



‘Non-stationarity’와 ‘Non-homogeneity’에 대한 짧은 의견



김 상 옥
강원대학교 토목공학과 교수
sukim70@kangwon.ac.kr



이 상 호
부경대학교 토목공학과 교수
peterlee@pknu.ac.kr

‘stationarity’가 가지고 있는 수학적 정의와는 다소 거리가 있는 것으로 보인다. 특히 ‘homogeneity’나 ‘non-homogeneity’와 같은 용어가 혼용되는 경우 내가 연구하려는 방향이 이 네 가지 용어 중 어떤 것 하나를 선택해야 하는 것인지 아니면 조합하여 선택해야 하는 것인지 혼란스러운 경우가 많다.

따라서 이 글에서는 이 네 가지 용어의 간략한 정의를 살펴봄으로써, 기후변화 또는 수문변화(hydrologic change)와 관련된 연구를 진행함에 있어 가장 적합한 용어를 선정하는 데 조금이나마 도움을 제공하고자 한다.

1. 들어가며

논문이나 공식적인 문서를 작성함에 있어 해당 분야에서 사용되고 있는 정확한 용어를 사용해야 하는 것은 너무나 당연한 일이므로, 공부의 시작은 용어를 익히는 것으로부터 시작된다고 해도 과언은 아닌 것 같다. 그러나 가끔(사실 그보다는 자주) 내가 사용하고 있는 용어가 옳은 지 그른 지를 판단하기 어려운 경우가 생기기 마련이고, 저자의 경우 최근 수공학 분야에서 흔히 접하게 되는 ‘stationarity(정상성)’와 ‘non-stationarity(비정상성)’라는 용어가 그 중 하나이다. 이 용어들은 특히 Milly et al.(2008)이 science에 게재한 “Stationarity is Dead: Whiter Water Management?” 이후 각종 기후변화 관련 논문의 제목에서 매우 사랑받고 있는데, 이들 논문의 대부분에서 이 용어가 나타내는 바는 주로 어떤 자료의 ‘change’를 대변하고 있는 것으로 보이며,

2. ‘stationarity’와 ‘non-stationarity’

정상성과 비정상성에 관한 수학적 정의는 Bendat and Piersol(2000)에서 찾아볼 수 있다. 특히 정상적이지 않은 모든 자료는 비정상적이므로 사실상 비정상성에 대한 수학적 정의보다는 정상성에 대한 수학적 정의를 명확히 이해하는 것이 중요하다.

우리가 가지고 있는 자료는 무작위적 현상을 나타내는 특정 표본함수(sample function) 중 하나이며, 따라서 주어진 자료는 발생 가능한 표본함수의 집합인 무작위과정(random process) 또는 추계적 과정(stochastic process)으로부터 추출된 것으로 볼 수 있다.

{ $x(t)$ }로 나타낼 수 있는 어떤 무작위 과정을 고려할 때(여기서, 기호 { }는 표본함수의 앙상블



(ensemble)을 나타냄, 특정 시간 t_1 에서의 평균과 $t_1 \sim t_1 + \tau$ 에서의 자기상관함수(autocorrelation function) $R_{xx}(t_1, t_1 + \tau)$ 는 식 (1) 및 식 (2)와 같다.

$$\mu_x(t_1) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x_k(t_1) \quad (\text{식 1})$$

$$R_{xx}(t_1, t_1 + \tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x_k(t_1)x_k(t_1 + \tau) \quad (\text{식 2})$$

정상성은 '약한 정상성(weak stationarity)'과 '강한 정상성(strong stationarity)'으로 구분해 볼 수 있다. 넓은 의미에서 자료가 정상적이라는 것은 자료가 약한 정상성을 가지고 있다는 것과 동일한 의미로 식 (1)과 식 (2)를 통해 계산된 $\mu_x(t_1)$ 과 $R_{xx}(t_1, t_1 + \tau)$ 이 t_1 이 변화함에 따라 변화하지 않는 경우에 해당된다. 즉 약한 정상성을 가지는 경우 평균은 시간에 상관없이 상수가 되고 자기상관함수는 시간의 차이에 의해서만 결정되어 다음의 식 (3)이 성립된다.

$$\mu_x(t_1) = \mu_x \ \& \ R_{xx}(t_1, t_1 + \tau) = R_{xx}(\tau) \quad (\text{식 3})$$

또한 좁은 의미에서 자료가 정상적이라는 것은 자료가 강한 정상성을 가지고 있다는 것과 동일한 의미로 $\mu_x(t_1)$ 과 $R_{xx}(t_1, t_1 + \tau)$ 이외에도 무작위 과정 $\{x(t)\}$ 의 고차 모멘트들도 t_1 이 변화함에 따라 변화되지 않는 경우를 말한다. 그러나 실제적인 문제에서는 약한 정상성의 검증(verification)만으로도 강한 정상성의 의미를 포함하는 경우가 많으므로, 일반적으로는 약한 정상성만을 검토하는 것만으로도 자료가 정상적이라고 표현한다. 위와 같이 정상성에 대한 검토는 주어진 자료의 앙상블 평균된 특성을 계산함으로써 판별이 가능하며, 이를 보다 일반적으로 표현하면 강한 정상성은 무작위 과정에 의해 발생된 앙상블이 같은 확률분포함수로부터 추출된 경우이고, 약한 정상성은 1차 및 2차 모

멘트들만 일치되는 경우를 말한다.

이와 같은 과정을 통해 정상성을 가지지 않은 경우 자료가 비정상적이라고 말할 수 있으며, 비정상적 자료는 각각 개별적인 특정 결정적(deterministic) 과정에 의해 추출된 자료이므로, 그 과정 자체를 무작위 과정으로만 표현하기는 불가능하다. $\{x(t)\}$ 가 비정상적 무작위 과정으로부터 파생되었다고 할 때, $x(t) = a(t)u(t)$ 가 대표적인 사례에 해당된다. 여기서 $u(t)$ 는 정상적 무작위 과정을 나타내는 표본함수이며, $a(t)$ 는 이 비정상적 무작위 과정을 지배하는 특정한 결정적 함수이다. 따라서 비정상적 무작위 과정을 다루기 위해서는 비정상성을 지배하는 특정 결정적 함수를 결정해야 한다는 매우 어려운 특성을 가지고 있다.

3. 'homogeneity'와 'non-homogeneity'

'homogeneity'와 'non-homogeneity'에 대해서는 그 적용분야에 따라 다양하게 사용되고 있다. 화학이나 재료관련 분야에서는 주로 '균질 또는 균일'과 '비균질 또는 비균일'로 번역되는 경우가 많으며, 수학에서는 '제차'와 '비제차'로 번역되어 사용된다. 수문학 분야에서는 어떻게 사용되고 있는지 선뜻 떠오르지 않는다. 다만 통계학분야에서는 이를 주로 '동질'과 '비동질'로 표현하고 있어, 통계학의 영향을 많이 받은 수문통계학 분야에서도 두 용어를 '동질'과 '비동질'로 사용하면 어느 정도 학문간 용어의 혼란을 방지할 수는 있을 것으로 보이므로 이 글에서도 같은 용어를 사용하였다.

'동질'과 '비동질'의 여부는 '정상성'과 '비정상성'의 개념보다 쉽게 설명할 수 있다. 이를 위해 포아송 과정(Poisson process)를 예로 들어보고자 한다. 포아송 과정은 큐잉(queueing) 이론에서 사용되는 가장 중요한 무작위 과정 중 하나로 배가 도착한다든지, 손님이 방문한다든지 하는 등의 무

작위 과정을 표출할 때 사용된다. 통계수문학 분야에서 가뭄이나 홍수의 발생을 포아송 과정을 사용하여 표현하는 경우가 많다.

포아송 과정은 주어진 구간에서 어떤 사건의 발생횟수가 포아송 분포를 따르며 겹치지 않는 구간에서 그 사건의 발생횟수가 서로 독립적일때의 확률과정을 나타낸 것이다. 즉 t 시간 동안에 어떤 사건이 n 회 발생할 확률은 포아송 과정에 따라 식 (4)와 같이 표현될 수 있다.

$$P\{N(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, n = 0, 1, 2, \dots \text{(식 4)}$$

여기서, λ 는 포아송 분포의 모수(parameter)로 어떤 사건의 특정 시간동안의 출현강도(arrival intensity)를 의미하며, 특히 포아송 과정에서 사건발생의 간격은 λ 를 모수로 하는 지수분포(exponential distribution)로 표현되는 특징이 있다.

위에서 설명한 포아송 과정을 보다 세밀히 구분하면 ‘동질 포아송 과정’과 ‘비동질 포아송 과정’으로 구분할 수 있는데, 앞에 설명한 λ 가 시간에 따라 변화되지 않는 상수로 취급되면 이를 ‘동질 포아송 과정’이라 하고 λ 가 시간에 따라 변화되어 $\lambda(t)$ 인 경우를 ‘비동질 포아송 과정’이라 한다.

즉 이로부터 특정자료를 나타내는 분포의 모수가 상수로 취급되는 경우는 ‘동질’한 분포로부터 추출된 자료라 할 수 있으며, 분포의 모수가 시간에 따라 변화되는 특정 함수로 주어지는 경우는 ‘비동질’한 분포로부터 추출된 자료라 표현할 수 있다.

4. 용어의 응용 및 수문분야에서의 적용

앞서 언급한 용어들을 아주 간단히 정리해보면 정상성은 확률분포함수가 시간에 따라 변화하지 않는 경우이고, 동질성은 확률분포함수의 모수가 시간에 따라 변화하지 않는 상수인 경우이며, 이들

이 변화하면 비정상성 및 비동질성이다. 이들을 조합하면 정상 동질, 정상 비동질, 비정상 동질, 비정상 비동질의 4가지 경우가 가능하다. 그러나 이 중 비정상 동질은 분포가 시간에 따라 변화되는 경우 모수의 종류도 함께 변화되므로 이론적으로 불가능하기 때문에 이를 제외한 3가지 경우만 가능하다.

이제 이 용어들을 사용해서 간단한 사례를 만들어 보자. 여러분들이 시간의 길이 100인 10개 세트의 양상블을 $N(10, 25)$ 인 정규분포로부터 추출하여 구성하였다면, 이 자료는 같은 분포 및 일정한 평균 및 분산으로부터 추출된 자료이므로 정상 동질성을 가진다고 할 수 있다. 또한 길이를 반으로 나누어 50개는 $N(10, 25)$ 으로부터 추출하고 나머지 50개는 $N(20, 25)$ 으로부터 추출했다면 정규분포의 모수가 시간에 따라 변화된 것이므로 이 자료는 정상 비동질성을 가진다고 할 수 있을 것이다. 마지막으로 50개는 $N(10, 25)$ 으로부터, 그리고 나머지 50개는 $LN(20, 25)$ 으로부터 추출하였다면 모든 것이 시간에 따라 변화되었으므로 이 자료는 비정상 비동질한 특성을 가지고 있다고 볼 수 있다. 그러나 이와 같은 비정상 비동질 특성은 앞서 언급한 바와 같이 엄격한 의미에서 비정상성은 모든 변화량을 특정 결정적 또는 물리적 모형을 사용해야 하므로 사실상 무작위 과정이 아니라고 볼 수 있으므로, 모멘트들이 의미가 없고 분석도 사실 불가능하다고 생각된다. 또한 반대로 특정 기간 동안의 강우량이나 유량자료 등이 주어졌을 때, 이를 어떤 범위에까지 통계적으로 또한 합리적으로 구분할 수 있는 지에 따라 주어진 자료의 특성도 찾아낼 수 있을 것이다.

이제 수문통계학 분야에 대해 간단히 언급해보고자 한다. 최근 ‘비정상성’이라는 용어가 사용된 빈도분석 또는 경향성 분석 등과 관련된 수문통계학적 논문을 살펴보면 사실 엄격한 의미에서 비정상 비동질한 자료를 활용하거나 분석했다기 보다는 정상 비동질성(즉, 시간에 따라 일정한 분포를 가지지만 그 분포의 모수가 시간에 따라 변화하는 확률분



포모형)에 관한 논문들이 대부분이다. 이 같은 용어의 잘못된 사용을 지적하는 논문(Koutsoyiannis, 2011)도 게재되어 있어 향후 용어의 정리가 필요한 대목이다. 이 논문의 시작이 "The start of wisdom is the visit of name."을 의미하는 희랍어 문구로 시작된다. 논문의 결론은 non-stationarity \neq change로 요약될 수 있을 것 같다.


사실 수문분야에서 강우량의 변화나 유량의 변화 등에 관한 논문은 최근 기후변화로 인한 주제 이외에도 예전부터 많이 다루어지던 주제이다. 단, 이때는 'non-stationary'가 관습적인 사랑을 받기 이전이어서 오히려 용어의 선정에 있어 정확했다. 예를 들어, Yevjevich and Jeng(1969)이 작성한 "Properties of non-homogeneous hydrologic series"에서는 인간의 개입으로 인한 유량의 변화나 기후로 인한 유량의 변화된 자료의 통계적 특성 변화를 'non-homogeneity'라고 정의한 후 기술하고 있어, 어떤 정의도 없는 'non-stationary'라는 용어의 사용보다 혼란이 없다. 또 다른 예로써 수문기상자료의 경향성 분석 특히 통계적 특성이 변화되는 변동점을 찾아내고 이로부터 시계열 자료를 구분함으로써 동질한 특성을 가지는 구간으로 나누는 작업을 기상학분야에서는 'detecting inhomogeneity'라 하며, 이를 위한 다양한 동질화 방법을 다룬 논문들이 많다 (Domonkos, 2011).

5. 나가며

과거에 누군가가 어떤 용어를 사용했다고 해서 또 현재 다른 분야에서 어떤 용어를 사용하고 있다고 해서 반드시 그 용어를 따라야 함을 언급하고자 글을 작성한 것은 아니다. 다만 같은 특성을 가진 현상을 표현함에 있어 다른 용어를 사용하고자 한다면, 그 이유는 명백하게 정의한 후 그 용어를 사용하는 것이 바람직하다는 것을 언급하고 싶었다. 글을 작성하면서 저자가 용어를 잘못 이해하고 있는 것은 아닌가 하는 의구심을 계속 가지게 되었고, 글을 마치는 지금도 그렇다. 어쩌면 이 글을 쓰기 위해 찾아 읽어 본 몇 가지의 글들 덕분에 오히려 더 복잡해졌는지도 모르겠다.

글을 접으면서 스스로에게 질문을 던져 본다. 기후가 바뀌어 정상성이 죽었다 하는데, 죽은 것이 정말 정상성인가? 정상성이 맞다면 비정상성이라는 것이 확률통계학적으로 분석 가능한 대상이 된다고 생각하는 것인가? 죽은 것이 동질성이라면 수문통계학자들은 그것이 죽을 수 있음을 아주 오래전부터 알고 있었던 것 아닌가? 이로부터 정말 새로운 틀이 필요한 것인지 아니면 연구 분야에만 머물러 있던 비동질성에 관한 문제를 실무에 적용하도록 하면 되는 것인지? 지속적인 고민과 연구 및 정리가 필요한 부분이라 생각된다.

감사의 글

본 기고문은 국토교통부 물관리연구사업의 연구비지원(11기술혁신C06)의 일환으로 작성되었습니다. 

● 참고문헌

1. Bendat, and J.S., Piersol, A.G. (2000). *Random data: analysis and measurement procedures: 3rd edition*, John Wiley & Sons, New York.
2. Domonkos, P. (2011). "Efficiency evaluation for detecting inhomogeneities by objective homonisation methods." *Theor. Appl. Climatol.* Vol. 105, pp. 455-467.
3. Koutsoyiannis, D. (2011). "Hurst-Kolmogorov dynamics and uncertainty." *Journal of the American Water Resources Association*, Vol. 47, No. 3, pp. 481-495.
4. Milly, P.C.D., Betancourt, J., Falkenmark, M., Hirsch, R.M., Kundzewicz, Z.W., Lettenmaier, D.P., and Stouffer, R.J. (2008). "Stationary is Dead: Whiter Water Management?" *Science*, 319, pp. 573-574.
5. Yevjevich, V., and Jeng, R.I. (1969). *Properties of non-homogeneous hydrologic series*. Hydrology papers No. 32, Colorado State University.