

전염병 확산과 통계물리학

2020년 9월 11일

고광일



코로나 바이러스^{COVID-19}의 시대이다. 가장 시급한 과제는 급한 불을 끄는 것이다. 이미 전 세계에서 수많은 과학자와 제약회사가 치료제와 백신 개발에 힘쓰고, 정부와 시민 모두 지역확산을 최소화하기 위해 함께 노력하고 있다. 그러나 한편 돌아해보면 새로운 변종 바이러스 및 세균에 의한 감염병의 유행은 정도의 차이가 있을지언정 최근까지도 거의 주기적으로 반복되었던 것 같다. 왜 이런 일이 일어난 것일까? 이번 유행은 무엇이 특별한가? 모두가 한 번쯤은 던져본 질문일 것이다.

전 세계의 과학자들도 마찬가지인 것 같다. 전 세계 다양한 분야의 연구자들이 코로나19의 유행에 자극을 받아 다양한 질문을 던지고 나름의 답을 담은 연구물을 그야말로 쏟아내고 있다. 2020년 1월부터 9월 초 현재까지 1년이 채 안 되는 기간 동안 의학 및 생물학 분야의 출판전논문^{preprint}의 인터넷 저장소인 medRxiv.org와 bioRxiv.org에 투고된 코로나19 관련 연구물은 도합 8500편이 넘는다.¹ 흥미로운 것은 얼핏 전통적인 감염병, 전염병 연구 분야와 직접적인 관련이 없어 보이는 물리학 및 수학 중심의 출판전논문 인터넷 저장소인 arXiv.org에 투고된 코로나19 관련 연구물의 수도 9월 초 현재 2200편에 육박한다는 점이다.² 물론 물리학자나 수학자가 직접 바이러스의 분자적 기작을 연구하고 백신을 개발하거나 개선된 임상적 조치를 보고하는 것은 아닐 것이다. 그러나 우리 모두가 던져봤을 법한 앞서 질문과 같은 의문은 물리학적이고 수학적인 탐구의 대상이 된다.

¹ 이 숫자는 계속 업데이트되고 있으며 [링크된 사이트](#)에서 확인할 수 있다.

² 이 숫자 역시 계속 업데이트되고 있으며 [링크된 사이트](#)에서 확인할 수 있다.

역사적으로 전염병 확산에 대한 최초의 수리적 연구는 천연두 확산과 예방접종의 효과에 대한 다니엘 베르누이 Daniel Bernoulli의 1766년 논문³으로 여겨진다. 이는 백신 vaccine이라는 용어의 기원이 된 에드워드 제너 Edward Jenner의 우두법 발견(1796년 논문 발표)과 동시대에 이뤄졌으며 바이러스의 발견보다 한 세기 이상 앞선 업적이다. 베르누이로부터 비롯한 전염병 확산에 대한 수리적 연구는 수리생물학 mathematical biology 또는 이론생물학 theoretical biology에서 여전히 활발히 연구되는 방대한 분야를 이룬다. 전염병 확산에 대한 수리생물학적 연구에 대한 더욱 자세한 역사와 소개는 해당 분야 전문가에게 맡기는 것이 도리일 것이다. 여기에서는 관심의 범위를 좁혀 전염병 확산과 통계물리학의 관계에 대해 간략하게 소개하고자 한다.

전염병 확산은 근본적으로 다체계 多體系; many-body system 문제이다. 작게는 특정 지역에서 넓게는 지구 전체에 이르는 영역 내에서 수많은 개개인이 서로 이동하며 직간접적으로 접촉함으로써 전염병은 전파된다. 또한 전염병 확산은 근본적으로 확률적인 probabilistic 또는 stochastic 문제이다. 전염 과정(바이러스의 전달)에 관여하는 요소는 너무도 많고 복잡해서 우리는 이 과정을 확률적으로 파악하는 것이 최선이다. (물론 우리는 이 확률을 어느 정도 적극적으로 조정할 수 있다. 다른 사람과 접촉할 때는 마스크를 쓰는 것이 좋을 것이다.) 통계물리학이 다루는 문제도 마찬가지다. 통계물리학은 서로 상호작용하는 매우 많은 수의 입자 무리를 다룬다. 이 문제가 흔히 접하는 전형적인 물리 문제와 어떻게 다른지 생각해 보자.



그림1 다니엘 베르누이

가장 전형적인 물리 문제 중 하나가 단단한 두 구슬의 충돌 문제이다. 충돌 전후 구슬의 궤적 문제는 구슬치기나 당구 시합을 해본 사람이라면 본능적으로, 그리고 고등학교 수준 이상의 물리 지식을 가진 사람이라면 수학적으로도 어느 정도 접근할 수 있는 문제이다. 그런데 구슬이 두 개가 아니라 열 개, 백 개, 아니 그 이상이라면? 게다가 구슬 사이에서 끌어당기거나 밀치는 것과 같은 힘이 작용한다면? 이제 문제는 명백히 다체계 문제가 되었다. 또한 구슬의 충돌 과정과 궤적은 너무도 복잡해져서 이를 가장 효율적으로 다루는 방법은 확률을 도입하는 것이다. 이렇게 확률을 도구로 사용하여 다체계를 기술하고 그 성질을 분석하는 물리학의 이론들을 통계역학^{statistical mechanics}이라 한다. 통계물리학은 통계역학 이론을 기반으로 다양한 다체계의 물리 현상을 연구하는 분야이다. 이러한 공통점을 매개로 전염병 확산은 통계물리학의 대상이 된다.

통계물리학은 전염병 확산에 대해 수리생물학 등 여타 학문분야에 비해 어떤 새로운 것을 알려줄까? 좀 더 직설적으로 묻자면, 통계물리학은 전염병 확산에 대해 무엇이 관심 있을까? 모든 학문 분야는 그 나름의 관점과 도구로 현상을 파악하고 분석하고 진단한다. 통계물리학의 주된 관심 중 하나는 '상전이^{相轉移}'이다. 다시 충돌하는 구슬 문제로 돌아가 보자. 그다지 실재할 것 같지 않은 서로 충돌하는 무수히 많은 구슬의 문제는 사실 기체의 모형^{model}으로 생각할 수 있다. 우리가 중고등학교에서 배우는 '이상기체'를 떠올리면 도움이 될 것이다. 기체를 이루는 분자는 무수히 많고(이 개수가 대략 아보가드로^{Amedeo Avogadro}의 수⁴, 즉 약 6×10^{23} 개 정도임을 우리는 배웠다) 매우 빠르고 복잡하게 움직이고 있어서 그 궤도를 정확히 기술하는 것이 실제로 불가능하다.

이러한 기체의 상태를 우리는 압력, 부피, 온도 등의 거시적^{macroscopic} 변수를 도입하여 표현한다. 거시적 변수란 기체를 이루는 분자 하나하나의 상태가 아니라 분자들의 모임으로서의 기체의 집단적 상태의 성질을 기술하는 변수를 말한다. 집단적 상태를 나타내는 변수들은 자연스럽게 집단의 통계적인 성질로서 표현된다. (그러한 이유로 통계역학, 통계물리학이라는 용어가 채택되었을 것이다.) 집단적 상태는 외부 조건에 따라 변화한다. 기체를 예로 들면, 열을 가하면(온도를 높이면) 팽창하고(부피가 증가하고) 반대로 열을 빼앗으면(온도를 낮추면) 수축한다.

⁴ 아보가드로 수와 관련해 HORIZON의 기존 기사 "[아보가드로 수-물질세계의 불연속성에 대하여](#)"를 참고할 수 있다.

⁵ 이러한 노력은 켄 윌슨^{Kenneth G. Wilson}의 1982년 노벨물리학상 수상으로 하나의 결실을 맺었다.

⁶ 이 전염과정 모형은 1차원의 접촉 과정^{contact process}라 불린다.

⁷ 감염자 비율이 0이라는 것의 의미는 아무도 감염된 적이 없다는 것이 아니라, 모두 치유된 상태라는 것이다. 접촉 과정 모형에서는, 다행스럽게도, 아무도 병으로 사망하지 않는다.

⁸ 여기서 제거란 감염 과정에 더 이상 참여하지 않는 것을 의미하며, 면역 상태와 사망 상태를 모두 포함한다.

이러한 변화는 대부분의 경우 연속적이고 점진적이다. 물론 집단적 상태의 변화가 때로는 매우 급격하게 일어나기도 한다. 기체의 온도를 끓는 점보다 더 낮추면 액체로 변화하는 것, 또는 온도를 더욱 낮춰 어느점 아래까지 낮추면 고체로 변화하는 것을 예로 들 수 있다. 이러한 집단적 상태의 급격한 변화를 상전이라고 한다. 하나의 상^相: phase에서 다른

상으로의 전이는 상호작용하는 다체계의 고유의 성질이다. 상전이에 대한 이론적인 이해는 20세기 통계물리학이 이루고자 한 가장 주요한 과제였다.⁵

전염병의 확산에 상전이의 관점이 적용될 수 있을까? 비현실적이지만 다음과 같이 단순한 전염과정 모형을 생각해보자. 학생들이 살고 있는 기숙사에 전염병이 돌고 있다. 기숙사는 한 층으로 되어있고 일렬로 붙어있는 각 방에 학생들이 살고 있다. 전염병은 전염성이 그리 강하지 않아, 감염된 사람은 바로 옆 방 학생에게만 일정한 감염율로 전염시키고, 동시에 감염된 사람들은 일정한 치유율로 자연치유된다.⁶ 이러한 전염 과정에서 나타나는 집단적 상태는 무엇일까? 먼저 감염율이 치유율에 비해 충분히 낮다면 전염병이 기숙사 전반에 퍼지지 못하고 자연소멸할 것이다. 이를 전염병 소멸 상태(disease-free state)라고 부르자. 반대로 감염율이 치유율에 비해 충분히 높다면 전염병이 기숙사 전반으로 확산될 것이다. 전염병 확산 상태(epidemic state)이다. 이 두 가지 집단적 상태를 전염 과정의 두 가지 상으로 생각할 수 있고 이 두 가지 상 사이의 변화가 상전이가 된다.

이 두 가지 상은 어떤 면에서 다른 상일까? 물론 두 상태는 감염자 비율로 쉽게 구분된다. 전염병 확산 상태는 이 비율이 0보다 큰 반면, 전염병 소멸 상태는 이 비율이 0이다.⁷ 한편 전염병 확산 상태는 '무언가가 일어나고 있는 상태'인 반면에 전염병이 소멸된 상태는 '아무 일도 일어나지 않는 상태'이다. 이런 이유로 통계물리학에서는 전자를 활성상(active phase), 후자를 비활성상(inactive phase)이라고 부른다. 후자는 한번 그 상태에 들어가면 다시 빠져나오지 못한다는 의미에서 흡수상(absorbing phase)이라고도 부른다.

또 하나의 고전적인 전염병 확산 모형은 1927년 윌리엄 커맥(William O. Kermack)과 앤더슨 맥켄드릭(Anderson G. McKendrick)이 제안한 모형으로까지 거슬러 올라가는 소위 SIR(susceptible-infected-removed) 모형이다. 이들은 집단을 취약군(S; susceptible), 감염군(I; infected), 그리고 제거군(R; removed)의 세 구획(compartment)으로 나누고 각 구획의 비율이 주어진 감염율과 제거율⁸에 따라 어떻게 변할지 탐구할 것을 제안하였다. 가장 간단한 SIR 유형의 전염과정은 다음과 같다. 감염군은 취약군을 만나면 일정한 감염율로 전염시킨다. 감염군은 일정한 제거율로 자발적으로 제거군으로 변화한다. 감염군의 비율이 0이되면 전염 과정은 종료된다. SIR 모형에서 최종적인 제거군의 비율은 전염 확산의 전체 과정 중 일어난 전체 전염 비율과 같으며 이런 이유로 전염병 확산의 척도로 이용할 수 있다. 최종적인 제거군의 비율을 알면 전염병 확산이 집단 전반에 걸친 것인지 아닌지 판정할 수 있을 것이다. 또한 이 비율이 감염율과 제거율과 같은 전염병의 특성은 물론 접촉 방식 및 빈도 등 집단의 특성과 어떤 관계가 있는지 알면 전염병 통제에도 도움이 될 것이다.

//

또 하나의 고전적인 전염병 확산 모형은,

1927년 윌리엄 커맥과 앤더슨 맥켄드릭이 제안한 모형으로까지 거슬러 올라가는 소위 SIR(susceptible-infected-removed) 모형이다.

//

최종적인 제거군의 비율은 어떻게 알 수 있을까? 1980년대 초반 페터 그라스버거^{Peter Grassberger}를 위시한 일련의 통계물리학자들은 이 문제에 흥미로운 해법을 제시하였다. 이 해법의 세부 사항은 약간의 수학적 기술을 필요로 하지만 기본적인 착상의 핵심은 이해하기 어렵지 않다. 이를 위해 단 한 명의 감염자(0번 환자^{patient zero})로부터 비롯한 전염병이 앞서 제시한 가장 간단한 SIR 모형에 의해 확산된 결과를 생각해보자. 전염병의 전파는 사람들 사이의 접촉을 통해 일어난다. 이러한 사람들 사이의 접촉은 사람들 사이의 네트워크를 이룬다.

물론 모든 접촉이 전염을 일으키는 것은 아니다. 감염자는 병이 낫거나 사망하기 전에(SIR 모형의 용어로는 '제거되기' 전에) 접촉자들에게 전염에 성공해야 한다. 성공 여부는 감염율과 제거율에 의해 확률적으로 결정된다. 가장 간단한 SIR 모형에서는 사람들 사이의 차이(성별, 나이, 건강상태 등)와 접촉들 사이의 차이(접촉 강도, 접촉 빈도, 접촉 시간 등)를 무시하므로 접촉이 전염으로 이어질 확률은 모든 접촉에 대해 동일하다. 이 확률값을 편의상 p 라고 하자. p 값은 감염율이 높아질수록, 그리고 제거율이 낮아질수록 높아질 것이다. (p 값이 감염율과 제거율에 의해 정확히 어떻게 결정되는지는 구체적인 SIR 모형의 상세 규칙에 따라 달라진다.)

이제 감염병 전파 과정을 다시 들여다보자. 0번 환자는 자신이 접촉한 사람들을 각각 확률 p 로 전염시킨다. 전염에 성공한 접촉에는 색칠을 하여 표시하자. 새로운 감염자들은 마찬가지로 각자가 접촉하는 취약군 사람들을 각각 확률 p 로 전염시킨다. 전염에 성공한 접촉에는 역시 색칠을 하여 표시한다. 이 과정은 더이상 새로운 감염자가 없을 때까지 계속된다. 전염 과정이 종료된 후, 제거군에 속하는 사람들은 모두 색칠된 접촉선으로 연결되어있다. 이제 최종적인 제거군의 비율은 0번 환자로부터 색칠된 접촉선을 따라 연결된 덩어리(이를 수학 용어로 연결 성분^{connected component}이라 한다)의 전체 네트워크에 대한 비율이 된다.

//

동역학 과정의 결과를 순수히 기하학적인 연결 구조 문제로 변환할 수 있게 된 것이다.

후자의 문제는 통계물리학과 확률론에서 널리 알려진 문제로 퍼콜레이션 문제라 불린다.

//

통계물리학자들이 간파한 것은 연결 덩어리의 통계적 성질을 알기 위해 굳이 전염과정을 시간상으로 따라가며 분석할 필요가 없다는 것이었다. 전염의 성공 여부는 오로지 확률적으로 결정되므로, 실제로 전염과정을 시간상으로 따라가며 형성된 제거군과 주어진 네트워크의 각각의 접촉선에 미리 확률 p 로 색칠을 해서 만들어지는 연결 덩어리는 겉으로 구

별되지 않고, 둘은 동일한 통계적 성질을 갖게 된다. 동역학 과정의 결과(전염병의 확률적 전파로 생성된 제거군)를 순수히 기하학적인 연결 구조 문제(확률적으로 색칠된 네트워크의 연결 덩어리)로 변환할 수 있게 된 것이다. 후자의 문제는 통계 물리학과 확률론에서 널리 알려진 문제로 퍼콜레이션(percolation) 문제라 불린다.⁹

바둑판 세상을 생각해보자. 바둑판의 인접한 꼭짓점을 잇는 선에 일정한 확률(그 값을 q 라 하자)로 색칠을 한다. 이제 바둑판의 한쪽 끝 변의 한 꼭짓점에서 시작하여 색칠된 길을 따라 마주 보는 끝 변까지 연결되는 길이 존재하는가 하는 것이 전형적인 퍼콜레이션 문제이다. q 값이 매우 작다면 색칠된 선이 너무 적어 연결된 길이 존재하지 못할 것이다. 반면에 q 값이 충분히 크다면 높은 확률로 연결된 길을 찾을 수 있다. 전자를 비연결상(unpercolated phase), 후자를 연결상(percolated phase)이라 부를 수 있겠다. 그리고 두 상 사이에 상전이가 존재한다. 퍼콜레이션 문제는 상전이가 나타나는 가장 간단한 문제 중 하나이다. 이제 확률적 전염병 확산 과정의 전염 확률 p 값과 퍼콜레이션 문제의 색칠 확률 q 값을 대응시키면 전염병 확산의 상전이가 퍼콜레이션 문제의 상전이가 서로 대응된다. 이 방법을 이용하여 통계물리학자들은 이를테면 바둑판 세상에서의 전염병 전파 현상의 상전이에 대한 새로운 해답을 제공할 수 있었다.

⁹ 당시 대표적 논문 중 하나의 제목이 <전염병 모형과 퍼콜레이션(Epidemic models and percolation)>이다

¹⁰ R. Pastor-Satorras and A. Vespignani, "척도없는 네트워크에서의 전염병 확산(Epidemic spreading in scale-free networks)," Physical Review Letters **86**, 3200 (2001). 이 논문의 전문은 arXiv에서도 읽을 수 있다; 라즐로 바라바시(Albert-László Barabási)의 대중과학서 『링크(Linked)』(동아시아, 2002)의 10장 <바이러스와 유행>을 참조해도 좋다.

그러나 우리는 바둑판 세상에서 살지 않는다. 이 당연한 사실을 과학자들은 약 20년 전이야 새삼 깨달았다. 더 정확히 말하자면, 사람과 사람 사이의 접촉과 같은 네트워크 구조가 전염병 확산과 같은 동역학 현상에 미치는 영향에 대한 구체적인 연구에 착수한 것이다. 바둑판 세상에서는 모든 사람이 똑같은 수의 다른 사람과 접촉한다. 실제 세상에는 소위 슈퍼전파자(superspreader)가 존재한다. 이들은 평균보다 훨씬 많은 수의 사람과 접촉하는, 네트워크 이론에서 말하는 허브(hub)에 해당한다. 허브의 존재는 전염병의 확산을 현저히 증폭시킨다. 그 결과 지극히 낮은 감염율로도 전염병 확산 상태(epidemic state)를 이끌어낼 수 있게 된다.

이를 수학적으로 규명한 것이 전염병 확산 모형 연구의 패러다임을 바꾼 최초의 주목할만한 연구 결과¹⁰로서 이후 복잡계 네트워크 이론 및 네트워크 과학 발전의 주요 기폭제가 되었다. 이 흐름의 중심에 역시 통계물리학자들이 있었다. 전염병 확산 모형 연구는 여전히 네트워크 과학의 주요 주제로 활발히 연구되고 있다.

코로나19의 유행은 인류 사회에 넓고 깊은 흔적을 남기고 있다. 그런데 코로나19의 유행은 사스(SARS), 메르스(MERS) 등 여타 코로나바이러스 감염증의 확산 과정과 근본적으로 다른가? 다르다면 그 차이점을 어떻게 통계물리학적 이론들에 반영해야 할까? 그리하여 과연 코로나19의 유행은 통계물리학에도 의미있는 흔적을 남길 것인가? 이제까지 쏟아진 2200여 편의 출판논문들 어딘가에 이미 그 해답이 있을까?