<< 확률의 역사 >>

고대 이래로 우연의 문제를 접해 온 것은 의심할 수 없으나, 15세기까지는 확률에 대한 수학적인 취급이 시작되지 않았다. 우연의 게임에 대한 최초의 인쇄물 중 하나는 1494년에 간행된 파촐리의 'Summa de Arithmetica'라는 책이다. 그는 게임이 중단되었을 때, 상금을 공정하게 분배하는―오늘날 '분배 문제'라고 하는―방법에 대해 적고 있다.

16세기에는 카르다노가 도박꾼을 위한 책을 썼는데, 이 책에서 한 개의 주사위를 사용하는 문제를 푸는 어떤 법칙에 대한 완전한 표를 제시하고 있다. 우연의 법칙을 수학화하려고 했던 이러한 초기의 시도에도 불구하고 수학으로서의 확률론의 탄생은, 브로신과 루이 14세의 조신이었던 드 메레가 주사위 문제와 분배 문제를 파스칼에게 제기했던 1654년으로 생각되고 있다. 파스칼은 이 문제를 페르마에게 전했고, 두 사람은 곧 그 문제를 해결했다. 두 사람의 연구가 확률의 수학적 이론을 세우는 데 있어서 중요한 고비가 되었던 것으로 간주되고 있다.

역사적으로 도박을 통해 확률, 통계이론을 체계화하기 시작한 것은 17세기부터이다. 당시 석학으로 꼽히던 파스칼은 도박사 친구인 드미어가 문의한 주사위 문제를 푸는 과정에서 "파스칼의 삼각형"으로 유명한 이항 전개 계수를 발견했다. 수학자 페르마와 서신교환을 통해 이를 정리했다. 수학자 폰 노이만과 물리학자 페르미는 확률과 통계이론에 쓰이는 반복실험(시뮬레이션)을 이용해 원자폭탄의 중성자충돌 확률을 계산해 냈다. 원자폭탄 개발로 이어진 이 시뮬레이션은 도박의 도시의 이름을 딴 "몬테카를로 방법"으로 명명했다.

<< 통계의 역사 >>

국가가 징세·징병, 다른 어떤 목적을 위해 호적이나 토지대장을 만들어 이것을 근거로 통계를 작성한 것은 예로부터 있었던 일이다. 고대 로마에서의 인구에 대한 신고조사(申告調査)는 '센서스'라 불렀으며, 오늘날까지 그 명칭이 전한다. 또 중국에는 전한말기(前漢末期) 이래 '호수인구수(戶數人口數)'라는 기록이 남아 있다. 그러나 통일국가가 무너지고 봉건적 분권화가 진행된 중세에는 신뢰할 만한 인구통계를 거의 남기지 못했다.

근대적 통계는 19세기 초의 유럽에서 성립되었으며, 이때부터 인구의 정확성이 구축되어 갔다. 정부통계의 개선에 크게 공헌한 사람은 L.A.J.케틀레로서 그는 1853년 제1회 국제통계회의를 개최하고 통계의 보급과 발전에 힘썼다. 19세기 중엽 선진국들의 통계제도는 정비되었으며, 통계의 대상도 인구·범죄에서 폭을 넓혀 산업·무역도 포함하게 되었고, 19세기 후반에는 특히 사회문제가 중요시되어 가계조사 등의 통계까지 등장하였으며, 20세기에 들어와 선진국들의 통계는 더욱 확충·강화되었다.

제1차 세계대전을 고비로 국가가 경제정책·사회보장정책 등 국민생활의 여러 부문에 걸쳐 정책적 개입을 다면화함에 따라 각종 통계가 필요하게 되었으며, 이와 같은 통계들은 국민경제 계산론·산업연관론 등 국민경제 전체를 포괄적으로 파악할 수 있는 하나의 방법론으로 확립됨으로써 통계의 체계가 특히 경제면을 중심으로 정비되어 갔다. 그러나 일부 사회주의 국가는 국민경제계산에 있어 상호 비교 가능성이 문제가 되고 있다. 또 일부 개발도상국 중에는 인구·중요물자 생산량 등 기초적 통계에 있어서도 국제적 신뢰를 얻지 못하는 나라도 있다.

통계는 재정부담이나 프라이버시 의식의 발달로 조사실시에 어려움이 더욱 증대되고 있으며, 이에 대비 앞으로의 통계는 조사의 개폐·신설과 컴퓨터 등 새로운 정보기술을 활용하는 방법을 모색하는 추세에 있다.

통계 統計 (statistics)

집단현상에 대한 구체적인 양적 기술(量的記述)을 반영하는 숫자.

특히 사회집단 또는 자연집단의 상황을 숫자로 나타낸 것이다. 예를 들어 서울 인구의 생계비, 한국 쌀 생산량의 추이, 추출검사한 제품 중의 불량품의 개수 등이 그것이다. 통계는 집단에 관한 것으로서, 어떤 사람의 재산이라든가 한라산의 높이 등, 어떤 개체에 관한 수적 기술은 아무리 구체적이더라도 통계는 아니다. 통계는 사회의 발전과 함께 발달해 왔는데, 오늘날의 사회생활과 과학은 통계 없이는 존재할 수 없다.

집단현상을 통계로 나타낼 때, 그 집단을 구성하는 각 개체를 통계단위 또는 단위라고 한다. 이 단위는 공통의 성질을 가지고 있는데, 이 공통의 성질을 표지(標識)라고 한다. 이를테면 한국의 인구를 구성하는 단위는 일정한 날짜와 시간에 한국에 살고 있는 사람이며, 이 조건이 표지가 된다. 이들 단위는 표지 이외의 점에서는 이질(異質)이다. 표지에는 남녀, 산업·직업 등 질적인 것과, 연령·소득금액 등 양적인 것이 있다. 질적인 표지의 통계를 속성통계(屬性統計), 양적인 표지의 통계를 변수통계(變數統計)라고 한다. 또, 집단의 성질에 따라 자연현상에 관한 자연통계와, 사회현상에 관한 사회통계로 나누어지는데, 자연통계는 기후통계(氣候統計)·생물통계(生物統計) 등으로, 사회통계는 경제통계·경영통계 등으로 세분할 수 있다.

또한, 국세조사(國勢調査)와 같이, 집단의 한 시점에 관한 것인 정태통계(靜態統計)와, 1년간의 출생수·사망수·공업생산 등과 같이 어떤 기간에 관한 동태통계(動態統計)로도 나누어진다. 이 밖에 집단의 전체에 걸치는 전수통계(全數統計)와, 일부분을 관찰한 부분통계로 나누는 수도 있는데, 전수통계는 비교적 소박한 기술적 수리(記述的數理)처리에 따른 방법으로 기술통계라고 불리며, 부분통계는 부분에서 전체에로의 추측기법(推測技法)을 포함하기 때문에 추측통계라고 한다.

통계를 이용하는 데는, 작성자·작성시기·작성방법·대상(단위표지)·대상의 존재장소 등에 관한 깊은 인식을 필요로 한다. 이 같은 모든 통계는 현실의 일정한 사회관계를 바탕으로, 조사자와 피조사자 사이에서 질문·응답이 행해지는 통계조사(統計調査)라는 특수한 과정을 거쳐 이루어지는데, 거기에는 상호협조와 이해에 따르는 대항관계가 작용한다. 또한 통계는 그 필요성과 작성능력이라는 점으로 보아, 그 대부분이 정부나 지방자치단체 등에 의한 관청 통계로 작성된다는 특성을 지닌다.

-----------------------------------------------------------------

통계,확률의 역사

1. 확률의 역사

확률에 관한 두 가지 수수께끼

확률의 역사는 곧 도박의 역사로 르네상스 시대부터 시작되었다. 당시 지중해 연안의 도시에는 일확천금을 꿈꾸는 상인들이 많이 모여들었는데 이들은 날씨가 나빠 출항하지 못할 때 심심함을 달래기 위하여 도박을 하곤 했다.

이때 사람들이 그 승률의 대소를 미리 알기 위해 수학자와 함께 연구하기 시작하면서 확률의 사상은 싹텄다. 그러다가 수학자 카르다노가 도박에 수학을 적용하여 이론적으로 연구하기 시작했다.

그러나 보다 본격적인 연구가 진행된 것은 17세기의 페르마와 파스칼에 의해서였다. 파스칼에게는 주사위 도박의 문제를 수학적으로 생각하여 늘 좋은 결과를 얻곤 했던 드 멜레라는 프랑스 친구가 있었다.

이 드 멜레가 주사위 도박에서의 알쏭달쏭한 두 가지 문제를 수학자인 파스칼에게 물었는데 앞의 두 문제가 ‘드 멜레의 수수께끼’로 잘 알려져 있다.

1. 첫 번째 문제

드 멜레는 다음과 같이 생각하였다.

'주사위 한 개를 4번 던지면 6의 눈이 적어도 한 번 나올 확률이 0.5보다 크다. 따라서 2개의 주사위를 던질 때는 눈이 나타나는 방법이 주사위 1개를 던질 때의 6배이므로 n=4 x 6=24(회)로 하면 던지는 사람에게 유리하다'

그런데 위와 같이 실제로 실행을 해 보니 24회로는 던진 사람에게 손해가 있었다. 그러면 과연 n을 얼마로 하여야 던지는 사람에게 유리할까?

이것이 드 멜레의 첫 번째 문제인데 수학적으로 계산해보자. 2개의 주사위를 n번 던져서 적어도 한 번 두 개 모두 6의 눈이 나올 확률은 1 - (35/36)n 이다. 따라서 1 - (35/36)n > 0.5 이 되면 던지는 사람에게 유리하다.

로그를 사용하여 이 식을 풀면 n>=25가 되어 n=24로는 손해보는 것이 당연하였다.

2. 두 번째 문제

앞의 2와 같은 내용의 편지를 받고 고민한 파스칼의 답변은 다음과 같다.

'다음 한 판을 더 해서 A가 이긴다면 A는 3번이긴 것이므로 64피스톨을 가지게 된다. 그러나 만약 B가 이긴다면 A도 2번, B도 2번이긴 셈이므로 비기게 되어 각각 32피스톨씩을 가지게 되어 있다.

나머지 32피스톨은 A나 B중 이기는 사람의 몫이 되겠지만 누가 이길지 모르고 A와 B중 이기는 사람의 몫이 되겠지만 누가 이길지 모르고 A와 B두 사람의 솜씨가 비슷하므로 이기거나 질 확률은 반반이다. 그러므로 A에게 32피스톨을 먼저 주고 그 나머지의 반인 16피스톨을 더 주면 된다. 결국 A는 48피스톨을, B는 16피스톨을 가지는 것이 가장 합리적이다.'

이 대답을 확률적 사고로 고쳐 표현해 보면 다음과 같다.

먼저 A가 이 게임에서 이길 확률을 구해 보자. A가 다음 번 게임에서 이길 확률도 질 확률도 0.5인데, 만약 다음 번 게임에서 A가 이기면 A는 B보다 먼저 3점을 따는 것이 되어 이 게임은 A의 승리가 된다.

만약, 다음 번 게임에서 A가 지면 그 시점에서는 A도 2점, B도 2점을 얻은 것이 되어 게임은 끝나지 않는다. 이 경우 A가 이 게임에서 이기려면 그 다음 게임에서 A가 이겨야 하는데 일어날 확률은 0.5 X 0.5 = 0.25 이다.

결국 A가 이 게임에서 이길 확률은 0.5 + 0.25 = 0.75 이고, B가 이길 확률은 0.25 이다.

따라서 내기 돈 64피스톨은

A: 64 0.75 = 48 , B = 64 0.25 = 16

으로 분배하는 것이 가장 합리적이다.

2. 통계학의 역사

오늘날 우리가 쓰고 있는 뜻의 통계학(Statistics)이라는 용어는 19세기 말에 확립된 것이다. 19세기 중엽까지의 통계의 의미는 정치적 현황에 관한 정보를 의미하는 것으로서, 국세조사와 같이 국가의 상태를 조사, 연구하는 것이 었다.

학문으로서의 통계학은 17세기에 이르러 독일, 영국, 프랑스 등에서 발생하였다.

독일에서는 콘링(Conring, H.; 1605～1681)과 아켄웰(Achenwell, G.; 1719～1772) 등이 대학에서 국세학(國勢學, Staatskunde)을 강의하였다. 이들은 정치, 경제, 토지, 인구 등 국가적인 상황을 계통적으로 기술하고, 국가가 어떤 국토와 어떤 민족으로 구성되어 있으며 얼마만큼의 부를 가지고 있는가를 정확히 파악하려 하였으므로 국세학파라고 부른다. 이 학파에서는 수치의 사용을 되도록 피하고 관념적으로 현상을 파악하려고 노력한 것이 특징이다.

한편, 영국에서는 그라운트(Graunt, J.; 1620～1674), 패티(Petty, W.;1623～1687) 등에 의하여 소위 정치산술학파(Political arithmetic)가 형성되었다. 그랜트는 1662년에 그의 저서 ≪사망표에 관한 자연적, 지역적 관찰≫에서 교회의 기록을 근거로 하여 처음으로 사망표를 작성하고, 남녀의 출생수, 결혼 상황 등을 기록, 정리, 비교함으로써 수량적 관찰에 입각한 사회 현상의 규칙성을 발견하는 방법을 설명하였다. 페티는 그랜트의 연구를 발전시켜 인구통계학을 만들고 전 지구상의 인구를 3억 6천만이라고 추정하였다. 이 학파는 국세학파와는 달리 수치에 기초를 두고, 이로부터 어떤 법칙을 발견하려고 노력한 것이 그 특징이다.

17세기에서 18세기에 걸쳐 프랑스에서는 파스칼(Pascal, B.; 1623～1662), 페르마(Fermat, P.; 1601～1665), 베르누이(Bernoulli, J.; 1654～1705), 드 무아브르(De Moivre, A.; 1667～1754), 라플라스(Laplace, P.S.; 1749～1827) 등에 의하여 확률론이 정립되어 우연히 일어나는 현상을 수학적으로 관찰, 처리하는 방법이 개발되었다.

국세학파 이론과 정치산술학파가 쓴 자료를 이용하여 확률론에 기초를 둔 통계학이 싹트기 시작한 것은 19세기 초였다.

벨기에의 케틀레(Quételet, L.A.J.; 1796～1874)는 확률론에 입각한 통계학을 과학으로서의 학문으로 체계화하여 근대통계학의 기틀을 마련하였다. 그 후 통계학은 이론 분야인 수리통계학과 응용 분야인 사회통계학으로 나뉘어 발달해 왔고, 특히 경제학, 생물학, 사회학 등에 많은 공헌을 하였다.

골턴(Galton, F.; 1822　～1911)은 다윈(Darwin, C.R.; 1809～1882)의 ≪종의 기원≫(Origin of Species)의 영향을 받아 유전학의 통계적 연구에 몰두하였고, 변수들 사이의 관계를 연구하여 현대통계학의 기초를 닦았다. 그는 아들의 키와 아버지의 키의 관계에서 아들의 키는 평균키로 회귀하려는 경향이 있음을 알아내었으며, 이것은 현대통계학의 큰 분야 중의 하나인 회귀분석의 시초가 되었다.

골턴의 통계적 사상은 그의 후계자 피어슨(Pearson, K.; 1857～1936)에 의하여 계승되었다. 런던 대학의 우생학의 교수로 있으면서 피어슨은 생물학, 우생학, 유전학 등을 통하여 통계적 연구 방법의 확립에 공헌하였고, 그의 연구결과로 오늘날의 기술통계학(descriptive statistics)의 기초가 확립되었다. 특히, 그는 1900년에 카이제곱 판정법을 도입하여 널리 보급하였고, 1901년에는 골턴 및 동물학자 웰턴(Walton, W.F.R.; 1860～1906)과 협력하여 잡지 ≪생물통계≫(Biometrica)를 창간하였다. 이것은 현존하는 학술 잡지 중에서는 가장 오래된 것이다.

이때까지의 통계학은 모두가 많은 양의 관측값을 얻을 수 있다는 것을 전제로 한 대표본 이론에 의한 추론이었다. 따라서, 많은 양의 자료를 얻을 수 없는 경우에는 그때까지의 이론으로는 추론이 어려웠다. 1906년에 피어슨의 제자인 고세트(Gosset, W.; 1876～1936)에 의하여 소표본의 이론인 현재의 -분포가 발견되었다. 고세트는 영국의 더불린 (Dublin) 맥주 회사에서 기사로 종사하면서 맥주의 효모균이나 밀의 수확량 연구에서 대표본을 얻을 수 없는 경우가 많아, 이를 해결하고자 소표본 이론의 연구에 착수하여 1908년에 스튜던트(Student)란 가명으로 -분포에 관한 논문을 Biometrica에 발표하였다. 이것을 20세기의 추측통계학(inferential statistics)의 발생으로 볼 수 있다.

런던 교외에 있는 농업시험장의 통계부장인 피셔(Fischer, R.A.; 1890～1962)는 소수의 자료의 통계처리의 필요성을 느끼고 과거의 문헌을 조사한 결과 위에서 말한 논문을 발견하였다. 실은,-분포라는 이름은 피셔가 Student의 -분포의 미비한 점을 보완한 후 그의 이름 자를 따서 붙인 것이다. 피셔는 농업시험에 통계적 방법을 적용시키는 연구를 계속하여 F-분포를 비롯하여 표본상관계수, 표본회귀계수 등의 많은 통계량의 분포를 유도하여 소표본에 기초를 둔 추론법을 확립하였으며, 분산분석법, 실험계획법 등을 창시하여 현대추측통계학의 기초를 확립하였다.

한편, K. 피어슨의 아들 E. 피어슨과 폰 노이만(von Neumann, J.; 1903～1957)은 현대적인 가설검정 이론을 확립하였다.

20세기의 확률론과 통계학 이 두 분야는 순수 수학 뿐 아니라 현대의 가장 두드러진 특징 중의 하나인 고속 컴퓨터의 역할에 힘입어 크게 발전하고 있다. 대량 생산에서의 품질 관리, 시장 조사, 여론 조사에 통계적 방법이 적용되고, OR(operation's research), 선형계획법(linear programming), 게임(game)의 이론, 정보 이론 등의 연구의 발달과 작업의 자동화(automation)등과 관련하여 행정관리, 경영 관리 등에서도 매우 중요한 역할을 하고 있다.

1. 확률의 역사적 배경

100만년전 인류가 탄생한 이래로 자연의 우연과 필연사이를 인간이 계속적으로 접하며 살아 왔으나 궁극적인 우연과 불확실에 대한 학문적인 접근은 이루어지지 않았었다. 최근 500여 년 전인 16세기에 이르러서야 구체적으로 확률계산에 대한 생각을 하게 되었다. 이 우연의 게임은 1494년의 파촐리(Pacioli, 1450-1520)가 지은 "summa de arithmetica" 라는 책에서 게임이 중단되었을 경우의 상금의 분배문제를 언급하고 있다.

또한, 우연의 사실을 법칙으로 수학화하려고 노력한 여러 사람들 중에 파스칼(Pascal, 1623-1662)은 친구의 부탁으로 주사위 문제와 분배의 문제를 1654년에 고려하고 숙고하였으며, 이 문제를 페르마(Fermat, 1601-1655)에게 전하고 이들 두 사람은 이 문제를 명쾌하게 해결하였다. 이 사건이 확률이 수학적 이론으로서 세워지는 것을 제기하게 된 결정적 계기로 보고 있다. 이들 두 사람의 연구는 당시에 확률에 대한 지대한 관심을 촉발했고, 이로 인해 확률론의 초기의 문제는 주로 우연의 게임의 결과에 모아졌으며, 확률에 대한 불충분한 정의로부터 야기되는 여러 가지 문제점이 있었으나 1700년을 지나면서 발전하기 시작했다.

1655년에 호이겐스(Huygens, 1629-1695)는 파스칼의 아이디어를 이용하여 확률에 관한 독자적인 논문을 처음 작성하게 된다. 그 후 베르누이(Bernoulii,1667-1748)에 의해 확률론 만을 다룬 저서가 만들어지고 드므와브르(DeMoivre, 1667-1754)와 오일러(Euler, 1707-1783), 라플라스(Laplace, 1749-1827), 가우스(Gauss, 1777-1855)등의 노력으로 확률론은 급속히 발전해 나갔다.

그러나 아직까지도 확률의 정의가 역시 불충분한 관계로 20세기에 들어와 수학자들의 연합된 노력의 결과로 1930년대에 출판된 콜모고로프(Kolmogorov)의 확률론의 기초라는 책에서 엄밀한 공리론적 토대위의 공리적 확률을 정의하기에 이른다.

2. 확률의 정의

"확률이란 무엇인가" 라고 질문을 받는다면 우리는 무엇이라고 말하겠는가 ? 이러한 쉽지 않은 질문의 대답으로서 다음과 같은 각각의 확률의 의미를 알아보자.

1) 고전적 확률(수학적 확률)

같은 조건하에서 여러 번 반복할 수 있는 어떤 시행에서 일어날 가능성이 있는 모든 결과를 원소로 하는 집합 S를 그 시행의 표본공간(sample space)이라고 한다. 이 때 어떤 사건이 일어날 가능성이 같은 정도로 기대되어(equally likely)질 때 이를 수치적 척도로서 고정하여 가정하는 것으로 라플라스(Laplace)는 다음과 같이 확률을 정의하였다.

N개의 근원사건으로 구성된 표본공간에서 각각의 근원사건이 일어날 가능성이 같은 정도일 때, m개의 근원사건으로 구성된 사건 A가 일어날 확률 P(A)는 P(A) = m/N 이다.

그러나 이 정의는 근원사건이 유한개인 경우만 적용가능하며 실제로 일어나는 문제에 이 개념을 도입하기가 어렵다.

2) 통계적 확률(경험적 확률)

오랜 시간을 두고 여러 번 통계적 시행을 반복하면 한 사건 A의 상대도수는 어떤 값에 가까워 질 것이다. 즉, 오랜 관찰 끝에는 일정한 패턴을 찾아 낼 수 있는 것처럼 상대도수의 극한으로 확률을 정의할 수 있다.

그러나 이 극한도 증명할 방법은 없다는 단점을 가지고 있다.

3) 공리적 확률

이 때문에 러시아 수학자 Kolmogorov는, 수학자들이 기하학에서 점과 선에 대한 개념을 탄생시키는 것과 같은 흡사한 과정으로 추상적 접근을 하게 되는데 이것이 다음과 같은 공리적 확률이다. 이는 오늘날의 확률공리로서 도입되어 확률이론을 정립하게 되었다. 이는 오랫동안 쌓아온 확률현상에 대한 경험적 인식을 이론적으로 뒷받침할 필요가 있었기 때문이다.

표본공간 S에서의 임의의 사건 A에 대하여

(1) 0≤ P(A) ≤1

(2) P(S)=1

(3) 서로 배반인 사건 A, B에 대하여

P(A ∪ B)=P(A) + P(B)

을 만족할 때, 이 P(A)를 사건 A의 확률이라고 한다.