

L, LH, LQ-모멘트의 비교와 GEV 분포의 매개변수 추정

Comparison of L, LH, LQ-moments and Parameter Estimation of GEV Distribution

이길성*, 진락선**

Kil Seong Lee, Lak Sun Jin

요 지

본 연구에서는 Probability Weighted Moments의 새로운 선형조합기법인 LQ-moments를 이용하여 GEV 분포의 매개변수를 추정하고 L, LH, LQ-moments를 사용하여 뉴욕주의 Donnattsburg에 위치한 Independence River의 홍수량을 빈도 해석하였다. LH, LQ-moments가 제시된 근본적인 이유는 L-moments가 극치값에 대해 지나치게 민감한 단점을 보완하기 위해서인데, 이번 연구의 결과에 의하면 오히려 LH, LQ-moments가 극치값에 대해 민감하게 반응하여 부정확한 결과가 도출되었다. 그러므로 항상 LH, LQ-moments가 L-moments의 대안이 될 수 있는 것은 아님을 알게 되었다. 그리고 수학적 유도에서 L, LH, LQ-moments는 좀더 쉽고 간편한 매개변수 추정을 위해 Probability Weighted Moments의 선형조합을 통해 고안되었다는 공통점을 가지고 있지만, 이 점을 제외한 나머지 부분의 수식 유도에서는 서로 많은 차이가 있어서 지역적인 특성과 확률분포형의 특성을 고려하여 L, LH, LQ-moments 중에서 선별 사용해야 할 것이다.

핵심용어: 빈도해석, 확률분포형, GEV 분포, L-moment, LH-moment, LQ-moment

1. 서 론

수리구조물의 설계, 운영, 관리 및 개·보수 측면에서 농업용수를 비롯한 각종 산업 용수의 개발을 위한 댐 및 저수지의 출구부 단면결정, 그리고 배수시설의 용량결정 등에 있어서 우선적으로 설계홍수량이 필수적으로 제시되어야만 한다. 설계홍수량을 제시하기 위해서는 적정 확률분포형의 선정과 선정된 분포함수의 매개변수를 추정하는 것이 중요한 일이다.

2. 이론적 배경

2.1 L-moments(Hosking, 1990)

누적분포함수(Cumulative Distribution Function)인 $F(x)$ 와 사분위함수(Quantile Function)인 $x(F)$ 에서 X 는 실수값을 가진 확률변수(Random Variable)라고 하고 $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ 은 크기가 n 인 무작위

* 정회원·서울대학교 지구환경시스템공학부 교수·E-mail: kilselee@snu.ac.kr

** 정회원·서울대학교 지구환경시스템공학부 석사과정·E-mail: ok7749@hanmail.net

표본의 Order Statistics라고 하자. 그러면 r 번째 L-moments는

$$\begin{aligned}\lambda_r &= r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} E(X_{r-k:r}), \quad r = 1, 2, \dots \\ E(X_{j:r}) &= \frac{r!}{(j-1)!(r-j)!} \int_0^1 x \{F(x)\}^{j-1} \{1-F(x)\}^{r-j} dF(x) \\ \lambda_1 &= E[X_{1:1}] = \int_0^1 x(F) dF \\ \lambda_2 &= \frac{1}{2} E[X_{2:2} - X_{1:2}] = \int_0^1 x(F)(2F-1) dF \\ \lambda_3 &= \frac{1}{3} E[X_{3:3} - 2X_{2:3} + X_{1:3}] = \int_0^1 x(F)(6F^2 - 6F + 1) dF \\ \lambda_4 &= \frac{1}{4} E[X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}] = \int_0^1 x(F)(20F^3 - 30F^2 + 12F - 1) dF.\end{aligned}$$

2.2 LH-moments(Wang, 1997)

LH-moments는 분포의 상위부분에 대한 특징을 보다 잘 나타내기 위해서 Wang(1997)에 의해 소개되었다. 제안된 LH-moments는 고차통계(Higher Order Statistics)의 선형 결합을 바탕으로 하고 있으며, $F(x) = \Pr(X \leq x)$ 을 가지는 m 크기의 표본으로부터 j 번째 보다 작은 변수의 기대값은 다음과 같다.

$$\begin{aligned}E[X_{j:m}] &= \frac{m!}{(j-1)!(m-j)!} \int_0^1 x(F) F^{j-1} (1-F)^{m-j} dF \\ \lambda_1^n &= E[X_{(n+1):(n+1)}] \\ \lambda_2^n &= \frac{1}{2} E[X_{(n+2):(n+2)} - X_{(n+1):(n+2)}] \\ \lambda_3^n &= \frac{1}{3} E[X_{(n+3):(n+3)} - 2X_{(n+2):(n+3)} + X_{(n+1):(n+3)}] \\ \lambda_4^n &= \frac{1}{4} E[X_{(n+4):(n+4)} - 3X_{(n+3):(n+4)} + 3X_{(n+2):(n+4)} - X_{(n+1):(n+4)}]\end{aligned}$$

여기서, λ_1^n 는 크기가 $n+1$ 인 표본에서 가장 큰 값에 대한 기대값으로서 분포의 위치(Measure of Location), λ_2^n 는 $n+2$ 크기를 가지는 표본에 대한 분포의 상위부분에 대한 퍼짐(Spreadness), λ_3^n 는 $n+3$ 크기의 표본인 분포의 상위부분에 대한 비대칭(Asymmetry), λ_4^n 는 $n+4$ 크기의 표본인 분포의 상위부분에 대한 첨두(Peakedness)를 나타낸다. 여기서 $n = 0$ 이라면, LH-moment는 Hosking(1990)의 L-moment와 같게 된다. n 이 증가할수록 LH-moments는 더욱더 분포의 상위부분과 큰 사상에 대한 특성을 반영한다.

2.3 LQ-moments(Mudholkar, 1998)

Quantile Function $Q_X(u) = F_X^{-1}(u)$ 을 가진 Continuous Distribution Function, $F_X(\cdot)$ 으로부터의 표본을 X_1, X_2, \dots, X_n 라고 하자. X 의 r 번째 LQ-moments는

$$\zeta_r = r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} \tau_{p,\alpha}(X_{r-k:r}), \quad r = 1, 2, \dots \quad (0 \leq \alpha \leq 1/2, \quad 0 \leq p \leq 1/2)$$

$$\tau_{p,\alpha}(X_{r-k:r}) = pQ_{X_{r-k:r}}(\alpha) + (1-2p)Q_{X_{r-k:r}}(1/2) + pQ_{X_{r-k:r}}(1-\alpha).$$

$$\hat{\zeta}_1 = \hat{\tau}_{p,\alpha}(X),$$

$$\hat{\zeta}_2 = \frac{1}{2} [\hat{\tau}_{p,\alpha}(X_{2:2}) - \hat{\tau}_{p,\alpha}(X_{1:2})],$$

$$\hat{\zeta}_3 = \frac{1}{3} [\hat{\tau}_{p,\alpha}(X_{3:3}) - 2\hat{\tau}_{p,\alpha}(X_{2:3}) + \hat{\tau}_{p,\alpha}(X_{1:3})],$$

$$\hat{\zeta}_4 = \frac{1}{4} [\hat{\tau}_{p,\alpha}(X_{4:4}) - 3\hat{\tau}_{p,\alpha}(X_{3:4}) + 3\hat{\tau}_{p,\alpha}(X_{2:4}) - \hat{\tau}_{p,\alpha}(X_{1:4})].$$

3. 연구 대상 선정 및 자료 검정

미국 지질조사국(U. S. Geological Survey)은 홍수량 자료를 오랜 기간 동안 여러 지점에서 관측하여 그 기록을 보유하고 있으며 자료의 수준 또한 양질이다. 본 연구에서는 뉴욕주의 Donnattsburg에 있는 Independence River(유역 면적은 88.7 mi², 하도경사는 38.9 ft/mile, 하천연장은 25.4 miles)의 홍수량 자료를 사용하였다. 그리고 빈도해석을 하기에 앞서 대상 자료를 도시하여 그 경향을 파악하는 것은 자료의 개략적인 성질을 예측하는데 많은 도움을 준다. 또한, 대상 자료가 무작위성(Randomness)을 갖고 있는가도 검토를 해야 한다. 자료의 일반적인 균형분포 상태에서 훨씬 위나 또는 아래로 격리되어 Outlier를 유발하는 Data Point들은 부적절한 통계학적 매개변수의 유도과 실제홍수량의 제시에 불확실성을 초래하므로, High Outlier 및 Low Outlier가 있음이 확인되는 Data Point들은 사용하지 말아야 한다.

4. GEV 분포의 매개변수 추정 및 GEV 분포의 적합도 검정

GEV(General Extreme Value)분포는 극값분포의 한계형태를 모르고 있을때, 기상자료의 가장 큰값들의 빈도분포를 동일하게 하기 위하여 Jenkinson(1955, 1969)에 의해 소개되었다. GEV분포의 누가분포함수(Cumulative Distribution Function)는

$$F(x) = \exp \left[- \left\{ 1 - \frac{k(x-\epsilon)}{\alpha} \right\}^{1/k} \right],$$

여기서 α 는 Scale Parameter이고 ϵ 은 Location Parameter, k 는 Shape Parameter이다. 임의의 확률분포에 대한 적합도 검정은 그 확률분포의 상대도수함수와 누가도수함수의 이론값과 표본값을 비교하여 그 정도를 판별하게 되는데, 검정 방법에는 Chi-square(χ^2) 검정, Kolmogorov-Smirnov(K-S) 검정, Cramer Von Mises 검정, Probability Plot Correlation Coefficient(PPCC) 검정 등이 있다. 본 연구에서는 전구간에 대한 적합도를 나타내는 Chi-square 검정과 각각의 소구간별 적합도 분석을 위한 Kolmogorov-Smirnov 검정을 이용하여 GEV분포의 적합성을 검증하였는데, 유의수준 5% 수준에서 적합성이 인정되었다.

5. 연구 결과

L, LH, LQ-moments를 이용하여 GEV 분포의 매개변수 추정 결과를 표 1에 나타내었다. 여기서 세가지 모멘트방법 모두 Shewness가 양의 값을 가지는 것으로 나타났는데, 이것은 PDF(Probability Density

Function)가 좌측으로 왜곡되어 있다는 것을 알 수 있다. Kurtosis 또한 양의 값을 가지며 L-moments에서부터 LQ-moments까지 점점 작아지고 있는데, 이것은 PDF의 첨두 형태가 점점 완만해지고 있음을 보여준다. 그리고 Shape Parameter(k)가 음의 값을 가지는데 이것은 EV2(Extreme Value Type 2)분포를 적용가능하다는 정보를 준다. 표 2에서는 모멘트방법에 따른 GEV 분포의 재현기간별 설계홍수량을 제시하고 있는데, 재현기간이 100년일 때를 제외하고는 모두 LQ-moments를 사용해서 설계홍수량을 산정했을 때 가장 큰값을 보이는데, 이것은 LQ-moments가 다른 모멘트들에 비해 GEV 분포에서 동일한 재현기간에서 과대한 설계홍수량을 제시한다는 것을 알 수 있다. 위에서 언급한 내용들의 이해를 돕기 위하여 표 1, 2의 결과값들을 이용하여 모멘트방법에 따른 PDF, CDF, 재현기간별 설계홍수량, 그리고 모멘트방법에 의해 추정된 설계 홍수량과 실측 홍수량을 그림 1부터 4까지 나타내었다. 그림 4에서 LH-moments를 보면 모멘트의 차수가 증가할수록 설계홍수량이 실측홍수량을 잘 따르지 않는데, LH-moments 이론에 의하면 모멘트의 차수가 증가할수록 설계 홍수량이 실측 홍수량을 잘 따르는 것으로 나와 있다. 본 연구에서 사용한 자료와 분포형에 의하면 이론과는 반대의 결과를 보여주고 있는데, 그 이유는 자료의 지역적 특성으로 인한 LH-moments의 차수의 선별문제와 GEV분포가 본 연구에서 사용한 모멘트들에 부적합하기 때문인 것으로 생각된다. 또한, LQ-moments도 그래프의 우측부분에서 과대산정되어 설계홍수량이 실측홍수량을 잘 따르지 않는 것으로 나타났다.

표 1. 모멘트별 GEV 분포의 매개변수 추정 결과

	Skewness	Kurtosis	ϵ	α	k
L-moment	0.3390	0.2320	1859.9355	610.9761	- 0.2500
LH1-moment	0.3809	0.2023	1858.6484	620.9618	- 0.2425
LH2-moment	0.3784	0.1838	1827.3755	673.7241	- 0.2087
LH3-moment	0.3688	0.1661	1773.5204	742.0319	- 0.1719
LH4-moment	0.3564	0.1469	1695.6207	826.8919	- 0.1328
LQ-moment	0.3834	0.1343	1806.1102	764.3824	- 0.3074

표 2. 모멘트법에 따른 GEV 분포의 재현기간별 설계 홍수량

T (Return Period)	Flood Discharge (cfs)					
	L	LH1	LH2	LH3	LH4	LQ
5 years	2971.838	2981.997	3013.9718	3043.182	3068.064	3262.734
10 years	3705.611	3717.284	3762.4743	3812.352	3864.38	3885.836
50 years	5898.371	5894.143	5887.3877	5898.9	5923.26	6070.834
100 years	7134.641	7111.099	7030.6757	6975.445	6938.963	7046.228
200 years	8600.831	8546.86	8348.1244	8184.574	8049.092	8784.616

6. 결 론

LH, LQ-moments가 제시된 근본적인 원인은 L-moments가 큰 표본값에 대하여 불충분한 가중을 주고 극치값에 대하여 지나치게 민감한 단점을 보완하기 위해서인데, 본 연구의 결과에 의하면 LH, LQ-moments가 자료의 극치부분에서 과대 산정되어 LH, LQ-moments가 항상 L-moments의 대안이 될 수는 없다는 것을 보여주었다. 본 연구의 대상 지점에서는 L, LH1-moments로 추정된 설계홍수량이 실측홍수량을 가장 잘 따르는 것으로 나타났으며, 지역적인 특성과 분포의 특성을 고려하여 L, LH, LQ-moments를 선별 사용해

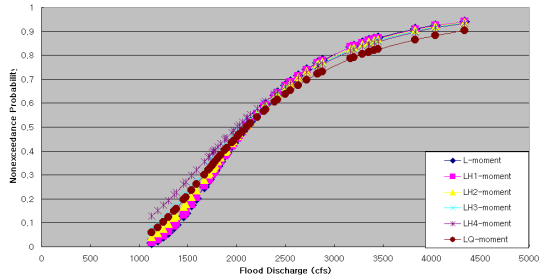


그림 1. 모멘트방법에 따른 PDF

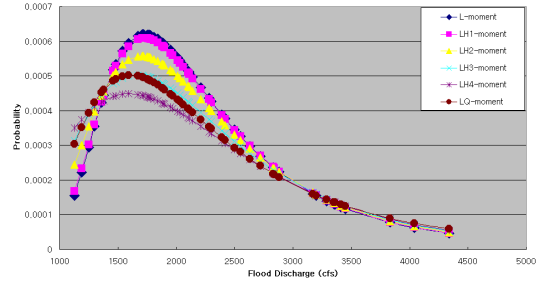


그림 2. 모멘트방법에 따른 CDF

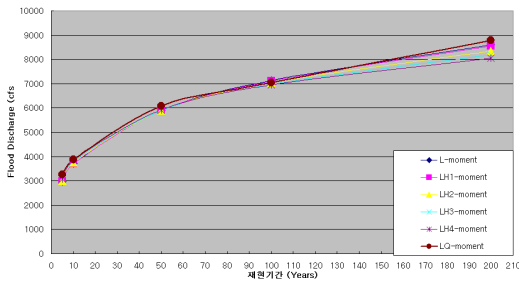


그림 3. 모멘트방법에 따른 재현기간별 설계 홍수량

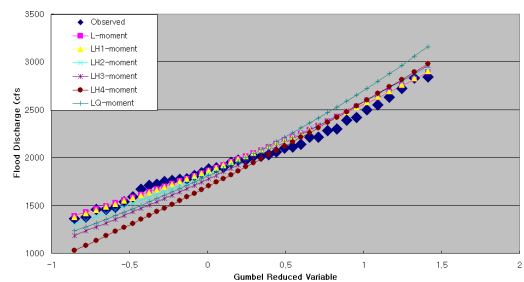


그림 4. 모멘트방법에 의해 추정된 설계 홍수량과 실측 홍수량

야 할 것이다. 본 연구에서는 빈도해석 지점을 한곳만 선정하였기 때문에 일반적으로 어떤 모멘트가 우수하다고 단정 할 수 없었다. 그러므로 향후 연구에는 여러 지점을 대상으로 빈도해석을 실시하고, GEV 분포를 제외한 다른 확률분포형들에 대해서도 L, LH, LQ-moments를 사용하여 매개변수를 추정해 보아야 할 것이다. 이렇게 다양하고 다각적인 빈도해석을 통해 산출된 결과값들을 이용하여 통계적으로 어떤 모멘트가 우수한 결과를 도출하는지 밝히고 지역적 특성에 맞는 확률분포형을 추천할 수 있도록 해야겠다.

감사의 글

본 연구는 21세기 프론티어 연구개발사업인 수자원의 지속적 확보기술개발사업단의 연구비지원(과제번호: 6-2-1)에 의해 수행되었습니다.

참고 문헌

1. Hosking, J. R. M. (1990). "L-moments: Analysis and estimation of distributions using linear combinations of order statistics." J. R. Stat. Soc. Ser. B., Vol. 52, No. 2, pp. 105 - 124.
2. Jenkinson, A. F. (1955). "The frequency distribution of the annual maximum (or minimum) value of meteorological elements." Q. J. Royal Meteorol. Soc., Vol. 81, pp. 158 - 171.
3. Jenkinson, A. F. (1969). "Estimation of maximum floods." World Meteorological Organization, Technical Note, No. 98, pp. 183 - 257.
4. Mudholkar, G. S. (1998). "LQ-moments: Analogs of L-moments." J. Stat. Plan., Vol. 71, pp. 191 - 208.
5. Wang, Q. J. (1997). "LH-moments for statistical analysis of extreme events." W. R. R., Vol. 33, No. 12, pp. 2841 - 2848.