Journal of Korean
Data & Information Science Society
2003, Vol. 14, No.4 pp. 799 ~803

On Estimation of Parameter of the Generalized Pareto Distribution[1)]

In Suk Lee[2)] · Sang Moon Kim[3)]

Abstarct

We obtained the best linear unbiased estimator for the scale parameter of the generalized Pareto Distribution by the order statistics.

주제어 : 일반화파레토분포, 최량선형비편향추정량, 순서통계량

Ⅰ. 서론

일반화파레토분포(Generalized Pareto Distribution : GPD)는 Pickands (1975)에 의해 처음으로 소개된 후 환경문제, 대기권의 오존 수준, 보험관계 및 신뢰성 문제 등에 많이 활용되는 확률분포다. 그러므로 GPD에 관계되는 연구가 많이 이루어지고 있다. 즉 Smith(1985)와 Grimshaw(1993)가 모수의 최우추정량에 대하여 연구하였고, Hosking과 Wallis(1987)가 적률을 이용한 모수와 분위수의 추정에 대한 연구를 하였으며 Lin과 Wang(2000)은 중도절단자료를 갖는 경우의 모수에 대한 최우추정량을 연구하였다.

한편 Balakrishnan과 Ahsanulah(1994)는 GPD의 l차 적률 및 곱의 적률에 대한 Recurrence 관계를 밝혔으며, Pawlas와 Szynal(1999, 2001)은 일반화 순서통계량의 l차 적률과 곱의 적률에 대한 Recurrence 관계를 밝혔다. 또한 Lloyd(1952)는 순서통계량을 이용하여 위치모수와 척도모수에 대한 일반화최소제곱 추정량을 구하였다.

그러므로 본 논문에서는 GPD의 순서통계량의 적률 및 곱의 적률과 Lloyd (1952)의 결과를 이용하여 지금까지 연구되지 않은 척도모수의 최량선형비편향추정량(best linear unbiased estimator : BLUE)을 n개의 통계량으로부터 선택한 $k(k < n)$개의

---

1) This research was supported by Kyungpook National University Research Fund 2002.
2) Professor, Department of Statistics, Kuungpook National University, Daegu, 702-701, Korea.
E-mail : islee@bh.knu.ac.kr
3) Department of Statistics, Kynugpook National University, Daegu, 702-701, Korea.

순서통계량으로 얻고자 한다.

Ⅱ. 본론

확률변수 X가 확률밀도함수

$$f(x)=\begin{cases}\frac{1}{\alpha}\left(1-\frac{\beta x}{\alpha}\right)^{1/\beta-1}, & \beta\neq 0\\ \frac{1}{\alpha}\exp\left(-\frac{x}{\alpha}\right), & \beta=0\end{cases} \tag{1}$$

를 가질 때 X를 일반화파레토분포(GPD)라 한다. (1)에서 $\beta\leq 0$이면 $0\leq x<\infty$이고, $\beta>0$이면 $0\leq x<\frac{\alpha}{\beta}$이며 α는 척도모수이고, β는 형상모수로 이 논문에서는 알고 있는 것으로 가정한다. 예컨데 $\beta=0$이면 X는 지수분포이고 $\beta=0.5$이면 삼각분포이며, $\beta=1$이면 균일분포이다. 지금 $Y_1<Y_2<\cdots<Y_n$을 (1)에서 추출한 확률표본 $X_1,X_2,\cdots,X_n$의 순서통계량이라 한다.

Y_i의 기대치와 Y_i와 Y_j의 곱의 기대치를 구하기 의하여 먼저 $\beta>0$인 경우를 생각한다. 그러면 Y_i의 확률밀도함수 $g_i(y_i)$는

$$g_i(y_i)=\frac{n!}{(i-1)!\,(n-i)!}\left(\frac{1}{\alpha}\right)$$
$$\times\left[1-\left(1-\frac{\beta y_i}{\alpha}\right)^{\frac{1}{\beta}}\right]^{i-1}\left(1-\frac{\beta y_i}{\alpha}\right)^{\frac{n-i-\beta+1}{\beta}},\quad 0<y_i<\frac{\alpha}{\beta}$$

와 같이 쉽게 얻을 수 있다. 그러므로 $\frac{\beta y_i}{\alpha}=z$로 변수변환하고 이항전개식을 이용하여 몇 번의 계산을 하면

$$E[Y_i^l]=\int_0^{\alpha/\beta}y_i^l g_i(y_i)dy_i$$
$$=\frac{n!}{(i-1)!\,(n-i)!}\left(\frac{\alpha}{\beta}\right)^l\sum_{h=0}^{l}(-1)^h\binom{l}{h}B(n-i+h\beta+1,i) \tag{2}$$

을 얻을 수 있다. 여기서 $B(a,b)$는 베타함수이다. 따라서

$$E[Y_i]=\frac{n!}{(i-1)!\,(n-i)!}\left(\frac{\alpha}{\beta}\right)\{B(n-i+1,i)-B(n-i+\beta+1,i)\}$$
$$=\frac{\alpha}{\beta}(1-C_i). \tag{3}$$

$$E[Y_i^2]=\frac{n!}{(i-1)!\,(n-i)!}\left(\frac{\alpha}{\beta}\right)^2\{B(n-i+1,i)$$

$$-2B(n-i+\beta+1,i)+B(n-i+2\beta+1,i)\}$$

$$=\left(\frac{\alpha}{\beta}\right)^2 D_i. \tag{4}$$

여기서 $C_i=\dfrac{\Gamma(n+1)\Gamma(n-i+\beta+1)}{\Gamma(n-i+1)\Gamma(n+\beta+1)}$ 이고,

$D_i=\dfrac{\Gamma(n+1)\Gamma(n-i+2\beta+1)}{\Gamma(n-i+1)\Gamma(n+2\beta+1)}$ 이다.

한편 Y_i와 Y_j의 주변결합확률밀도함수 $g_{ij}(y_i, y_j)$도 아래와 같이 쉽게 구할 수 있다.

$$g_{ij}(y_i, y_j)=\frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}\left(\frac{1}{\alpha^2}\right)\left[1-\left(1-\frac{\beta y_i}{\alpha}\right)^{\frac{1}{\beta}}\right]^{i-1}$$

$$\times\left(1-\frac{\beta y_i}{\alpha}\right)^{\frac{1}{\beta}-1}\left[\left(1-\frac{\beta y_i}{\alpha}\right)^{\frac{1}{\beta}}-\left(1-\frac{\beta y_j}{\alpha}\right)^{\frac{1}{\beta}}\right]^{j-i-1}$$

$$\times\left(1-\frac{\beta y_j}{\alpha}\right)^{\frac{n-j-\beta+1}{\beta}}, \quad 0<y_i<y_j<\frac{\alpha}{\beta}.$$

따라서 (2)의 계산에서와 같이 몇 번의 변수변환과 이항전개식을 이용하여 조금 지루한 계산을 하면

$$E[Y_i^l Y_j^m]=\int_0^{\alpha/\beta} y_i^l\left[\int_{y_i}^{\alpha/\beta} y_j^m g_{ij}(y_i, y_j)dy_j\right]dy_i$$

$$=\frac{n!}{(i-1)!\,(j-i-1)!\,(n-i)!}\left(\frac{\alpha}{\beta}\right)^{l+m}$$

$$\times\sum_{h=0}^{l}\sum_{g=0}^{m}(-1)^{h+g}\binom{l}{h}\binom{m}{g}$$

$$\times B(n-j+g\beta+1, j-i)B(n-i+(h+g)\beta+1, i)$$

을 얻을 수 있다. 그러므로

$$E[Y_iY_j]=\frac{n!}{(i-1)!\,(j-i-1)!\,(n-i)!}\left(\frac{\alpha}{\beta}\right)^2\{B(n-j+\beta+1, j-i)$$

$$\times[B(n-i+2\beta+1,i)-B(n-i+\beta+1,i)]$$

$$+B(n-j+1,j-i)[B(n-i+1,i)-B(n-i+\beta+1,i)]\}$$

$$=\left(\frac{\alpha}{\beta}\right)^2\left\{D_i\frac{\Gamma(n-j+\beta+1)}{\Gamma(n-i+\beta+1)}\right\} \tag{5}$$

다음에는 $\beta<0$인 경우를 생각한다. 그러면 $0<y_i<y_j<\infty$이다. 그러므로 이 경우에서도 $E[Y_i^l]$와 $E[Y_i^lY_j^m]$은 앞에서 계산된 것과 똑 같은 결과를 얻을 수 있고, 그러므로 (3)에서 (5)까지의 관계가 똑 같이 성립한다. 따라서 (3)에서 (5)까지의 결과를

이용하면 다음과 같은 Y_i와 Y_j의 공분산을 얻는다.

$$COV(Y_i, Y_j) = \left(\frac{\alpha}{\beta}\right)^2 \left\{\frac{C_j D_i}{C_i} - C_i C_j\right\}. \tag{6}$$

지금 $U_i = \frac{Y_i}{\alpha}$ 이라 하면 (3)에서

$$E(U_i) = (1/\beta)\{1 - C_i\}, \quad i = 1, 2, \cdots, n, \tag{7}$$

이고, (6)에서

$$V_{ij} = COV(U_i, U_j) = (1/\beta^2)\left\{\frac{C_j D_i}{C_i} - C_i C_j\right\}, \quad i \le j; \quad i, j = 1, 2, \cdots, n, \tag{8}$$

임을 알 수 있다. 또한 $Y_1 < Y_2 < \cdots < Y_n$에서 임의로 선택한 k개의 순서통계량을 백터 $Y^* = (Y_{n_1}, Y_{n_2}, \cdots, Y_{n_k})'$로 표시한다. 여기서 $Y_{n_1} < Y_{n_2} < \cdots < Y_{n_k}$ $(1 \le n_1 \le n_2 \le \cdots \le n_k \le n)$이다. 또한 그에 대응하는 $U_{n_1} < U_{n_2} < \cdots < U_{n_k}$를 백터 $U^* = (U_{n_1}, U_{n_2}, \cdots, U n_k)'$로, $\mu_U = (E[U_{n_1}], E[U_{n_2}], \cdots, E[U_{n_k}])'$를 U^*의 평균백터로 표시하며, U^*의 분산공분산행렬을 $V^* = \{V_{ij}\}_{k \times k}$라 한다. 그러면 k개의 순서통계량 $Y_{n_1} < Y_{n_2} < \cdots < Y_{n_k}$에 기초한 α의 최량선형비편향추정량 $\tilde{\alpha}$는 위 결과 (7)과 (8)을 Lloyd(1952)의 일반화최소제곱추정량에 적용하면 형상모수 β를 알 때

$$\tilde{\alpha} = \frac{\mu_U' V^{*-1} Y^*}{\mu_U' V^{*-1} \mu_U} \tag{9}$$

이다. 또한 $Var(\tilde{\alpha}) = \frac{\alpha^2}{\mu_U' V^{*-1} \mu_U}$ 이다.

Ⅱ. 결론

(9)에서 우리는 형상모수 β를 알고 있을 때 크기가 n인 확률표본에서 임의로 선택된 k개의 순서통계량을 이용한 척도모수 α의 선형비편향추정량들 중에서 최소분산을 갖는 추정량인 $\tilde{\alpha}$을 얻었다. 그러므로 주어진 형상모수 β값에 따른 여러 가지 확률분포에 대하여 임의의 중도절단표본에서도 α의 최량선형비편향추정량으로 (9)의 결과를 적용할 수 있다.

참 고 문 헌

1. Balakishnan, N and Ahsanullah, M., (1994), Recurrence relations for single and product moments of record values from generalized Pareto distribution, *Communications in Statistics-Theory and Methods*, 23(10), 2841-2852.
2. Grimshaw, S.D., (1993), Computing maximum likelihood estimators for the generalized Pareto distribution, *Tecnometrics,* 35, 185-191.
3. Hosking, J. R. M. and Wallis, J. R., (1987), Parameter and quantile estimation for the generalized Pareto distribution, *Tecnometrics*, 29, 339-349.
4. Lin, C. T. and Wang, W. Y., (2000), Estimation for the generalized Pareto distribution with censored data, *Communications in Statistics -Simulation and Computation*, 29(4), 1183-1213.
5. Lloyd, E. H., (1952), Least-squares estimation of location and scale parameters using order statistics, *Biometrics*, 39, 88-95.
6. Pawlas P. and Szynal D., (1999), Recurrence relations for single and product moments of record values from generalized Pareto and Burr distribution, *Communications in Statistics-Theory and Methods*, 28(7), 1699-1709.
7. Pawlas P. and Szynal D., (2001), Recurrence relations for single and product moments of generalized order statistics from Pareto, generalized Pareto and Burr distribution, *Communications in Statistics-Theory and Methods*, 30(4), 739-749.
8. Pickands, J., (1975), Statistical inference using extreme order statistics, *The Annals of Statistics*, 3, 119-131.
9. Smith, R. L., (1985), Maximum likelihood estimation in class of nonregular cases, *Biometrika*, 72, 67-90.

[2003년 7월 접수, 2003년 10월 채택]