

Objective Bayesian Estimation of Two-Parameter Pareto Distribution

Young Sook Son^{a,1}

^aDepartment of Statistics, Chonnam National University

(Received July 15, 2013; Revised August 23, 2013; Accepted October 15, 2013)

Abstract

An objective Bayesian estimation procedure of the two-parameter Pareto distribution is presented under the reference prior and the noninformative prior. Bayesian estimators are obtained by Gibbs sampling. The steps to generate parameters in the Gibbs sampler are from the shape parameter of the gamma distribution and then the scale parameter by the adaptive rejection sampling algorithm. A numerical study shows that the proposed objective Bayesian estimation outperforms other estimations in simulated bias and mean squared error.

Keywords: 2-parameter Pareto distribution, L-moment estimation, maximum likelihood estimation, noninformative prior, reference prior, objective Bayesian estimation, Gibbs sampling, adaptive rejection sampling.

1. 서론

Pareto (1897)에 의해 처음 소개되었던 파레토분포(Pareto distribution)는 인구 규모, 주가 변동, 보험 손실, 보험금 청구, 소득, 급여, 인터넷 트래픽, 최대 풍속, 최대 강우량, 지진 혹은 산불과 같은 자연 재해의 규모, 유전 지역에서 오일 매장량 등과 같이 인구, 사회, 경제, 자연 환경적 현상 등 많은 분야의 자료에 적용되는 확률모형으로 중요한 역할을 해 왔다 (Crovella 등, 1998; Harris 등, 2000; Smith, 2003; Embrechts 등, 1997; Jackson과 Kagan, 1999; Cumming, 2001; Arnold, 1983; Arnold와 Press, 1989). 파레토분포는 지수분포와 모양은 비슷하나 꼬리가 더 두터우므로 극대값의 발생이 상대적으로 더 많이 발생하는 자료의 적용에 유용하다. 2-모수 파레토 pdf(probability density function)는 다음과 같이 정의된다.

$$f(x) = \frac{\alpha\beta^\alpha}{x^{\alpha+1}}, \quad x \geq \beta, \quad \alpha, \beta > 0, \quad (1.1)$$

여기서 α 는 분포의 모양을 결정하는 형태(shape) 모수이고, β 는 척도(scale) 모수이다. 분포가 시작하는 위치를 나타낸다는 의미에서 β 를 위치(location) 모수라고도 한다. 식 (1.1)처럼 정의되는 파레토분포를 $\text{Pareto}(\alpha, \beta)$ 분포라 나타내자. Figure 1.1은 모수 α, β 의 여러 값에 따른 $\text{Pareto}(\alpha, \beta)$ pdf를 보

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(NRF-2011-0022864).

¹Professor, Department of Statistics, Chonnam National university, 300 Yongbong-Dong, Buk-Gu, Kwangju 500-757, Korea. E-mail: ysson@jnu.ac.kr

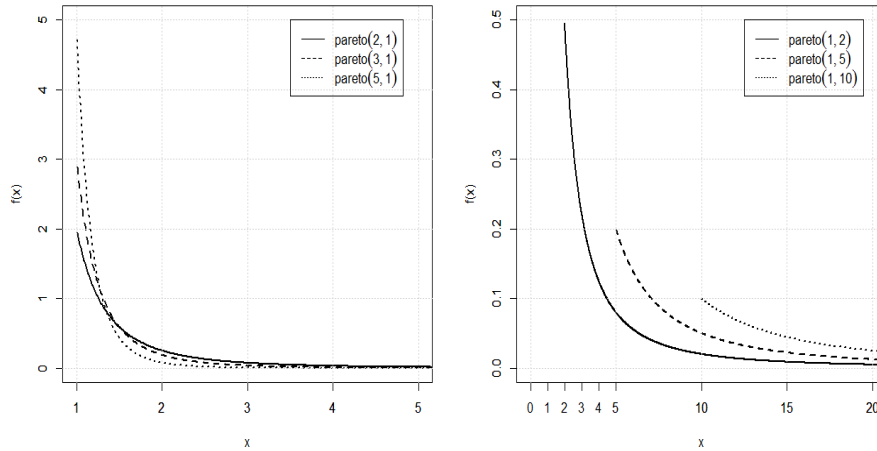


Figure 1.1. Pareto pdf for various α and β

여준다. 형태 모수 α 는 분포의 산포(dispersion)를 나타내는 척도로서 α 가 작을수록 분포의 꼬리가 두터워지며 산포가 커진다. 반대로 α 가 커지면 β 에서 확률 1을 가지는 퇴화분포(degenerate distribution)가 된다.

파레토분포는 Pickands (1975)가 제안한 일반화 파레토분포의 특별한 형태이다. 2-모수 파레토분포가 하한(lower limit)만 갖는 반면에 일반화 파레토분포는 상한(upper limit)도 허용한다. 그동안 2-모수 파레토분포의 모수 추정에 관한 많은 연구가 있어 왔다. Singh와 Guo (1995)는 전통적인 추정 방법으로 적률법, 확률가중법, 최대엔트로피법, 최우추정법의 성능을 평가하는 모의실험을 수행하였다. RMSE(root mean square error)의 기준 하에서 적률법은 다른 세 개의 추정법보다 현저하게 큰 RMSE 값을 주었으며 소표본 및 대표본에서 최우추정법이 가장 우수하였다. 일반적으로 최우추정법을 선호하는 대표적인 이유는 확률분포가 정규조건(regularity condition)을 만족하는 경우에 최우추정량이 점근 정규성을 가지므로 대표본에서 모수의 함수에 대한 구간추정을 가능하게 한다. 일반적으로 확률분포의 지지(support)가 모수에 종속하는 경우에는 정규조건을 만족시키지 못한다 (Casella와 Berger, 2002). 파레토분포가 이에 해당한다. 이에 비해서 깁스샘플링(Gibbs sampling)에 의한 베이지안 추론에서는 결합사후분포로부터 주변사후분포를 구할 때 적분 계산에 의하지 않고 깁스샘플러(Gibbs sampler)에서 모수를 생성하여 주변사후분포를 얻을 수 있으며, 생성된 모수들의 표본으로부터 모수 뿐만 아니라 모수들의 함수에 대한 추정량의 분포도 쉽게 얻을 수 있다. 따라서 모수의 함수들에 대한 구간 추정도 가능하다. Arnold와 Press (1989)는 확률표본 및 중도절단표본의 경우에 공액사전분포(conjugate prior)의 가정 하에서 베이지안 추정을 제안하였다. Berger (2006)는 베이지안 추론에서 사전분포에 대한 정보가 부족할 때 주관적 사전분포의 사용은 좋지 않은 추론 결과를 줄 수 있다고 지적하였다. 사전분포에 대한 정보가 없을 때 디폴트사전분포(default prior)라고 하는 무정보사전분포(noninformative prior) 하에서 수행하는 객관적 베이지안 추론(objective Bayesian inference)이 대안이 될 수 있다고 하였다. Fu 등 (2012)은 전진적인 유형 II 중도절단 자료(progressive Type II censored data)에 대하여 디폴트사전분포인 준거사전분포(reference prior)하에서 객관적 베이지안 추론을 제안하였다.

본 연구에서는 Fu 등 (2012)이 사용한 준거사전분포의 가정 하에서 2-모수 파레토분포에 대한 객관적 베이지안 모수추정 절차를 제안한다. 제안된 베이지안 추정법은 모의실험과 자료분석을 통하여 L-적률 추정법 및 최우추정법과 비교 평가된다.

2. 고전적 추정방법

본 절에서는 고전적 추정방법으로서 L-적률추정법 및 최우추정법에 의하여 2-모수 파레토분포의 모수 α 및 β 에 대한 추정량을 계산한다. $X = \{X_1, X_2, \dots, X_n\}$ 를 식 (1.1)의 $\text{Pareto}(\alpha, \beta)$ 분포로부터 추출한 확률표본이라 하고, $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ 를 만족하는 $X_{(j)}$ 를 j 번째 순서통계량이라고 하자.

2.1. L-적률추정법

$\text{Pareto}(\alpha, \beta)$ 분포의 cdf(cumulative distribution function)는 다음과 같다.

$$F(x) = 1 - \left(\frac{\beta}{x}\right)^\alpha, \quad x \geq \beta, \alpha, \beta > 0. \quad (2.1)$$

Greenwood 등 (1979)이 처음 제안하여 수문학 분야에서 널리 사용되어 온 확률가중적률(probability weighted moment) 추정량은 식 (2.2)로 부터, 그리고 Hosking (1990)이 제안한 L-적률 추정량은 식 (2.3)으로 부터 얻을 수 있다. L-적률은 순서통계량을 사용하여 분포의 위치, 산포, 왜도, 첨도를 정의할 수 있으며 기존의 적률에 비하여 분포의 특징을 잘 나타내주어 매우 유용하게 사용되어 오고 있다.

$$\alpha_s = E[X\{1 - F(X)\}^s] = \sum_{r=0}^s \binom{s}{r} (-1)^r \beta_r, \quad s = 0, 1, 2, \dots, \quad (2.2)$$

$$\beta_r = E[X\{F(X)\}^r], \quad r = 0, 1, 2, \dots \quad (2.3)$$

식 (2.3)에 식 (2.1)의 cdf를 대입한 적분계산은 쉽지 않다. 따라서 α_s 를 통하여 β_r 를 구한다. 식 (2.2)로부터 다음을 얻는다.

$$\alpha_s = \frac{\alpha\beta}{\alpha(s+1) - 1}, \quad \alpha > \frac{1}{s+1}, \quad \alpha_0 = \beta_0, \quad \alpha_1 = \beta_0 - \beta_1. \quad (2.4)$$

$\text{Pareto}(\alpha, \beta)$ 분포는 모수의 개수가 두 개이므로 다음과 같이 처음 두 개의 L-적률 λ_1 과 λ_2 가 필요하다.

$$\lambda_1 = \beta_0, \quad \lambda_2 = 2\beta_1 - \beta_0. \quad (2.5)$$

식 (2.4)와 식 (2.5)의 관계에서 모 L-적률 λ_1, λ_2 를 표본 L-적률 l_1, l_2 로 각각 대체하여 얻어진 α 와 β 에 대한 L-적률추정량(L-moment estimator; LME)은 다음과 같다.

$$\hat{\alpha}_{LME} = \frac{l_1 + l_2}{2l_2}, \quad \hat{\beta}_{LME} = \frac{(\hat{\alpha}_{LME} - 1)l_1}{\hat{\alpha}_{LME}},$$

여기서 $l_1 = 1/n \sum_{j=1}^n X_{(j)}, l_2 = 1/\{n(n-1)\} \sum_{j=1}^n (2j - n - 1)X_{(j)}$.

2.2. 최우추정법

$\text{Pareto}(\alpha, \beta)$ 분포의 우도함수(likelihood function),

$$L(\alpha, \beta|X) \propto \alpha^n \beta^{n\alpha} \left(\prod_{j=1}^n X_j\right)^{-\alpha} \quad (2.6)$$

를 최대로 해주는 최우추정량(maximum likelihood estimator; MLE)은 다음과 같다.

$$\hat{\alpha}_{MLE} = \frac{n}{\sum_{j=1}^n \ln X_j - n \ln X_{(1)}}, \quad \hat{\beta}_{MLE} = X_{(1)}. \quad (2.7)$$

3. 베이지안 추정

3.1. 공액사전분포 하에서 베이지안 추정

본 절에서는 Arnold와 Press (1983)가 제안한 공액사전분포 가정 하에서의 베이지안 모수추정 결과를 소개한다. 식 (3.1)과 같이 정의되는 $\text{Pareto}(\alpha, \beta)$ 분포에 대한 공액사전분포는 Lwin (1972)에 의해 처음 제안되었고 Arnold와 Press (1983)에 의해 일반화되었다.

$$\alpha \sim \text{Gamma}(a_0, b_0), \quad \beta^{-1}|\alpha \sim \text{Pareto}(\alpha c_0, d_0), \quad (3.1)$$

여기서 α 의 사전 pdf는 $\pi(\alpha) \propto \alpha^{a_0-1} \exp(-b_0\alpha)$ 이고, a_0, b_0, c_0, d_0 는 양의 값을 가지는 초모수(hyperparameter)이다. 공액사전분포의 식 (3.1) 하에서 주변사후분포는 다음과 같이 얻어진다..

$$\alpha|X \sim \text{Gamma}(a_0 + n, b_0 + \bar{v}), \quad \beta^{-1}|\alpha, X \sim \text{Pareto}(\alpha(c_0 + n), \min(d_0, X_{(1)})), \quad (3.2)$$

여기서 $\bar{v} = \sum_{j=1}^n \ln X_j + c_0 \ln d_0 - (n + c_0) \ln(\min(d_0, X_{(1)}))$.

식 (3.2)로 부터 α 및 β 의 각 사후최빈값(posterior mode)을 공액사전분포 하에서의 베이지안 추정량으로 하였을 때 다음과 같이 얻어진다.

$$\hat{\alpha}_{conjugate} = \frac{a_0 + n - 1}{b_0 + \bar{v}}, \quad \hat{\beta}_{conjugate} = \min(d_0, X_{(1)}). \quad (3.3)$$

3.2. 무정보 사전분포 하에서 베이지안 추정

본 절에서는 Fu 등 (2012)이 제안한 객관적 사전분포로서 다음의 준거사전분포(reference prior)의 가정 하에서 객관적 베이지안 모수추정 절차를 제안한다.

$$\pi(\alpha, \beta) \propto \frac{1}{\alpha\beta}. \quad (3.4)$$

결합사후분포는 우도함수 식 (2.6)과 사전분포 식 (3.4)를 곱하여 다음과 같이 얻어진다.

$$\pi(\alpha, \beta|X) \propto \alpha^{n-1} \beta^{n\alpha-1} \left(\prod_{j=1}^n X_j \right)^{-\alpha}. \quad (3.5)$$

사전분포 식 (3.4)는 부적절 사전분포(improper prior)이다. 부적절 사전분포를 사용하는 경우에 사후분포가 pdf로서 부적절한 경우에는 베이지안 추정이 불가능하다. 그러나 다음에 의해 식 (3.5)에서 정의되는 사후분포는 pdf로서 적절함을 알 수 있다.

$$\begin{aligned} \int_0^\infty \int_0^{x_{(1)}} \pi(\alpha, \beta|X) d\beta d\alpha &= \int_0^\infty \frac{1}{n} \alpha^{(n-1)-1} \exp \left[-\alpha \left\{ \sum_{j=1}^n \ln x_j - n \ln x_{(1)} \right\} \right] d\alpha \\ &= \frac{\Gamma(n-1)}{n \left\{ \sum_{j=1}^n \ln x_j - n \ln x_{(1)} \right\}^{n-1}} < \infty. \end{aligned}$$

식 (3.5)의 결합사후분포로부터 완전 조건부사후분포는 다음과 같이 정의된다.

$$\pi(\alpha|\beta, X) \propto \alpha^{n-1} \exp \left[-\alpha \left\{ \sum_{j=1}^n \ln x_j - n \ln \beta \right\} \right]. \quad (3.6)$$

$$\pi(\beta|\alpha, X) \propto \beta^{n\alpha-1}. \quad (3.7)$$

Initial step: Adopt $\beta^{(0)}$ as the initial value of β .

Replication step: for $i = 0, 1, 2, \dots, I$,

- Generate $\alpha^{(i+1)} \sim \text{Gamma} \left(n, \sum_{j=1}^n \ln x_j - n \ln \beta^{(i)} \right)$.
- Generate $\beta^{(i+1)}$ by ARS algorithm given $\alpha^{(i+1)}$.

Figure 3.1. Gibbs Sampler for generating α and β

기각표집법(rejection sampling)은 pdf로부터 확률변수를 생성하는 일반적인 방법이다. 기각표집법을 사용하여 정규화상수 위까지만 명시되어 있는, 즉, pdf의 kernel만 주어진 kernel pdf로부터 변수를 생성하고자 하는 경우에 필요한 kernel pdf의 덮개함수(envelope function)를 찾는 것은 현실적으로 어렵다. Gilks와 Wild (1992)가 제안한 적응기각표집법(adaptive rejection sampling; ARS)은 kernel pdf의 로그-오목성(log-concavity)을 이용하여 kernel pdf의 로그함수에 접선을 연결하여 만든 조각적 선덮개함수(piecewise linear envelope function)를 사용한다.

식 (3.6)은 Gamma pdf의 커널이며, 식 (3.7)에서 $n > 1/\alpha$ 인 경우에 $\partial^2 \ln \pi(\beta|\alpha, X)/\partial \beta^2 = -(n\alpha - 1)/\beta^2 < 0$ 가 되어 $\pi(\beta|\alpha, X)$ 는 로그-오목성을 갖는다. $n > 1/\alpha$ 는 현실적으로 만족되는 조건이다. 따라서 깁스샘플러를 Figure 3.1과 같이 구성한다.

4. 수치분석

본 연구에서 제안한 객관적 베이지안 모수추정법의 평가를 위하여 모의실험과 자료분석을 수행하였다. 모든 계산은 R package(version 2.15.3)를 사용하여 이루어졌다. 파레토, 감마 확률변수 생성을 위하여 rpareto 및 rgamma 함수를 각각 사용하였고 ARS 알고리즘을 구현하기 위하여 ars.new 함수를 사용하였다. 또한 HPD(highest posterior density) interval을 구하기 위하여 HPDinterval 함수를 사용하였다. HPDinterval 함수는 경험적 cdf로부터 지정된 확률을 가지는 모든 가능한 구간 중에서 가장 길이가 짧은 구간을 HPD interval로 계산한다.

4.1. 모의실험

척도모수 β 는 척도불변성(scale invariance)을 가지므로 $\beta = 1$ 로 놓는다. 그리고 $\alpha = 0.5, 1, 3, 5, 10$ 인 $\text{Pareto}(\alpha, \beta)$ 로부터 표본의 크기가 $n = 15, 30, 100$ 인 자료세트를 각각 1,000회 반복 생성하여 모의 실험을 하였다. $n > 2$ 이므로 ARS 알고리즘의 사용이 가능하다. 깁스샘플링은 11,000회 반복 수행하여 처음 1,000회를 버리고 나머지 10,000회 생성된 α 및 β 값들로 사후분포를 구성하였다. 시험적인 모의실험에서 α 및 β 의 사후분포는 각각 대칭 및 좌비대칭 분포의 형태를 가지고 있었다. 따라서 α 및 β 의 각 사후분포의 평균과 최빈값을 베이지안 추정량으로 사용하기로 하였다. Table 4.1에는 L-적률추정량(LME), 최우추정량(MLE), 준거사전분포 가정 하의 객관적 베이지안 추정량(BAYES)들의 bias와 MSE(mean squared error)가 제시되어 있다. 모의실험 결과는 다음과 같이 요약된다. 첫째, 모형 및 표본의 크기에 관계없이 bias와 MSE의 기준 하에서 BAYES, MLE, LME 순으로 우수하였다. BAYES는 소표본($n = 15$)에서 MLE보다 더 우수하고 대표본($n = 100$)에서는 비슷하다. 둘째, α 의 값이 커질수록 분포의 산포가 커지므로 α 추정치들의 bias와 MSE도 점점 커진다. 셋째, 표본의 크기가 커짐에 따라 bias와 MSE는 점점 작아진다.

Table 4.1. Simulated bias and MSE for five Pareto(α, β) models

estimation		$n = 15$		$n = 30$		$n = 100$	
method		bias	MSE	bias	MSE	bias	MSE
$\alpha = 0.5$	LME	0.6139	0.3947	0.5608	0.3188	0.5251	0.2764
	MLE	0.0830	0.0356	0.0313	0.0109	0.0116	0.0028
	BAYES	0.0442	0.0270	0.0136	0.0095	0.0065	0.0027
$\beta = 1.0$	LME	0.1582	0.0555	0.0699	0.0104	0.0208	0.0009
	MLE	0.1593	0.0557	0.0698	0.0104	0.0208	0.0009
	BAYES	0.1293	0.0456	0.0556	0.0085	0.0167	0.0007
$\alpha = 1.0$	LME	0.4513	0.3153	0.3513	0.1741	0.2414	0.0780
	MLE	0.1497	0.1312	0.0696	0.0454	0.0203	0.0108
	BAYES	0.0730	0.1002	0.0340	0.0391	0.0100	0.0103
$\beta = 1.0$	LME	0.0439	0.0107	0.0229	0.0026	0.0069	0.0004
	MLE	0.0695	0.0105	0.0332	0.0021	0.0102	0.0002
	BAYES	0.0551	0.0085	0.0262	0.0017	0.0081	0.0002
$\alpha = 3.0$	LME	0.4796	1.5096	0.2646	0.6271	0.1000	0.1793
	MLE	0.4433	1.1654	0.2081	0.4195	0.0667	0.1040
	BAYES	0.2140	0.8909	0.1013	0.3619	0.0361	0.0987
$\beta = 1.0$	LME	-0.0065	0.0024	-0.0087	0.0011	-0.0077	0.0004
	MLE	0.0232	0.0010	0.0111	0.0002	0.0033	0.0000
	BAYES	0.0185	0.0008	0.0088	0.0002	0.0026	0.0000
$\alpha = 5.0$	LME	0.6899	4.0487	0.3538	1.8020	0.1299	0.4605
	MLE	0.7970	3.3483	0.3667	1.2686	0.1199	0.2955
	BAYES	0.4096	2.5315	0.1878	1.0951	0.0688	0.2805
$\beta = 1.0$	LME	-0.0054	0.0009	-0.0061	0.0004	-0.0042	0.0001
	MLE	0.0140	0.0004	0.0068	0.0001	0.0021	0.0000
	BAYES	0.0112	0.0003	0.0054	0.0001	0.0017	0.0000
$\alpha = 10.0$	LME	1.2918	16.0761	0.5647	5.4183	0.1927	1.6072
	MLE	1.6380	14.4762	0.7329	4.4283	0.2215	1.1762
	BAYES	0.8630	11.0176	0.3756	3.7810	0.1192	1.1203
$\beta = 1.0$	LME	-0.0020	0.0002	-0.0027	0.0001	-0.0019	0.0000
	MLE	0.0071	0.0001	0.0032	0.0000	0.0010	0.0000
	BAYES	0.0057	0.0001	0.0025	0.0000	0.0008	0.0000

Table 4.2. Annual wage data(unit: in multiples of 100 U.S. dollars)

112	154	119	108	112	156	123	103	115	107	125	119	128	132	107
151	103	104	116	140	108	105	158	104	119	111	101	157	112	115

4.2. 자료분석

Table 4.2의 자료는 Arnold와 Press (1989)가 공액사전분포의 가정 하에서 제안한 베이지안 추정 절차에 의해 Pareto(α, β)분포에 적합하였던 연간 임금(annual wage) 자료이다. 이 자료의 확률히스토그램인 Figure 4.1을 보면 꼬리부분에 과도하게 자료가 분포하고 있음을 볼 수 있다. 그들이 가정한 공액사전분포의 초모수는 식 (3.1)에서 $a_0 = 6.02$, $b_0 = 3.17$, $c_0 = 10.197$, $d_0 = 103.866$ 이다. 이렇게 가정된 공액사전분포 하에서 베이지안 추정치는 식 (3.3)에 의해 $\hat{\alpha}_{conjugate} = 4.1081$, $\hat{\beta}_{conjugate} = 101$ 로 얻어졌다.

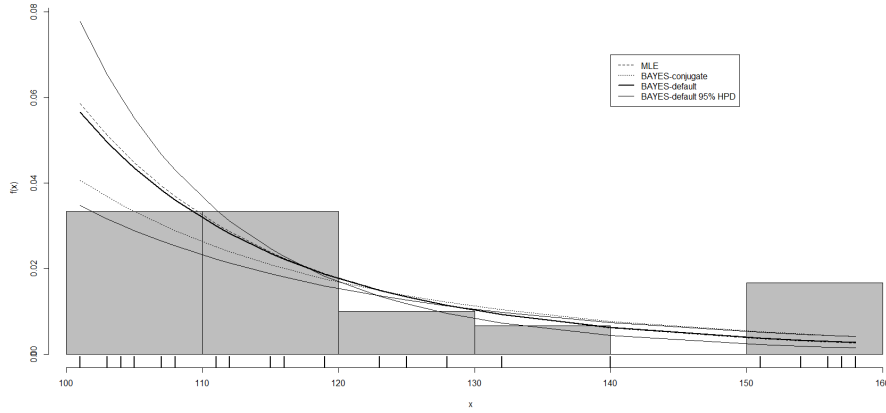


Figure 4.1. Pareto density fitted to annual wage data

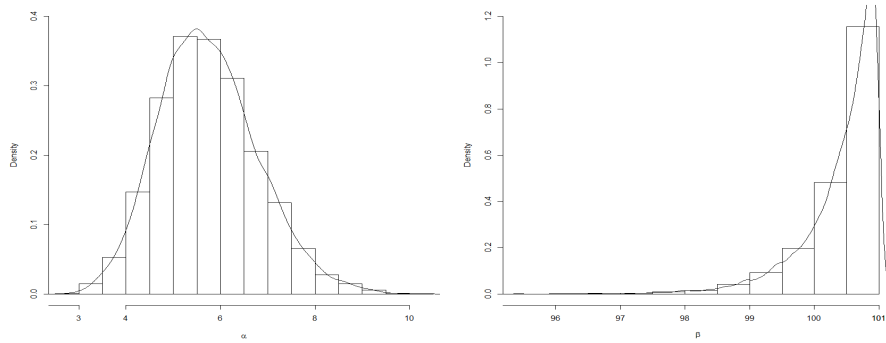


Figure 4.2. Posterior distribution of α and β

Table 4.3. Descriptive summary for the posterior distribution of α and β

parameter	mean	SD	median	mode	95% HPD interval	
					[lower,	upper]
α	5.7478	1.0538	5.6718	5.4873	3.7529	7.8529
β	100.4121	0.6067	100.6035	100.8880	99.2001	101.0000

본 연구에서 제안한 객관적 베이زي안 모수추정을 위하여 Figure 3.1의 Gibbs sampler를 $I = 11,000$ 회 반복 수행하여 처음 1,000회를 버리고 나머지 10,000회 반복 생성한 α 와 β 값들로 사후분포를 구성하였다. 초기값 $\beta^{(0)}$ 는 $\hat{\beta}^{LME}$ 를 사용하였다. Figure 4.2에서 보여주는 α 및 β 의 사후분포로부터 각 모수의 베이زي안 추정치로서 α 에 대해서는 사후평균값을 β 에 대해서는 사후최빈값을 사용하기로 한다. Table 4.3은 α 및 β 에 대한 사후분포의 기술통계분석 결과를 보여준다. 깃스샘플링의 매 반복에서 표집된 α 및 β 의 시계열 도표인 Figure 4.3에서 이 시계열들은 각각 세로축에 수평 참조선으로 표시된 α 의 사후 평균값과 β 의 사후최빈값으로 수렴하고 있음을 알 수 있다. 또한 α 및 β 의 각 시계열에 대한 제1차 자기상관계수는 각각 0.044 및 0.043으로 깃스샘플들이 상호 관련성이 없음을 알 수 있다.

Table 4.4는 모수 α 와 β 에 대한 각 추정치와 각 추정된 모형에 대한 적합도 검정으로서 χ^2 -test와 KS-test(Kolmogrov Smirnov test)의 p -value를 제시하고 있고, Figure 4.4는 Q-Q plot(quantile-quantile

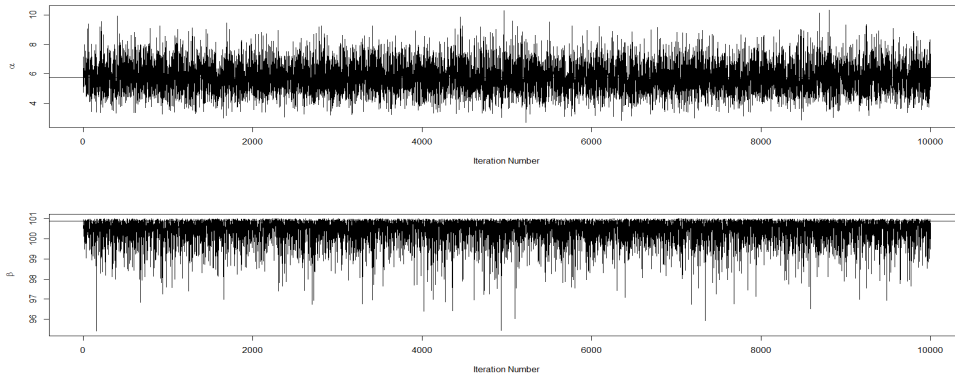


Figure 4.3. Gibbs sampling iterations plot for α and β

Table 4.4. Results fitted to annual wage data

estimation method	parameter estimate		goodness of fit test(p - value)	
	α	β	χ^2 -test	K-S test
LME	6.6589	101	0.2844	0.6601
MLE	5.9179	101	0.5034	0.9506
BAYES-conjugate	4.1081	101	0.4442	0.3076
BAYES-default	5.7478	100.89	0.5713	0.9709

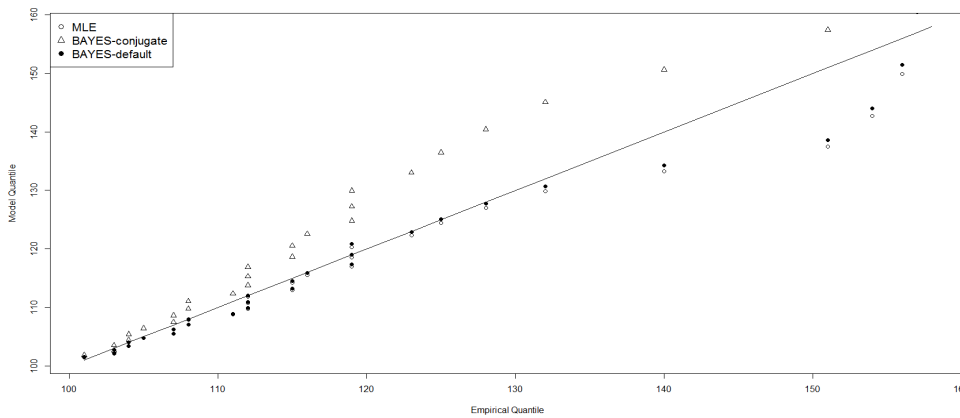


Figure 4.4. Q-Q plot

plot)을 그리고 Figure 4.5는 P-P plot(probability-probability plot)을 보여준다. 이상과 같은 모형 적합 결과로 부터 객관적 베이지안 추정이 최우추정보다 다소 더 적합한 그러나 대체로 비슷한 적합도를 보여주고 있으며, 공액사전분포 하의 베이지안 추정은 이들 보다 적합도가 떨어짐을 알 수 있다. 원 자료의 확률 히스토그램과 최우추정 및 베이지안추정에 의해 추정된 파레토 pdf를 동시에 보여주고 있는 Figure 4.1에서 꼬리부분에 과도하게 분포되어 있는 자료는 Figure 4.4의 QQ-plot에서 극대 분위수들이 직선에서 상당히 이탈한 결과와 일치하고 있음을 알 수 있다.

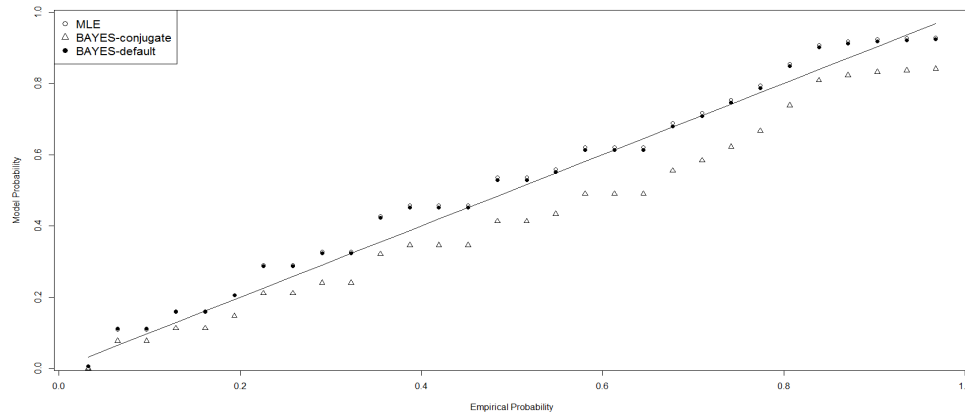


Figure 4.5. P-P plot

5. 결과

본 연구에서는 2-모수 파레토분포에 대해 무정보사전분포인 준거사전분포의 가정 하에서 객관적 베이저안 모수추정 절차를 제안하였다. 베이저안 추정은 깁스샘플링에 의해서 수행된다. 깁스샘플러에서 형태모수는 감마분포로부터 생성되고, 척도모수는 Gilks와 Wild (1992)의 ARS 알고리즘에 의해 생성된다. 제안된 베이저안 모수추정 절차는 모의실험과 자료분석에서 기존의 추정방법들인 L-적률추정법, 최우추정법, 그리고 공액사전분포 하의 주관적 베이저안 모수추정법과 비교된다. 제안된 객관적 베이저안 추정법이 최우추정법보다 다소 더 좋은 결과를 보여 주었다. 본 논문에 모의실험 결과를 제시하지는 않았으나 주관적 베이저안 추정의 경우에 공액사전분포의 초모수 선택의 주관적 판단에 따라 모형 적합 결과는 매우 좌우되며 오히려 오차가 크게 발생하게 하는 결과를 준다.

References

- Arnold, B. (1983). *Pareto Distributions*, Fairland, MD: International Co-operative Publishing House.
- Arnold, B. C. and Press, S. J. (1983). Bayesian inference for Pareto populations, *Journal of Econometrics*, **21**, 287–306.
- Arnold, B. C. and Press, S. J. (1989). Bayesian estimation and prediction for Pareto data, *Journal of the American Statistical Association*, **84**, 1079–1084.
- Berger, J. O. (2006). The case for objective Bayesian analysis (with discussion), *Bayesian Analysis*, **1**, 385–402.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference*, Duxbury Press.
- Crovella, M. E., Taqqu, M. S. and Bestavros, A. (1998). Heavy-Tailed Probability Distributions in the World Wide Web, *A Practical Guide to Heavy Tails*, Adler, R. J., Feldman, R. E., and Taqqu, M. S., Editors, Birkhauser, Boston, MA, 3–27.
- Cumming, S. G. (2001). A parametric model of the fire-size distribution, *Canadian Journal of Forest Research*, **31**, 1297–1303.
- Embrechts, P., Kluppelberg, C. and Mikosch, T. (1997). *Modeling Extremal Events*, Springer, New York.
- Fu, J., Xu, A. and Tang, Y. (2012). Objective Bayesian analysis of Pareto distribution under progressive Type-II censoring, *Statistics and Probability Letters*, **82**, 1829–1836.
- Gilks, W. R. and Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling, *Applied Statistics*, **41**, 337–348.

- Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form, *Water Resources Research*, **15**, 1049–1054.
- Harris, C. M., Brill, P. H. and Fischer, M. J. (2000). Internet-type queues with power-tailed interarrival times and computational methods for their analysis, *INFORMS Journal on Computing*, **12**, 261–271.
- Hosking, J. R. M. (1990). L-Moments: Analysis and estimation of distributions using linear combinations of order statistics, *Journal of the Royal Statistical Society: Series B*, **52**, 105–124.
- Jackson, D. D. and Kagan, Y. Y. (1999). Testable earthquake forecasts for 1999, *Seismological Research Letters*, **70**, 393–403.
- Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, 2nd ed., Wiley, NY.
- Lwin, T. (1972). Estimation of the tail of the Paretian law, *Scandinavian Actuarial Journal*, **55**, 170–178.
- Pareto, V. (1897). *Cours d'Économie Politique*, F.Pichou, Paris.
- Pickands, III, J. (1975). Statistical inference using extreme order statistics, *The Annals of Statistics*, **3**, 119–131.
- Singh, V. P. and Guo, H. (1995). Parameter estimations for 2-parameter Pareto distribution by Pome, *Water Resources Management*, **9**, 81–93.
- Smith, R. L. (2003). *Extreme Values in Finance, Telecom-Munications and the Environment*, Finkenstadt, B. and Rootzen, H., Editors, Chapman and Hall/CRC Press, London.

2-모수 파레토분포의 객관적 베이저안 추정

손영숙^{a,1}

^a전남대학교 통계학과

(2013년 7월 15일 접수, 2013년 8월 23일 수정, 2013년 10월 15일 채택)

요약

본 연구에서는 2-모수 파레토분포에 대해 무정보사전분포인 준거사전분포의 가정 하에서 객관적 베이저안 모수추정 절차를 제안하였다. 베이저안 추정은 깃스샘플링에 의해서 수행된다. 깃스샘플러에서 모수생성하는 방법은 형태모수는 감마분포로부터 생성하고 척도모수는 적응기각표집 알고리즘에 의해 생성한다. 제안된 베이저안 모수추정 절차는 모의실험과 자료분석에서 기존의 추정방법들인 L-적률추정법, 최우추정법, 공액사전분포 하의 주관적 베이저안 모수추정법과 비교된다.

주요용어: 2-모수 파레토분포, L-적률추정법, 최우추정법, 무정보사전분포, 준거사전분포, 객관적 베이저안 추정, 깃스샘플링, 적응기각표집.

본 연구는 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임 (NRF-2011-0022864).

¹(500-757) 광주광역시 북구 용봉동 300번지, 전남대학교, 통계학과. E-mail: ysson@jnu.ac.kr