

특별연재

現代를 變化시킨 20대發明·發見〈5〉

數의 裁判

아이언 해킹

피어슨이 개발한 ‘카이제곱검정’은 그 자체로 본다면 하나의 사소한 사건이었으나 우리의 숫자 세계를 해석하는 방법에서 하나의 전환을 구획하는 신호가 되었다. 오늘날 아이디어를 정책수립가들과 일반에게 제시하는 하나의 표준방법이 될 수 있는 것이다.

「칼 피어슨」의 ‘카이 제곱검정(chisquare test)’은 이론과 현실간의 적합성을 측정하여 새로운 종류의 의사결정방법을 예고했다.

오늘의 세계는 통계로 넘쳐 흐른다. 의학의 공포와 개가를 알릴 때마다 복잡한 가망성의 분석도가 붙어 다닌다. 암이며 심장병이며 후천성면역결핍증(AIDS)을 생각할 때 우리가 실제로 알고 있는 것보다 더 많은 확률에 관한 이야기를 듣고 있다. 날마다 퍼붓는 통계의 세례는 단순한 계산문제가 아니라 불확실성시대와 맞서서 추정과 결정을 내리기 위한 문제와 관련 된다. 통학버스나 조기의 가석방허가가 미치는 영향을 연구·조사하지 않고 학교나 교도소에 관한 정책을 바꿀 수 없다. 금융시장, 취중운전, 가정생활, 고에너지물리학 심지어는 이상인간세포까지 모두 자료분석실험의 대상이 되고 있다.

이런 현상은 1900년 ‘칼 피어슨’(Karl Pearson, 1857, ~ 1936)이 이론적인 가설이 실제의 관측결과와 얼마나 들어맞는가 측정하기 위한 방법인 ‘적합도에 관한 카이제곱검정’을 발표하였을 때부터 시작되었다. 그 기본적인 아이디어는 매우 간단하다. 가령 주사위가 6개의 면 중 어느 한면이든지 똑 같은 빈도로 구른다고 가정해 본다. 6백번을 굴린다고 하자. 6면이 나

타날 빈도는 모두 같다고 생각된다. 그것은 단순히 우연한 것이라고 할 수 있을까? 주사위가 공평하다는 가설이 실제 데이터와 얼마나 잘 맞아 들어갈까? 이 이론이 가장 잘 맞아 들어간다고 하면 그 결과는 6백번 굴릴때 각 면이 나타나는 빈도는 1백번이 되어야 한다. 그러나 실제로는 이 비율은 언제나 다르다. 그것은 주사위가 공평하다고해도 여러번 던지는 가운데 우연이라는 요소가 있기 때문이다. 이론과 6백회의 관측결과간의 차이를 측정할 수 있는 방법은 없을까? 피어슨의 ‘카이제곱검정’법은 이론과 데이터를 조화시키는 하나의 방법을 제공한다.

‘카이제곱검정’은 가설과 데이터에 쓰일 수 있다. 여기서 관측결과는 통계학자들이 ‘셀(cell)’이라고 부르는 추상적인 카테고리 속에 들어가게 된다.

예컨대 콜레라의 어떤 치료법이 쓸모없을 것인가를 가려내는 실험을 할 때 환자는 4개의 ‘셀’로 나눈다. 곧 치료하여 회복하는 것, 치료하되 사망하는 것, 치료하지 않고 회복하는 것, 그리고 치료하지 않고 사망하는 것 등이다. 만약에

치료가 쓸모없는 것이라면 치료를 받은 환자와 받지 않은 환자간의 회복률에는 아무 차이도 기대할 수 없다. 그러나 우연과 제어할 수 없는 변수가 언제나 어떤 차이를 가져오게 한다. 피어슨의 검정은 치료가 쓸모 없다는 가설이 얼마나 관측결과와 들어 맞는가 하는 것을 알리면서 이런 것을 계산해 놓는다.

피어슨보다 앞서 이와 같은 종류의 측정법이 있었다. 그러나 피어슨의 공헌은 어떤 학문분야에도 적용할 수 있는 표준검정법을 제공했다는 것을 들 수 있다. 그는 연구소와 학생과 그의 저널인 '바이오메트리카(Biometrika)'와의 작업을 통해 이것을 제도화했다. 피어슨의 검정



칼 피어슨

차 례

〈1900 - 1919〉

- (1) 플라스틱이 바꾼 세계
- (2) IQ 테스트와 함정
- (3) 아인슈타인의 멋진해
- (4) 혈액형발견이 구제한 듯한 인명
- (5) 수의 재판
- (6) 휴지통에서 나온 진공관
- (7) 식량증산의 길을 트 잡종옥수수
- (8) 진공소제기에서 揚力을 얻은 현대 항공술

〈1920 - 1939〉

- (9) 醫學의 第2 革命을 가져온 抗生劑
- (10) 人類의 뿌리를 밝힌 胎童의 어린이 頭蓋骨
- (11) 核融合과 두개의 날
- (12) 밝혀진 宇宙開闢의 수수께끼
- (13) 短命의 DDT와 環境運動
- (14) 20世紀의 얼굴, TV가 나오기 까지

〈1940 ~ 1959〉

- (15) 人口의 均衡을 잡은 避妊藥
- (16) 나치스의 暗號풀기로 출발한 컴퓨터
- (17) 精神疾患을 구제한 클로르프로마진
- (18) 20世紀 産業의 쌀, 半導體
- (19) 2 重나사선이 펼친 新世界
- (20) 레이저가 연 光産業

은 널리 알려진 수학에 근거를 두었다. 또 이것은 실용적이었다. 사람들이 손으로 계산하던 시절에 이 검정을 응용하는 산수(셈법)가 간단해야 한다는 것은 기본적인 조건이었다. 이 검정은 피어슨의 학생들이 계산하여 그의 저널에 발표한 표준표만 있으면 되었다. '카이제곱검정'은 그 자체로 본다면 하나의 사소한 사건이었으나 우리의 숫자 세계를 해석하는 방법에서 하나의 전환을 구획하는 신호가 되었다. 오늘날 아이디어를 정책수립가들과 일반에게 제시하는 하나의 표준방법이 될 수 있는 것이다.

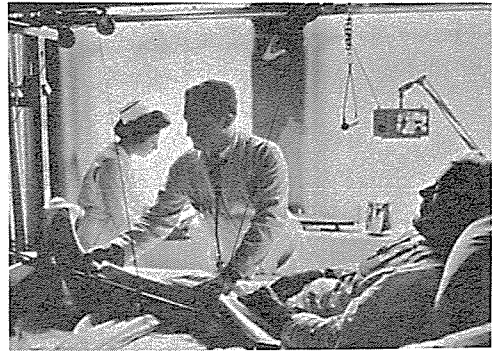
피어슨은 온갖 종류의 프로젝트를 동시에 해결하는 놀라운 정력의 빅토리아시대인의 한 사람이었다. 온건한 사회주의자이며 강력한 여권 지지자인 그는 또 열렬한 우생학신봉자였기 때문에 언론에서 나쁜 평을 받고 있다. 그는 유럽 중류계층을 선별적으로 번식시킨다면 문명의 쇠퇴를 막는데 도움을 줄 것이라고 생각했다. 이

런저런 면에서 그는 또 다른 한 사람의 정력적인 빅토리아시대인인 프란시스 골턴 (Francis Galton)<sup>1)</sup>의 정신적인 후계자였다. 우리는 폐암의 발생과 하루에 피운 담배개피 수와 같은 2개의 변수간의 연관성을 수량화하는 방법인 현대 상관이론의 공을 골턴에게 돌리고 있다.

골턴의 상관론은 1880년대에 유전연구의 일부로서 비롯되었다. 그는 ‘훌륭한’ 가족은 재능을 물려 주되 ‘나쁜’ 가족은 악을 물려준다는 그의 가설을 설명하고 측정하는 방법이 필요했다. 그동안이야 어떻든간에 그의 방법은 어떤 한쌍의 변수에 대해서도 적용할 수 있다. 부유한 그는 런던의 유니버시티 칼리지 (University College)에 우생학강좌를 후원했으며 그곳에서 우생학연구에 통계가 응용되었다. 피어슨이 최초의 수학통계교수로 불릴 수 있게 된 곳은 바로 그곳이었다.

19세기초에 가우스(Johann Karl Friedrich Gauss, 1777~1855)<sup>2)</sup>와 라플라스(Pierre Simon de Laplace, 1749~1827)<sup>3)</sup>는 오차곡선의 이론을 확립했다. 이 낮익는 종모양의 곡선은 예컨대 A에서 F에 이르는 학생의 등급은 이 곡선에 준하도록 되어 있다. 19세기 중엽에 이르자 키, 무게, 지능, 내성등 거의 모든 인간과 생물학적인 현상은 이 모양의 곡선으로 설명할 수 있게 되었다. 피어슨 자신은 이 곡선에서 흔한 영어이름을 붙였는데 우리는 이것을 정규곡선<sup>4)</sup>이라고 부른다. 그것은 한때 정규적인 그림이라고 생각하고 있었기 때문이다.

사람들은 경쟁대상이 없는 하나의 곡선으로 거의 모든 현상을 설명할 수 있다고 생각했으므로 이의 적응성 문제는 선뜻 생기지 않았다. 그것은 시중에 있는 유일한 곡선이었기 때문에 이 곡선은 적용되어야 했던 것이다. 그러나 피어슨과 여러 다른 사람들은 이것은 터무니 없는 공상이라는 것을 깨닫게 되었다. 그는 경쟁대상이 되는 곡선들을 고안해냈다. 그 중에는 왼쪽에 굽은 곡선도 있었고 오른쪽에 가는 곡선도 있었으며 대칭형인 종의 곡선보다는 파도의 물마루와 같이 생긴 곡선도 있었다. 이것은 키의 분



어떤 환자가 새로운 실험치료를 받을 것인가 결정하는데 카이제곱 검정은 흔히 편견없는 조증자구실을 한다.

포나 또는 1마일을 달리는 능력이나 그밖의 데이터형에 대해 더 잘 어울렸던 것이다. 이윽고 그는 이 곡선들이 데이터와 얼마나 잘 들어맞을까 진지하게 반문할 수 있었다. 인구가 머리카리나 또는 하루의 각연 개피수와 같은 분리된 ‘셀’로 자연스럽게 나뉘지는 경우는 대개 쉽게 이해하지만 ‘카이제곱검정’은 표준화된 방법으로 다른 곡선에도 적용할 수 있다.

다른 검정방법이 곧 뒤를 따랐다. 제1차 세계대전후 휘셔(R. A. Fisher)가 최소한 영어사용 세계에서는 이 분야를 지배하게 되었다. 그는 통계적인 구상에서는 뛰어나게 풍부한 배진을 창출했다. 그는 또 로담스티드에서 실험농장을 지도했는데 그곳은 비료, 양계용 사료 등을 통계적으로 매일 시험하는 일 외에도 전체적인 통계방법론을 위한 부양장 구실을 했다.

통계방법론은 조용하게 발전한 것은 아니다. 칼 피어슨의 두 사람의 가장 우수한 제자는 그의 아들인 이건(Egon)과 그의 연구소의 젊은 루마니아 태생의 제자인 제르지 네이만(Jerzy Neyman)이었다. 휘셔가 칼 피어슨의 자리를 빼앗았다고 생각되는 분야에서 네이만과 E. S. 피어슨의 아이디어가 휘셔의 아이디어와 대치되었다. 그러나 이런 일들은 논쟁없이 이루어진 것은 아니다. 휘셔는 피어슨의 이론이 과학적인 추론과는 아무 관계도 없는 것이라고 야

유했다. 네이만은 이에 대해 격렬하게 응수했다. 그러나 이 논쟁의 알맹이 보다는 결과적으로 생긴 제도와 실행이 더욱 중요했다. 예컨대 네이만은 버클리로 옮겨 그곳에서 지도한 결과 버클리는 오랫동안 미국의 가장 우수한 통계학교가 되었다.

통계학은 아직도 기초에 대한 완전한 합의를 본 안정된 학문분야로 들어 갈 나이는 되지 못했다. 여기서 언급한 모든 통계학자들은 확률의 관건은 서로 다른 종류의 사건이 발생하는 비교빈도에 달려 있다고 추정했다. 이것은 기상대가 내일 비가 올 수 있는 기회는 90퍼센트라고 말할 때와 같은 공산을 말한다. 그러나 이것은 무슨 뜻을 갖는 것일까? 확률이라는 것은 주관적인 신념도와 관련되고 있고 또 주관적인 접근책만이 타당한 확신도를 주기때문에 일부에서는 이에 대해 아무말도 하지 않는다. 영국의 램지(F. P. Ramsey), 이탈리아의 휘네트(Bruno de Finetti), 미국의 새비지(L. J. Savage)에게서 나온 연구는 이런 주관성을 진지한 과학적인 접근방법으로 만들었다. 오늘날 우리는 이에 대해 정력적으로 때로는 맹렬하게 의견

을 달리하고 있다. 그러나 첫번째 원칙에 대한 싸움은 여러 경쟁적인 방법을 대규모로 적용하는 것보다 중요하지 못한 것이다.

통계에 관한 이야기는 현재 일상생활의 여러 분야로 침입하여 이따금 주관성에 대한 거짓된 인상까지 만들어 낸다. 정책수립자들은 모든 위험과 희망을 측정하기 위해 숫자를 요구한다.

이리하여 미국 원자력규제위원회의 '핵로안전성연구(1975년)', 는 여러가지의 위험에 대해 확률을 부여했다. 일부 비평가들은 이것이 바로 무지를 인정하는 책임을 회피하는 하나의 방법이라고 주장하고 있다. 다른 비평가들은 어떤 통계적인 감각으로 많은 데이터 뭉치에 '적응'시키기 위해 복잡한 모델들을 날이갈수록 더 많이 사용하게 된다고 지적하고 있다. 결국 이 모델들은 실존하는 세계와 아무관계가 없는 환상의 세계가 될 수 있는 것이다.

종건 나쁘건 통계의 추론은 전혀 새로운 형의 논법을 제공했다. 조용한 통계학자들은 새로운 사실이나 기술적인 발전을 발견한 것이 아니라 우리의 의견을 추론하고 실험하며 형성하는 방법을 바꿈으로써 우리의 세계를 바꿔 놓았다.

(註)

1) 골턴(Sir Francis Galton, 1822~1911) : 영국의 인류학자며 우생학의 창시자. 「천재와 유전」(1869)과 「영국의 과학자」(1874)에서 대과학자의 혈연관계나 과학자의 질의 분포를 문제시하여 과학의 사회학적 연구의 선구자가 되었다.

2) 가우스(Johann Karl Friedrich Gauss, 1777~1855) : 독일의 수학자. Braunschweig의 가난한 집안에서 태어났으나 어려서부터 수학에 뛰어난 재능을 나타내어 Ferdinand공의 후원으로 1795~98년 Göttingen 대학에서 배울 수 있었다. 1799년 Helmstedt대학에서 학위를 얻은 뒤 1807년까지 Ferdinand공의 보호아래서 수학을 연구했으며 1807년 이래 별세할때까지 Göttingen대학교수겸 천문대장을 지냈다. 청년시절에는 정수론에 몰두했고 순수수학에서는 학위논문에서 이른바 대수학의 기본정리를 증명하고 1812년의 논문으로 초기하급수를 논했으며 1827년에는 곡면론을 전개하는 등 대수학, 해석, 기하학 등 여러 방면에서 큰 공헌을 했다.

1821~23년 논문에서 최소 2승법의 이론을 세우고 통계에서의 가우스 분포의 의의를 강조했다.

3) 라플라스(Pierre Simon de Laplace, 1749~1827) : 프랑스의 수학자·천문학자. Beaumont의 가난한 농가에 태어났다. 어렸을 때 부터 남다른 재능을 갖고 있어 고향의 Ecole militaire에서, 수학을 가르쳤으며 18세 때 Paris에서 J Le Rd' Alembert에게 인정을 받았다. Napoléon I 밑에서 내상에 임명되어 백작이 되었으나 왕정복귀를 맞아 후작이 되어 정치가로서는 무절조한 사람이었다. 도량형위원으로서 미터법 제정에 참가했다. 해석수학에 뛰어난 재주를 갖고 있어 특히 천체역학이나 확률론에 재치있게 응용했다.

4) 가우스 분포(Gaussian distribution) : 정규분포(normal distribution)라고도 함. 자연현상·사회현상에 흔히 볼 수 있는 우연사상. 평균치 근처에 많이 생기고 그곳에서 떨어지면서 적어지는 분포를 말함.