

- I. 통계학의 기초(Basic Statistics)
- II. 빈도해석(비매개변수적 방법)(Nonparametric Frequency Analysis)
빈도해석(매개변수적 방법) (Parametric Frequency Analysis)
- III. 검정방법(Various Tests)
- IV. 자료의 경향 및 변동 측정방법(Detection of Changes and Trend in Data)
- V. 결측치보완 및 자료확충방법(Filling in Missing Data and Extension of Records)

1. 머리말

지난번 강좌에서는 확률현상, 임의사상, 확률변수, 확률밀도함수와 누가분포함수와 모집단과 표본에 대한 모멘트, 확률가중모멘트, L-모멘트의 정의 및 여러 가지 통계치 등 통계학의 기초에 대해 간단하게 설명하였으며, 연속되는 이번 강좌에서는 강우량 및 홍수량 같은 자료의 빈도해석에 대한 내용을 다루기로 한다. 이를 위해 경험적인 방법과 확률분포형을 이용한 방법에 대해 간단히 설명하기로 한다.

2. 빈도해석

수공구조물을 설계할 때 강우나 홍수자료에 대한 빈도해석이 필요하게 된다. 이를 위해 본 장에서는 경험적인 빈도해석 및 확률분포형을 사용한 빈도해

석 방법에 대해 설명하기로 한다.

2.1 경험적인 빈도해석 (Empirical frequency analysis)

경험적인 빈도해석은 자료의 특성에 따라 전자료를 사용하는 경우와 자료를 군집화하는 경우 등 2가지로 구분하여 빈도해석을 실시할 수 있다.

2.1.1 전자료를 이용한 빈도해석(frequency analysis of ungrouped data)

연최대홍수량, 연평균유출량, 연최소유출량 등과 같은 연자료에 대해서는 일반적으로 자료를 군집화하지 않고 전체 자료를 도시공식(plotting position formula)을 사용하여 빈도해석을 실시한다. 여기서는 자료기간 중 전자료가 존재하는 경우, 자료기간 중 0의 값을 갖거나 자료를 기준값(censoring level)으로 구분하는 경우, 자료기간 중 일부자료가 없는 경우, 역사적 정보(historical information)를 이용한 경우에 대해서 설명하기로 한다.

(1) 전자료가 존재하는 경우(complete records)

기록기간이 N인 임의의 자료 Y_1, Y_2, Y_N 에 대해 빈도해석을 하는 절차는 다음과 같다.

① 자료 Y_1, Y_2, \dots, Y_N 을 크기순으로 다음과 같이 $X_1 \geq X_2 \geq X_3 \dots X_{N-1} \geq X_N$ 정렬한다(여기서 X_1 은 자료 중 최대값이며, X_N 은 최소값을 의미한다).

② 크기순으로 정렬된 임의의 X_i 에 대한 빈도를

* 연세대학교 토목공학과 조교수

표 2.1 도시공식별 상수와 적용 확률분포형

Plotting Position Formula	α and β	Reference	Remarks
$\frac{i}{N}$	$\alpha=0.0$ and $\beta=0.0$	California(1953)	Kolmogorov-Smirnov test
$\frac{i}{N+1}$	$\alpha=0.0$ and $\beta=1.0$	Weibull(1939)	uniform
$\frac{i-0.375}{N+0.25}$	$\alpha=3/8$ and $\beta=1.0$	Blom(1958)	gamma, normal lognormal
$\frac{i-0.4}{N+0.20}$	$\alpha=0.4$ and $\beta=1.0$	Cunnane(1978)	best compromise, GEV
$\frac{i-0.44}{N+0.12}$	$\alpha=0.44$ and $\beta=1.0$	Gringorten(1963)	GEV, Gumbel, Weibull, exponential
$\frac{i-0.5}{N}$	$\alpha=0.5$ and $\beta=1.0$	Hazen(1914)	traditional choice
$\frac{i-1.0}{N}$	$\alpha=1.0$ and $\beta=2.0$	California(1953)	
$\frac{i-0.31}{N+0.38}$	$\alpha=0.31$ and $\beta=1.0$	Beard(1943)	
$\frac{i-0.3}{N+0.4}$	$\alpha=0.3$ and $\beta=1.0$	Chegodayev(1955) or Benard and Bos-Levenbach(1953)	
$\frac{i-0.25}{N+0.5}$	$\alpha=0.25$ and $\beta=10.0$	Adamowski(1981)	

계산한다.

절차 ②에서 빈도를 계산하기 위하여 수문학 분야에서는 i/N 또는 $i/(N+1)$ 과 같은 도시공식이 널리 사용되어 왔으나 Cunnane(1978)이 기존의 여러 가지 도시공식을 다음과 같은 일반식으로 제안하였다.

$$P(i) = \frac{(i-\alpha)}{(N+\beta-2\alpha)} \quad (2.1)$$

여기서 N 은 자료기간, i 는 자료의 크기순서, α , β 는 사용하는 확률분포형에 따라 결정되는 상수이며, $P(i)$ 는 i 번째 큰 값 X_i 에 상응하는 초과확률(exceedance probability)이다. 표 2.1에 각각의 도시공식과 이에 해당하는 상수 α , β , 그리고 적용 확률분포형을 나타내었다.

표 2.1에 주어진 도시공식별 차이점을 알아보기 위하여 자료기간 $N=100$ 인 경우에 대한 초과확률과 재현기간을 표 2.2에 나타내었다. 표 2.2에서 $P(1)$ 과 $P(N)$ 은 각각 최대값과 최소값에 대한 초과확률이며 $T(1)$ 과 $T(N)$ 은 이에 대응하는 재현

기간을 나타낸다. 표에서 보는바와같이 최대값에 대한 재현기간은 도시공식별로 큰 차이를 보이고 있으며, 작은 값으로 갈수록 차이가 없음을 알 수 있다. 참고적으로 초과확률 p 와 재현기간 T 의 관계식은 다음과 같이 주어진다.

$$T = \frac{1}{p} = \frac{1}{(1-q)} \quad (2.2)$$

여기서 $q(=1-p)$ 는 비초과확률(nonexceedance probability)이다.

(2) 자료기간 중 0의 값을 갖거나 자료를 기준값으로 구분하는 경우

건조한 지역에서는 계절에 따라 유출량이 전혀 없는 경우가 발생할 수 있으며, 이런 경우의 유출량값은 0이 된다(그림 2.1 참고). 또한 하천 수위가 수위표 관측지점의 관측 가능한 값보다 낮거나(그림 2.2), 이상홍수 등에 의해 관측 가능한 값보다 큰 수위를 갖게 되는 경우에는(그림 2.3) 앞에서 설명한 도시공식이 아닌 다른 도시공식을 이용

표 2.2 자료기간 N=100인 경우 도시공식별 초과확률 및 재현기간의 비교

Plotting Position Formula	N=100			
	P(1)	T(1)	P(N)	T(N)
$\frac{i}{N}$	0.0100	100.0	1.0000	1.00
$\frac{i}{N+1}$	0.0099	101.0	1.9901	1.01
$\frac{i-0.375}{N+0.25}$	0.0062	160.4	0.9938	1.01
$\frac{i-0.4}{N+0.20}$	0.0060	167.0	1.9940	1.01
$\frac{i-0.44}{N+0.12}$	0.0056	178.8	0.9944	1.01
$\frac{i-0.5}{N}$	0.0050	200.0	0.9950	1.01
$\frac{i-1.0}{N}$	0.0000	∞	0.9900	1.01
$\frac{i-0.31}{N+0.38}$	0.0069	145.5	0.9931	1.01
$\frac{i-0.3}{N+0.4}$	0.0070	143.4	0.9930	1.01
$\frac{i-0.25}{N+0.5}$	0.0075	134.0	0.9925	1.01

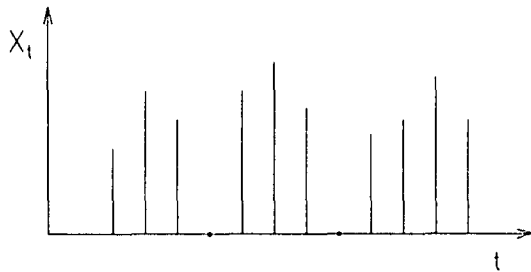


그림 2.1 관측값중 0의 값을 갖는 자료

하여 빈도해석을 실시하여야 한다. 그림 2.2와 2.3에서 보는바와같이 관측 가능한 값을 가지고 빈도 해석을 하기 위하여 임의의 값(x_0)으로 자료를 구분해야 하는데 이때 임의의 값을 기준값(censoring level)이라고 한다.

기준값은 임의의 값으로 정의할 수 있기 때문에 0의 값을 갖는 경우와 관측 가능한 값보다 낮은 자료를 갖는 경우는 동일한 과정을 통해 다음과 같이 빈도해석을 할 수 있다. 자료기간이 N인 관측자료 X_1, X_2, \dots, X_N 에서 임의의 기준값 x_0 보다 큰 값을

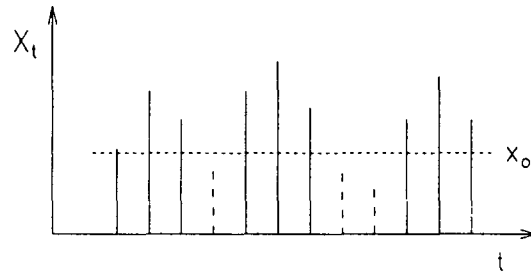


그림 2.2 관측 가능한 값보다 낮은 자료를 갖는 경우

갖는 자료수를 e이라 하면, 기준값과 같거나 작은 값을 갖는 자료수는 $N-e$ 가 된다. 이 경우에 사용하는 도시공식은 다음과 같이 주어진다(Jennings와 Benson, 1969; IACWD, 1982).

$$P(i) = \frac{i-\alpha}{(e+1-2\alpha)} \left(\frac{e}{N} \right) \quad (2.3)$$

여기서 i는 자료의 크기 순서이며 α 는 적용분포형에 따라 결정되는 계수이다.

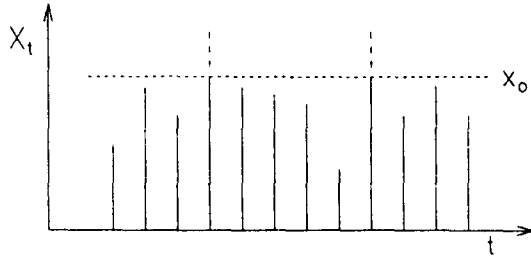


그림 2.3 관측 가능한 값보다 큰 자료를 갖는 경우

그림 2.3과 같이 관측 가능한 값보다 큰 경우에서는 위에서 설명한 방법과 다른 공식을 이용하여 빈도해석을 실시하여야 한다. 이 경우의 도시공식은 다음과 같이 주어진다(Salas and Fernandez, 1995).

$$P(i) = \frac{e}{N} + \left(1 - \frac{e}{N}\right) \frac{i - e - \alpha}{N - e + 1 - 2\alpha} \quad (2.4)$$

여기서 e 는 기준값보다 큰 자료수, i 는 자료의 크기순서, α 는 적용분포형에 따라 결정되는 계수이다.

(3) 자료기간 중 결측값이 있는 경우

수문자료 중에는 결측값이 있는 경우가 있으며, 우리나라에서는 1940년대 말에서 1950년대 중간 기간에 자료가 결측된 경우를 많이 볼 수 있다. 이와 같은 경우에 결측값이 있는 기간을 제외하고 관측된 값만을 가지고 빈도해석을 하는 오류를 범해서는 안되며, 가능한 한 주변 지역에 있는 한 지점 또는 여러 지점의 자료를 이용하여 자료를 보완한 뒤에 빈도해석을 실시하여야 한다(수문자료를 보완하는 방법은 다음에 설명하기로 한다). 일단 자료가 보완된 후에도 두 가지 방법을 이용하여 해석할 수 있다. 즉 보완된 자료를 실제로 관측된 값으로 가정하고 식 (2.1)을 이용하여 빈도해석을 실시하는 방법이 있으며, 이를 구한 다음 보완된 자료를 제외하고 실제 관측된 값과 이에 해당하는 빈도만을 사용하는 방법이 있다.

(4) 역사적 정보(historical information)를 이용한 경우

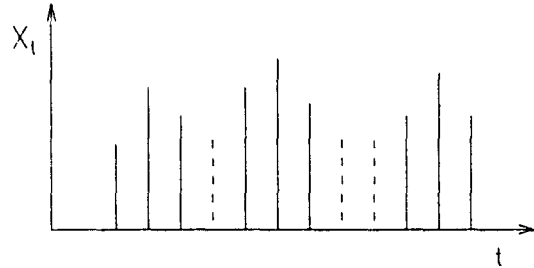


그림 2.4 관측기간중 결측값을 갖는 경우

앞에서 설명한 방법들은 체계적인 관측이 시작된 후의 자료들을 이용하여 빈도해석을 실시하는 방법이지만, 일반적으로 우리가 사용하는 수문자료는 관측기간이 짧으므로 빈도해석을 할 때에 정확도에 문제가 발생할 수가 있다. 이러한 문제점을 해결하기 위한 한가지 방법으로 제안된 것이 역사적 정보를 이용한 빈도해석 방법이다. 홍수량의 예를 들면, 과거의 홍수흔적, 말로 전해져 온 역사적 근거, 지형학적 또는 식물학적 증거 등에 의해 과거 어떤 시점에서의 홍수량을 추정할 수 있다. 만약 이 홍수량과 시점이 정확하다면 이를 이용하여 홍수 빈도해석을 할 수 있으며 빈도 해석시 정확도를 높일 수 있는 것이다.

역사적 정보를 이용한 빈도해석 방법은 그림 2.5a와 2.5b에 나타낸 바와 같이 크게 Type I과 Type II 두 가지로 나눌 수 있다(Condie 와 Lee, 1982; Salas와 Fernandez, 1995). Type I의 기준화된 자료는 기준값 이상의 체계적인 자료(systematic records)와 역사적 정보에 의해 이루어진다. Type I의 경우 기준값 이상의 값을 갖는 자료수는 확률변수(random variable)가 된다. Type II의 경우에는 전 자료기간 중에서 최대값을 갖는 자료수가 정해진 뒤에 기준화된 자료가 결정되기 때문에, 기준값이 확률변수가 된다. 예로 그림 2.5b에서는 3개의 최대값을 기준화된 자료로 정했을 때의 기준값을 나타내고 있다.

그림 2.5a에서 n 은 전체 자료기간, s 와 h 는 각각 체계적인 자료의 관측기간과 역사적 자료기간을 나타낸다. 먼저 Type I의 경우 e 와 e' 은 각각 체계적인 자료 관측기간과 역사적 자료기간에서 기준값보다 큰 자료의 수를 나타내며, k 는 e 와 e' 을 합한 자료수로, 실제로 전체자료기간 n 동안에 우

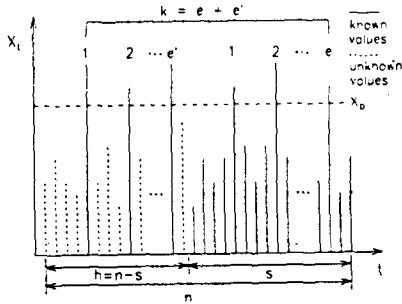


그림 2.5a Type I 기준화된 자료(x_0 고정)

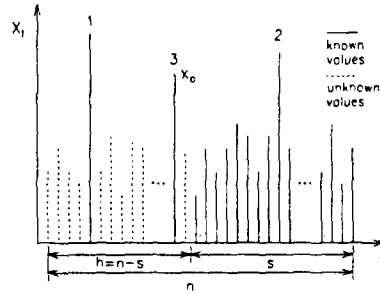


그림 2.5b Type II 기준화된 자료($k=3$ 으로 고정된 경우)

리가 정확한 값을 알고 있는 자료수 g 는 $s+k-e$ 로 표시된다. Type I의 경우에 사용되는 도시공식 여러 가지가 제안되어 있다. 이 중 IACWD(1982)가 제안한 식은 다음과 같다.

$$P(i) = \frac{1}{n+1} \quad i=1, \dots, k \quad (2.5)$$

$$= \frac{k+0.5}{n+1} + \frac{n-k}{n+1} \frac{i-k-0.5}{s-e} \quad i=k+1, \dots, g$$

또한 Hirsch 와 Stedinger(1986) 및 Hirsch(1987)는 다음과 같이 제안하기도 하였다.

$$P(i) = \frac{i-\alpha}{k+1-2\alpha} \left(\frac{k}{n}\right) \quad i=1, \dots, k \quad (2.6)$$

$$= \frac{k}{n} + \left(1 - \frac{k}{n}\right) \frac{i-k-\alpha}{s-e+1-2\alpha} \quad i=k+1, \dots, g$$

한편 Salas와 Fernandez(1995)는 Type II에 대해 다음과 같은 식을 제안하였다.

$$P(i) = \frac{i-\alpha}{k+1-2\alpha} \left(\frac{k}{n}\right) \quad i=1, \dots, k \quad (2.7)$$

$$= P_k + (1-P_k) \frac{i-k-\alpha}{s-e+1-2\alpha} \quad i=k+1, \dots, g$$

여기서

$$P_k = \left(\frac{k-\alpha}{k+1-2\alpha}\right) \left(\frac{k}{n}\right) \quad (2.8)$$

식 (2.6)과 (2.7)에서 k 번째 큰 값에 대해서는

식이 동일한 것을 알 수 있으며, 나머지에 대해서만 서로 다른 값을 갖는 것을 알 수 있다.

【예제】한강 인도교지점의 1918년부터 1991년까지의 연최대홍수량은 그림 2.6 과 같다. 그림에서 보는바와같이 1941년부터 1951년까지 11년간의 결측값을 가지고 있다. 인도교지점의 경우 자기상관계수(autocorrelation coefficient)가 거의 0에 가깝고 결측된 자료의 기간이 연속적으로 11년간이나 결측되어 AR(autoregressive)모형 등을 이용한 결측값 보완에 문제점이 있으며, 가장 많은 연관성을 갖고 있다고 볼 수 있는 고안지점의 자료도 비슷한 기간(1941-1953)에 결측값을 갖고 있으므로 자료 보완에 어려움이 있다. 따라서, 여기서는 결측된 자료를 보충하는 대신 역사적 홍수정보를 이용한 빈도해석 기법을 이용하여 인도교지점에 대한 빈도해석을 실시하여 결과를 비교하기로 한다. 22,000cms를 임의의 기준값으로 가정하자. 그리고, 1952년부터 기록된 자료를 체계적인 자료로, 1952년 이전 자료에서는 기준값 이상의 홍수량과 시기만 알려져 있고 그 이하의 자료는 알려져 있지 않다고 가정하자. 그림에서 보는바와같이 기준값 22,000cms 보다 큰 홍수량의 수는 9개($k=9$)로 역사적 홍수정보에서 3개 ($e'=3$), 그리고 체계적인 자료기간에서는 6개 ($e=6$)임을 알 수 있다. 표 2.3은 기준값이 22,000cms 일 때의 IACWD(1982)와 Hirsch(1987)가 제안한 도시공식을 사용하여 구한 재현기간을 나타내고 있다. 표 2.3에서 기준값 이상에서는 IACWD에 의한 값보다는 Hirsch에 의한 재현기간이 크게 나타남을 알

표 2.3 역사적 홍수정보와 Cunnane의 도시공식을 이용한 확률홍수량(기준값 22,000 cms)

순위	홍수량 (cms)	IACWD		Hirsch		순위	홍수량 (cms)	Cunnane	
		CDF	재현기간	CDF	재현기간			CDF	재현기간
1	32986	.0133	75.0	.008	121.6	1	32986	.009	105.3
2	32972	.0266	37.5	.021	46.8	2	32972	.025	39.5
3	28836	.0400	25.0	.034	28.9	3	28836	.041	24.3
4	27409	.0533	18.7	.047	20.9	4	27409	.057	17.5
5	26431	.0666	15.0	.060	16.4	5	26431	.072	13.7
6	24430	.0800	12.5	.073	13.5	6	24430	.088	11.2
7	23075	.0933	10.7	.087	11.4	7	23075	.104	9.5
8	22373	.1066	9.3	.100	9.9	8	22373	.120	8.3
9	22338	.1200	8.3	.113	8.8	9	22338	.136	7.3
10	19400	.1394	7.1	.137	7.2	10	21053	.151	6.5
11	19049	.1649	6.0	.163	6.1	11	20813	.167	5.9
12	18875	.1903	5.2	.188	5.2	12	19399	.183	5.4
13	17518	.2158	4.6	.214	4.6	13	19269	.199	5.0
15	16365	.2413	4.1	.240	4.1	14	19049	.215	4.6
15	15922	.2668	3.7	.265	3.7	15	18875	.231	4.3
.
.
.

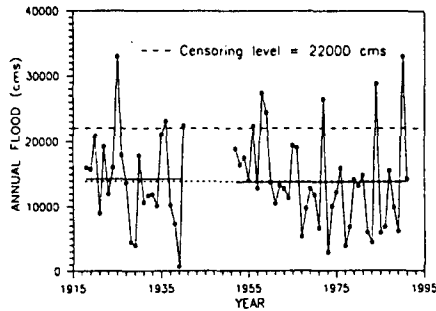


그림 2.6 인도교지점의 연도별 홍수자료

수 있다. 일반적인 도시공식을 사용할 때에는 결측치가 없어야 하나, 결측값을 제외한 전 자료에 대해 Cunnane의 도시공식을 이용하여 구한 재현기간을 참고적으로 표에 나타내었다. 전 자료기간은 총 74년이나 11년간의 결측치를 뺀 63년간의 자료를 가지고 산정된 값이므로 실제로는 표에 나타난 값보다 큰 재현기간을 가져야 한다. 그러므로, Cunnane의 공식에 의한 재현기간보다 작은 값을 갖는 IACWD에 의한 재현기간은 과소 추정된 것이라 할 수 있으며, 크게 산정된 Hirsch의 공식은 합리적이라 할 수 있다.

선정된 확률분포형에 의한 확률홍수량과 역사적

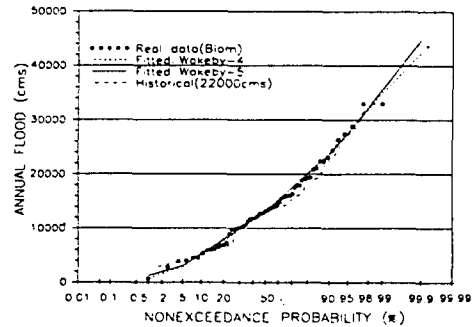


그림 2.7 인도교지점의 확률홍수량(정규확률지)

홍수정보를 이용한 Hirsch의 방법을 비교할 때 일반적으로 확률분포형에 의한 확률홍수량이 약간 큰 것으로 나타났으며, 재현기간 100년인 경우 Hirsch공식에 의한 홍수량이 전반적으로 작게 나온 것은 1 순위와 2 순위에 해당하는 홍수량 값이 거의 일치하기 때문으로, 이것은 도시공식 사용시의 단점이라 할 수 있다.

그림 2.7은 Blom의 도시공식을 이용한 경험적인 빈도곡선, 인도교지점의 홍수량자료에 비교적 잘 맞는 것으로 알려진 Wakeby분포에 의해 적합된 재현기간별 확률홍수량, 그리고 기준값에 의한 역사적 홍수자료를 이용하여 구한 확률홍수량을 정

규확률지에 나타낸 것이다(서규우 등, 1995; 김경덕 등, 1996b).

2.1.2 군집화된 자료를 이용한 빈도해석(frequency analysis of grouped data)

일반적으로 자료기간이 긴 경우에는 자료에 따라 몇 개의 적절한 계급구간(class interval)으로 나누어 빈도해석을 실시할 수 있다. 여기서 구해지는 주상도(histogram) 또는 경험적 확률밀도함수(empirical probability density function)는 대상 자료에 적합하다고 가정한 확률분포형으로부터 얻어지는 확률밀도함수의 적합성을 판단하는 기준을 제공할 수 있다.

자료를 군집화하여 빈도해석을 실시함으로써 주상도, 경험적 확률밀도함수, 경험적 누가분포함수를 구하는 절차는 아래와 같다.

- ① 자료를 작은 값부터 크기순으로 재정렬한다.
- ② 아래의 경험식 중 하나를 택하여 계급구간의 수(N_c)를 결정한다.

$$N_c = 1 + 3.322 \log_{10}(N) \quad (2.9)$$

$$N_c = 5 \log_{10}(N) \quad (2.10)$$

여기서 N 은 자료기간으로 일반적으로 Panofsky 와 Brier(1968)가 제안한 식 (2.10) 보다는 Sturges(1926)가 제안한 식 (2.9)를 많이 사용한다.

- ③ 다음 식을 이용하여 계급구간(ΔX)을 정한다.

$$\Delta X = \frac{X_{\max} - X_{\min}}{(N_c - 1)} \quad (2.11)$$

여기서 X_{\max} 과 X_{\min} 은 자료의 최대값과 최소값을 의미한다.

- ④ 계급구간의 중간값(class mark)인 X_i' ($i=1, 2, \dots, N_c$)을 구한다. 여기서 $X_1' = X_{\min}$ 이고 $X_{N_c}' = X_{\max}$ 이며, i 번째 계급구간의 상한값(upper class limit)과 하한값(lower class limit)은 $[X_i' - \Delta X, X_i' + \Delta X]$ 에 의해 구한다.

- ⑤ i 번째 계급구간 안에 들어가는 자료수를 계산

하여 절대도수(absolute frequency) n_i 를 계산한다.

- ⑥ 각 계급구간에서의 상대빈도 $f(i)$ 와 상대 누가빈도 $F(i)$ 를 다음 식을 이용하여 구한다.

$$f(i) = \frac{n_i}{N} \quad i=1, \dots, N_c \quad (2.12)$$

$$F(j) = \sum_{i=1}^j f(i) = \sum_{i=1}^j \left(\frac{n_i}{N} \right) \quad j=1, \dots, N_c \quad (2.13)$$

- ⑦ i 번째 경험적 확률밀도함수 $\hat{f}(i)$ 와 j 번째 경험적 누가분포함수 $\hat{F}(j)$ 를 식 (2.14)와 (2.15)를 이용하여 구한다.

$$\hat{f}(i) = \frac{f(i)}{\Delta X} \quad (2.14)$$

$$\hat{F}(j) = \sum_{i=1}^j \hat{f}(i) \Delta X \quad (2.15)$$

위에서 언급한 빈도해석 과정에서 계급구간 수나 구간은 임의로 정할 수 있으며, 이 경우에는 ②, ③, ④ 번 절차를 생략할 수 있다. 또한 경험적 확률밀도함수와 누가분포함수를 도출할 때는 다음과 같은 사항에 유의하여야 한다. 즉, 경험적 확률밀도함수를 도출할 때는 계급구간의 중간값(X_i')에 $\hat{f}(i)$ 를 표시하며, 경험적 누가분포함수는 각 계급구간의 상한값($X_i' + \Delta X/2$)에 이에 해당하는 $\hat{F}(j)$ 를 표시한다.

【예제】 어떤 지점의 연최대홍수량이 표 2.4와 같이 주어졌다. 자료를 군집화하여 빈도해석을 하라.

자료기간: $N=66$ (1895~1960)

- ① 주어진 홍수자료를 크기순으로 정렬한다.
- ② 본 예제에서는 계급구간의 수를 임의로 $N_c=7$ 로 가정한다.

- ③, ④ 각 계급구간의 하한값을 자료로부터 15.0, 30.0, 45.0, 60.0, 75.0, 90.0, 105.0로 가정한다(표 2.5의 1, 3칸 참고). 따라서 $\Delta X=15$ 이고, 계급구간의 중간값(X_i')은 표 2.5의 2칸과 같다.

- ⑤ 각 계급구간별 절대도수는 4칸과 같다(그림 2.8a의 주상도 참고).

표 2.4 연최대홍수량 자료

연도	연최대 홍수량									
1891					47.3	54.4	87.2	65.7	91.5	53.5
1901	67.8	70.0	66.9	34.7	58.0	47.0	66.3	80.9	80.0	52.3
1911	58.0	67.2	115.0	46.1	52.4	94.3	111.0	71.7	96.1	92.5
1921	34.1	69.0	73.4	99.1	79.2	62.6	93.7	68.7	80.1	32.3
1931	43.1	77.0	53.6	70.8	89.4	62.6	112.0	44.0	84.3	45.0
1941	28.4	46.0	80.4	55.0	72.9	71.2	46.8	84.1	61.3	87.1
1951	70.5	77.7	44.2	20.6	85.0	82.9	88.7	60.2	40.3	50.5

표 2.5 균집화된 자료의 빈도해석 결과

class interval			absolute frequency	relative frequency	empirical density function $f(x)$	empirical cumulative dist. func. $F(x)$
lower class limit	class mark (X_c')	upper class limit				
15.0	22.5	30.0	2.0	0.0303	0.00202	0.0303
30.0	37.5	45.0	8.0	0.1212	0.00808	0.1515
45.0	52.5	60.0	14.0	0.2121	0.01414	0.3636
60.0	67.5	75.0	18.0	0.2727	0.01818	0.6363
75.0	82.5	90.0	15.0	0.2273	0.01515	0.8636
90.0	97.5	105.0	6.0	0.0909	0.00606	0.9545
105.0	112.5	120.0	3.0	0.0455	0.00303	1.0000

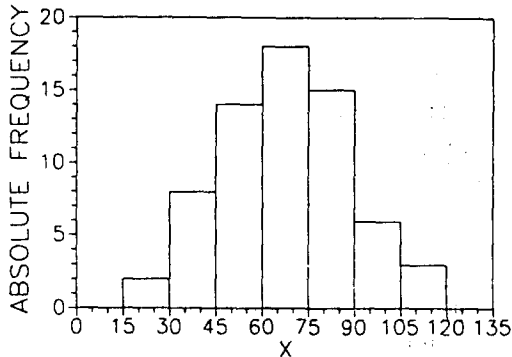


그림 2.8a 주상도(histogram)의 예

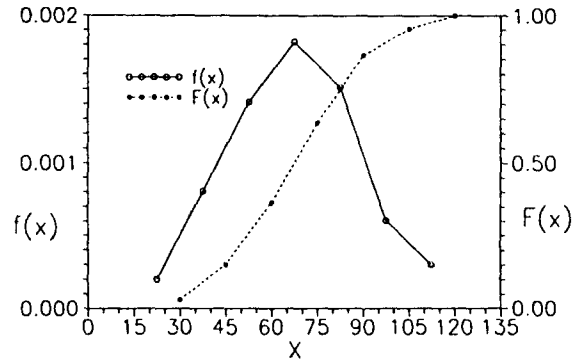


그림 2.8b. 경험적 확률밀도함수와 누적분포함수의 예

⑥ 각 계급구간별 상대도수를 5칸에 나타내었다.

⑦ 각 계급구간별 경험적 확률밀도함수와 누적분포함수는 각각 6, 7칸과 같으며 이를 그림으로 표시한 것이 그림 2.8b와 같다. 참고적으로 확률밀도함수와 누적분포함수의 표시 위치에 주의하여야 한다.

2.2 확률분포형을 이용한 빈도해석

앞 절에서 설명한 방법은 자료값과 자료수만을 이용하여 빈도해석을 하는 것으로 비모수적 방법

(non-parametric method)이라고도 한다. 본 절에서는 대상 자료에 대해 특정 확률분포형을 가정하고 모멘트법, 확률가중모멘트법, 최우도법, L-모멘트법 등 여러 가지 모수 추정방법을 이용하여 모수를 추정하여 빈도해석을 하는 방법에 대해 간단하게 설명하고자 한다.

2.2.1 확률분포형

일반적으로 수공학 및 환경공학분야에서 많이 사용하는 확률분포형을 표 2.6에 나타내었다(김경덕 등, 1996a). 표 2.6에서 x_0 , y_0 및 m 로 표시되는

표 2.6 확률분포형의 확률밀도함수 또는 누적분포함수

확률분포형	확률밀도함수 또는 누적분포함수
Gamma	$f(x) = \frac{1}{ \alpha \Gamma(\beta)} \left[\frac{x-x_0}{\alpha} \right]^{\beta-1} \exp\left[-\frac{x-x_0}{\alpha}\right]$
GEV	$f(x) = \frac{1}{\alpha} \left[1 - \frac{\beta(x-x_0)}{\alpha} \right]^{(1/\beta)-1}$
Gumbel	$f(x) = \frac{1}{\alpha} \exp\left\{ -\frac{(x-x_0)}{\alpha} - \exp\left[-\frac{(x-x_0)}{\alpha}\right] \right\}$
Log-Gumbel	$F(x) = \exp\left[-\left(\frac{\theta-x_0}{x-x_0}\right)^\beta\right]$
Lognormal	$f(x) = \frac{1}{\sqrt{2\pi}(x-x_0)\sigma_y} \exp\left[-\frac{1}{2}\left[\frac{\ln(x-x_0)-\mu_y}{\sigma_y}\right]^2\right]$
Log-Pearson type III	$f(x) = \frac{1}{ \alpha \Gamma(\beta)x} \left[\frac{\ln(x)-y_0}{\alpha} \right]^{\beta-1} \exp\left[-\frac{\ln(x)-y_0}{\alpha}\right]$
Wakeby	$x = m + a[1-(1-F)^b] - c[1-(1-F)^d]$
Weibull	$f(x) = \frac{\beta}{\alpha} \left[\frac{x-x_0}{\alpha} \right]^{\beta-1} \exp\left\{ -\left[\frac{x-x_0}{\alpha} \right]^\beta \right\}$

주) $\Gamma(\beta) = \int_0^{\infty} z^{\beta-1} e^{-z} dz$: gamma 함수

표 2.7 확률분포형의 확률변수 범위 및 모수 적합성 조건

확률분포형	확률변수의 범위 및 모수 적합성 조건
Gamma	$\alpha > 0$ 일 때 $x_0 < x < \infty$ $\alpha < 0$ 일 때 $-\infty < x < x_0, \beta < 0$
GEV	$\beta = 0$ 일 때 GEV-1: $-\infty < x < \infty$ $\beta < 0$ 일 때 GEV-2: $x_0 + \alpha/\beta < x < \infty$ $\beta > 0$ 일 때 GEV-3: $-\infty < x < x_0 + \alpha/\beta$
Gumbel	$-\infty < x < \infty$
Log-Gumbel	$x_0 < x < \infty, \theta > x, \beta > 0$
Lognormal	$x_0 < x < \infty$
Log-Pearson type III	$\alpha > 0$ 일 때 $\exp(y_0) \leq x < \infty$ $\alpha < 0$ 일 때 $-\infty \leq x < \exp(y_0)$
Wakeby	$b+d > 0$ 또는 $b=cd=d=0$ $ab=0$ 이면 $b=0, cd=0$ 이면 $d=0,$ $cd \geq 0, ab+cd \geq 0, b > -1$ 이고 $d \geq 1$
Weibull	$x_0 < x < \infty, \alpha > 0, \beta > 0$

위치모수(location parameter)의 값이 0 또는 임의의 상수로 주어지게 되면 해당 확률분포형의 모수 수는 하나 줄어들게 된다. 예를 들어 GEV분포인 경우 x_0 (location parameter), α (scale parameter), β (shape parameter) 등 3개의 모수를 갖지만 x_0 가 0 또는 임의의 상수값을 갖게 되면 2개의 모수를 갖는 GEV분포가 된다.

선정된 확률분포형으로부터 모수를 추정할 때에는 추정된 모수가 모수의 적합성 조건을 만족하는지 여부를 판정하여야 한다. 즉 어떤 모수 추정방법

법에 의해 모수가 추정되었다 하더라도 적합성 조건을 만족하지 못하면 적용 확률분포형은 대상 자료에 적합하지 않다고 할 수 있다. 표 2.6에 표시된 각각의 확률분포형에 대한 확률변수의 범위 및 모수 적합성 조건을 표 2.7에 나타내었다.

2.2.2 모수 추정방법

일반적으로 빈도해석시 많이 사용하는 모수 추정 방법으로는 모멘트법(method of moments), 확률가중모멘트법(method of probability weighted

moments), 최우도법(method of maximum likelihood)과 L-모멘트법(method of L-moments) 등이 사용되고 있다.

여기서 각각의 확률분포형에 대해 모멘트법, 확률가중모멘트법, 최우도법을 이용하여 모수를 추정하는 방법은 그 내용이 방대하기 때문에 본 강좌에서는 그 자세한 내용은 다루지 않고, 그간의 많은 연구 결과를 종합한 문헌에 대해서만 간단히 언급하기로 한다(김경덕 등, 1996c; 허준행, 1996; Salas 등, 1996). 확률가중모멘트의 선형조합으로 구성된 L-모멘트를 이용하여 모수를 추정하는 L-모멘트법에 대해서도 최근에 많은 연구가 진행되고 있으므로 이에 관한 연구 결과에 대해서만 간단히 언급하기로 한다(Hosknig, 1986; 1990; 1991a; 1991b).

2.3 맺음말

본 강좌에서는 수공학 분야에서 많이 사용하는 빈도해석 방법에 대해서 간단히 설명하였다. 확률분포형을 선정하여 모수를 추정하는 방법에 대해서는 지면관계상 참고문헌으로 대체하였으며, 경험적인 방법과 확률분포형을 이용한 빈도해석 방법 외에도, 빈도계수(frequency factor)를 이용한 방법(Chow, 1951)이 재현기간별 확률홍수량 산정시에 널리 사용되고 있다.

다음 강좌에서는 여러 가지 검정 방법에 대해 설명하기로 한다.

참 고 문 헌

서규우, 김경덕, 허준행, 조원철 (1995). “한강 인도교 지점에서의 홍수빈도해석(I)-자료해석 및 적정 확률분포형 선정-”, 대한토목학회 논문집, 제15권 제6호, pp. 1697-1709.

김경덕, 허준행, 조원철 (1996a). “연최대 강우자료의 적정 확률분포형 선정에 관한 연구”, 대한토목학회 논문집(계재예정).

김경덕, 서규우, 허준행, 조원철 (1996b). “한강 인도교지점에서의 홍수빈도해석(II)-확률홍수량 산정-”, 대한토목학회 논문집, 제16권 제II-1호, pp.23-32.

김경덕, 배덕효, 허준행, 조원철 (1996c). “매개변수 추정방법에 따른 강우자료의 분리효과”, 한국수자원학회지, 제29권 제1호, pp. 129-140.

허준행 (1996). “확률가중모멘트법을 이용한 매개변수 추정과 적용”, '95분과위원회 연구결과 보고서, 한국수자원학회, pp. 1-1~1-90.

Adamowski, K. (1981). “Plotting formula for flood frequency”, Water Resources Bulletin, American Water Resources Association, Vol. 17, No. 2, pp. 197-202.

Beard, L.R. (1943). “Statistical analysis in hydrology”, Transaction, ASCE, Paper No. 2201, pp. 1110-1121.

Benard, A. and Bos-Levenbach, E.C. (1953). “The plotting of observations on probability paper”, (translated from Dutch), Statistica Nederlandica, Vol. 7, pp. 163-173.

Blom, G. (1958). Statistical estimates and transformed beta variables, Wiley, New York. California Department of Public Works (1953). Flow in California Streams, California Dept. Public Works, Div. Eng. Irrig., Bulletin 5.

Chegodayev (1964). Quoted in “Formulas for the calculation of the confidence of hydrologic quantities” by A.G. Alekseyev, in Handbook of applied hydrology edited by V.T. Chow, McGraw-Hill.

Chow, V.T. (1951). “A general formula for hydrologic frequency analysis”, Transactions, American Geophysical Union, Vol. 32, pp. 231-237.

Condie, R. and Lee, K. A. (1982). “Flood frequency analysis with historic information”, Jour. Hydrology, Vol. 58, pp. 47-61.

Cunnane, C. (1978). “Unbiased plotting positions—A review”, Jour. Hydrology, Vol. 37, pp. 205-222.

Gringorten, I.I. (1963). “A plotting rule for extreme probability paper”, Jour. Geophysical Research, Vol. 63, No. 3, pp. 813-814.

Hazen, A. (1914). “Storage to be provided in impounding reservoirs for municipal water supply”, Transaction, ASCE, Paper No. 1308,

- pp. 1539-1640.
- Hirsch, R.M. and Stedinger, J.R. (1986). "Plotting position for historical floods and their precision", *Water Resources Research*, Vol. 23, No. 4, pp. 715-727.
- Hirsch, R.M. (1987). "Probability plotting position formulas for flood records with historical information", *Jour. Hydrology*, Vol. 96, pp. 185-199.
- Hosking, J. R. M. (1986). "The theory of probability weighted moments", Research Report RC-12210, IBM T. J. Watson Research Center, Yorktown, New York.
- Hosking, J.R.M. (1990). "L-moments: Analysis and estimation of distributions using linear combinations of order statistics", *Jour. Royal Statistical Society, Series B*, Vol. 52, No. 2, pp. 105-124.
- Hosking, J.R.M. (1991a). "Approximations for use in constructing L-moment ratio diagrams", Research Report, RC-16635, IBM T. J. Watson Research Center, Yorktown, New York.
- Hosking, J.R.M.(1991b). "Fortran routines for use with the method of L-moments", Research Report RC-17097, IBM T. J. Watson Research Center, Yorktown Height, New York.
- Interagency Advisory Committee on Water Data (IACWD), (1982). "Guidelines for determining flood flow frequency", Bulletin #17B, Hydrology Committee, U.S. Geological Survey, Reston, VA.
- Jennings, M.E. and Benson, M.A. (1969). "Frequency curves for annual flood series with some zero events or incomplete data", *Water Resources Research*, Vol. 5, No. 1, pp. 276-280.
- Panofsky, H.A. and Brier, G.W. (1968). *Some Applications of statistics to meteorology*, Pennsylvania State University Press, University Park.
- Salas, J.D. and Fernandez, B. (1995). "Plotting Position Formulas Based on Systematic Historical and Paleoflood Records", submitted to *Jour. Hydrology*.
- Salas, J.D., Smith, R.A., Tabios, G, and Heo, Jun-Haeng. (1996). *Statistical computer techniques in water resources and environmental engineering*, Forthcoming Book.
- Sturges, H.A. (1926). "The choice of a class interval", *Jour. American Statistical Association*, Vol. 21, pp. 65-66.
- Weibull, W. (1939). "A statistical theory of the strength of material." *Ingeni rs Vetenskaps Akademiens Handigar*, No. 151, Stockholm, Sweden.