The Korean Communications
in Statistics Vol. 12 No. 3, 2005
pp. 729-742

Bayesian Survival Estimation of Pareto Distribution of the Second Kind Based on Type II Censored Data

Dal Ho Kim[1], Woo Dong Lee[2], and Sang Gil Kang[3]

Abstract

In this paper, we discuss the propriety of the various noninformative priors for the Pareto distribution. The reference prior, Jeffreys prior and ad hoc noninformative prior which is used in several literatures will be introduced and showed that which prior gives the proper posterior distribution.

The reference prior and Jeffreys prior give a proper posterior distribution, but ad hoc noninformative prior which is proportional to reciprocal of the parameters does not give a proper posterior.

To compute survival function, we use the well-known approximation method proposed by Lindley (1980) and Tireney and Kadane (1986). And two methods are compared by simulation. A real data example is given to illustrate our methodology.

Keywords : Survival function, Noninformative prior, Lindley approximation, Tireney-Kadane approximation

1. 서론

파레토분포는 1897년 소득의 분포를 설명하기 위하여 소개되었다. 최근에 파레토분포의 변형된 모형을 여러 학자들이 연구하였다. 대표적인 연구로 Davis와 Feldstein (1979), Cohen과 Whitten (1988), Grimshaw (1993) 등이 있다. 이 논문에서 소개되는 제2종 파레토분포는 Lomax의 분포 혹은 Pearson의 제4종분포로 알려져 있으며, Bain과 Engelhardt (1992)는 의학통계학분야에서 심장이식 후 생존시간과 wire의 결점과 결점사이의 길이를 설명하는데 이용하였다. Dyer (1981)은 거대기업의 생산라인에 종사하는 노동자의 연봉을 연구하는데 이용하였고, Lomax (1954)는 기업도산자료를 설명하는데 이 분포를 이용하였다. 그리고, 본 논문과 유사한 연구로 Howlader와 Hossain (2002)이 제2종 중단자료를 이용하여 생존함수에 대한 추정문제를 다루었다.

제2종 중단자료는 공학이나 의학에서 흔히 관찰되는 자료의 형태인데, 처음부터 끝까지 모든 개

1) Associate Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea
E-mail : dalkim@knu.ac.kr
2) Associate Professor, Department of Asset Management, Daegu Haany University, Kyungpook 712-715, Korea
3) Assistant Professor, Department of Applied Statistics, Sangji University, Kwangwon 220-702, Korea

체의 생존시간을 관찰하는 것이 아니라 시간과 경비절약을 위하여 사전에 생존시간을 관찰한 개체의 수를 정하고 그 개수만큼 관측된다면 실험을 중단하는 경우에 얻을 수 있다.

베이지안 통계추론에서 사전분포의 역할은 매우 크다. 사전분포의 사용은 주관적인 측면이 있으나, 현재 Jeffreys' prior, reference prior, probability matching prior 등 객관적인 사전분포 (objective prior)가 여러 통계모형에서 개발되었거나 개발 중에 있다. 특히, 사전정보가 거의 없는 경우에 이용할 수 있는 noninformative prior는 흔히 부적절 분포 (improper distribution)로서 사후분포가 부적절 분포가 되는 경우가 빈번히 일어난다. 그러므로 부적절 사전분포를 적용한 경우, 반드시 사후분포의 적절성을 밝히는 것은 매우 중요한 문제임이 분명하다. 이러한 점에서 앞서 언급한 Howlader와 Hossain (2002)의 논문에서 사용한 사전분포는 부적절 사후분포를 유도함을 본 논문에서 밝힌다.

제2종 중단자료를 이용한 베이지안 추론에서 완전자료 (complete sample)를 가정하고 유도된 사전분포를 사용하는 것은 분명 문제점이 있을 것으로 예상된다. 이와 관련한 연구결과로 Santis, Mortera와 Nardi (2001)는 지수분포와 로그정규분포에서 완전자료에 대한 Jeffreys 사전분포와 제1종 중단모형 (type I censorship model), 점진적 제1종 중단모형 (progressive type I censorship model), 임의중단모형 (random censorship model)하에서의 Jeffreys 사전분포를 유도하고, 모수에 대한 포함확률을 계산하여 비교하였다. 그 결과, 중단모형에서의 Jeffreys 사전분포가 frequentist 포함확률측면에서 더 바람직한 결과를 보인다는 것을 모의실험을 통하여 증명하였다. 그리고, Sweeting (2001)은 frequentist 포함확률을 점근적으로 만족하는 자료의 형태에 의존하는 data-dependent 사전분포를 제안하였는데, 그가 제안한 사전분포도 제1종 중단자료인 경우에는 frequentist 포함확률을 만족하나 제2종 중단자료인 경우는 만족하지 않음을 지적하였다.

이 논문에서는 파레토분포로부터 추출된 확률표본이 제2종 중단되었을 때, 생존함수에 대한 베이지안 추정문제를 다룬다. 앞서 언급한 바, 제2종 중단모형인 경우, frequentist 포함확률을 점근적으로 만족하는 사전분포의 개발에 어려움이 있으므로, Kim, Lee 와 Kang (2004)에서 제안된 완전자료인 경우에 제안된 사전분포를 이용하여 제2종 중단된 자료의 베이지안 분석에 이용한다. 특히, Howlader와 Hossain (2002)의 논문에서 사용된 사전분포는 부적절 사후분포를 유도함을 증명하고, Kim, Lee 와 Kang (2004)의 사전분포는 적절 사후분포를 유도함을 증명한다. 2절에서는 파레토모형에서의 사전분포를 소개하고, 3절에서는 사후분포의 적절성을 밝히고, 4절에서는 실제자료를 이용한 분석사례와 모의실험을 통하여 생존함수의 추정에 있어서 어떤 사전분포가 더 적절하며, 어떤 근사방법이 더 정확한지를 비교한다.

2. 무정보적 사전분포와 사후분포의 적절성

제2종 파레토분포는 정칙 (regular)이고, 연속인 분포이며, 확률밀도함수는 다음과 같다.

$$f(x \mid \lambda, \beta)=\frac{\lambda}{\beta}\left(1+\frac{x}{\beta}\right)^{-(\lambda+1)}, x>0, \lambda>0, \beta>0 \tag{1}$$

여기에서 β는 척도모수 (scale parameter)이며 λ는 형상모수 (shape parameter)이다. 경제학 분야에서는 형상모수인 λ의 추정에 많은 관심을 가진다. 이 형상모수는 소득의 불평등의 정도를 나타내기 때문이다. 위의 확률밀도함수 (1)로부터 이 논문에서 관심을 가지는 생존함수는

$$S(t)=P(X\geq t)=\left(1+\frac{t}{\beta}\right)^{-\lambda},\ t>0$$

임을 알 수 있다.

파레토분포를 하는 n개의 표본을 시간과 경비문제 혹은 불가피한 상황에 의하여 제2종 중단실험을 한다고 가정해 보자. 만약 실험중인 n개의 개체 중 $r(\leq n)$개를 관찰하고 실험을 중단하는 경우, 관찰된 표본은

$$x_{(1)}\leq x_{(2)}\leq\cdots\leq x_{(r)}$$

이 될 것이며, $x_{(i)}$는 i번째 순서통계량을 나타낸다. 이 표본을 이용한 모수 β와 λ에 대한 우도함수(likelihood function)는 다음과 같다.

$$L(\lambda,\beta\mid \mathbf{x})=\left(1+\frac{x_{(r)}}{\beta}\right)^{-\lambda(n-r)}\left(\frac{\lambda}{\beta}\right)^{r}\prod_{i=1}^{r}\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)},\lambda>0,\ \beta>0 \tag{2}$$

여기에서 $\mathbf{x}=(x_{(1)},x_{(2)},\cdots,x_{(r)})$이다.

Kim, Lee 와 Kang (2004)은 완전한 표본이 주어진 경우, 제2종 파레토분포의 모수에 대한 무정보적 사전분포 (noninformative priors)를 개발하였다. 이 논문에서 Jeffreys' 사전분포는

$$\pi_J(\lambda,\beta)\propto\frac{1}{\beta(1+\lambda)\sqrt{\lambda(\lambda+2)}},\lambda>0,\beta>0, \tag{3}$$

이며, λ가 관심모수 (parameter of interest)인 경우, reference 사전분포는

$$\pi_R(\lambda,\beta)\propto\frac{1}{\beta\lambda(1+\lambda)},\lambda>0,\ \beta>0, \tag{4}$$

이다. Howlader와 Hossain (2002)은 생존함수의 베이지안 추정문제를 다루었는데 그들은 특정한 모수 값에 편중되지 않는다는 이유로 아래의 사전분포를 사용하였다. 그들의 사전분포는 다음과 같다.

$$\pi_{HH}(\lambda,\beta)\propto\frac{1}{\lambda\beta},\lambda>0,\beta>0\,. \tag{5}$$

이제, 우도함수 (2)와 사전분포 (3), (4), (5)에 의한 사후분포의 적절성을 조사하기로 한다. 먼저, Howlader와 Hossain (2002)이 사용했던 사전분포 (5)를 적용하였을 때, 사후분포의 적절성을 조사해보자. 그 결과는 아래의 정리1과 같다.

정리 1. $\pi_{HH}(\lambda,\beta)$에 의한 사후분포는 부적절 분포이다.

증명. 우도함수 (2)와 사전분포 (5)에 의한 사후분포는 다음과 같다.

$$\begin{aligned}\pi_{HH}(\lambda,\beta\mid \mathbf{x}) &\propto \lambda^{r}\beta^{-r}\left(1+\frac{x_{(r)}}{\beta}\right)^{-\lambda(n-r)}\prod_{i=1}^{r}\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)}\lambda^{-1}\beta^{-1}\\ &\propto \lambda^{r-1}\beta^{-(r+1)}\left(1+\frac{x_{(r)}}{\beta}\right)^{-\lambda(n-r)}\prod_{i=1}^{r}\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)}.\end{aligned}$$

그리고,

$$\prod_{i=1}^{r}\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)}\geq\left(1+\frac{x_{(r)}}{\beta}\right)^{-r(\lambda+1)}$$

인 관계가 성립하므로, 다음의 관계가 성립한다.

$$\pi_{HH}(\lambda, \beta \mid \boldsymbol{x}) \geq \lambda^{r-1}\beta^{-(r+1)}\left(1+\frac{x_{(r)}}{\beta}\right)^{-(\lambda n+r)}.$$

위의 식의 오른쪽 항에 대하여 β에 대한 적분값을 구해보면,

$$\int_0^\infty \lambda^{r-1}\beta^{-(r+1)}\left(1+\frac{x_{(r)}}{\beta}\right)^{-(\lambda n+r)} d\beta = \lambda^{r-1}\int_0^\infty \beta^{\lambda n-1}(\beta + x_{(r)})^{-(\lambda n+r)} d\beta,$$

이 된다. 위의 식 오른쪽 항의 적분값을 $y=\dfrac{\beta}{\beta + x_{(r)}}$ 로 변환하여 계산하면,

$$\int_0^\infty \lambda^{r-1}\beta^{-(r+1)}\left(1+\frac{x_{(r)}}{\beta}\right)^{-(\lambda n+r)} d\beta = \lambda^{r-1}x_{(r)}^{-r}\frac{\Gamma(\lambda n)\Gamma(r)}{\Gamma(\lambda n+r)}.$$

위의 오른쪽 식에서 상수항을 제외하고, $\Gamma(\lambda n+r) = (\lambda n+r-1)(\lambda n+r-2)\cdots\lambda n\Gamma(\lambda n)$인 성질을 이용하여, λ에 대하여 적분하면 다음과 같다.

$$\begin{aligned}
\int_0^\infty \lambda^{r-1}\frac{\Gamma(\lambda n)}{\Gamma(\lambda n+r)} d\lambda &= \int_0^\infty \frac{\lambda^{r-1}}{(\lambda n+r-1)(\lambda n+r-2)\cdots\lambda n} d\lambda \\
&\geq \int_0^\infty \frac{\lambda^{r-1}}{(\lambda n+r-1)^r} d\lambda \\
&= n^r\int_0^\infty \lambda^{r-1}(\lambda+d)^{-r} d\lambda,
\end{aligned}$$

여기에서 $d=(r-1)/n$이다. 위의 식의 마지막 적분을 $z=\lambda/(\lambda+d)$로 변환하여 적분하면

$$\int_0^1 z^{r-1}(1-z)^{-1}dz = \infty,$$

이므로 사후분포는 부적절 분포이다.

정리 2. (3)에 주어진 Jeffreys 사전분포에 의한 사후분포는 적절분포이다.

증명. (3)에 주어진 Jeffreys 사전분포와 (2)에 주어진 우도함수를 이용하면 λ와 β의 우도함수는 다음과 같다.

$$\begin{aligned}
\pi_J(\lambda, \beta \mid \boldsymbol{x}) &\propto \left(1+\frac{x_{(r)}}{\beta}\right)^{-\lambda(n-r)}\left(\frac{\lambda}{\beta}\right)^r\prod_{i=1}^r\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)}\frac{1}{\beta(1+\lambda)\sqrt{\lambda(\lambda+2)}} \\
&\propto \lambda^{r-1/2}(1+\lambda)^{-1}(2+\lambda)^{-1/2}\beta^{-(r+1)}\left(1+\frac{x_{(r)}}{\beta}\right)^{-\lambda(n-r)}\prod_{i=1}^r\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)} \\
&\leq \lambda^{r-1/2}(1+\lambda)^{-1}(2+\lambda)^{-1/2}\beta^{-(r+1)}\left(1+\frac{x_{(1)}}{\beta}\right)^{-(n\lambda+r)},
\end{aligned}$$

위의 식에서 마지막 부등식은 $x_{(r)}$과 $x_{(i)}, i=1,2,\cdots,r$ 대신에 $x_{(1)}$을 대입한 경우이다. 이 마지막 식을 β에 대하여 적분하면,

$$\begin{aligned}
&\lambda^{r-1/2}(1+\lambda)^{-1}(2+\lambda)^{-1/2}\int_0^\infty \beta^{-(r+1)}\left(1+\frac{x_{(1)}}{\beta}\right)^{-(n\lambda+r)} d\beta \\
&= x_{(1)}^{r}\lambda^{r-1/2}(1+\lambda)^{-1}(2+\lambda)^{-1/2}B(n\lambda, r),
\end{aligned}$$

이다. 여기에서 $B(n\lambda, r)=\dfrac{\Gamma(n\lambda)\Gamma(r)}{\Gamma(n\lambda+r)}$ 이다. 위의 오른쪽 식을 λ에 대하여 정리하면

$$x_{(1)}^{r}\lambda^{r-1/2}(1+\lambda)^{-1}(2+\lambda)^{-1/2}\frac{\Gamma(n\lambda)\Gamma(r)}{\Gamma(n\lambda+r)}$$

$$= x_{(1)}^{r}\Gamma(r)\lambda^{r-1/2}(1+\lambda)^{-1}(2+\lambda)^{-1/2}\frac{\Gamma(n\lambda)}{(n\lambda+r-1)(n\lambda+r-2)\cdots n\lambda\Gamma(n\lambda)}$$
$$\le x_{(1)}^{r}\Gamma(r)\lambda^{r-1/2}(1+\lambda)^{-1}(2+\lambda)^{-1/2}(n\lambda)^{-r}$$

이 된다. 그리고 위의 마지막 식을 λ에 대하여 적분하면

$$\int_0^\infty \lambda^{-1/2}(1+\lambda)^{-1}(2+\lambda)^{-1/2}d\lambda = \frac{\pi}{2} < \infty,$$

이다. 그러므로, Jeffreys 사전분포를 사용하면 사후분포는 적절성을 가짐을 알 수 있다.

마지막으로 λ가 관심모수이고, β가 장애모수인 경우에 개발된 reference 사전분포 (4)에 의한 사후분포의 적절성을 증명한다.

정리3. reference 사전분포 (4)에 의한 사후분포는 $r>1$인 경우, 적절사후분포이다.

증명. 사전분포 (4)와 우도함수 (2)를 이용한 사후분포는 다음과 같다.

$$\pi_f(\lambda,\beta\mid \boldsymbol{x}) \propto \left(1+\frac{x_{(r)}}{\beta}\right)^{-\lambda(n-r)}\left(\frac{\lambda}{\beta}\right)^r\prod_{i=1}^r\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)}\frac{1}{\beta(1+\lambda)\lambda}$$
$$\propto \lambda^{r-1}(1+\lambda)^{-1}\beta^{-(r+1)}\left(1+\frac{x_{(r)}}{\beta}\right)^{-\lambda(n-r)}\prod_{i=1}^r\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)}$$
$$\le \lambda^{r-1}(1+\lambda)^{-1}\beta^{-(r+1)}\left(1+\frac{x_{(1)}}{\beta}\right)^{-(n\lambda+r)}.$$

위의 식을 β에 대해 적분하여 λ에 대하여 정리하면

$$x_{(1)}^{r}\Gamma(r)\lambda^{r-1}(1+\lambda)^{-1}\frac{\Gamma(n\lambda)}{(n\lambda+r-1)(n\lambda+r-2)\cdots n\lambda\Gamma(n\lambda)}$$

이 된다. 그리고,

$$\frac{1}{(n\lambda+r-1)(n\lambda+r-2)\cdots(n\lambda+1)n\lambda} \le \frac{1}{(n\lambda+1)^{r-1}n\lambda} \le \frac{1}{(\lambda+1)^{r-1}n\lambda}$$

인 관계가 성립하므로, 이를 이용하여 λ에 대하여 정리하면 다음과 같다.

$$\lambda^{r-2}(\lambda+1)^{-r}.$$

이를 적분하면 λ에 대해 적분하면, $r>1$인 경우

$$\int_0^\infty \frac{\lambda^{r-2}}{(\lambda+1)^r}d\lambda = \frac{1}{r-1} < \infty.$$

정리 1,2,3으로부터 Jeffreys 사전분포와 λ가 관심모수인 경우의 reference 사전분포는 적절 사후분포를 유도함을 알았고, 정리 1로부터 Howlader와 Hossain (2002)이 사용한 사전분포는 부적절사전분포를 유도함을 알았다.

3. 생존함수의 추정

이제 적절 사전분포를 유도하는 사전분포 (3)과 (4)를 이용한 생존함수의 베이지안 추정문제를 고려해 보자. 생존함수의 베이지안 추정량을 구하기 위해서는 2차원 적분이 필요하다. 수치해석적인 방법으로 추정량을 계산할 수도 있다. 그러나 이 논문에서는 추정량에 대한 계산을 위하여 널리 알려진 근사방법인 Lindley (1980)과 Tireney 와 Kadane (1986)을 이용하려고 한다.

제곱오차손실함수 (squared error loss function)를 고려하고 사전분포가 $\pi(\lambda,\beta)$라고 한다면, 생존함수에

대한 베이지안 추정량은 다음과 같다.

$$\widehat{S(t)} = \frac{\int_0^\infty \int_0^\infty \left(1+\frac{t}{\beta}\right)^{-\lambda}\left(1+\frac{x_{(r)}}{\beta}\right)^{-\lambda(n-r)}\left(\frac{\lambda}{\beta}\right)^r \prod_{i=1}^r\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)} \pi(\lambda,\beta)d\lambda d\beta}{\int_0^\infty \int_0^\infty \left(1+\frac{x_{(r)}}{\beta}\right)^{-\lambda(n-r)}\left(\frac{\lambda}{\beta}\right)^r \prod_{i=1}^r\left(1+\frac{x_{(i)}}{\beta}\right)^{-(\lambda+1)} \pi(\lambda,\beta)d\lambda d\beta}. \tag{6}$$

Lindley (1980)가 제안한 근사방법을 적용하기 위하여

$$\boldsymbol{\theta}=(\theta_1,\theta_2)=(\lambda,\beta),\ \Lambda(\boldsymbol{\theta})=\log L(\boldsymbol{\theta}\mid \boldsymbol{x})+\log\pi(\boldsymbol{\theta}),\ \phi(\boldsymbol{\theta})=S(t)=\left(1+\frac{t}{\beta}\right)^{-\lambda},$$

$$\rho(\boldsymbol{\theta})=\log\pi(\boldsymbol{\theta}),$$

$$\rho_i\equiv\rho_i(\boldsymbol{\theta})=\frac{\partial\log\pi(\boldsymbol{\theta})}{\partial\theta_i},\quad \rho_{ij}\equiv\rho_{ij}(\boldsymbol{\theta})=\frac{\partial^2\log\pi(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j},\quad \rho_{ijk}\equiv\rho_{ijk}(\boldsymbol{\theta})=\frac{\partial^3\log\pi(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j\partial\theta_k}$$

로 두자. 그리고,

$$\phi_i\equiv\phi_i(\boldsymbol{\theta})=\frac{\partial\phi(\boldsymbol{\theta})}{\partial\theta_i},\ \phi_{ij}\equiv\phi_{ij}(\boldsymbol{\theta})=\frac{\partial^2\phi(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j},\ \Lambda_{ij}\equiv\Lambda_{ij}(\boldsymbol{\theta})=\frac{\partial^2\Lambda(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j},$$

$$\Lambda_{ijk}\equiv\Lambda_{ijk}(\boldsymbol{\theta})=\frac{\partial^3\Lambda(\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j\partial\theta_k},\quad \tau_{ij}\equiv\tau_{ij}(\boldsymbol{\theta})=-(\Lambda_{ij})^{-1}_{2\times 2}$$

라고 하면, (6)에 대한 Lindley 근사값은

$$\begin{aligned}\widehat{S(t)}^L &= \phi(\hat{\boldsymbol{\theta}})+\frac{1}{2}\sum_{i=1}^2\sum_{j=1}^2\phi_{ij}(\hat{\boldsymbol{\theta}})\tau_{ij}(\hat{\boldsymbol{\theta}}) \\ &+ \frac{1}{2}\sum_{i=1}^2\sum_{j=1}^2\sum_{k=1}^2\sum_{l=1}^2\Lambda_{ijk}(\hat{\boldsymbol{\theta}})\phi_l(\hat{\boldsymbol{\theta}})\tau_{ij}(\hat{\boldsymbol{\theta}})\tau_{kl}(\hat{\boldsymbol{\theta}}),\end{aligned} \tag{7}$$

이 된다. 여기에서 $\hat{\boldsymbol{\theta}}$는 일반화 최우추정량 (generalized maximum likelihood estimator), 즉, $\boldsymbol{\theta}$의 사후분포에서 $\boldsymbol{\theta}$에 대한 최우추정량이다. 이를 계산하기 위하여 좀 더 구체적으로 계산하면 다음과 같다.

$$\phi_1=-\left(1+\frac{t}{\beta}\right)^{-\lambda}\log\left(1+\frac{t}{\beta}\right),\ \phi_2=\frac{\lambda t}{\beta^2}\left(1+\frac{t}{\beta}\right)^{-(\lambda+1)},$$

$$\phi_{11}=\left(1+\frac{t}{\beta}\right)^{-\lambda}\left[\log\left(1+\frac{t}{\beta}\right)\right]^2,\ \phi_{12}=\phi_{21}=\left(1+\frac{t}{\beta}\right)^{-(\lambda+1)}\frac{t}{\beta^2}\left[1-\lambda\log\left(1+\frac{t}{\beta}\right)\right],$$

그리고

$$\phi_{22}=\left(1+\frac{t}{\beta}\right)^{-(\lambda+1)}\frac{\lambda t}{\beta^2}\ \frac{\lambda t-2\beta-t}{\beta(\beta+t)}$$

이다. 또한,

$$\Lambda_1=-(n-r)\log\left(1+\frac{x_{(r)}}{\beta}\right)+\frac{r}{\lambda}-\sum_{i=1}^r\log\left(1+\frac{x_{(i)}}{\beta}\right)+\rho_1,$$

$$\Lambda_2=\frac{\lambda(n-r)x_{(r)}}{\beta^2+\beta x_{(r)}}-\frac{r}{\beta}+\sum_{i=1}^r\frac{(\lambda+1)x_{(i)}}{\beta^2+\beta x_{(i)}}+\rho_2$$

$$\Lambda_{11}=-\frac{r}{\lambda^2}+\rho_{11},\ \Lambda_{12}=\frac{(n-r)x_{(r)}}{\beta^2+\beta x_{(r)}}+\sum_{i=1}^r\frac{x_{(i)}}{\beta^2+\beta x_{(i)}}+\rho_{12},$$

$$\Lambda_{22} = \frac{r}{\beta^2} - \frac{\lambda(n-r)x_{(r)}(2\beta+x_{(r)})}{(\beta^2+\beta x_{(r)})^2} - (\lambda+1)\sum_{i=1}^{r}\frac{x_{(i)}(2\beta+x_{(i)})}{(\beta^2+\beta x_{(i)})^2} + \rho_{22}$$

$$\Lambda_{111} = \frac{2r}{\lambda^3} + \rho_{111},\ \ \Lambda_{112} = \rho_{112},$$

$$\Lambda_{122} = -\frac{(n-r)x_{(r)}(2\beta+x_{(r)})}{(\beta^2+\beta x_{(r)})^2} - \sum_{i=1}^{r}\frac{x_{(i)}(2\beta+x_{(i)})}{(\beta^2+\beta x_{(i)})^2} + \rho_{122},$$

$$\begin{aligned}\Lambda_{222} &= -\frac{2r}{\beta^3} + 2\lambda(n-r)x_{(r)}\frac{(\beta^2+\beta x_{(r)})(3\beta^2+3\beta x_{(r)}+x^2_{(r)})}{(\beta^2+\beta x_{(r)})^4} \\ &\quad + 2(\lambda+1)\sum_{i=1}^{r}\frac{x_{(i)}(\beta^2+\beta x_{(i)})(3\beta^2+3\beta x_{(i)}+x^2_{(i)})}{(\beta^2+\beta x_{(i)})^4} + \rho_{222}\end{aligned}$$

이다. 참고로 (7)을 계산하기 위하여 필수적인 $\hat{\boldsymbol{\theta}}$는 $\Lambda_1=0$, $\Lambda_2=0$을 만족하는 해이다.

다음으로 Tierney와 Kadane (1986)이 제안한 방법을 소개한다. Lindley 근사법은 많은 량의 계산이 필요한 반면, 이 방법은 계산량도 적고 정확한 근사값을 제공하는 것으로 알려져 있다. 이 근사방법을 적용하기 이하여 다음의 기호를 정의한다.

$$l(\boldsymbol{\theta}) \equiv \Lambda(\boldsymbol{\theta}) = \rho(\boldsymbol{\theta}) + \log L(\boldsymbol{\theta} \mid \boldsymbol{x}),$$

$$l^*(\boldsymbol{\theta}) \equiv \log(\phi(\boldsymbol{\theta})) + \Lambda(\boldsymbol{\theta}) = \log(\phi(\boldsymbol{\theta})) + \rho(\boldsymbol{\theta}) + \log L(\boldsymbol{\theta} \mid \boldsymbol{x})$$

라 두자. 그러면 (6)식은

$$E\{\phi(\boldsymbol{\theta}) \mid \boldsymbol{x}\} = \frac{\int \exp(l^*(\boldsymbol{\theta}))d\boldsymbol{\theta}}{\int \exp(l(\boldsymbol{\theta}))d\boldsymbol{\theta}}$$

과 같이 된다. Tirney와 Kadane은 윗식에 대해 분자와 분모 각각에 대해 근사방법을 고려했는데, 근사값은 다음과 같다.

$$\widehat{S(t)}^{TK} = \left(\frac{\Sigma^*}{\Sigma}\right)^{\frac{1}{2}} \exp[(l^*(\hat{\boldsymbol{\theta}}^*) - l(\hat{\boldsymbol{\theta}}))] \tag{9}$$

여기에서 $\hat{\boldsymbol{\theta}}^*$와 $\hat{\boldsymbol{\theta}}$는 각각 식 (8)에서 l^*와 l을 최대로 하는 값이다. 그리고

$$\Sigma^* = \begin{vmatrix} -\frac{\partial^2 l^*}{\partial\lambda^2} & -\frac{\partial^2 l^*}{\partial\lambda\partial\beta} \\ -\frac{\partial^2 l^*}{\partial\lambda\partial\beta} & -\frac{\partial^2 l^*}{\partial\beta^2} \end{vmatrix},\ \Sigma = \begin{vmatrix} -\frac{\partial^2 l}{\partial\lambda^2} & -\frac{\partial^2 l}{\partial\lambda\partial\beta} \\ -\frac{\partial^2 l}{\partial\lambda\partial\beta} & -\frac{\partial^2 l}{\partial\beta^2} \end{vmatrix} = \begin{vmatrix} -\Lambda_{11} & -\Lambda_{12} \\ -\Lambda_{12} & -\Lambda_{22} \end{vmatrix}$$

이다. 그리고,

$$l^* = \log\phi(\theta) + \Lambda = -\lambda\log\left(1+\frac{t}{\beta}\right) + \Lambda$$

이다. 위의 식으로부터

$$\frac{\partial l^*}{\partial\lambda} = -\log\left(1+\frac{t}{\beta}\right) + \Lambda_1,\ \ \frac{\partial l^*}{\partial\beta} = \frac{\lambda t}{\beta^2+\beta t} + \Lambda_2$$

임을 알 수 있고, 이 두식을 이용하여 l^*를 최대로 하는 λ와 β를 찾을 수 있다. 그리고,

$$\frac{\partial^2 l^*}{\partial \lambda^2} = \Lambda_{11}, \quad \frac{\partial^2 l^*}{\partial \lambda \partial \beta} = \frac{t}{\beta^2 + \beta t} + \Lambda_{12}, \quad \frac{\partial^2 l^*}{\partial \beta^2} = -\frac{\lambda t(2\beta + t)}{(\beta^2 + \beta t)^2} + \Lambda_{22}$$

이다.

4. 예제 및 모의실험

4절에서는 3절에서 소개한 두 가지 근사법들을 2절에서 제안된 사전분포들을 적용하여 실제 자료를 분석하는 사례를 보이고, 모의실험을 통하여 근사방법과 사전분포들을 추정된 편의와 평균제곱오차 측면에서 비교·연구한다.

예제. 다음의 자료는 Dyer (1981)에 의하여 분석된 자료이다. 이 자료는 미국의 대기업 생산라인에 근무하는 30명의 직원 연봉이다. Arnold와 Press (1989)은 이 자료를 공액사전분포와 독립사전분포를 적용하여 분석하였다. 자료는 다음과 같다.

1년 임금(단위 :100 U.S. $)
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.

위의 자료를 Howlader와 Hossain (2002)은 $r=28$인 경우 ($x_{(28)}=156$)로 가정하여 생존함수에 대한 추정값을 계산하였다 (Howlader와 Hossain (2002)의 Table 1 참고). 여기에서도 같은 가정을 사용하여 분석해 보기로 하자. 표1에서 Exact는 식(6)에 대하여 이차원 수치적분을 적용한 결과이다.

표1에서 알 수 있듯이, Tierney와 Kadane (1986) 근사법(T-K 근사법)과 Lindley 근사법이 모두 Exact 값에 가깝다는 것을 알 수 있다. 그러나 Lindley의 근사방법이 좀 더 정확한 근사값을 제공한다는 것을 알 수 있다.

<표 1> 임금자료에 대한 생존함수의 추정값

	Jeffreys prior			Reference prior		
Time	Lindley	T-K	Exact	Lindley	T-K	Exact
20.000	0.85182	0.85766	0.84695	0.85195	0.85774	0.84755
40.000	0.72952	0.73767	0.72285	0.72991	0.73798	0.72395
60.000	0.62782	0.63642	0.62123	0.62851	0.63704	0.62272
100.000	0.47108	0.47824	0.46741	0.47232	0.47957	0.46942
150.000	0.33612	0.34058	0.33750	0.33786	0.34270	0.33983
200.000	0.24492	0.24712	0.25066	0.24694	0.24976	0.25306

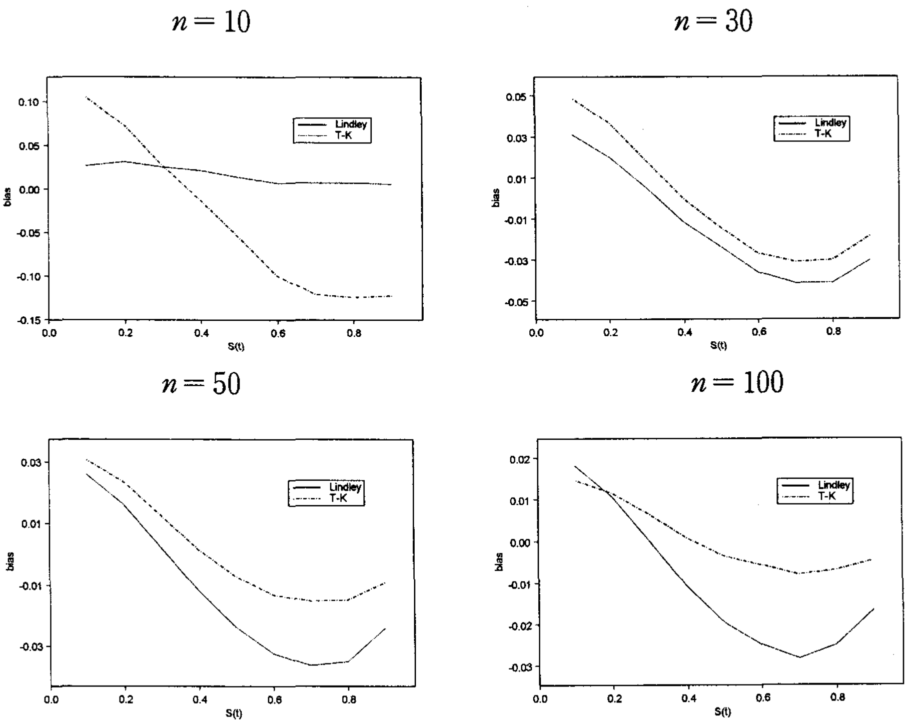
다음으로 난수를 이용한 모의실험을 통하여 생존함수의 추정량에 대한 근사방법들을 비교하고자 한다. 크기가 10, 30, 50, 100인 표본을 5000번 반복적으로 생성시켰으며, 이때 모수의 값은 각각 $\lambda=1.5$ 그리고 $\beta=2,4$로 가정하였고 중단비율을 10%로 고정시켜서 추정량들의 편의와 평

균제곱오차를 추정하였다. 그 결과는 그림1-그림8에서 보는 것과 같다.

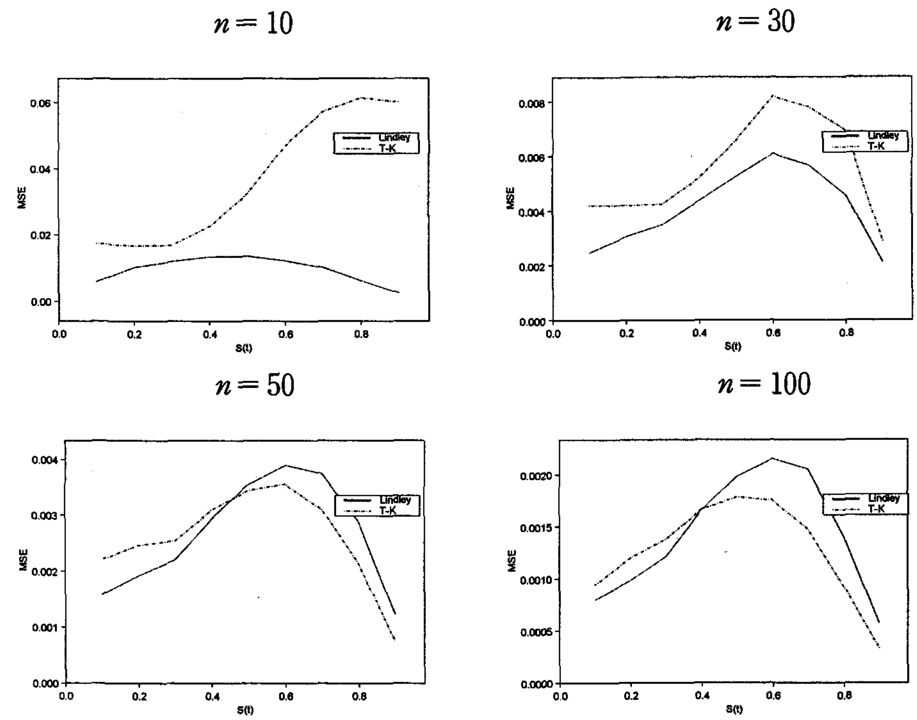
추정된 편의의 측면에서 표본의 수가 작을 때는 Lindley 근사법이 T-K 근사법에 비해 적은 편의를 가지나, 표본의 수가 커짐에 따라 T-K 근사법의 편의가 생존함수의 값이 커질 때 더 우수하다는 것을 알 수 있다. 그리고, 평균제곱오차의 측면에서도 표본의 수가 작을 때는 Lindley 근사법이 더 작은 값을 가졌다. 그리고 표본의 수가 커지면서 생존함수의 값이 클 때, 근소한 차이지만 T-K 방법이 우수하다는 것을 알 수 있다. 이러한 경향은 가정된 모수값과 사전분포의 종류에 관계없이 비슷한 경향을 나타내었다.

5. 결론

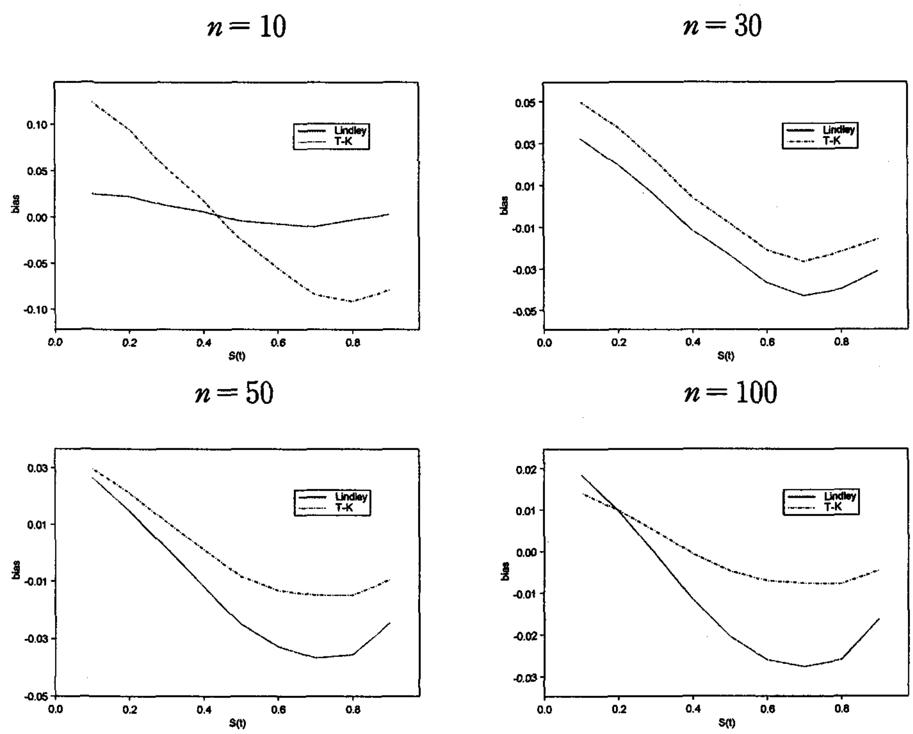
본 논문에서는 제2종 중도절단자료를 이용하여 파레토분포의 생존함수에 대한 베이지안 추정문제를 다루었다. Howlader와 Hossain (2002)의 연구에서 제안된 사전분포를 사용하면 부적절 사후분포를 가짐을 증명하였고, Jeffreys 사전분포와 λ가 관심모수인 경우의 reference 사전분포를 이용하여 생존함수에 대한 베이지안 추정을 제안하였다. 그리고 Lindley 근사법과 Tireney와 Kadane 근사방법을 이용하여 생존함수에 대한 추정량을 계산하여 이를 비교하였다. 실제의 자료를 이용하여 분석사례를 보였으며, 모의실험을 통하여 근사방법을 비교하였다. 모의실험의 결과, 표본의 수가 적을 때는 Lindley 근사법이 우수하다는 것을 알았다.



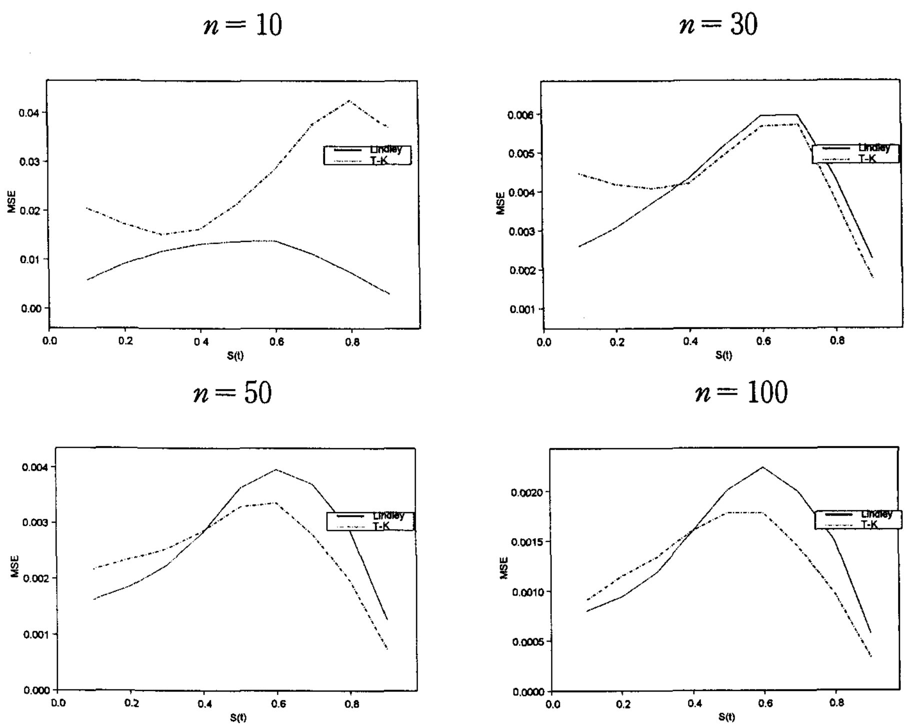
<그림 1> Jeffreys 사전분포 하에서의 추정된 편의(bias)($\lambda=1.5$, $\beta=2$)



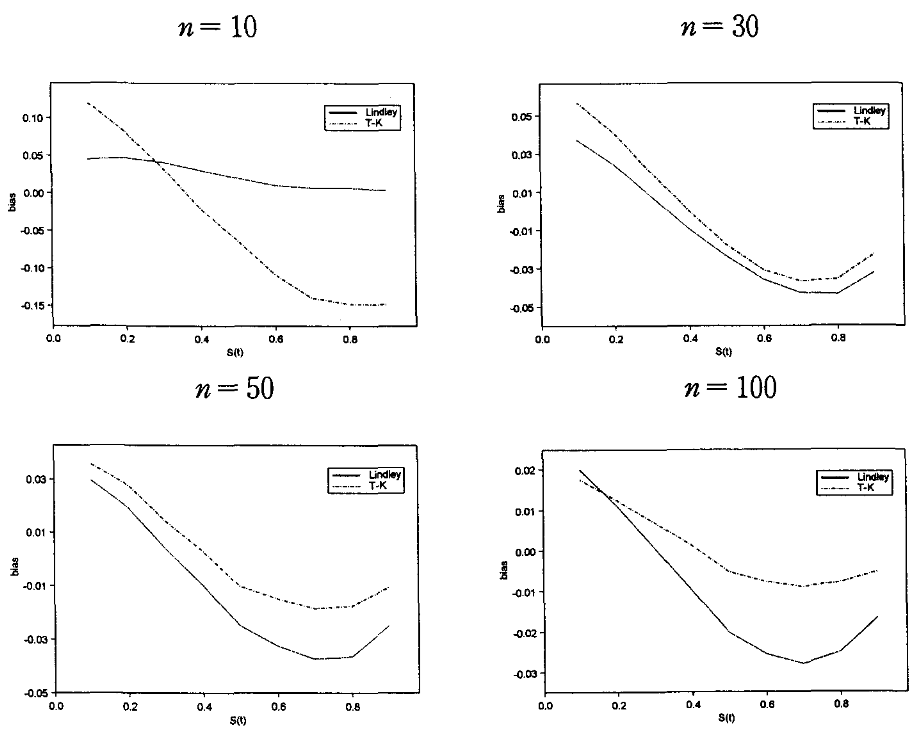
<그림 2> Jeffreys 사전분포 하에서의 추정된 평균제곱오차(MSE)($\lambda=1.5$, $\beta=2$)



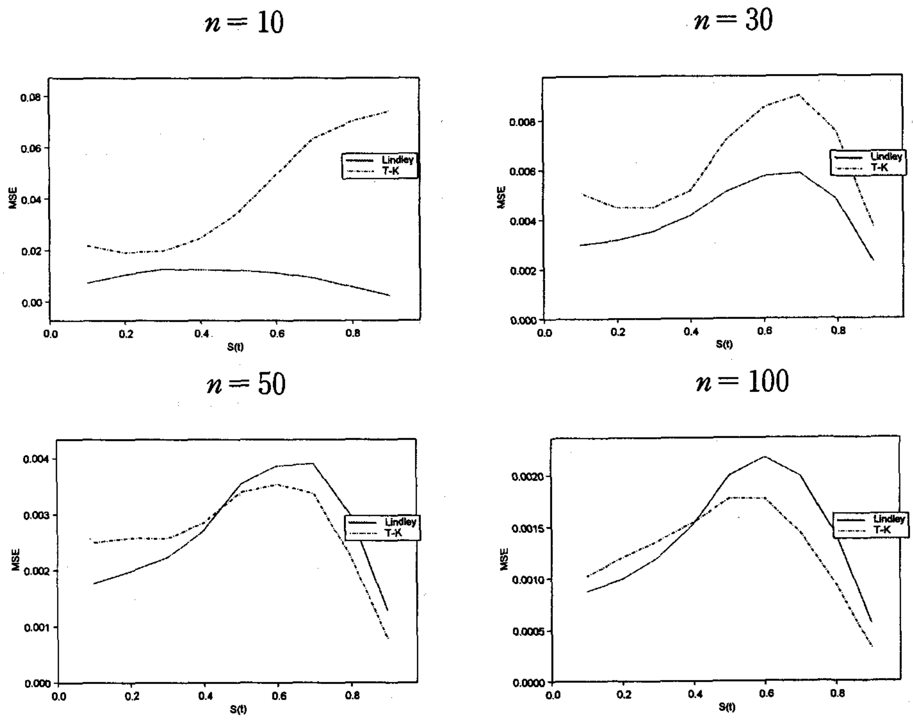
<그림 3> Jeffreys 사전분포 하에서의 추정된 편의(bias)($\lambda=1.5$, $\beta=4$)



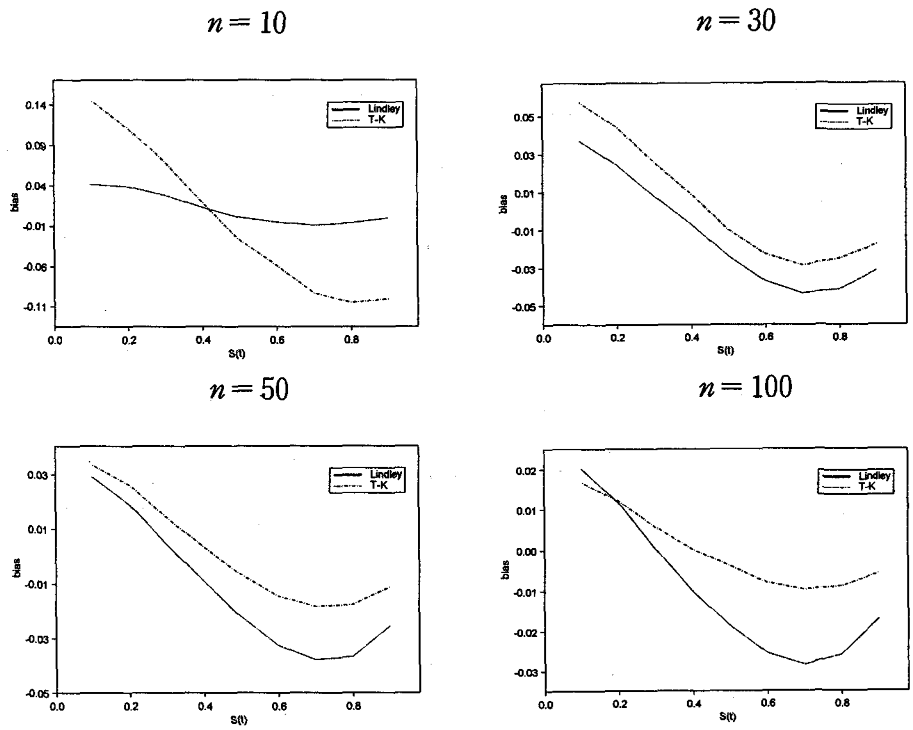
<그림 4> Jeffreys 사전분포 하에서의 추정된 평균제곱오차(MSE)($\lambda=1.5$, $\beta=4$)



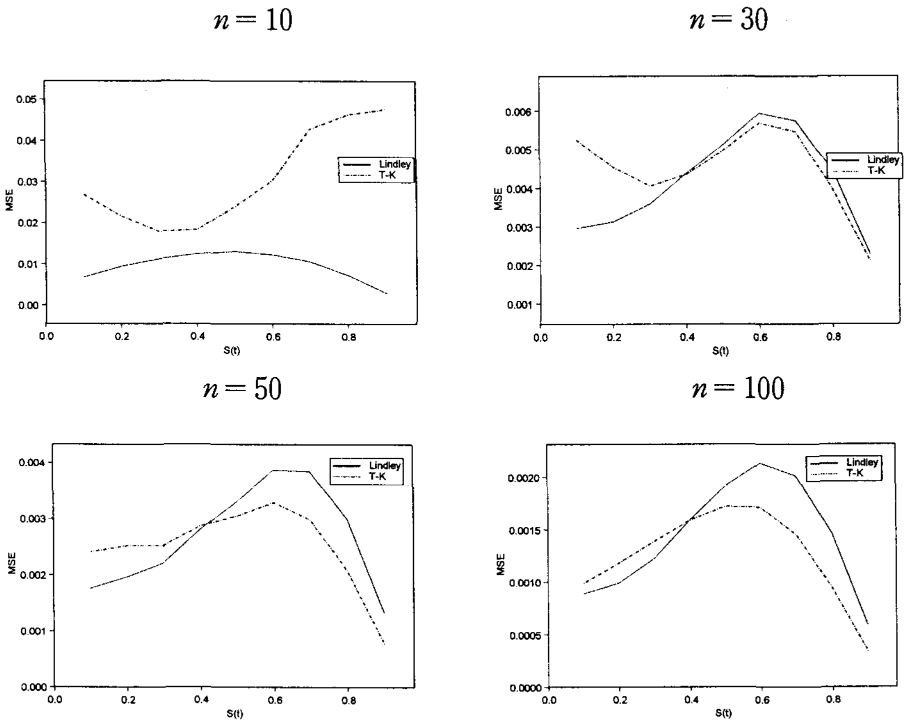
<그림 5> Reference 사전분포 하에서의 추정된 편의(bias)($\lambda=1.5$, $\beta=2$)



<그림 6> Reference 사전분포 하에서의 추정된 평균제곱오차(MSE)($\lambda=1.5$, $\beta=2$)



<그림 7> Reference 사전분포 하에서의 추정된 편의(bias)($\lambda=1.5$, $\beta=4$)



<그림 8> Reference 사전분포하에서의 추정된 평균제곱오차(MSE)($\lambda=1.5$, $\beta=4$)

References

[1] Arnold, B.C. and Press, S.J. (1989). Bayesian estimation and prediction for Pareto data, *Journal of the American Statistical Association*, Vol. 84, 1079-1084.

[2] Bain, L.J. and Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics*, 2nd Edition. PWS-KENT Publishing company, Boston, Massachusett.

[3] Cohen, A.C. and Whitten, B.J. (1988). *Parameter Estimation in Reliability and Life Span Models*, Marcel Dekker, Inc. New York.

[4] Davis, T.H. and Feldstein, L.M. (1979). The generalized Pareto law as a model for Progressively censored survival data, *Biometrika*, Vol. 66 (2), 299-306.

[5] Dyer, D. (1981). Structural Probability Bounds for the Strong Pareto Law, *Canadian Journal of Statistics*, 9, 71-77.

[6] Grimshaw, D.S. (1993). Computing Maximum Likelihood Estimates for the Generalized Pareto Distribution. *Technometrics*, 35, (2), 185-191.

[7] Howlader, H. A. and Hossain, A. M. (2002). Bayesian Survival Estimation of Pareto Distribution of the Second Kind based on Failure-Censored Data, *Computational Statistics & Data Analysis*, 38, 301-314.

[8] Kim, D. H., Lee, W. D. and Kang, S. G. (2004). Noninformative Priors for Pareto Distribution, *Preprint*.

[9] Lindley, D.V. (1980). *Approximate Bayesian Methods, in Bayesian Statistics,* Bernardo, J.M., De Groot, M.H., Lindley, D.V. and smith, A.F.M., eds. Valencia, Spain: Valencia Press, 223-245.

[10] Lomax, K.S. (1954). Business failures. Another example of the analysis of failure data, *Journal of the American Statistical Association*, Vol. 49, 847-852.

[11] Santis, F.D, Mortera, J. and Nardi, A. (2001). Jeffreys priors for survival models with censored data, *Journal of Statistical Planning and Inference*, Vol. 99 (2), 193-209.

[12] Sweeting, T.J. (2001). Coverage Probability Bias, Objective Bayes and the Likelihood Principle. *Biomerika*, 88, 657-675.

[13] Tireney, L. and Kadane, J. (1986). Accurate approximations for posterior moments and marginals, *Journal of the American Statistical Association*, Vol. 81, 82-86.

[2005년 7월 접수, 2005년 10월 채택]