

- 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)

머리말

수문학, 수자원공학 뿐만 아니라 거의 모든 공학 분야에서 통계학은 널리 사용되고 있다. 통계학의 광범위한 모든 내용을 설명하는 것은 불가능하므로, 본 강좌에서는 물 분야에서 많이 사용되고 응용되는 일부분에 대해서 소개하기로 한다. 이를 위하여 통계학의 기초적인 개념을 간단하게 소개하고, 이를 바탕으로 수문학 또는 수자원 분야에서 어떻게 응용되고 있는지 실례를 들어 설명하고자 한다.

먼저, 통계학의 기초 편에서는 확률현상, 임의사상, 확률변수, 확률밀도함수와 누적분포함수에 대해 설명하고, 모집단과 표본에 대한 모멘트, 확률가중모멘트, L-모멘트의 정의와 여러 가지 통계치에 대해 정의한 뒤 Weibull 분포를 예로 들어 설명한다. 또한 비선형확률함수의 모멘트를 근사적으

로 구하는 방법을 소개하고 Manning 공식을 예로 설명하고자 한다.

연속되는 강좌에서는 강우나 홍수 빈도해석시 사용하는 비매개변수적 방법과 매개변수적 방법에 대해서 설명하고, 통계분석시 일반적으로 사용되는 여러 가지 검정방법에 대해서 검정목적에 따른 검정방법을 소개한다. 그리고 우리가 수문자료를 해석할 때 일반적으로 직면하게 되는 자료의 일관성 문제에서 시간에 따른 수문자료의 변화나 경향성을 분석하는 방법에 대해서 설명하며, 특히 결측치가 있는 경우 자료를 어떻게 보완하고 자료수가 짧은 경우 자료를 어떻게 확충할 것인가에 대해서 소개하기로 한다.

1. 통계학의 기초

1.1 확률현상(Random phenomena)

일반적으로 우리가 수자원 또는 수문 현상을 언급할 때 random이란 용어를 사용하게 되는데 이는 어떤 현상이 불확실성을 가지고 일어난다는 것을 의미한다. 다시 말하면 어떤 자연 현상이 미래에 발생할 것이라고 확실성을 가지고 예측할 수 없다는 것이다. 예를 들어 강우 현상의 발생에 대해 생각해 보자. 내일 비가 올 것인지 아닌지, 어느 정도의 비가 내릴지 확실성을 가지고 예측할 수는 없는 것이다. 다른 예로, 어떤 지점에서의 유출량에 대해서 생각해 보자. 다음 달의 평균 유출량이 얼마나 될지 또는 다음 해의 첨두 유출량이 얼마나 될지 확실성을 가지고 예측할 수 있는 사람은 아무

* 연세대학교 토목공학과 조교수

도 없는 것이다. 위에서 언급한 강우 현상이나 유출 등과 같이 확실성을 가지고 예측할 수 없는 현상을 우리는 확률현상(random phenomena)이라고 말한다. 그러나 확률현상 중의 하나인 일유출량은 일부 물리적인 또는 결정론적(physical or deterministic) 법칙과 일부 확률론적인(probabilistic or stochastic) 법칙을 병용하여 설명될 수 있다. 즉 결정론적인 부분과 확률론적인 부분으로 구성되어 있다고 할 수 있는 것이다. 이와 같은 수문학적인 현상을 넓은 의미에서의 확률현상(random phenomena in broad sense)이라 한다. 일반적으로 수문사상의 모든 요소들은 넓은 의미에서의 확률현상이라 할 수 있다.

1.2 임의 사상(Random Events)

임의 사상을 설명하는데 있어서는 먼저 표본공간(sample space)의 개념을 이해하는 것이 중요하다. 간단한 예로 특정 일에 비가 내리는 현상에 대해 고려해 보자. 이 현상에는 2개의 사상이 가능하다. 즉 비가 내리거나(R) 또는 내리지 않는(N) 2가지 사상이 가능한 것이다. 그러므로 표본공간 Φ 는 다음과 같이 주어지게 된다.

$$\Phi = \{R, N\}$$

다른 예로 강우량에 대해 생각해 보면 다음과 같이 나타낼 수 있다. 즉 사상 (1) : 일강우량이 20 mm 이하, 사상 (2) : 일강우량이 20~40mm, 사상 (3) : 일강우량이 40~60mm, 사상 (4) : 일강우량이 60mm 이상이라 하면, 이 경우 4가지의 사상이 가능하며 표본공간은 $\Phi = \{(1), (2), (3), (4)\}$ 이 된다. 또 다른 예로 저수지에서의 저류량을 예로 들 수 있다. 즉 사상 (1) : 저수지 수위가 최저수위 이하일 때의 저류량, 사상 (2) : 저수지 수위가 최저수위에서 상시만수위인 경우의 저류량, 사상 (3) : 저수지 수위가 상시만수위 이상인 경우의 저류량 등 3가지의 사상으로 구분하면 표본공간은 $\Phi = \{(1), (2), (3)\}$ 와 같이 주어지게 된다.

위 예제는 1차원적인 예로, 2차원인 경우도 생각해 볼 수 있다. 즉 2개의 인접 지점을 생각해 보자. 각각의 지점에서 일정 수위를 넘는 유출이 발

생하는 경우를 사상 “F”라 하고, 넘지 않는 경우를 사상 “N”이라 하면 가능한 결과는 {FF, FN, NF, NN}과 같이 4가지이며 이것이 바로 표본공간이 된다. 여기서 FF 사상은 두지점 모두 일정 수위를 넘는 유출이 발생한 것을 의미한다. 이상과 같이 어떤 모집단(population) 으로부터 발생 가능한 결과들로 이루어진 것이 표본공간이며, 불확실성을 가지고 발생하는 현상을 임의 사상이라고 한다.

1.3 확률변수(Random Variable)

전 절에서 언급한 사상이나 표본공간은 기호적으로 표시한 것으로 수문현상을 이해하는 데 도움을 줄 수 있으나 수학적으로 표시하는데에는 어려움이 있다. 수학적으로 사상을 표현하기 위하여 사상 또는 결과를 어떤 값이나 구간 X에 대응시킬 필요가 있으며, 여기서 어떤 값이나 구간 X를 확률변수라 한다. 앞 절에서 언급한 일강우량의 예를 들어보자. 어떤 사상 A가 $A = [(1)(2)]$ 로 주어지면 이것은 $[0 < X < 40(\text{mm})]$ 와 같은 의미를 갖게 된다. 즉 사상 A를 수학적으로 구간으로 표시한 X는 확률변수인 것이다. 이러한 확률변수는 이산형(discrete) 또는 연속형(continuous)을 갖게 된다. 예로 일강우량의 경우 확률변수 X는 0에서 ∞ 의 값을 갖는 연속형이며, 어떤 특정 월(예, 8월)의 강우일수 같은 경우에는 이산형 확률변수를 갖게 된다. 즉, 확률변수 X는 1에서 31 사이의 값중 하나의 값을 갖게 되는 것이다. 경우에 따라서 확률변수는 이산형과 연속형이 같이 존재하는 혼합형(mixed)으로 표시될 수도 있다. 저수지의 예에서 보면 보다 현실적으로 5가지의 사상이 가능하다. 사상 (1) : 최저수위에서의 저수량, 사상 (2) : 상시만수에서의 저수량, 사상 (3) : 최저수위 이하의 저수량, 사상 (4) : 최저수위와 상시만수위 사이의 저수량, 사상 (5) : 상시만수위 이상의 저수량, 이와 같은 경우는 이산형 확률변수로 사상 (1)과 (2)를 연속형 확률변수로 사상 (3), (4), (5)를 갖는 혼합형 확률변수이다.

1.4 확률밀도함수와 누가분포함수(Probability Density Function & Cumulative Distribution

Function)

위에서 언급한 확률변수의 분포 형태를 나타내는 데 사용되는 것이 확률밀도함수와 누가분포함수이며, 이 두 가지 함수를 나타내는데에는 확률변수에 따라 크게 두 가지 종류로 나눌 수 있다. 즉 이산형(discrete) 함수와 연속형(continuous) 함수이다.

먼저 이산형 분포함수인 경우에 대해서 설명하면 다음과 같다. $x_1, x_2, \dots, x_n, \dots$ 와 같이 이산형 값을 갖는 확률변수 X 의 확률밀도함수는 다음과 같이 정의된다(Mood 등, 1974).

$$f_x(x) = P[X=x_j] \text{ if } x=x_j, j=1,2,\dots,n,\dots \quad (1.1)$$

$$=0 \quad \text{if } x \neq x_j$$

또한 이와 같이 정의된 확률밀도함수는 다음의 조건을 만족해야 한다.

$$f(x) > 0 \text{ for } j=1,2,\dots \quad (1.2a)$$

$$f_x(x) = 0 \text{ for } x \neq x_j \quad (1.2b)$$

$$\sum_x f_x(x_j) = 1 \quad (1.2c)$$

식 (1.2)에서 보는바와 같이 확률변수 X 가 이산형 값중 하나와 일치하면 0보다 큰 값의 확률을 가지며, 일치하지 않는 경우에는 0 확률을 갖게 된다. 그리고, 모든 이산형 값에 대한 확률을 더하면 확률 1의 값을 갖게 된다. 식(1.1)에 정의된 확률밀도함수는 영어로 probability mass function, discrete frequency function, probability function, discrete density function 등 다양하게 불리고 있다 (Mood 등, 1974).

이산형 확률변수 X 의 누가분포함수는 식 (1.3)과 같이 정의되며 식 (1.4)의 조건을 만족한다 (Montgomery 와 Runger, 1994).

$$F_x(x) = P[X \leq x] = \sum_{x_i \leq x} f_x(x_i) \quad (1.3)$$

$$0 \leq F_x(x) \leq 1 \quad (1.4a)$$

$$F_x(x) \leq F_x(y) \text{ if } x \leq y \quad (1.4b)$$

연속형 분포형인 경우의 확률밀도함수 $f_x(x)$ 는 다음 식 (1.5)를 만족한다.

$$f_x(x) \geq 0 \quad (1.5a)$$

$$\int_{-\infty}^{\infty} f_x(x) dx = 1 \quad (1.5b)$$

누가분포함수는 다음 식과 같이 주어진다.

$$F_x(x) = P[-\infty < X < x] = \int_{-\infty}^x f_x(z) dz \quad (1.6)$$

으로부터 다음의 관계식이 성립함을 알 수 있다.

$$f_x(x) = \frac{dF_x(x)}{dx} \quad (1.7)$$

이상에서 보는바와 같이 확률밀도함수는 0보다 크거나 같은 값을 갖고, 누가분포함수는 0과 1 사이의 값을 갖게 된다. 연속형 확률변수인 경우의 확률밀도함수와 누가분포함수의 관계를 그림 1에 나타내었다. 그림에서 임의의 값 x 까지의 확률밀도함수의 면적은 누가분포함수의 종기에 해당하는 값이 된다. 그러므로, 확률밀도함수 전체 면적은 1의 값을 갖게 되며 식 (1.5b)를 만족하게 된다.

[예제] $x > 0$ 이고 $\lambda > 0$ 일 때 임의의 함수 $f(x) = e^{-\lambda x}$ 가 확률밀도함수인가?

먼저 $f(x)$ 가 확률밀도함수가 되기 위해서는 식 (1.5)의 조건을 만족해야 한다. 임의의 함수 $f(x)$ 가 식 (1.5a)의 조건을 만족하는 것은 명백한 사실임을 알 수 있다. 다음으로 식 (1.5b)의 조건을 만족하는지 알아보기 위하여 임의의 함수 $f(x)$ 를 적분하면 다음과 같다.

$$\int_0^{\infty} e^{-\lambda x} dx = -\left(\frac{1}{\lambda}\right) \int_0^{\infty} -\lambda e^{-\lambda x} dx \quad (1.8)$$

여기서 $\int -\lambda e^{-\lambda x} dx = e^{-\lambda x}$ 이므로 식 (1.8)은 다음과 같이 다시 정리할 수 있다.

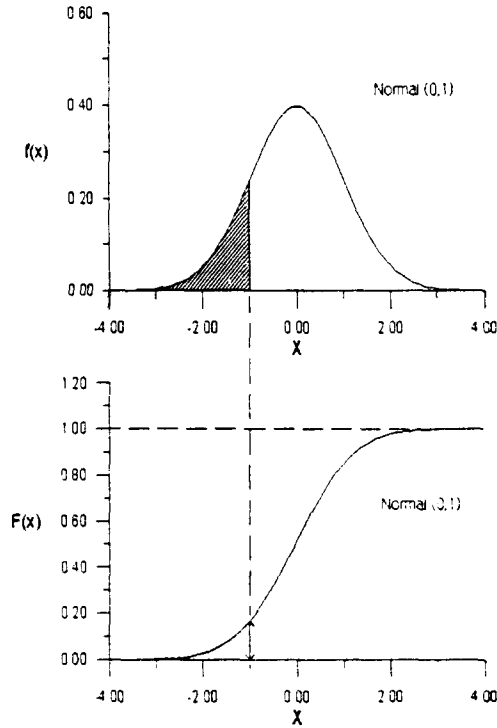


그림 1. 연속형 확률변수의 확률밀도함수와 누적분포함수의 관계도

$$\int_0^{\infty} e^{-\lambda x} dx = -\left(\frac{1}{\lambda}\right) \left[\lim_{x \rightarrow \infty} e^{-\lambda x} - \lim_{x \rightarrow 0} e^{-\lambda x} \right] = \frac{1}{\lambda} \neq 1 \quad (1.9)$$

식 (1.9)에서 보듯이 같이 함수 $f(x)$ 는 식 (1.5b)의 조건을 만족하지 못하므로 확률밀도함수가 아니며, 이 임의의 함수가 확률밀도함수가 되기 위해서는 다음과 같이 함수를 변형시키면 된다.

$$f(x) = \lambda e^{-\lambda x} \quad (1.10)$$

식 (1.10)에 정의된 함수는 매개변수 λ 를 갖는 지수분포함수(exponential distribution)임을 알 수 있다.

1.5 모멘트

모멘트(moment)는 모집단과 표본에 대한 모멘트로 나눌 수 있으며 다음과 같다.

1.5.1 모집단 모멘트(Population Moments)

모멘트

확률밀도함수 $f(x)$ 를 갖는 확률변수 X 의 함수 $g(X)$ 에 대한 r 차 모멘트는 다음과 같이 정의된다 (Mood 등, 1974).

$$E[g'(X)] = \int_{-\infty}^{\infty} g'(x)f(x)dx \quad (1.11)$$

또한 $g(X) = X$ 라고 가정하면 X 에 대한 r 차 모멘트는 다음 식과 같이 주어진다.

$$\mu_r' = E[X^r] = \int_{-\infty}^{\infty} x^r f(x)dx \quad (1.12)$$

식 (1.12)에서 $r=1$ 인 경우 1차 모멘트라고 하며 X 의 모집단 평균 (population mean)이 된다.

위에서 언급한 모멘트는 원점에 대한 모멘트(r^{th} moment about the origin or r^{th} raw moment)이며, $g(X)=X-\mu$ 라 가정하면 평균에 대한 r 차 모멘트(r^{th} moment about the mean or r^{th} central moment)는 다음 식과 같이 정의된다.

$$\mu_r = E[(X-\mu)^r] = \int_{-\infty}^{\infty} (x-\mu)^r f(x) dx \quad (1.13)$$

식 (1.13)에서 $r=1$ 인 경우 $\mu_1=0$ 가 되며, $r=2$ 인 경우 즉

$$\begin{aligned} \mu_2 &= E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx \\ &= \text{Var}(x) = \sigma^2 \end{aligned} \quad (1.14)$$

인 경우 평균에 대한 2차 모멘트로 확률변수 X 의 분산 (variance) 이다.

위에서 언급한 평균과 분산 이외에 수문학에서 많이 사용하는 통계치로는 변동계수(coefficient of variation), 왜곡도계수(coefficient of skewness), 첨예도계수(coefficient of kurtosis)등이 있으며 다음과 같이 정의된다.

$$\eta = \sigma/\mu \quad (1.15)$$

$$\gamma = \frac{E[(X-\mu)^3]}{\sigma^3} \quad (1.16)$$

$$\chi = \frac{E[(X-\mu)^4]}{\sigma^4} \quad (1.17)$$

대칭형 확률분포형인 경우에 왜곡도계수는 0의 값을 가지며, 정규분포의 첨예도계수는 3이므로 각 분포형의 첨예도계수의 값이 3보다 큰 값이나 작은 값이나에 따라 정규분포의 확률밀도함수보다 첨예한지 평평한지를 알 수 있다.

간단한 예로서, 3개의 매개변수를 갖는 Weibull 분포에 대해 모멘트를 구하는 과정은 아래와 같다. Weibull분포의 누가분포함수와 확률밀도함수는 다음과 같이 주어진다 (Weibull, 1951).

$$F(x) = 1 - \exp\left[-\left(\frac{x-x_0}{\alpha}\right)^\beta\right], \quad x_0 \leq x < \infty \quad (1.18)$$

$$f(x) = \frac{\beta}{\alpha} \left(\frac{x-x_0}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x-x_0}{\alpha}\right)^\beta\right] \quad (1.19)$$

여기서 x_0, α, β 는 각각 위치매개변수(location parameter), 규모매개변수(scale parameter), 형상매개변수(shape parameter)로 $\alpha > 0, \beta > 0$ 조건을 만족해야 한다. 원점에 대한 r 차 모멘트는 식 (1.12)로부터 다음과 같이 구할 수 있다.

$$\mu_r' = E[X^r] = \int_{x_0}^{\infty} X^r f(x) dx \quad (1.20)$$

$$= \int_{x_0}^{\infty} x^r \frac{\beta}{\alpha} \left(\frac{x-x_0}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x-x_0}{\alpha}\right)^\beta\right] dx$$

식 (1.20)에서 $r=1$ 인 경우 즉 평균은 $\mu_1' = x_0 + \alpha \Gamma(1+1/\beta)$ 이 된다.

확률가중모멘트(Probability weighted moments)

누가분포함수 $F(x)$ 를 갖는 확률변수 X 에 대한 확률가중모멘트(PWM)의 일반식은 식 (1.21)과 같이 주어진다(Greenwood 등, 1979; Landwehr 등, 1979).

$$M_{p,r,s} = E[X^p F^r(x) \{1-F(x)\}^s] \quad (1.21)$$

여기서 p, r, s 는 정수이고, $M_{p,0,0}$ 이면 원점에 대한 기존의 p 차 모멘트와 동일한 값을 나타낸다. 즉 $M_{p,0,0} = E[X^p] = \mu_p'$ 이므로 식 (1.12)와 같은 형태가 된다. 매개변수 추정에 있어서 특별히 관심의 대상이 되는 것은 $p=1, s$ 또는 r 이 0 또는 음이 아닌 정수일 때이다. 모집단의 확률가중모멘트(population PWM)는 식 (1.22), (1.23)과 같이 두 가지 형태로 나타낼 수 있다.

$$M_{1,r,0} = E[X F^r(x)] \equiv B_r, \quad p=1, r, s=0 \quad (1.22)$$

$$\begin{aligned} M_{1,0,s} &= E[X \{1-F(x)\}^s] \equiv B_s', \\ &p=1, r=0, s \end{aligned} \quad (1.23)$$

식 (1.22)와 (1.23)중 어느 확률가중모멘트를 사용할 것인가 선택은 선정된 분포형에 따라 매개변수 추정에 편리한 것을 사용한다. Lognormal, gamma, log-Pearson type III, GEV, Gumbel, log-Gumbel 분포는 식 (1.22)를 이용하고, Weibull, Wakeby 분포의 경우는 식 (1.23)을 이용하여 매개변수를 추정하는 것이 편리하다.

3개의 매개변수를 갖는 Weibull 분포의 모집단 확률가중모멘트를 구하는 경우에는 앞에서 설명한 바와 같이 식 (1.23)을 사용하는 것이 유도상 편리하므로, s차 확률가중모멘트를 구하기 위하여 식 (1.18)과 (1.19)를 식 (1.23)에 대입하면

$$B'_s = M_{1,0,s} = E\{X[1-F(x)]^s\} \\ = \int_{x_0}^{\infty} \{x[1-F(x)]^s\}f(x)dx \quad (1.24) \\ = \int_{x_0}^{\infty} x e^{-s(\frac{x-x_0}{\alpha})^\beta} \frac{\beta}{\alpha} \left(\frac{x-x_0}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x-x_0}{\alpha}\right)^\beta} dx$$

여기서 $y = (x-x_0)/\alpha$ 라고 가정하면 식 (1.24)는 다음과 같이 쓸 수 있다.

$$B'_s = \int_0^{\infty} (x_0 + \alpha y^{1/\beta}) e^{-(s+1)y^\beta} dy \\ = \frac{x_0}{s+1} + \frac{\alpha}{(s+1)^{1+1/\beta}} \Gamma(1+1/\beta) \quad (1.25)$$

여기서 $\Gamma(\cdot)$ 는 gamma 함수를 나타낸다. 식 (1.25)를 이용하여 3개의 매개변수를 갖는 Weibull 분포의 확률가중모멘트를 정리하면 다음과 같이 나타낼 수 있다.

$$B'_0 = x_0 + \alpha \Gamma(1+1/\beta) \quad (1.26a)$$

$$B'_1 = \frac{1}{2} [x_0 + 2^{-1/\beta} \alpha \Gamma(1+1/\beta)] \quad (1.26b)$$

$$B'_2 = \frac{1}{3} [x_0 + 3^{-1/\beta} \alpha \Gamma(1+1/\beta)] \quad (1.26c)$$

$$B'_3 = \frac{1}{4} [x_0 + 4^{-1/\beta} \alpha \Gamma(1+1/\beta)] \quad (1.26d)$$

L-모멘트(L-moment)

최근에 거론되고 있는 L-모멘트 방법에 의한 모집단의 추정량 (population L-moment) (Hosking, 1986; 1990)은 다음 식과 같이 주어진다.

$$\lambda_{N+1} = (-1)^N \sum_{k=0}^N p^*_{N,k} B'_k = \sum_{k=0}^N p^*_{N,k} B_k \quad (1.27)$$

여기서 B'_k 과 B_k 는 식 (1.22)와 (1.23)에 정의되어 있고

$$p^*_{N,k} = (-1)^{N-k} \binom{N}{K} \binom{N+k}{k} \quad (1.28)$$

참고적으로, 모집단의 L-모멘트와 확률가중모멘트의 관계는 다음과 같이 주어진다 (Hosking, 1986).

$$\lambda_1 = B'_0 = B_0 \quad (1.29a)$$

$$\lambda_2 = B'_0 - 2B'_1 = 2B_1 - B_0 \quad (1.29b)$$

$$\lambda_3 = B'_0 - 6B'_1 + 6B'_2 = 6B_2 - 6B_1 + B_0 \quad (1.29c)$$

$$\lambda_4 = B'_0 - 12B'_1 + 30B'_2 - 20B'_3 \\ = 20B_3 - 30B_2 + 12B_1 + B_0 \quad (1.29d)$$

위 관계식에서 보는 바와 같이 L-모멘트 추정량은 확률가중모멘트의 선형조합이며, 확률가중모멘트법에 의한 매개변수 추정치와 L-모멘트법에 의한 추정치는 동일한 값을 갖는다. 이러한 관계를 이용하여 3개의 매개변수를 갖는 Weibull 분포에 대한 L-모멘트를 구할 수 있다. 즉 식 (1.26)의 결과를 식 (1.29)에 대입하면 Weibull 분포의 모집단 L-모멘트는 다음과 같이 주어진다.

$$\lambda_1 = B'_0 = x_0 + \alpha \Gamma(1+1/\beta) \quad (1.30a)$$

$$\lambda_2 = B'_0 - 2B'_1 = (1-2^{-1/\beta})$$

$$\alpha \Gamma(1+1/\beta) \quad (1.30b)$$

$$\lambda_3 = B'_0 - 6B'_1 + 6B'_2 = (1-3 \cdot 2^{-1/\beta} + 2 \cdot 3^{-1/\beta}) \\ \alpha \Gamma(1+1/\beta) \quad (1.30c)$$

$$\lambda_4 = B'_0 - 12B'_1 + 30B'_2 - 20B'_3 = (1-6 \cdot 2^{-1/\beta} \\ + 10 \cdot 3^{-1/\beta} - 5 \cdot 4^{-1/\beta}) \alpha \Gamma(1+1/\beta) \quad (1.30d)$$

일반 모멘트와 마찬가지로, L-모멘트의 경우에도 L-변동계수(L-coefficient of variation), L-왜곡도계수(L-skewness), L-첨예도계수(L-kurtosis)를 가지며, 각각 다음 식과 같이 정의된다 (Hosking, 1986).

$$\tau_2 = \frac{\lambda_2}{\lambda_1} \quad (1.31)$$

$$\tau_3 = \frac{\lambda_3}{\lambda_2} \quad (1.32)$$

$$\tau_4 = \frac{\lambda_4}{\lambda_2} \quad (1.33)$$

여기서 $\lambda_1, \lambda_2, \lambda_3$ 와 λ_4 는 식 (1.30)에 정의되어 있으며, 참고적으로, λ_2, τ_3 와 τ_4 의 범위는 다음과 같이 주어진다(Hosking, 1990).

$$\lambda_2 > 0 \quad (1.34)$$

$$-1 < \tau_3 < 1 \quad (1.35)$$

$$\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 \leq 1 \quad (1.36)$$

1.5.2 표본 모멘트(Sample moments)

모멘트

표본자료 x_1, \dots, x_N 의 원점에 대한 r차 표본 모멘트는 다음과 같이 정의된다.

$$\bar{\mu}_r' = \frac{1}{N} \sum_{i=1}^N X_i^r \quad (1.37)$$

여기서 $r=1$ 일 때의 모멘트를 $\bar{\mu}_1'$ 또는 \bar{X} 로 표시하며 표본의 평균값이다. 이와 마찬가지로 평균에 대한 r차 표본 모멘트는 다음 식과 같이 주어진다.

$$\bar{\mu}_r = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^r \quad (1.38)$$

위 식 (1.37)과 (1.38)은 일반적인 통계해석에 가장 많이 쓰이는 식이다.

확률가중모멘트

식 (1.22)와 (1.23)에 정의된 모집단 확률가중 모멘트에 대한 각각의 불편 확률가중모멘트(unbiased PWM moments)는 다음 식과 같이 주어진다 (Hosking, 1986).

$$\hat{B}_r = \frac{1}{N} \sum_{j=1}^N x_j \frac{(j-1)(j-2)\cdots(j-r)}{(N-1)(N-2)\cdots(N-r)}, \quad r > 1 \quad (1.39)$$

$$\hat{B}_r' = \frac{1}{N} \sum_{j=1}^N x_j \frac{(N-1)!(N-s-1)}{(N-j-s)!(N-1)!}, \quad s \geq 0 \quad (1.40)$$

여기서 $x_1 \leq x_2 \leq \dots \leq x_N$ 인 순서 통계량이다. 식 (1.39)와 (1.40)에서 보는바와 같이 $r=0, s=0$ 인 경우에는 $\hat{B}_0 = \hat{B}_0' = \bar{X}$ 로 표본자료의 평균값과 같다. 참고적으로 식 (1.39)와 (1.40)은 서로 다른 가중치를 갖는다. 식 (1.39)의 경우 가장 큰 자료에 대해 가장 큰 가중치 (즉 $j=N$ 일 때 가중치는 1이다)를 곱하며, 값이 작은 자료에는 작은 가중치를 주는 것을 알 수 있다($j=1$ 일 때에는 가중치가 0이다). 식 (1.40)의 경우에는 이와 정반대의 가중치를 주게 된다. 그러나, 추정치 \hat{B}_r 과 \hat{B}_s' 간의 관계는 식 (1.29)와 같이 주어지므로, 추정치 \hat{B}_r 과 \hat{B}_s' 중 한가지를 알고 있는 경우에도 다른 추정량을 이용하여 확률가중모멘트법에 의한 매개변수 추정치를 구할 수 있다.

L-모멘트

r차 표본 L-모멘트는 다음 식과 같이 주어진다.

$$l_r = \binom{N}{r}^{-1} \sum_{1 \leq i_1} \sum_{i_1 \leq i_2 \leq \dots} \dots \sum_{\dots \leq i_{r-1} \leq N} r^{-1} \sum_{k=0}^{r-1} (-1)^k \binom{r-1}{k} x_{i_r - k} \quad (1.41)$$

앞에서 설명한 바와 같이 L-모멘트는 PWM의 선형조합이므로 이 관계식을 이용하여 r차 표본 L-모멘트를 다음과 같이 구할 수 있다.

$$l_r = \sum_{k=0}^{r-1} p_{r-1, k}^* \hat{B}_k = (-1)^{r-1} \sum_{k=0}^{r-1} p_{r-1, k}^* \hat{B}_k' \quad (1.42)$$

여기서 \hat{B}_k 와 \hat{B}_k' 는 식 (1.39)와 (1.40)에 주어

진 값을 사용하며 $p_{r-1, k}^*$ 는 식 (1.28)에서 구할 수 있다.

일반적으로 L-모멘트나 확률가중모멘트는 일반적인 모멘트보다 작은 자료수에 의한 영향을 덜 받으며, 안정적이고 (robustic), 간단한 장점을 가지고 있다.

1.6 비선형확률함수의 모멘트(Moment of nonlinear functions of random variables)

일반적으로 수자원 및 수문학에서 비선형 확률변수를 취급하게 되는 경우가 발생하게 된다. 확률변수 X로 주어진 확률함수에서 새롭게 정의되는 확률변수 Y의 확률함수를 찾게 되는 경우가 발생하게 되는데 ($Y=g(X)$), 이 경우 정확한 확률함수를 구할 수는 있지만 확률변수 X와 Y의 관계식이 비선형인 경우에는 이를 구하기가 어려운 경우가 종종 발생하게 된다. 이런 경우 확률변수 X의 평균 및 분산을 알고 있으면, 확률변수 Y의 평균 및 분산을 다음과 같이 근사적으로 구할 수 있다.

평균 μ 와 공분산 $Cov(X_i, X_j)$ 를 갖는 확률변수 X_1, X_2, \dots, X_n 의 함수를 $g(X_1, \dots, X_n)$ 라 하고 a_1, \dots, a_n 에 대해 Taylor급수 전개를 하면 다음과 같다.

$$g(X_1, \dots, X_n) \approx g(a_1, \dots, a_n) + \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i} \right) (X_i - a_i) \quad (1.43)$$

여기에 기대값을 취하면 다음 식과 같이 표시할 수 있다.

$$E[g(X_1, \dots, X_n)] \approx g(a_1, \dots, a_n) + \sum_{i=1}^n \left(\frac{\partial g}{\partial X_i} \right) [E(X_i) - a_i] \quad (1.44)$$

여기서 $a_i = \mu_i$ 라 하면 식 (1.44)는 다음과 같이 표현된다.

$$E[g(X_1, \dots, X_n)] \approx g(\mu_1, \dots, \mu_n) \quad (1.45)$$

또한 $g(X_1, \dots, X_n)$ 의 분산은 다음 식과 같이 주어진다.

$$\begin{aligned} \text{Var}[g(X_1, \dots, X_n)] &\approx \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial g}{\partial X_i} \right)_{\mu} \left(\frac{\partial g}{\partial X_j} \right)_{\mu} \\ &\text{Cov}(X_i, X_j) \end{aligned} \quad (1.46)$$

위에서 설명한 내용을 Manning공식을 이용한 수로 설계시 어떻게 응용되는가 살펴보자. 즉 수로 설계시 조도계수와 수로바닥경사가 수심 Y에 어떠한 영향을 미치는가 살펴보기로 한다. Manning공식은 다음과 같이 주어진다.

$$Q = \frac{1}{n} AR^{2/3} S^{1/2} \quad (1.47)$$

여기서, Q는 설계유량, n은 조도계수, R은 동수반경, S는 수로바닥경사이다.

우리가 수로를 설계할 때 Q, 단면형상, S 등의 값은 알고 있으며, 조도계수는 수로에 따라 임의의 값을 선택하게 된다. 그러므로 수로 설계시 구하는 값은 수심 Y가 된다. 그러나 수로 설계시 Manning공식의 일부 인자는 불확실성을 갖게 된다. 즉 조도계수와 수로바닥경사의 경우 정확한 값이 아니라 수로의 특성에 따라 변하기 때문에 불확실성을 내포하게 된다. 그러므로 문제는 과연 조도계수와 수로바닥경사의 불확실성이 수심 Y에 어떠한 영향을 미치는가 하는 것이다. 여기서 수심 Y가 조도계수 n과 수로바닥경사 S의 비선형함수 $Y=g(n, S)$ 라고 하면 식 (1.45)를 이용하여 다음과 같이 나타낼 수 있다.

$$Q = \frac{1}{E[n]} (AR^{2/3})_{\mu} \sqrt{E(S)} \quad (1.48)$$

여기서 $(AR^{2/3})_{\mu}$ 는 괄호 안의 값을 기대값으로 계산하는 것을 의미한다. 식 (1.48)로부터

$$(AR^{2/3})_{\mu} = \frac{QE(n)}{\sqrt{E(S)}} \quad (1.49)$$

로 주어지며 식 (1.49)의 해가 E[Y]가 되는 것이다.

조도계수와 수로바닥경사가 상관성이 없다고 가정하면 수심에 대한 분산은 식 (1.46)을 이용하여 다음과 같이 표시할 수 있다.

$$\text{Var}(Y) = \left(\frac{\partial Y}{\partial n}\right)^2 \text{Var}(n) + \left(\frac{\partial Y}{\partial S}\right)^2 \text{Var}(S) \quad (1.50)$$

또한 Manning 공식으로부터

$$n = \left(\frac{S^{1/2}}{Q}\right) AR^{2/3} \quad (1.51)$$

$$\begin{aligned} \frac{\partial n}{\partial Y} &= \left(\frac{S^{1/2}}{Q}\right) \frac{d(AR^{2/3})}{dY} \\ &= \left(\frac{S^{1/2}}{Q}\right) AR^{2/3} \left(\frac{2}{3R} \frac{dR}{dY} + \frac{1}{A} \frac{dA}{dY}\right) \end{aligned} \quad (1.52)$$

이고 $E(n) = \frac{[E(S)]^{1/2}}{Q} (AR^{2/3})_{\mu}$ 이므로

$$\left(\frac{\partial n}{\partial Y}\right)_{\mu} = E(n) \left(\frac{2}{3R} \frac{dR}{dY} + \frac{1}{A} \frac{dA}{dY}\right)_{\mu} \quad (1.53)$$

이와 마찬가지로 수로바닥경사에 대해서는 다음과 같이 나타낼 수 있다.

$$S^{1/2} = \frac{Qn}{AR^{2/3}} = Qn(A^{-1}R^{-2/3}) \quad (1.54)$$

$$\frac{\partial S}{\partial Y} = -2S \left(\frac{2}{3R} \frac{dR}{dY} + \frac{1}{A} \frac{dA}{dY}\right) \quad (1.55)$$

$$\left(\frac{\partial S}{\partial Y}\right)_{\mu} = -2E(S) \left(\frac{2}{3R} \frac{dR}{dY} + \frac{1}{A} \frac{dA}{dY}\right)_{\mu} \quad (1.56)$$

식 (1.53)과 (1.56)을 (1.50)에 대입하면 수심에 대한 분산은 다음 식과 같이 주어진다.

$$\text{Var}(Y) = \frac{\text{Var}(n)}{E(n)^2 \left(\frac{2}{3R} \frac{dR}{dY} + \frac{1}{A} \frac{dA}{dY}\right)_{\mu}^2}$$

$$+ \frac{(1/4)\text{Var}(S)}{E(S)^2 \left(\frac{2}{3R} \frac{dR}{dY} + \frac{1}{A} \frac{dA}{dY}\right)_{\mu}^2} \quad (1.57a)$$

또는

$$\text{Var}(Y) = \frac{CV(n)^2 + 0.25CV(S)^2}{\left(\frac{2}{3R} \frac{dR}{dY} + \frac{1}{A} \frac{dA}{dY}\right)_{\mu}^2} \quad (1.57b)$$

여기서 CV는 변동계수(coefficient of variation)를 의미하며, $[CV(Y)]^2 = \text{Var}(Y)/[E(Y)]^2$ 이다.

[예제] 하폭 B=15m, E(S)=0.012, CV(S)=0.05, E(n)=0.03, CV(n)=0.1의 값을 갖는 직사각형 수로에서 설계유량 Q=120m³/s이라고 가정하자. 이 수로에서 E(Y)와 CV(Y)은 얼마인가. 또 수심 Y가 정규분포를 갖는다면 여유고(free board)는 얼마나 될 것인가.

직사각형 단면에서 A=BY, R=BY/(B+2Y) 이므로 dA/dY=B, dR/dY=[B/(B+2Y)]².

식 (1.49)로부터

$$(AR^{2/3})_{\mu} = \frac{QE(n)}{\sqrt{E(S)}} = \frac{120 \times 0.03}{\sqrt{0.012}} = 32.86 \quad (1.58)$$

$$E(A) = BE(Y) = 15E(Y) \quad (1.59a)$$

$$E(R) = \frac{BE(Y)}{B+2E(Y)} = \frac{15E(Y)}{15+2E(Y)} \quad (1.59b)$$

이므로 식 (1.59)를 식 (1.58)의 좌변에 대입하여 풀면 E(Y)=1.74(m)가 된다. 또한

$$\begin{aligned} \left(\frac{2}{3R} \frac{dR}{dY} + \frac{1}{A} \frac{dA}{dY}\right)_{\mu} &= \left(\frac{5B+6Y}{3Y(B+2Y)}\right)_{\mu} \\ &= \frac{5 \times 15 + 6 \times 1.74}{3 \times 1.74 \times (15 + 2 \times 1.74)} = 0.886 \end{aligned}$$

이므로 수심 Y에 대한 분산은 식 (1.57)로부터 다음과 같이 주어지게 된다.

$$\text{Var}(Y) = \frac{(0.1)^2 + 0.25(0.05)^2}{(0.886)^2} = 0.0135(\text{m}^2)$$

또는 $\sigma(Y) = 0.116(\text{m})$

그러므로, $\text{CV}(Y) = \sigma(Y)/E[Y] = 0.067$ 이 된다. 여기서 여유고를 표준편차의 2배 ($0.12 \times 2 = 0.24\text{m}$)라고 하고 수심 Y 가 정규분포라고 하면 월류될 확률은 2.3%가 된다.

참 고 문 헌

- 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*, Vol. 15, No. 5, pp. 1049~1054.
- Hoksing, J. R. M.(1986). "The theory of probability weightd moments.", *Research Report RC 12210, IBM T. J. Weatson Research Center, Yorktown Height, New York*.
- Hosking, J.R.(1990). "*L-moments Analysis and estimation of distribution using linear combinations of order statistics.*", *Journal of Royal Statistical Society*, Vol. 52, No. 1, pp. 105~124.
- Landwehr J. M., Matalas N. C., and Wallis J. R. (1979a). "Probability weighted moments compared with same traditional techniques in estimating Gumbel parameters and quantiles.", *Water Resources Research*, Vol. 15, No. 5, pp. 1055~1064.
- Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979b). "Estimation of parameters and quantiles of Wakeby distributions: 1. Known lower bounds.", *Water Resources Research*, Vol. 15, No. 6, pp. 1361~1372.
- Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979c). "Estimation of parameters and quantiles of Wakeby distributions: 2. Unknown lower bounds.", *Water Resources Research*, Vol. 15, No. 6, pp. 1373~1379.
- Montgomery, D. C. and Runger, G. C. (1994). *Applied Statistics and Probability for Engineers*. John Wiley, New York.
- Mood, A. M., Graybill, F. A., and Boes, D. C. (1974). *Introduction to the Theory of Statistics*, McGraw-Hill.
- Weibull, W.(1951). "Statistical distribution function of wide application.", *Jour. Appl. Mech.*, ASME, pp. 293~297.