

현대과학의 낯선 패러다임 ‘확률과 통계’

글 | 이상욱 _ 한양대학교 철학과 교수 dappled@hanyang.ac.kr

지난 호에 과학문화의 세 측면을 언급하면서 대중의 과학이해를 높여려는 활동으로서의 과학문화와 과학적 세계관과 현대과학이론이 전제하거나 함축하는 바에 대한 사상적 탐색으로서의 과학문화 모두 중요하다는 점을 강조했다. 어떤 경우에는 이 두 측면이 매우 밀접하게 연결되어 있을 수 있는데 이번 호에서 다룰 확률과 통계의 개념적 낯설음이 바로 그것이다.

우리나라에서 중등교육을 정상적으로 마친 사람이라면 누구나 확률과 통계에 대한 기초지식을 가지고 있다고 할 수 있다. 그러나 실제적인 교육과정에서 구체적인 문제풀이에 몰두하다보면 학생들은 확률이 세계에 대해 어떤 존재론적 가정을 하는지, 신문이나 방송매체에서 거의 매일 쏟아져 나오는 통계적 주장을 어떻게 이해해야 하는지에 대해 깊이 생각해볼 기회를 갖지 못하기 쉽다. 필자는 확률과 통계가 현대 과학에서 차지하는 중요성을 고려할 때 이에 대해 올바르게 이해하는 일과 근대 이후 새롭게 등장하여 이러한 개념의 배경을 이루었던 과학적 사유의 틀이 얼마나 낯선 것이었는지를 검토하는 일이 훌륭한 과학문화의 실천 예로 여겨질 수 있다고 생각한다.

평균값 제시할 때 분산·표준편차도 함께 제시해야

구체적인 예를 들어 생각해보자. 최근 자녀의 지능이 엄마와 아빠 중에서 엄마 지능에 더 영향을 많이 받는다는 연구결과가 발표되어 화제가 된 적이 있다. 필자가 수업시간에 이 예를 언급했을 때 거의 대부분 학생들이 알고 있었던 정도였다. 그런데 문제는 대부분의 학생이 이 연구결과가 의미하는 바를 잘못 이해하고 있다는 사실이었다.

연구결과는 확률적인 진술을 포함하는 통계적 주장의 형태로 제시되었고, 자녀의 (선천적) 지능에 대해 엄마의 지능이 아빠의 지능보다 약간 더 높은 비율로 영향을 끼친다는 내용이었다. 하지만 수업시간에는 대조를 분명하게 하기 위해 엄마지능이 70, 아빠지능이 30의 비율로 아이지능에 영향을 끼친다는 가상적 주장을 제시하고 학생들이 이 주장을 어떻게 이해하는지 알아보았다. 대강 예

상은 했지만, 거의 모든 학생이 새로 태어나는 ‘각각의’ 아이의 지능에 대해 만약 그 아이의 지능이 100이라면 그 중 70은 엄마로부터 오고 30은 아빠로부터 온다고 이해했다.

물론 이는 확률적 진술이 포함된 통계적 주장을 개별 인과적 주장으로 파악하는 전형적인 오해였다. 주어진 주장에 대한 정확한 이해는 대강 다음과 같을 것이다. 새로 태어나는 아이의 집단에 속하는 ‘모든’ 아이의 지능에 대해 만약 그 아이의 지능이 100이라면 ‘평균적으로’ 지능의 70은 엄마로부터 오고 30은 아빠로부터 온다. 재미있는 점은 이상의 두 표현을 들려주었을 때 그게 무슨 차이가 있는냐는 질문을 하는 학생들이 많았다는 사실이다.

간단하게 두 진술의 차이를 알 수 있는 방법은 둘째 진술은 한 아이의 지능이 평균적 경향과는 반대로 아빠의 지능에서 70을 취하고 엄마의 지능에서 30을 취한 경우를 허용한다는 점에 주목하는 것이다. 물론 이런 상황은 첫째 진술이 참이라면 불가능하다. 이처럼 확률적 주장은 집단 전체의 평균값에 대해서만 말할 뿐 그 집단의 구성원이 어떻게 분포되어 있는지에 대해서는 아무런 제한 조건을 주지 않는다. 예를 들어, 성인 남자의 키의 평균이 170cm라고 할 때 모든 성인 남자의 키가 170cm이어서 평균이 170일 수도 있지만, 성인 남자의 키가 키 70cm에서 270cm까지 고르게 분포해 있어서 일 수도 있다. 평균값 이외에도 분포의 전체 형태를 보여주는 분산이나 표준편차가 필요한 이유가 여기에 있다.

두 경우를 혼동하는 것은 지능을 포함하여 대부분의 인과적 속성에 대한 과학적 진술이 진공 속에서의 물체의 질점에 대한 힘과 속도의 관계처럼 단일한 원인에 의해 결과가 확정적으로 주어지는 것이 아니라 수많은 요인의 의해 복합적으로 영향을 받는다는 점을 이해하지 못하고 있는 것이다. 이렇게 수많은 요인에 의해 특정 속성이 복합적으로 영향을 받을 때는 특정 개인이 그 속성을 어떻게 가질 것인지를 진공 속의 물체의 속도처럼 확정적으로 결정하는 것은 불가능하다. 우리가 그 많은 요인에 대한 인과법칙을 다 알고 있지 못하는데다가 여러 요인이 한꺼번에 작용할 때 개별적인 인과법칙을 어떻게 결합시켜야 하는지도 잘 모르기 때문이다.

그러므로 이런 상황을 연구하는 한 가지 유력한 과학적 방법은 연구주제를 대상이 아니라 몇 가지 변수로 규정되는 대상의 모임이나 사람의 집단이 갖는 특정 속성의 분포함수로 바꾸고 그 분포함수를 연구하는 것이다. 물리학의 통계역학이 개별 물 분자가 아니라 온도와 압력 정도의 간단한 변수로 규정되는, 엄청난 숫자의 물 분자 집단을 연구하거나, 사회통계학이 소득수준, 가정환경, 주거환경 등의 몇 가지 변수로 규정되는 수많은 가족의 집단에서 자녀들의 학업성취도의 분포를 연구하는 것이 이에 해당된다. 두 경우의 차이점은 이렇게 개별 대상에 대한 직접적인 연구의 어려움에서 집단으로 옮겨간 다음에 다시 개별 대상에 집단에 대한 연구결과를 적용할 때 나타난다. 물 분자의 경우에 비해 가족은 분석에 포함되지 않은 변수 때문에 집단의 결론이 개별 가족에 성립하지 않을 확률이 높고 열전도도와 같은 열역학적 속성의 분포에 비해 학업성취도의 분포는 훨씬 더 넓게 퍼져 있는 형태를 갖는 경향이 있다.

근대적인 확률적 과학관에 낯선 대중의 직관

보다 대중적인 영향력이 큰 ‘흡연은 폐암을 일으킨다’는 주장에 대해 생각해보자. 이는 ‘흡연을 하는 사람이 흡연을 하지 않는 사람보다 폐암에 걸릴 확률이 높다’는 말과 정확히 같은 말이다. 분명히 ‘일으킨다’는 일상적인 어휘가 사용되었는데 과학적 의미는 확률이 자동적으로 들어간다는 이 사실을 일반적으로 사람들은 잘 이해하지 못한다. 우리들은 워낙 고전적인 인과론에 익숙해져 있어서 확률적 원인-결과 관계는 낯설 수밖에 없다. 손으로 책상을 밀면 책상이 움직이는 현상처럼 우리에게 ‘일으킨다’는 것은 ‘틀림없이’ 일으킨다는 의미로 이해된다. 하지만 자연과학, 공학, 사회과학, 인문학을 통틀어서 압도적인 다수의 인과관계는 확률적이다. 그래서 우리는 일반적으로 A가 B의 원인이다. 혹은 A가 B를 일으킨다는 주장은 거의 항상 A가 B의 확률을 높인다. 즉 A가 있을 때 A가 없을 때보다 B가 발생할 확률이 높다고 이해해야 한다. 이 점을 제대로 이해하지 못하면 ‘나는 담배를 매일 두 갑씩 20년이나 피워왔는데도 폐암에 걸리지 않았다’는 자랑이나 ‘나는 일생 동안 담배 근처에도 가지 않았는데 폐암에 걸렸다’는 푸념이 ‘담배가 폐암을 일으킨다’에 대한 반례라고 생각하게 된다. 이토록 우리의 직관은 너무나 고전적이어서 근대적인 확률적 과학관에 익숙지 않은 것이다.

일단 이렇게 과학적 진술의 다수를 확률적으로 파악한다 하더라도 문제는 여전히 남는다. 흡연이 폐암발생 확률을 높인다는 말이

지니는 의미는 통계적으로 유효한 대응책을 생각하는 사람들에게는 매우 유용할 수 있다. 어떤 방법으로도 연구집단 전체에서 흡연을 억제할 수 있는 정책을 시행하면 ‘평균적으로’ 폐암발생은 줄어들 것이기 때문이다. 하지만 담배를 피우는 개별적인 사람에게는 생활습관이나 건강과 관련된 다른 요인이 폐암발생과 관련하여 더욱 결정적일 수 있다. 우유는 아동 성장에 도움이 된다는 주장은 개인성이 높은 진술이다. 그러나 우유에 대한 지독한 알레르기를 가진 아이에게는 이 진술이 성립하지 않는다. 그럼에도 불구하고 ‘평균적으로’ 우유와 아동성장 사이의 상관관계는 여전히 성립되고 우리는 이 연관관계를 우리 나라 아동의 성장을 돕는 정책을 입안하는 과정에서 사용할 수 있다. 이러한 정책의 타당성은 우유 알레르기가 있는 아동에게는 이 ‘평균적’ 진술이 적용되지 않는다는 사실과 동시에 성립한다는 점을 이해하는 것이 중요하다.

문제는 평균적으로 성립되는 확률진술을 개별사례에 적용하는 것이 적절한지에 영향을 미치는 알레르기의 유무와 같은 요인을 항상 모두 알고 있는 것이 불가능하다는 점이다. 그러므로 집단에 대해 타당한 확률적 진술이라도 그 집단에 속한 특정 개인에게 적용할 때는 집단의 평균적 경향에 간섭하는 추가적 인과요인을 그 개인이 갖고 있는지를 꼭 살펴야 한다. 지금은 타계한 유명한 진화생물학자 스티븐 제이 굴드는 한때 중피종이라는 희귀한 악성종양에 걸렸다는 판정과 이 병의 중간값 생존율이 8개월이라는 말을 듣고 절망했다고 한다. 그러나 그는 진화생물학자로서 자신의 평소 신념, 즉 집단의 경향성을 한 개인에게 일어날 확률과 동일시해서는 안 된다는 신념을 자신의 병에 대해 적용하여 자신에게 특수한 추가적 요인을 찾게 되었다. 굴드는 중피종의 분포가 생존가능기간이 긴 오른쪽 꼬리가 매우 긴 형태를 가지고 있음을 알아차렸다. 그리고 자신이 오른쪽 꼬리 부근에 위치한 개인일 가능성이 높다는 근거, 즉 아직은 젊고 의료환경이 좋은 곳에서 살고 있으며 조기에 발견했고 병과 싸워 이기려는 투지가 넘친다는 사실 등에 입각하여 자신은 8개월보다는 훨씬 높은 생존가능성이 있다고 결론 내렸다. 실제로 굴드는 이런 추론에 근거하여 상당기간 활발한 저술활동을 벌이다가 다른 질병으로 사망했다. ⑤



글쓰이는 서울대학교 물리학과 및 동대학원을 졸업했다. 서울대학교 과학사 및 과학철학 협동과정 박사 수료(과학철학 전공) 후 런던대학교에서 철학박사 학위를 받았으며 런던정경대학교 철학과 교수를 지냈다.