 따라서 Fisher 정보행렬은 Eq. (15)를 이용하여 다음과 같이 나타낼 수 있으며, Fisher 정

보행렬과 관측정보행렬은 $I(\theta) = E[J(\theta)]$ 의 관계가 성립한다.

$$I(\theta) = \begin{bmatrix} E\left\{-\frac{\partial^2 \log L}{\partial \epsilon^2}\right\} & E\left\{-\frac{\partial^2 \log L}{\partial \epsilon \partial \alpha}\right\} & E\left\{-\frac{\partial^2 \log L}{\partial \epsilon \partial \beta}\right\} \\ & E\left\{-\frac{\partial^2 \log L}{\partial \alpha^2}\right\} & E\left\{-\frac{\partial^2 \log L}{\partial \alpha \partial \beta}\right\} \\ & & E\left\{-\frac{\partial^2 \log L}{\partial \beta^2}\right\} \end{bmatrix} \quad (16)$$

여기서, Fisher 정보행렬의 각 항인 이계 편미분 도함수의 기대값은 신흥준 등 (2007)의 부록 B에 수록되어 있다.

3. 모의실험을 통한 추정값의 분산-공분산

GL분포에 대한 Fisher 정보행렬 $I(\theta)$ 와 관측정보행렬 $J(\theta)$ 를 사용하여 quantile과 매개변수에 대한 분산을 비교하기 위해서 Prescott and Walden (1983)의 실험과정에 따라 표본크기 10, 20, 30, 50, 80, 100, 200, 1000에 대해서 10,000번의 반복을 통해 모의실험을 수행하였다. 위치매개변수와 크기매개변수는 각각 $x_0 = 0$ 과 $\alpha = 1$ 로 고정하였으며, 형상매개변수는 여러 가지 모양의 분포를 표현하기 위해서 -0.45부터 +0.45까지 0.05 단위로 변화시켜 각 세트별로 총 10개의 형상매개변수를 변화시켰다. 각 표본 세트에서 반복계산을 통해 최우도법에 의한 매개변수가 결정되며, 각 표본의 최우도 추정값에서 Fisher 정보행렬과 관측정보행렬의 역행렬이 결정되었다. 위 모의과정을 토대로 다음과 같은 항목을 비교하였다.

(1) 매개변수의 분산-공분산 행렬

- ① ML 매개변수 10,000개 세트의 분산과 공분산
- ② Fisher 정보행렬의 역행렬 평균
- ③ 관측정보행렬의 역행렬 평균
- ④ 특정 형상매개변수에서의 매개변수에 대한 근사적 분산과 공분산

(2) 재현기간 T=10, 100, 1000년일 때의 quantile에 대한 분산

- ① ML quantile 10,000개 세트의 분산
- ② Fisher 정보행렬을 이용한 quantile에 대한 분산의 평균
- ③ 관측정보행렬을 이용한 quantile에 대한 분산의 평균

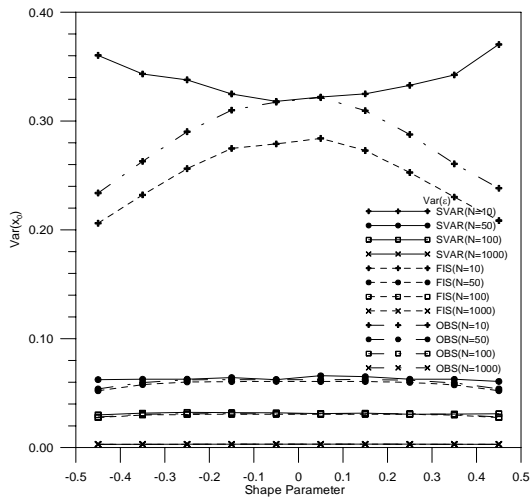
④ 특정 형상매개변수에서의 quantile에 대한 근사적 분산

매개변수에 대한 분산-공분산 추정 모의결과는 Fig. 1에 나타내었으며, 분량의 제한으로 인하여 표본크기 20, 30, 80, 200에 대한 경우는 수록하지 않았다. Figure 1에서 “SVAR”는 모의로 얻어진 결과로 실제값을 나타내며, “FIS”는 Fisher 정보행렬을 이용한 결과, “OBS”는 관측정보행렬을 이용한 결과를 나타낸다. 위의 결과를 비교해보면 위치매개변수의 경우 모든 경우에 대해서 관측정보행렬에 의한 결과가 Fisher 정보행렬에 의한 결과보다 실제값에 더 근접하는 결과를 보이며, 크기매개변수에 대한 분산과 크기매개변수와 형상매개변수에 대한 공분산은 관측정보행렬과 Fisher 정보행렬에 의한 결과가 거의 비슷한 값을 가진다. 한편 형상매개변수의 분산, 위치매개변수와 크기매개변수의 공분산, 위치매개변수와 형상매개변수의 공분산에 대한 결과는 표본크기가 큰 경우에는 관측정보행렬에 의한 결과가 실제값에 더 가까우나, 표본 크기가 작은 경우에는 Fisher 정보행렬에 의한 결과가 실제값에 더 가까운 것을 알 수 있다. 하지만 표본크기가 큰 경우에도 관측정보행렬에 의한 결과와 Fisher 정보행렬에 의한 결과의 차이가 거의 발생하지 않았다.

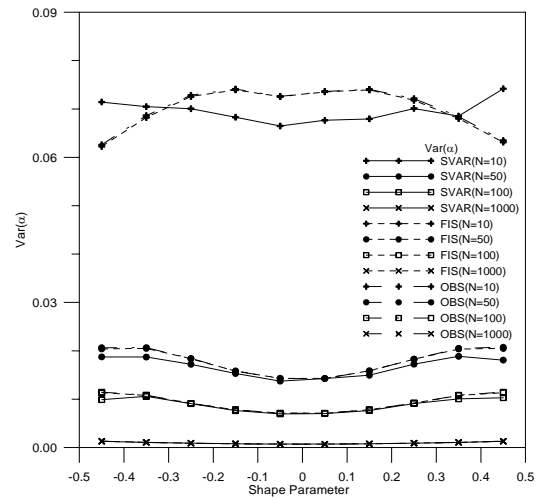
Quantile의 분산에 관한 모의결과는 Tables 1~4 및 Fig. 2에 나타내었다. Fig. 1에서와 마찬가지로 Tables 1~4의 “SVAR”는 모의로 얻어진 실제값, “FIS”는 Fisher 정보행렬을 이용한 결과, “OBS”는 관측정보행렬을 이용한 결과를 의미한다. 표본크기가 10~50일 경우 Fisher 정보행렬로부터 얻어진 quantile에 대한 분산은 관측정보행렬에 의한 결과보다 실제값에 더 가까우나, 표본 크기가 100보다 큰 경우 형상매개변수가 양쪽 끝 범위(± 0.5)에 가까운 경우를 제외하고는 표본 크기가 증가할수록 관측정보행렬에 의한 결과가 실제값에 더 근접하는 것을 알 수 있다.

특히 N=1000인 경우에는 관측정보행렬의 역행렬 평균값이 Fisher 정보행렬의 역행렬 평균값보다 모의결과에 의한 실제 분산값에 더 근접하는 값을 가지나, 두 정보행렬에 의한 차이는 거의 없다. 한편 표본 크기가 80~100이고 양으로 왜곡된 자료의 경우(즉, 음의 형상매개변수를 가지고 있는 경우) Fisher 정보행렬에 의한 결과가 실제에 더 가까우며, 음으로 왜곡된 자료의 경우(즉, 양의 형상매개변수를 가지고 있는 경우)에도 두 정보행렬에 의한 차이는 표본 크기가 작은 경우보다 상대적으로 훨씬 작다고 할 수 있다.

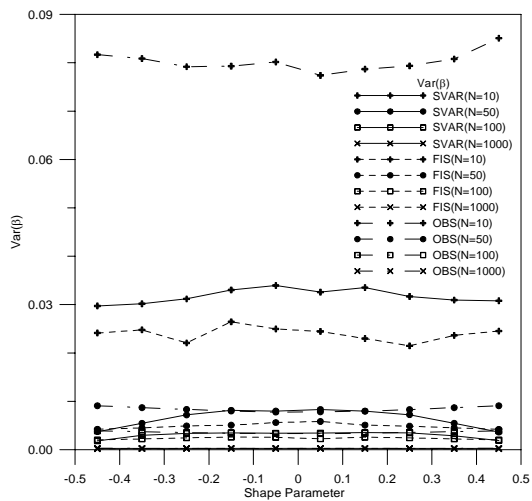
그러므로 실제 자료들의 경우 일반적으로 양으로 왜



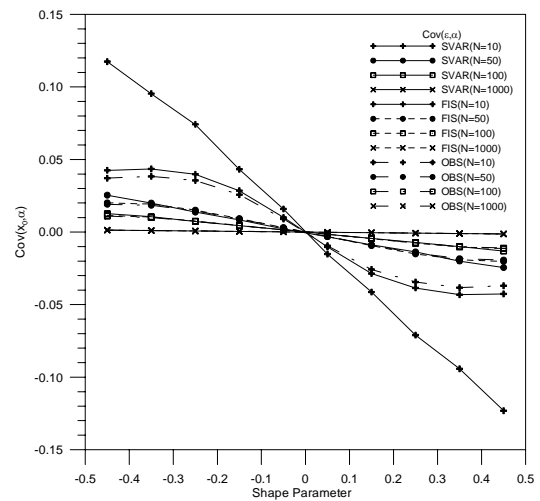
(a) $Var(\epsilon)$



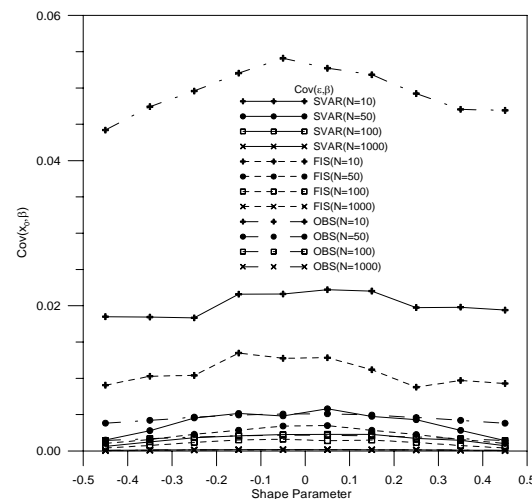
(b) $Var(\alpha)$



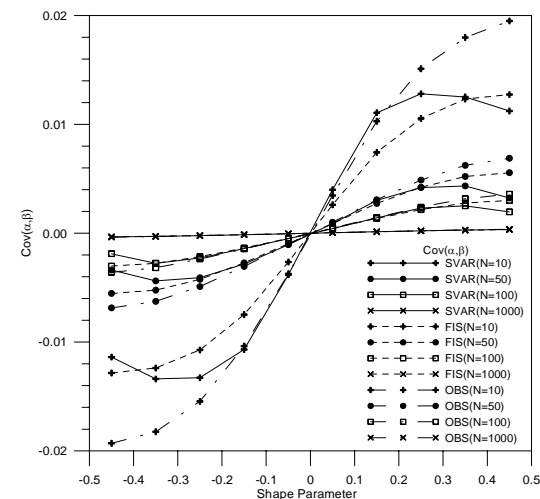
(c) $Var(\beta)$



(d) $Cov(\epsilon, \alpha)$



(e) $Cov(\epsilon, \beta)$



(f) $Cov(\alpha, \beta)$

Fig. 1. Comparisons of Variances-Covariances of Parameters for $N=10, 50, 100,$ and 1000

곡되어 있으며 자료의 크기도 거의 대부분 100보다 작기 때문에, 표본 크기가 100보다 작을 경우 최우도법에 의한 quantile의 분산을 추정하기 위해서는 관측정보행

렬보다 Fisher 정보행렬의 사용이 더 적절할 것으로 보인다.

Table 1. Comparison of Variance of Quantile with T=10, 100, and 1000 and N=10

Sample Size	Return Period	Shape Parameter	Var(XT)			②-①	③-①
			SVAR	FIS	OBS		
			①	②	③		
10	10	-0.45	1.84	1.50	1.95	-0.34	0.11
		-0.35	1.71	1.57	2.01	-0.14	0.30
		-0.25	1.55	1.54	1.93	-0.02	0.38
		-0.15	1.23	1.34	1.67	0.10	0.43
		-0.05	0.89	1.02	1.28	0.13	0.39
		0.05	0.59	0.75	0.94	0.15	0.34
		0.15	0.39	0.54	0.66	0.15	0.27
		0.25	0.26	0.40	0.48	0.14	0.22
		0.35	0.18	0.31	0.37	0.13	0.20
	0.45	0.13	0.26	0.30	0.14	0.18	
	100	-0.45	17.36	26.26	56.33	8.91	38.98
		-0.35	16.90	26.01	55.23	9.11	38.33
		-0.25	15.84	23.98	49.07	8.14	33.23
		-0.15	12.31	18.66	38.75	6.34	26.44
		-0.05	7.88	12.02	27.53	4.13	19.64
		0.05	4.56	7.21	17.71	2.65	13.15
		0.15	2.76	4.48	11.11	1.72	8.35
		0.25	1.85	3.06	6.66	1.21	4.81
		0.35	1.43	2.14	4.72	0.72	3.30
	0.45	1.20	1.78	3.36	0.58	2.16	
	1000	-0.45	151.35	326.97	843.87	175.62	692.53
		-0.35	143.88	307.29	805.05	163.41	661.17
		-0.25	132.56	270.14	670.64	137.58	538.08
		-0.15	96.01	189.64	491.17	93.63	395.15
		-0.05	55.68	107.91	323.17	52.22	267.49
		0.05	29.74	56.30	190.10	26.56	160.36
		0.15	17.83	32.24	107.73	14.42	89.91
0.25		12.19	21.08	47.57	8.89	35.38	
0.35		10.59	15.71	23.89	5.12	13.30	
0.45	8.46	12.49	16.27	4.03	7.80		

Table 2. Comparison of Variance of Quantile with T=10, 100, and 1000 and N=50

Sample Size	Return Period	Shape Parameter	Var(XT)			②-①	③-①
			SVAR	FIS	OBS		
			①	②	③		
50	10	-0.45	0.60	0.68	0.78	0.08	0.17
		-0.35	0.56	0.62	0.70	0.06	0.14
		-0.25	0.44	0.47	0.52	0.03	0.08
		-0.15	0.31	0.31	0.34	0.01	0.03
		-0.05	0.19	0.20	0.21	0.01	0.02
		0.05	0.12	0.12	0.12	0.00	0.00
		0.15	0.07	0.07	0.08	0.00	0.00
		0.25	0.05	0.04	0.05	0.00	0.00
		0.35	0.03	0.03	0.03	0.00	0.00
		0.45	0.02	0.02	0.02	0.00	0.00
	100	-0.45	10.75	15.69	23.76	4.95	13.01
		-0.35	10.75	12.65	18.43	1.90	7.68
		-0.25	7.78	7.70	10.79	-0.08	3.01
		-0.15	4.30	3.86	5.11	-0.44	0.82
		-0.05	1.86	1.77	2.15	-0.09	0.29
		0.05	0.82	0.78	0.92	-0.03	0.11
		0.15	0.33	0.29	0.39	-0.05	0.05
		0.25	0.14	0.12	0.17	-0.02	0.03
		0.35	0.06	0.05	0.08	-0.01	0.03
		0.45	0.02	0.02	0.05	0.00	0.02
	1000	-0.45	147.39	243.37	420.59	95.97	273.20
		-0.35	139.19	172.89	287.18	33.70	147.99
		-0.25	86.11	85.06	136.58	-1.05	50.47
		-0.15	37.02	32.25	48.82	-4.77	11.80
		-0.05	11.52	10.69	14.33	-0.83	2.81
		0.05	3.77	3.60	4.57	-0.16	0.80
		0.15	1.16	0.95	1.43	-0.21	0.27
		0.25	0.37	0.30	0.50	-0.08	0.12
		0.35	0.12	0.10	0.20	-0.03	0.08
		0.45	0.04	0.04	0.10	-0.01	0.06

Table 3. Comparison of Variance of Quantile with T=10, 100, and 1000 and N=100

Sample Size	Return Period	Shape Parameter	Var(XT)			②-①	③-①	
			SVAR	FIS	OBS			
			①	②	③			
100	10	-0.45	0.34	0.39	0.44	0.06	0.10	
		-0.35	0.33	0.33	0.36	0.00	0.03	
		-0.25	0.23	0.23	0.24	0.00	0.01	
		-0.15	0.15	0.15	0.15	0.00	0.00	
		-0.05	0.09	0.09	0.10	0.00	0.00	
		0.05	0.06	0.06	0.06	0.00	0.00	
		0.15	0.04	0.04	0.04	0.00	0.00	
		0.25	0.02	0.02	0.02	0.00	0.00	
		0.35	0.01	0.01	0.01	0.00	0.00	
	0.45	0.01	0.01	0.01	0.00	0.00		
	100	100	-0.45	7.04	9.67	13.53	2.63	6.50
			-0.35	6.98	6.77	9.01	-0.21	2.03
			-0.25	3.89	3.51	4.36	-0.39	0.46
			-0.15	1.84	1.65	1.94	-0.19	0.10
			-0.05	0.80	0.77	0.85	-0.03	0.05
			0.05	0.35	0.33	0.37	-0.02	0.01
			0.15	0.15	0.13	0.16	-0.02	0.01
			0.25	0.06	0.05	0.07	-0.01	0.00
			0.35	0.02	0.02	0.03	-0.01	0.00
	0.45	0.01	0.01	0.01	0.00	0.01		
	1000	1000	-0.45	108.70	158.26	248.18	49.56	139.47
			-0.35	94.70	91.38	135.16	-3.32	40.46
			-0.25	40.01	34.91	47.36	-5.11	7.35
			-0.15	13.78	11.88	14.88	-1.91	1.10
			-0.05	4.30	4.09	4.67	-0.20	0.37
			0.05	1.40	1.34	1.50	-0.06	0.10
			0.15	0.47	0.39	0.50	-0.08	0.03
0.25			0.16	0.12	0.17	-0.04	0.02	
0.35			0.05	0.03	0.06	-0.02	0.01	
0.45	0.02	0.01	0.03	-0.01	0.01			

Table 4. Comparison of Variance of Quantile with T=10, 100, and 1000 and N=100

Sample Size	Return Period	Shape Parameter	Var(XT)			②-①	③-①
			SVAR	FIS	OBS		
			①	②	③		
1000	10	-0.45	0.05	0.05	0.05	0.00	0.00
		-0.35	0.03	0.03	0.03	0.00	0.00
		-0.25	0.02	0.02	0.02	0.00	0.00
		-0.15	0.01	0.01	0.01	0.00	0.00
		-0.05	0.01	0.01	0.01	0.00	0.00
		0.05	0.01	0.01	0.01	0.00	0.00
		0.15	0.00	0.00	0.00	0.00	0.00
		0.25	0.00	0.00	0.00	0.00	0.00
		0.35	0.00	0.00	0.00	0.00	0.00
	0.45	0.00	0.00	0.00	0.00	0.00	
	100	-0.45	1.41	1.26	1.53	-0.15	0.12
		-0.35	0.71	0.63	0.70	-0.08	-0.01
		-0.25	0.33	0.30	0.32	-0.02	-0.01
		-0.15	0.15	0.14	0.15	0.00	0.00
		-0.05	0.07	0.07	0.07	0.00	0.00
		0.05	0.03	0.03	0.03	0.00	0.00
		0.15	0.01	0.01	0.01	0.00	0.00
		0.25	0.01	0.00	0.01	0.00	0.00
		0.35	0.00	0.00	0.00	0.00	0.00
	0.45	0.00	0.00	0.00	0.00	0.00	
	1000	-0.45	25.93	21.93	28.74	-3.99	2.81
		-0.35	9.08	7.78	9.03	-1.30	-0.05
		-0.25	2.93	2.68	2.87	-0.25	-0.05
		-0.15	0.95	0.92	0.96	-0.03	0.01
		-0.05	0.33	0.34	0.33	0.01	0.00
		0.05	0.11	0.10	0.11	-0.01	0.00
		0.15	0.04	0.04	0.04	0.00	0.00
0.25		0.01	0.01	0.01	0.00	0.00	
0.35		0.00	0.00	0.00	0.00	0.00	
0.45	0.00	0.00	0.00	0.00	0.00		

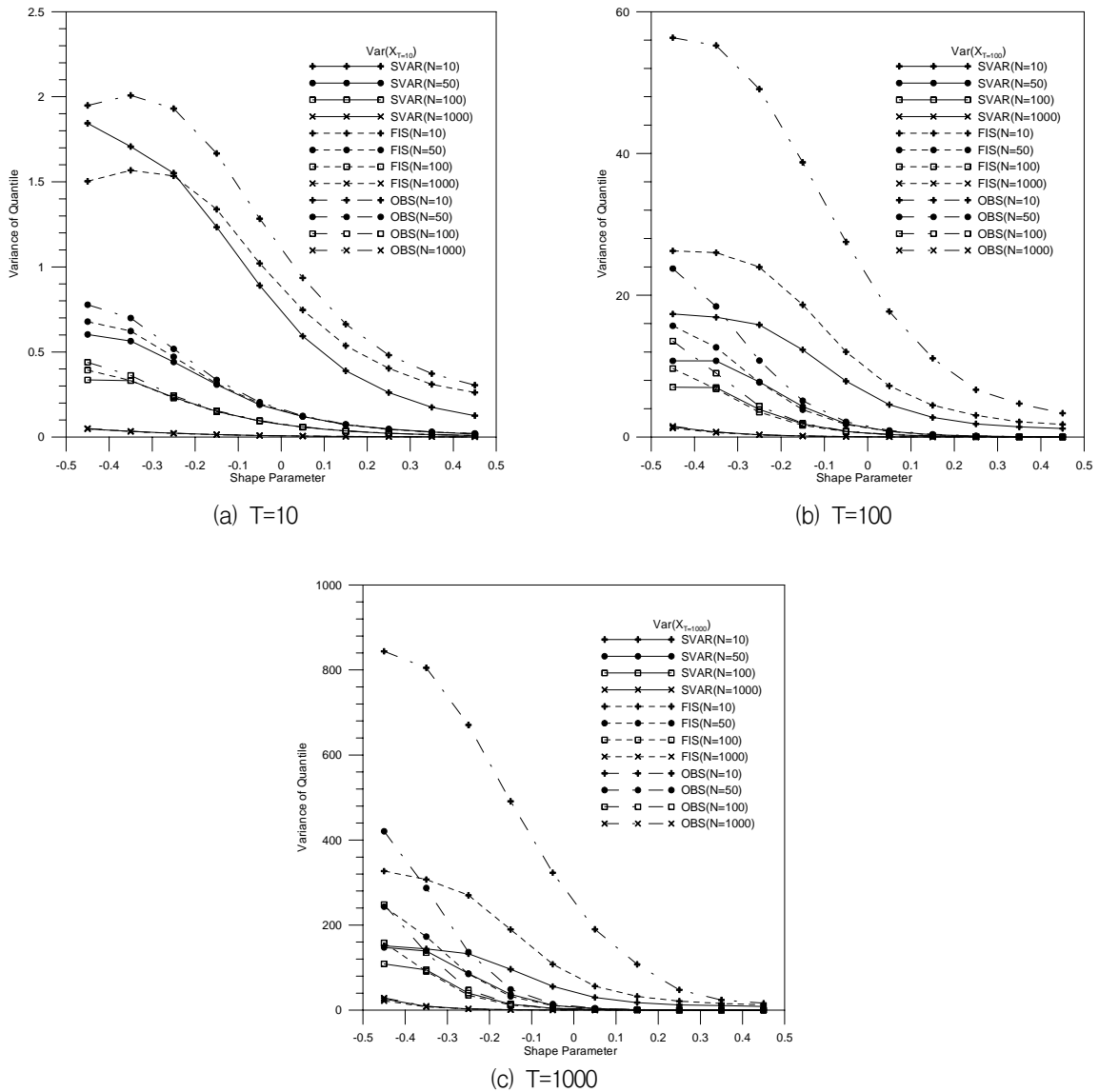


Fig. 2. Comparison of Variance of Quantile for T=10, 100, and 1000

4. 결론

본 논문은 일반적으로 관측정보행렬을 사용한 매개변수 및 quantile의 분산-공분산 추정이 Fisher 정보행렬을 이용한 결과보다 더 우수하다는 기존의 연구결과가 GL 분포에도 적용되는지 살펴보기 위한 연구이다. 아직 GL 분포에 대한 분산-공분산 추정 시 관측정보행렬의 사용에 관한 구체적인 연구결과가 발표된 적이 없기 때문에 본 논문에서는 모의실험을 통해 GL 분포의 정보행렬에 따른 분산-공분산 추정의 정확도를 비교해 보았다. 모의 결과 GL 분포의 매개변수에 대한 분산-공분산은 기존 문헌의 타 분포에 대한 연구 결과와 대체적으로 비슷한 결과를 보이나, 기존의 연구 결과와 달리 형상매개변수의 분산, 위치매개변수와 크기매개변

수의 공분산, 위치매개변수와 형상매개변수의 공분산에 대한 결과는 표본 크기가 작은 경우 Fisher 정보행렬에 의한 결과가 실제값에 더 가까운 것을 알 수 있다. 따라서 기존 문헌의 연구결과에서는 표본의 크기가 클 경우 관측정보행렬이 매개변수 추정시 동시에 추정되어 계산 시간도 훨씬 단축되며 Fisher 정보행렬의 정확도와 차이도 거의 없어 관측정보행렬의 사용이 추천되었으나, 현재는 계산기의 발달로 인한 계산시간의 단축으로 인하여 관측정보행렬에 의한 매개변수의 분산-공분산의 추정에 큰 장점이 없는 상황이다. 반면, quantile에 대한 분산의 모의결과는 $N \leq 100$ 인 경우 최우도법에 의한 quantile의 분산을 추정하기 위해서는, 관측정보행렬보다 Fisher 정보행렬의 사용이 더 적절할 것으로 나타났다. 따라서 Rao and Hamed (2000)에 의해 주장된 바와

같이 GL 분포에서 매개변수 및 quantile의 분산-공분산을 추정하기 위해 관측정보행렬을 사용하는 것은 잘못된 것이라고 판단되며, GL 분포의 최우도 추정값에 대한 불확실성 추정에는 Fisher 정보행렬의 사용이 적절한 것이라고 판단된다.

감사의 글

이 연구는 소방방재청 자연재해저감기술개발사업 [과제번호 : NEMA-07-NH-02] 연구비 지원으로 수행 되었으며 이에 감사드립니다.

참 고 문 헌

신홍준, 김수영, 허준행 (2007). "Generalized Logistic 분포형의 신뢰구간 추정을 위한 점근 분산 유도: I. 점근 분산식의 유도." **대한토목학회 논문집**, 대한토목학회, 제27권, 제3B호, pp. 331-338.

Ahmad, M.I., Sinclair, C.D., and Werritty, A. (1988). "Log-Logistic Flood Frequency Analysis with Historical Information." *Journal of Hydrology*, Vol. 98, pp. 205-224.

Cohen, A.C. (1965). "Maximum Likelihood Estimation in the Weibull Distribution." *Technometrics*, Vol. 15, No. 4, pp. 415-428.

Efron, B. and Hinkley, D.V. (1978). "Assessing the Accuracy of the Maximum Likelihood Estimator: Observed versus Expected Fisher Information." *Biometrika*, Vol. 65, No. 3, pp. 457-487.

Hinkley, D.V. (1978). "Likelihood Inference about Location and Scale Parameters." *Biometrika*, Vol. 65, No. 2, pp. 253-261.

Hosking, J.R.M. and Wallis, J.R. (1997). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, New York.

Lemon, G.H. (1975). "Maximum Likelihood Estimation for the Three Parameter Weibull Distribution based on Censored Samples." *Technometrics*, Vol. 17, No. 2, pp. 247-254.

Prescott, P. and Walden, A.T. (1983). "Maximum Likelihood Estimation of the Parameters of the Three Parameter Generalized Extreme Value Distribution from Censored Samples." *J. Statist. Comput. Simul.*, Vol. 16, pp. 241-250.

Rao, A.R. and Hamed, K.H. (2000). *Flood Frequency Analysis*, CRC Press, Florida.

(논문번호:09-20/접수:2009.02.11/심사완료:2009.03.25)