

Periodization in the History of Statistics¹⁾

Jae Keun Jo²⁾

Abstract

The history of statistics from the mid-seventeenth to the early twentieth century is considered and a scheme of periodization is proposed. In the first period(1650-1750), named 'the age of probability' in this paper, concept of probability emerged, and in the second period(1750-1820), named 'the age of error theory', statistical techniques such as the least square method are developed by astronomers and geodesists. Their techniques are supported theoretically by mathematicians like Laplace and Gauss in that period. The third period(1820-1880) is called 'the age of statistics(as a plural noun)' since statistical data played prominent roles in social sciences such as sociology, psychology. Finally the last period(1880-), called 'the age of statistics(as a singular noun)', the discipline of statistics came to maturity both in theory and application.

Keywords : history of statistics and probability, periodization, social sciences, determinism.

1. 서론

지난 1980년대 이후 확률과 통계학의 역사 (이 글에서는 확률보다는 통계학의 역사에 더 중점을 두고 있으므로 두 가지 역사를 구분하지 않고 모두 '통계학의 역사(혹은 통계학사)'라고 부를 것이다) 분야에서 이루어진 연구들은 양과 질의 측면에서 모두 주목할 만한 것이었다. 그 연구들이 특히 질적인 면에서 두드러져 보이는 이유에는 과학사를 전공하는 학자들 뿐 아니라 과학철학, 사회과학, 그리고 통계학 전공자들이 다수 통계학사 연구에 참여하게 되면서 그들 사이에 긴장관계 또는 보완관계가 가능해졌다는 점도 포함될 것이다. 당연히 것처럼 여러 학자들에 의해 깊이 있으면서도 다양한 통계학사 연구가 이루어지기 시작했다는 사실은 통계학 자체를 위해서도 매우 고무적인 일이다. 하지만 그 과정에서 통계학사 연구는 불가피하게도 점점 전문적인 모습을 갖게 되었고 그에 따라 연구결과들이 정작 통계학을 전공하는 사람에게조차 상당히 읽기 어려운 것이 된 것 또한 사실이다. 말할 필요도 없이 통계학사 역시 하나의 역사인 만큼 제대로 읽기 위해서는 근, 현대 서양사에 대한 공부도 필요할 뿐 아니라 자연과학의 역사, 사회과학 일반의 역사, 과학철학과 그 역사에 대한 공부도 필요할 것이다. 사실 그와 같은 공부는 통계학 전공자들에게 쉽지 않

1) This Research was supported by Kyungshung University Research Grants in 2002

2) Professor, Department of Statistical Information, Kyungshung University, 608-736, Pusan, Korea.
E-mail: jkjo@star.ks.ac.kr.

은 일이다. 더구나 수입 학문의 하나로 통계학을 공부하는 한국의 전공자에게 그러한 작업은 더욱 부담스러운 일일 수밖에 없는 것이, 그들에게는 이미 서구인들이 진행해온 서구 통계학의 역사를 나름대로 주체적으로 읽어내기 위한 시각을 마련해야하는 과제까지 덧붙여 요구되기 때문이다. 이 글의 바탕에는 바로 이와 같은 문제의식이 깔려 있다.

하지만 그러한 거창한 문제의식에 바탕을 둔 연구 작업은 차후의 장기적인 과제로 넘길 수밖에 없고, 대신 이 짧은 글에서는 통계학사를 공부하다 만난 아주 소박한 질문, 그리고 그 질문들에 대한 소박한 답을 제시하는데 그칠 것이다. 질문들은 한 마디로 말해서 ‘통계학사 연구들이 대상으로 삼는 시대들’에 대한 것이다. 예컨대 왜 Anders Hald는 통계학과 확률의 역사를 주제로 삼은 그의 두 책(Hald, 1990; 1998)을 하필 1750년대를 기준으로 나누었을까? 또 Daston(1988)은 왜 ‘고전적 확률’의 시기를 1650년대부터 1840년대까지로 잡았을까? 그리고 Lorenz Krüger를 비롯한 빌레펠트 그룹(Bielefeld Group)의 학자들은 왜 ‘확률혁명’이 일어난 시기를 1800년대부터 1930년대 사이로 잡았을까(Krüger et al. 1987a, 1987b)? 한편 Porter(1986)는 왜 ‘통계적 사고의 발흥’을 주제로 삼은 책에서 1820년부터 1900년 사이의 시기를 연구했을까? 한편 통계학의 전반적인 내용이 아니라 통계학의 많은 분야 가운데 특정 분야의 역사를 연구한 성과물들이 1990년대 이후 나오기 시작했는데 가령 ‘선형관계 적합’의 역사를 연구한 Farebrother(1999)는 왜 1750년대부터 1900년대까지의 시기를 대상으로 택했을까? 그리고 시계열분석의 역사를 연구한 Klein(1997)은 왜 1662년부터 1938년 사이의 역사에 주목했을까?

여기서 예로 든 경우에서 먼저 알 수 있는 점은 통계학사를 연구한 학자들이 연구 대상으로 택한 시기들이 서로 매우 다르다는 점이다. 그런데 조금 더 자세히 보면 몇몇 시점은 서로 다른 연구에서 공통적으로 나타나기도 한다는 사실도 알 수 있다. 가령 Porter(1986)와 Stigler(1986) 그리고 Farebrother(1999)는 통계학사를 각자 다른 시각에서 다른 내용을 중심으로 연구하면서 왜 한결같이 1900년이라는 특정 시점에서 책을 끝냈을까? 그리고 Hald(1998), Klein(1997), Krüger et al.(1987a, 1987b)은 왜 일제히 1930년대라는 시기를 연구의 끝으로 잡았을까? 그리고 1750년은 왜 Hald(1990, 1998)의 두 책을 나누는 해인 동시에 Farebrother(1999)의 책이 시작하는 해인가? 각 연구가 다루는 전체 시기만을 대상으로 한 위의 질문들은 각 연구의 안으로 들어가서 더 세부적인 시기를 대상으로 하면 훨씬 더 많은 질문으로 이어진다. 예컨대 꼭 같이 1986년에 출판된 Stigler(1986)와 Porter(1986)의 책은 그 제목에서 드러나듯 모두 1900년까지가 대상이다. 그런데 Stigler의 경우에는 라플라스(Pierre Simon Laplace)의 연구를 매우 길게 다루었을 뿐 아니라 모두 3부로 이루어진 책에서 제 1부의 제목을 아예 ‘1827년 이전까지의 천문학과 측지학에서의 수리통계학 발달사’라고 붙이기까지 했는데 Stigler가 통계학사에서 하나의 분수령으로 삼고 있는 1827년은 다름 아닌 바로 라플라스가 죽은 해이다. 반면, Porter의 경우 아예 라플라스에 대한 내용을 제외하고 1820년대부터 책을 시작하였다. 즉 같은 시기에 나온 두 권의 통계학사 책이 강조해서 다루는 시기에서부터 서로 크게 다른 것이다.

사실 연구 대상 시대를 정하는 문제는 연구자의 주된 관심분야와 통계학에 대한 견해에 따라 예민한 논쟁의 소재가 될 수도 있다. 그리고 이런 질문에 답하는 방법은 하나가 아닐 것이다. 여기서는 그러한 논쟁을 위한 하나의 준비 단계로 시대 순서에 따른 일종의 통계학사 연대기를 만들어보려 한다. 이미 어느 해 무슨 날에 누가 태어나고, 어떤 연구가 발표되고, 누가 죽었는가를 정리한 연대기 표는 Johnson and Kotz(1997, 주요 연구 결과에 대해서는 pp. xix-xxii, 주요 인물에 대해서는 pp. 400-403), 또는 19세기말부터 20세기말을 대상으로 한 것으로 Salsburg(2001, pp. 313-316), 그리고 통계학 용어에 대한 것으로는 Walker(1929, pp. 175-189), David and

Edwards(2001, pp. 209-246)등이 있는데 물론 이 모두가 통계학사 연구에 필수적인 기초자료임에 틀림없다. 하지만 그러한 표에서는 개별적인 인물이나 연구결과에 관련된 시기는 확인할 수 있겠지만 통계학사의 큰 줄기를 파악하기는 어렵다. 대신 여기서는 앞에서 거론한 통계학사들을 중심으로 각 연구에서 역사적인 분기점으로 제시하고 있는 시점에 대해 왜 그 시점이 선택되었는지 살펴보고 다른 연구에서 제시된 시점들과 비교해볼 것이다. 이 글에서 위와 같은 연대기의 역할은 단지 중요한 연도를 나열하고 설명하는데 그치지 않는다. 이 글의 더 중요한 몫은 그 연대기를 바탕으로 17세기 중반 이후부터 20세기 초반까지의 통계학사에 대한 '시대구분'(periodization)을 시도한 것이다. 이 글에서 제안될 시대구분은 통계학의 역사를 모두 네 부분으로 나누는 것인데, 각 시대의 이름은 '확률의 시대, 오차이론의 시대, 통계의 시대, 통계학의 시대'라고 붙여 보았다. 지금까지 통계학사를 대상으로 명시적으로 '시대구분'의 문제를 연구한 경우는 거의 없었던 듯하므로 이 글의 시도가 통계학사를 어떤 특정한 시기나 특정한 주제 또는 분야에 국한해서 살피기보다는 거칠게나마 전체적으로 조망하는데 도움이 될 수 있으리라 희망해본다.

2. 본론

많은 통계학사 연구들이 그러하듯 이 글에서도 통계학사, 더 구체적으로는 17세기 중엽부터 20세기 초반 무렵까지의 통계학사가 논의의 대상인데, 앞서 서론에서 밝혔듯 여러 통계학사 연구자들이 근, 현대 통계학의 역사를 어떤 기준으로 나누어 서술하고 있는지 살펴보는 것이 이 글의 우선적인 목적인데 이 글에서 살펴볼 연도는 1650, 1662, 1750, 1820, 1827, 1830, 1840, 1900, 1930, 1938년 등의 시점들이다. 어떤 해들은(가령 1662, 1827, 1938) 단일한 한 해를 지칭하는 반면 나머지는 그 전후 몇 년간, 또는 우리가 70년대, 80년대라고 할 때처럼 10년 정도의 기간을 일컫는다. 이들은 모두 통계학의 역사를 다룬 책들에서 책의 제목이나 서문, 혹은 책 안에 포함된 큰 갈래의 제목에 등장하는 것들이다. 즉 이 해들이 모두 통계학의 역사를 연구하는 사람들이 시대를 구분하는 분수령으로 찍은 시점들이라는 말이다. 그런데 이상한 점은 몇몇 연구들의 경우 그 연도들이 왜 중요한지 충분히 설명하지 않고 있다는 것이다. 그래서 우리는 이 글을 통해 어떤 해가(또는 시기가) 중요한 것이고 어떤 것이 단지 편의상 설정된 것인지를 가려서 살피고, 이를 바탕으로 여러 통계학사 연구자들 사이의 차이도 엿볼까 한다.

서론에서 제시한 바대로 이 글의 본문은 모두 넷으로 구분한 시대에 따라 나뉘어져 있는데 사실 통계학의 역사에서 시대구분이라는 주제는 많은 관심을 끌지 못했다. 쉽게 짐작할 수 있듯 사실 대부분의 통계학사가 다루는 시기 자체가 여러 세기에 걸친 오랜 시간이 아니라 17세기 또는 18세기 이후의 역사에 한정된 것이기 때문에, 굳이 따로 통계학사를 몇 개의 '시대'로 나누기까지 할 필요가 적었다는 것이 가장 으뜸가는 이유일 것이다. 하지만 다시 생각해보면 그렇다고 해서 삼, 사백 년이라는 시간이 짧다고만 할 수도 없는 데다 그 시기를 몇 개로 나누어서 살피는 것이 무의미할 정도로 통계학의 역사가 한결같았을 리는 더욱 없을 터이다. 이 글에서 붙인 각 시대의 이름은 '확률의 시대, 오차이론의 시대, 통계의 시대, 통계학의 시대'인데 무엇보다 이 시기들 사이의 관계를 서로 단절된 것으로 볼 필요는 없겠다. 과학사에서 새로운 학설이 옛 학설을 대체하는 과정이 흔히 그러하듯, 각 시기는 그 앞의 시기를 배척하는 것이 아니고 앞 시기의 성과들을 감싸면서 새로운 영역으로 외연을 넓혀 가는 과정이라고 보는 것이 타당할 것이다. 즉 확률의 시대에서 오차이론의 시대로 옮겨갔다고 해서 확률이론 연구가 중단되거나 쇠퇴하였다는 뜻은 전혀 아

니고 먼저 시작된 확률 연구가 지속되는 한편으로 이전에 없었던 오차이론 연구라는 새로운 분야가 등장하면서 이전의 확률이론을 새롭게 적용하려는 시도가 활발해지고 그 과정에서 통계학의 영역이 크게 확장되었다는 뜻으로 보면 되겠다.

2.1. 확률의 시대 (1650-1750)

동서양을 막론하고 아주 일찍부터 주사위던지기과 같은 놀이나 도박이 행해졌고(David, 1962, pp. 1-12; 허명희, 1994), 일찍이 14세기에는 상당한 규모의 항해보험도 있었지만(Daston, 1988, pp. 116-125), 17세기 중반이 되기 전까지는 수학적 확률(mathematical probability)이라고 할 만한 것이 등장하지 않았다는 데는 학자들 사이에 큰 이견이 없는 듯하고, 결국 많은 연구자(Hald, 1990; Hacking, 1975; Daston, 1988)들이 17세기 중반이라는 시기를 확률과 통계학의 역사를 쓰는 시발점으로 선택한 것은 자연스러운 결과였을 것이다. 통계학자로서 1750년 이전까지의 확률과 통계학의 역사를 연구한 Anders Hald는 이 시기에 서로 영향을 주고받으면서 성장한 것으로 첫째 확률 이론 및 우연의 게임(도박), 둘째 천문학과 인구학에서의 통계학, 셋째 생명보험 수학이라는 세 가지를 들었다(Hald, 1990, p. v). 당시 천문학의 경우 관측 데이터로부터 관계식을 추정할 경우들이 몇 가지 있었으나(Hald, 1990, pp. 144-168) 통계학적 추정과는 아직 거리가 먼 것들이었다. 그리고 정부가 판매하는 연금(annuity)과 생명 보험(life insurance) 분야에서는 이 시기 동안 몇 가지 생명표(life table)가 만들어졌을 뿐 아니라 18세기 전반기에는 드 무아브르(Abraham de Moivre)와 심프슨(Thomas Simpson)의 주목할 만한 수학적 연구가 나오기도 하였다. 하지만 드 무아브르 자신조차 바로 그 자신의 확률 연구를 연금을 연구하는데 활용하지 않았던 데서 알 수 있듯 연금과 보험 수학 분야는 아직 기초적인 생존 확률을 구하는 단계에 머물러 있었다(Stigler, 1986, p. 85; Hald, 1990, pp. 116-141, pp. 508-547). 따라서 통계학의 역사에서 이 시기에 가장 주목할 점은 확률 개념이 탄생했다는 점이라고 할 수 있다. 이 글에서 이 시기를 ‘확률의 시대’라고 부르는 또 하나의 이유는 17세기 후반 당시 탄생한 확률이 18세기 초반에 접어들면서 야콥 베르누이(Jakob Bernoulli), 드 무아브르, 심프슨 등에 의해 추정 문제를 비롯한 통계학 문제와 만나기 시작했지만 통계학적 추론을 위한 확률이론(예컨대 역확률)은 아직 나타나지 않고 있었기 때문이다(Stigler, 1986, pp. 62-98).

이 시기에 확률이 탄생하였다는 말이 뜻하는 바는 수학적인 면에서 볼 때 비로소 17세기 중반에 이르러 도박에서의 경우의 수, 그리고 몫 나누기와 같은 문제를 전문적인 수학자들이 체계적으로 다루게 되었다는 것이다(물론 도박은 그보다 수천 년 전부터 거의 모든 곳에서 성행했었다). 한편 이 시기 확률의 역사는 비수학적인 통계학사 연구에서도 중요한 주제로서 과학사학자인 Lorraine Daston은 1650년대에 시작된 확률 연구를 ‘고전적 확률’(classical probability)의 시작이라고 부르고 있다. 그녀는 1650년대부터 1840년대까지 거의 200년에 걸친 확률 연구를 관념(ideas)의 역사라는 측면에서 ‘합리성’(reasonableness), 또는 ‘상식’(common sense)을 수학적으로 표현하려는 근대적인 시도라고 해석하였다(Daston, 1988). Hacking(1975), Daston(1988)등에 따르면 이 시기 이전에 이미 오늘날 우리가 주관적, 객관적 확률(subjective and objective probability)이라고 표현하는 두 종류의 확률이 존재했었는데 그 가운데 상대도수로 표현되는 객관적 확률은 드 무아브르의 책제목 *The Doctrine of Chances*(1718)에서도 나타나듯 보통 ‘chance’라고 불렸었고, 주관적 확률은 수량화되지 않은 인식론적인 것으로서 ‘probability’라고 불렸던 것으로 오랫동안 그 두 가지는 별개의 것이었다가 비로소 18세기 초에 probability가 chance와 마찬가지로 수학자들의 관심

대상이 되기 시작하였다는 것이다. 위와 같은 이유 때문에 이 시기를 확률의 역사를 중심으로 볼 때 1708년을 기준으로 이 시기를 다시 둘로 구분할 수 있다. 먼저 1650년대부터 1708년 이전까지는 순열조합이론을 가지고 도박에서 생기는 여러 문제에 답하려했었던 시기였다면 1708년 이후는 이항분포와 관련된 극한 정리들이 등장하는 한편 그 이론들이 도박만이 아니고 (제한적인 것이기는 했지만) 관습적으로 이루어지던 통계학적 추정을 이론적으로 뒷받침하는 작업과 연결되기 시작한 시기라 할 수 있겠다.

1650년대-1708: 앞에서 ‘확률의 시기’를 둘로 구분할 수 있다고 하였는데 1708년 이전까지 약 반세기 가운데 주목해야할 시기는 1654년부터 처음 10년의 기간이다. 1654년이 확률의 역사에서 중요한 이유는 ‘두 사람이 일련의 게임으로 이루어진 도박을 하다가 중간에 그만두어야 할 경우 각자의 몫을 어떻게 나눌 것인가’라는 문제(보통 ‘problem of points’ 또는 ‘division problem’이라고 불린다)에 대해 파스칼(B. Pascal)과 페르마(P. Fermat)가 편지를 주고받은 해가 바로 1654년이었기(Todhunter, 1865, pp. 7-21; Hald, 1990, 54-64) 때문이다. 이어서 1657년에 호이헌스(C. Huygens)가 도박과 관련한 ‘수학적 기대값’에 대한 연구를 내놓았는데(Hald, 1990, 65-80) 이것이 바로 확률 이론을 처음으로 문헌을 통해 발표한 최초의 사례로 간주되는 글이다. 여기서 우리는 다음과 같은 문제들을 제기할 수 있다:

- (1) 1660년을 전후한 시기에 등장한 확률은 어떤 것이었는가?
- (2) 왜 그때 확률이 등장했는가?
- (3) 이때 등장했던 확률에 대한 연구가 겨우 10년 남짓 계속되다 18세기초에 이르는 거의 반 세기동안 중단되었던 이유는 무엇인가?

Todhunter(1865) 이후 대부분의 통계학자들은 (1)에 대해서는 상세히 다루고 있지만 (2), (3)에 대해서는 별다른 언급을 하지 않고 있다. 먼저 첫 번째 문제부터 살펴보고 하나씩 간단하게 답을 제시해보자. 최초로 확률을 연구한 사람들이 했던 일은 도박 문제를 중심으로 순열조합이론을 써서 여러 가지 사건의 경우의 수를 구하고 이를 가지고 기대값을 얻는 것이었다. 당시 이러한 ‘확률’은 ‘probability’가 아니고 ‘chance’라고 불렸었는데 그들이 연구한 것은 이항분포, 그리고 다항분포에 바탕을 둔 여러 사건의 확률과 기대값이었다. 그러므로 당시 연구자들에게 확률은 데이터 분석이나 통계학적 추론과는 일단 무관한 것이었다. 하지만 이항분포에 대한 그들의 연구는 반 세기 뒤인 18세기 초 드 무아브르에 의해 1712년의 포아송 근사와 1733년의 정규분포 근사를 유도하는 단계로 이어진다(Hald, 1990, pp. 213-216, pp. 485-489). 물론 이러한 분포들 역시 처음 탄생한 이후 18세기가 거의 다 가도록 통계학적 추론과는 무관한 상태에 머물러 있게 된다(Hald, 1990, pp. 489-495). 우리가 여기서 확인할 수 있는 사실은 수학자들이 개발한 확률 이론과 정작 그 이론의 통계학적 활용 사이에는 상당한 시차가 있었다는 점이다.

다음으로 ‘왜 1650년대에 확률이 탄생하였는가?’라는 질문에 대한 단순한 답은 수학의 역사에서 바로 당시에 순열조합이론이 일정한 틀을 갖추었기 때문에 이를 확률 계산에 이용할 수 있었다는 사실이다(Hald, 1990, pp. 42-54). 한편 우리는 파스칼과 페르마 그리고 호이헌스의 작업이 도박이라는 특수한 문제를 대상으로 하였다는 사실에도 주목하게 되는데, 이로 인해 Grattan-Guinness(1997, p. 284-285)와 같은 수학의 역사, 또는 Maistrov(1974, pp. 7-8)가 열거한 여러 통계학 연구들에서 도박이야말로 확률의 모태였다는 주장을 쉽게 찾아볼 수 있다. 반대로 도박에서 확률이 유래했다는 주장을 반박하는 연구자도 있는데 Loenid E. Maistrov가 대표적이다. 그는 도박의 역사는 거의 인류 역사의 시작과 동시에 시작되었기 때문에 17세기 중반이라는 특정

시기의 도박만을 강조하는 것은 확률이 여러 학문의 틀과 문제들과의 연관성 속에서 탄생했던 역사 전체를 보지 못하는 오류라고 비판하였다. 대신 그는 도박과 같이 편리하기는 하지만 너무 단순한 동기보다는 자본주의 경제의 발달사 속에서 확률과 통계학이 출현하는 과정을 주목해야 한다고 주장하였다(Maistrov, 1974, pp. 3-15). 한편 Ian Hacking은 과학철학자의 입장에서 확률과 관련된 관념들이 형성되는 과정을 강조하였다(Hacking, 1975). 도박을 강조하는 연구들에는 당시 도박이 성행했었다는 아주 짧은 언급이 있을 뿐인데 확률의 탄생 역사를 연구하기 위해서는 당시 유럽 각국 사회의 모습, 당시 사람들의 사고방식과 일상적인 생활상에 대한 세밀한 연구가 병행해야 할 것으로 보인다.

마지막으로 우리는 이 시기에 확률이 탄생했다는 역사만을 기억하지 말고 겨우 십 년 정도의 짧은 기간 동안 당대의 대수학자들이 연구하기 시작했던 확률 연구가 시작에 그친 채 계속 이어지지 못하고 꽤 오랫동안 중단상태에 빠지고 말았다는 사실에도 주목해야 한다. 통계학사를 연구하는 입장과는 달리 일반 과학사에서 17세기 후반에서 18세기 초에 이르는 시기는 '확률의 시대'가 아니고 '뉴턴의 시대'였다. 또한 수학의 역사에서 본다면 이 시기는 뉴턴과 라이프니츠의 '미적분학(calculus)의 시대'였다. 바야흐로 「프린키피아」(*Philosophiae Naturalis Principia Mathematica*, 1687)로 대표되는 뉴턴의 시대가 열린 것이다. 확률이 탄생한 1650년대는 아직 1642년 생인 뉴턴이 등장하기 이전이었는데 겨우 갓 태어난 확률은 뉴턴으로 말미암아 수학 연구자들의 관심 밖으로 밀려나고 말았다(뉴턴은, 나아가 당시의 영국인들 대부분은, 확률 연구에 거의 아무런 관심도 없었던 것으로 전해진다(David, 1962, pp. 123-129)). 그러다가 확률 연구가 다시 시작된 것은 18세기 초인데, 이는 확률이 탄생한지 무려 반세기나 지난 뒤의 일로서 이번에는 몽모(P. R. Montmort)와 베르누이 일가의 사람들로 대표되는 대륙의 학자들뿐만 아니라 드 무아브르와 심프슨처럼 영국에 있던 학자들도 일정한 역할을 하게 된다.

한편 유럽 대륙에서 수학적 확률이 태어난 것과 거의 같은 시기 영국의 그랜트(John Graunt)는 런던 교구에서 작성한 사망기록을 연구하여 *Observations upon the Bills of Mortality* (1662)를 출판했다 (원래 이름은 *Natural and Political Observations mentioned in a following Index, and made upon the Bills of Mortality*라는 긴 이름이다). 그는 사망원인에 따른 사망률, 도시와 시골, 성별 사망률 등을 비교하고 잔여기대수명을 계산하여 생명표(life table)도 만들었는데 이와 같은 작업은 대륙의 확률 연구와 아무런 상관없이 이루어진 것이었다. 그랜트의 작업이 있게 된 역사적 배경은 전염병에서 찾을 수 있다. 일찍이 1340년대 말 대규모 흑사병이 유행한 이래 약 400년 동안 기근, 전쟁과 함께 전염병은 유럽에서 주요 사망원인 중에 하나였다. 그에 대비하여 영국 런던에서는 일찍이 1530년대부터 전염병으로 인한 사망자 수를 교구의 성직자들이 매주 보고하도록 했는데 그 방대한 규모의 자료를 통계적으로 정리하고 분석한 최초의 사람이 그랜트였다(Hald, 1990, pp. 82-85). 그의 책에는 앞서 대륙에서 등장한 확률이론을 이용한 분석은 들어있지 않았지만 나름대로 후대에 중요한 영향을 미쳤다. Andrs Hald는 통계학사에서 *Bills of Mortality*가 중요한 이유를 세 가지로 정리해서 제시한 바 있다. 첫째, 정치경제학(political arithmetic), 둘째, 통계적 비율의 안정성에 대한 검증(testing the stability of statistical ratios) 그리고 마지막으로 기대수명과 생존확률 계산(calculation of expectations of life and survivorship probabilities)이 그것들이다(Hald, 1990, p. 6). 그랜트의 책은 출판 직후인 1660년대 중반 영국에서 전염병이 대규모로 유행하면서 더욱 주목받게 되었고 나아가 교구차원의 조사가 아닌 정부 차원에서의 공식적인 통계 기관의 필요성을 인식하게 만들었다. 그랜트로 인해 1662년이라는 해는 사회 통계의 역사에서 매우 중요하며, 또한 그가 시간에 따른 자료를 분석했다는 점에서 시계열분석의 시초로도 간주된

다(Klein, 1997).

1708-1750년대: 18세기 초에 들면서 백 년 뒤인 19세기 초 라플라스의 *Théorie analytique des probabilités*(1812)가 나올 때까지 확률 이론의 교과서 역할을 하게 되는 책들이 잇달아 등장했는데 몽모의 *Essay d'Analyse sur les Jeux de Hazard*(1708), 야콥 베르누이의 *Ars Conjectandi*(1713), 드 무아브르의 *Doctrine of Chance*와 같은 책들이 바로 그것이다. 그 가운데 야콥 베르누이의 *Ars Conjectandi*(Art of Conjecture, 즉 「추측술」이라는 뜻의 라틴어 제목이다)는 확률이론 연구서 가운데 확률과 통계학의 역사에서 주목받는 거의 최초의 것이라 할 수 있다. 이 책은 17세기의 확률 연구에 대해서는 아예 본문에서 언급조차 하지 않은 Stigler(1986)의 책에서까지 꽤 상세히 소개된 것으로, 베르누이의 책이 통계학사에서 중요한 이유로 두 가지를 들 수 있다. 먼저 이 책에서 ‘확률’의 개념이 확대되었다는 점이다. 베르누이가 정의한 확률은 ‘degree of certainty’, 즉 인간 지식의 한계로 인해 완전히 확실하게 알 수 없을 경우 불완전하게나마 알 수 있는 정도를 뜻했는데 이로 말미암아 베르누이는 최초로 ‘주관적 확률’(subjective probability)을 명시적으로 정의한 사람으로 일컬어지게 되었다. 두 번째는 확률 이론의 역사에서 최초로 극한 정리(저 유명한 베르누이의 weak law of large numbers)를 증명한 점이다. 이 정리는 확률 이론의 역사에서도 중요하지만 (물론 Stigler(1986, pp. 77-78)가 지적하듯 오늘날의 것과 차이가 있고 분명한 한계가 있다) 구간추정, 역확률 등 후세의 통계학과의 연관성이라는 측면에서도 매우 중요한 정리로 평가되기도 한다(Hald, 1990, pp. 245-254). 베르누이의 책이 나오고 나서 극한 정리에 대한 연구는 드 무아브르의 이항분포에 대한 극한분포 연구로 이어졌고 그의 책 *Doctrine of Chance*에서 정규분포 밀도함수로 결실을 맺게 된다(1718에 초판, 1738년에 제2판, 그리고 1756년에 제3판이 나왔는데 정규분포가 처음으로 등장한 것은 제2판이다). 그런데 당시의 확률 연구 역시 여전히 도박으로부터 자유롭지 못했던 모양인지 그 책의 맨 첫 문장이라는 것이 ‘이런 책이 도박을 부추긴다는 비난도 있지만 꼭 그런 것은 아니며 도리어 도박의 득실을 알려줌으로써 도박을 막는 역할을 할 수 있을 것’라고 되어 있을 정도이다(de Moivre, 1756, 페이지 번호가 붙어있지 않은 헌사). 전체 통계학사를 기준으로 볼 때, 결과적으로 야콥 베르누이와 드 무아브르의 연구는 데이터의 배후에 숨은 미지의 확률에 대한 확률적인 추론보다는, (참값은 이미 알고 있다고 가정하고) 데이터에서 계산하는 비(ratio)나 승수(odds) 등에 대한 확률적인 추론만을 생각한데다 이항분포라는 특정 분포에만 집중적으로 매달렸으므로 통계학적 추론을 위해서는 분명한 한계를 갖는 것들이었다(Stigler, 1986, pp. 85-88).

2.2. 오차이론의 시대 (1750-1820)

먼저 1750년대라는 시기에 대해 알아보자. Anders Hald는 자신의 두 책 Hald(1990, 1998)을 나누는 경계로 1750년대를 택했는데, 그가 1750년대를 선택한 이유는 다음 세 가지이다(Hald, 1990, pp. 7-8). 먼저 이 시기 이후의 확률 연구는 베이즈와 라플라스의 역확률 연구를 필두로 이전과 다른 모습을 보여준다; 두 번째로 1750년대 이후 오차이론과 추정 분야에서 통계학이 수학 이론과 결합한다; 마지막으로 1750년경부터 더 정확한 수명 관찰이 가능해지면서 생명보험이 발달하기 시작했다는 점이 그것들이다. 또한 Hald는 통계학의 역사에서 지금까지 모수통계학적 추론(parametric statistical inference)에서 있었던 세 번의 ‘혁명’(revolutions)으로 라플라스의 역확률을 이용한 추론(1774), 라플라스와 가우스의 최소제곱법과 중심극한정리에 대한 연구(1809-1812), 그

리고 마지막으로 가능도함수(likelihood function)에 대한 R. A. Fisher의 연구(1922)를 (Hald, 1998, pp. 1-8) 제시한 바 있는데 지금 우리가 살펴볼 시기는 그 세 차례의 혁명 가운데 처음 두 혁명이 있었던 시기에 해당한다. 한편 '선형관계 적합'의 역사는 마이어(Tobias Mayer), 보스코비치(Roger Boscovich), 라플라스를 거쳐 1805년 르장드르(Adrien Marie Legendre)의 최소제곱법에 이르게 되는데, 최소제곱법을 포함하여 여러 가지 선형모형의 역사를 연구한 Farebrother(1999)의 책이 1750년부터 시작되는 이유는 두 가지이다. 먼저 1750년은 마이어의 연구가 나왔던 해(Farebrother, 1999, pp. 9-17)이자 크래머(Gabriel Cramer)가 오늘날 Cramer's rule이라고 불리는 선형방정식을 푸는 방법을 발표한 해였다(Farebrother, 1999, pp. 3-4).

먼저 그 이전의 시기와 관련해서 본다면 이 시기는 라플라스로 대표되는 대수학자들이 확률 이론을 크게 발달시킨 시기이다. 라플라스는 1827년에 죽었지만 그의 연구는 상대적으로 확률 이론 연구의 정체기라 할 수 있는 19세기말까지 확률 연구를 지배했다고 일컬어진다(Maistrov, 1974, pp. 158-161). 그와 같은 확률 연구는 이미 앞 시기에 시작된 것이었던 반면, 데이터를 다루는 통계학 연구는 이 시기에 비로소 새롭게 시작되었고 기존의 확률 이론 연구 성과들과 결합하여 큰 진전을 이루게 된다. 이 시대는 확률의 역사에서는 라플라스와 가우스로 대표되는 수학자들이 중심극한정리를 비롯한 뚜렷한 연구들을 남겼던 시기인 한편 통계학의 역사에서는 천문학과 측지학을 중심으로 오차를 설명하는 통계학적 방법이 등장한 시기이다. 특히 수리 통계학의 역사에서는 라플라스에 의해 두 가지 방법, 즉 그 값을 알든 모르든 관심 대상인 모수를 고정된 값으로 보는 방법(direct probability)과 모수를 확률 변수로 간주하는 방법(inverse probability)을 이용한 추론 과정이 개발된 시기이다. 1750년에서 1805년 사이의 역사는 Hald(1998)의 제1, 2부, Stigler(1986)의 제 1, 2, 3장에서 상세히 다루고 있는 내용으로 정규분포, 산술평균의 분포, 유의성검증, 최대가능도추정법, 데이터에의 방정식 적합, 그리고 베이즈를 시작으로 주로 1770년대 이후 라플라스의 역확률 연구와 같은 주제가 거기에 포함된다. 여기서 열거한 여러 사항으로만 언뜻 판단하면 19세기가 시작되기 전에 이미 통계학에서 중요한 것들이 거의 다 태어난 것처럼 보일 정도이다. 그런데 실은 그렇지 않았다. 당시의 역사에서 우리가 주목할 점은 어떤 이론이나 방법이 단지 '태어났다'는 사실의 역사뿐이 아니고 '그것들이 당시에 했던 역할이 실제로 과연 무엇이었는가'라는 것이다. 알고 보면 당시에는 정규분포만 하더라도 단지 이항분포의 극한분포로만 알려져 있었으며, 최소제곱법이 개발되어 널리 쓰이고 있었지만 왜 좋은 방법인가에 대한 이론적인 배경은 전혀 마련되어있지 않은 상태였다. 또한 당시의 역확률 역시 단지 모수를 고정된 값으로 보지 않았을 뿐 사전분포는 모두 균등분포(uniform distribution)에 국한되어 있었을 뿐더러 역확률을 이용해서 추정했을 때 언제 산술평균을 추정량으로 얻게 될지도 알려져 있지 않았다. 통계학적 추론을 위한 좋은 도구들이 거의 다 마련되어 있는데도 아직 껴지 못한 상태였던 것이다. 통계학사에서 1809년부터 1820년대 사이의 기간 동안 라플라스와 가우스가 했던 역할이 결정적으로 중요한 이유는 바로 18세기를 지나 19세기 초에 접어들 당시 확률과 통계학이 가졌던 이와 같은 본질적인 한계 때문이다.

이처럼 통계학과 확률이론에서 지극히 중요한 많은 연구가 등장했던 시대를 이 글에서 간단하게 '오차이론의 시대'라고 부르려는 이유는 다음과 같다. 먼저 당시 과학계 전반의 지적 분위기에 '결정론'(determinism)이 차지했던 비중을 언급해야겠다. 세상의 모든 것들은 인간의 지적인 능력이 못 미쳐서 확실히 알 수 없을 뿐 사실은 엄밀한 인과법칙을 따른다는 결정론적 사고는 특히 뉴턴이 등장한 이후 과학계 전체에 널리 퍼져 있었다. 당연히 라플라스를 비롯한 확률을 연구하던 사람들도 예외는 아니었으니, 그들의 목표는 측정을 통해 단 하나의 참값을 알아내는 것이었다.

따라서 그들은 측정으로 얻은 데이터를 참값에 오차가 덧붙은 것으로 생각했으므로 참값을 정확히 찾아내기 위해서는 그 오차를 설명해내야만 했다. 바로 그 목적을 위해 여러 가지 분포들이 만들어지고 최소제곱법을 비롯한 통계적 추정방법과 확률이론이 개발되었던 것이다. 이와 같은 시대적 배경을 이해하는 것은 ‘오차이론이 왜 천문학과 측지학에서 발달하였는가?’라는 질문에 답하는데도 필요하다. 당시 지구의 모양, 크기를 알아보려는 측지학은 당연히 과학의 주류분야라기보다는 항해와 같은 목적을 위해 시급히 해결해야 할 실제적인 과제 가운데 하나였다. 예컨대 바다에서 경도를 재는 문제는 엄청난 현상금이 걸렸던 문제였는데 마이어도 그 문제에 몰두한 결과 상금을 나눠 가진 사람 가운데 하나였다(Farebrother, 1999, p. 11). 그리고 측지학은 프랑스 과학이 유럽의 중심이었던 18세기 후반부터 19세기 초까지 프랑스의 의욕적인 탐사 및 연구 활동과 분리해서 볼 수 없다. 18세기 중 후반 여러 차례에 걸쳐 경도의 길이를 재기 위한 탐사대가 파견되었고 그 결과를 보스코비치, 라플라스, 그리고 르장드르가 연구하는 과정에서 최소제곱법을 비롯한 추정방법이 개발되었던 것이다.

우리가 여기서 잠깐 통계학 내부뿐 아니라 그 주변을 둘러볼 때 통계학사에서 별로 다루어지지 않았던 중요한 문제 가운데 하나를 만나게 된다. 그 문제는 이 시기 최소제곱법을 비롯한 통계학적 추정 방법이 개발되는 과정에서 오차를 설명하려는 통계학적 시도가 과연 다른 학문들과 어떤 관계에 있었는지에 대한 것이다. 먼저 18세기 후반 과학의 전반적인 면모를 잠깐 살펴보자. 과학사 연구자들이 흔히 ‘과학혁명’(scientific revolution)이 있었다고 일컫는 시기는 16, 17세기인데 그 중에서 더 중요하게 평가되는 시기는 17세기이다. 앞서 보았듯 비록 17세기에 확률이 탄생했다고는 하나 그 역할은 지극히 제한된 것이었으며 천문학과 측지학에서 통계학적 추정 방법이 개발된 시기는 18세기 후반 이후의 일이었으므로 확률과 통계학은 과학혁명에는 아무런 기여도 못했던 셈이다. 그리고 과학혁명을 대표하는 분야가 천문학, 역학, 생리학, 그리고 수학이었다고 할 때(김영식, 1984), 당연히 그 수학에는 확률과 통계학이 포함되어 있지 않았다. 오늘날의 통계학 전공자들은 측정이 이루어진 모든 곳에서 통계학이 중요한 역할을 하였다고 믿는 경향이 있는 듯 한데 근대 과학 전체로 범위를 넓혀볼 때 측정이 중요한 역할을 한 분야에서도 통계학의 역할은 상당히 제한적이었던 셈이다. 요컨대 1820년대까지 대단히 중요하고 많은 확률, 통계학 이론이 만들어졌지만 여러 학문 분야에서의 그 활용도는 아직 매우 제한적이었다. 또한 그러한 이론을 만드는데 주요한 역할을 한 사람이 모두 당대 최고 수준의 수학자들이었기는 했지만 그들 가운데 적어도 ‘통계학자’라 할 만한 사람은 아직 아무도 없었다. 이 시대와 다음 시대를 가르는 가장 두드러진 차이가 바로 여기, 즉 통계학사에서 수학 바깥에 있는 사람들의 역할 여부에 있다.

2.3. 통계의 시대 (1820-1880)

지금까지 이 글에서 가장 많이 언급된 글은 Anders Hald의 두 책(Hald, 1990, 1998)이었다. 선부른 판단인지 모르겠지만 확률의 역사와 수리통계학의 역사를 연구한 책 가운데 그의 두 책을 넘어설 수 있는 책은 이전에도 없었고 아마 앞으로도 나오기 어려울 것이다. 하지만 19세기 초를 지나고 나면 우리가 자주 언급해야 할 통계학사 연구서는 다른 책으로 옮겨간다. 그 이전과 19세기 말과 비교했을 때 확률과 통계학 이론 분야의 연구가 상대적으로 지지부진했음에도 불구하고 이 시기부터 통계학은 바야흐로 수학자들만의(넓게 잡는다면 천문학자를 위시한 자연과학자들만의) 작업이 아니게 된다는 면에서 이 시기는 주목받아 마땅하다. 따라서 앞 시기까지의 역사에 비해 이 시기의 역사는 통계학 바깥의 연구자들로부터 더 많은 주목을 받았는데 그 결과물로는

Porter(1986), Krüger et al. (1987a, 1987b), Hacking(1990), Desrosières(1998)등이 대표적이다. 이들 가운데 가장 ‘거시적으로’ 이 시대 통계학사를 정리한 것은 ‘확률혁명’(probabilistic revolution)을 제목으로 한 Krüger et al. (1987a, 1987b)의 연구라고 할 수 있는데 그들이 확률혁명이 있었던 때로 논의한 시기는 1800년대부터 1930년대까지 약 130여 년에 이르는 기간이다. 그 중 시작시점인 1800년대는 확률과 통계학이 비로소 나름대로의 틀을 갖춘 시기로 선택되었고(Cohen, 1987, p. 37) 연구자들이 ‘확률혁명’이라고 부르는 내용적 변화는 케틀레(Adolphe Quetelet)로부터 시작하고 있으므로, 실제 논의의 시작은 케틀레가 활동하는 1820년대부터라고 볼 수 있다. 사실 그 책에 실린 다른 글에서 Ian Hacking은 확률혁명의 시기를 구체적으로 1823년(Joseph Fourier의 주도로 파리와 세느 현 지역의 통계조사보고서가 출판된 해이다)부터 1936년까지라고 더 명시적으로 제시하기도 하였다(Hacking, 1987, p. 48).

‘확률혁명’이라는 이름은 말할 필요도 없이 ‘과학혁명’에서 그 이름을 빌려 온 것이다. 과학혁명은 과학사에서 대단히 중요한 주제 가운데 하나였는데(김영식, 1984) 과학사학자인 I. Bernard Cohen은 전체 과학을 변혁시키는 ‘과학혁명’(scientific revolution-대문자로 Scientific Revolution이라고 쓸 때는 대개 16, 17세기의 과학혁명을 뜻한다), 그리고 어떤 한 분야에서 커다란 변화를 가져오는 ‘과학에서의 혁명’(revolution in science), 마지막으로 혁명보다는 점진적인 변화를 뜻하는 ‘진화’(evolution)라는 세 가지를 구분하면서 다음과 같이 주장했다: (i) 17세기 중반에 확률은 새로운 것이 탄생한 것이므로 기존의 틀에 급격한 변화를 초래한 혁명도 아니고, 또한 기존의 틀을 점진적으로 바꾼 진화라고 볼 수도 없으므로 Hacking(1975)의 표현을 빌려 ‘출현’(emergence)이라 부를 수 있으며, (ii) 약 130여 년 동안 확률이론에서는 큰 발전이나 변화가 없었으나 그 이론이 통계학에 적용되는 과정에서는 혁명이라 부를 만한 사례들이 있었으며, (iii) 그보다 더 주목할 것은 계량적인 통계학적 사고가 자연과학과 사회과학 모두에서 큰 변혁을 일으켰는데 이를 ‘활용에 의한 혁명’(revolution by application)이라 부를 수 있다(Cohen, 1987, p. 37). 한편 Ian Hacking은 쿤(Thomas Kuhn)이 말하는 제2의 과학혁명(second scientific revolution)과 과학혁명의 관계를 탐색하면서 쿤이 말하는 제2의 과학혁명의 시기는 1800년부터 1850년 사이인데 첫 번째 과학혁명이 부르주아 계급의 부상과 관련이 있다면 두 번째 혁명은 자본주의의 발흥과 산업혁명과 깊은 관련을 갖는 것으로서 이 시기에는 전례 없이 축정이 중요한 역할을 하였으며 온갖 것을 수량화해서 측정하기 시작하였다고 주장하였다.

사실 이 시기 확률과 통계학이 활용되었던 분야의 폭을 고려하면, 이전 시기의 확률 연구나 오차 이론은 사실 매우 좁은 영역에 국한된 역할만 했었다고 볼 수 있다. 그러다가 19세기 초반에 이르러 유럽 각 나라에서는 지역별, 나라별로 통계조사가 활발하게 실시되기 시작했고 무엇보다 그 결과들이 인쇄물의 형태로 대대적으로 널리 발표되기 시작했다(Hacking, 1990). 거기에는 국민국가(national state) 형성이라는 시대적인 분위기가 크게 작용했는데 이 시기가 주목받는 데에는 당시 각국 정부의 통계업무를 맡은 사람들이 단지 통계자료를 조사, 수집하고 발표하는 데 그치지 않고 그 자료에 확률이론을 적용하려는 시도를 하였다는 점이다. 벨기에의 케틀레(Adolphe Quetelet)가 19세기 후반 동안에 했던 작업은 그 가운데에서 가장 두드러진 것이었는데 무엇보다 어떤 사회 전체를 대표하는 ‘평균적인 사람’(average man)이라는 그의 발명품은 1835년에 발표된 이후 대단한 호응을 받았다. 또한 케틀레로 인해 이전까지의 출생, 사망, 혼인과 같은 현상뿐 아니라 범죄, 재판, 자살 등으로 통계학의 범위를 크게 확장되었을 뿐만 아니라 이 시대에는 여러 나라의 통계조사방법을 비교, 토론하기 위한 국제적인 모임까지 처음으로 탄생하게 되었다. 그 기구가 바로 케틀레의 주도로 시작된 ISC(International Statistical Congress)이다. 이 기구는 오늘날

ISI의 진신에 해당하는데 케틀레의 모국인 벨기에 브뤼셀에서 그 첫 회의가 열린 해는 1853년이였다.

이처럼 통계학의 활용 범위를 크게 확대시킨 그의 공은 지금도 부정하기 어렵다. 하지만 본질적으로 케틀레의 시도는 Stigler(1986)가 잘 지적했듯이 성공적인 것은 아니었는데, 그의 결정적인 한계는 확률 이론이나 통계학에 대한 지식의 부족에 기인한다기보다는 결정론적 세계관이라는 시대적 한계 때문으로 보아야한다. 즉 그는 ‘평균적인 사람’에서 잘 드러나듯 하나의 이상적인 중심을 지나치게 강조하고 그 중심과 일치하지 않는 결과들을 모두 오차라고 보았던 것이다. 천문학에서는 그러한 관점이 문제가 아니었으나 사람들이 모여 이루어지는 사회에서 하나의 전범만을 내세우고 나머지 모든 사람을 오차적인 인간으로 볼 수는 없는 일이었다. 이러한 케틀레의 한계는 사람에 대한 개별 측정 결과를 어떤 모범적인 값에 오차가 덧붙은 것으로 보는 대신 그 관측값 자체에 주목하게 되는 골턴(F. Galton)의 시각과 가장 분명한 대조를 보이는 것으로서 골턴이 나타나려면 평균적인 사람이 나타난 시기보다 반세기 뒤인 19세기말까지 기다려야 했다(Hilts, 1973). 여기서도 우리는 수학적 이론이나 통계학 내부의 변화보다는 걸음으로 잘 드러나지 않는 그 시대의 관념이 학문의 전반적인 흐름을 결정하는데 큰 역할을 했던 사례를 볼 수 있다. 결국 이 시대는 비록 후대에 ‘통계에 대한 열정의 시대’(Westergaard, 1932)라고 일컬어질 정도로 통계 조사가 활발했던 시대였지만 확률 이론과 통계자료 사이의 관계가 제대로 연결되지 못했었다는 면에서 ‘통계학’의 시대라기보다는 ‘통계’의 시대라고 부르는 편이 나올 듯하다. 확률 이론이 데이터분석에까지 활용되어 데이터에서 정보를 추론해내는 작업이 본격적으로 이루어지려면 19세기말 영국에서 비로소 ‘통계학의 시대’가 열리기까지 기다려야 했다. 아래에서는 이 기간 가운데 자주 언급되는 몇 시기를 살펴보자.

1820-30년대: 이 시기는 여러 통계학사에서 주목받는 시기이다. Stigler(1986)는 라플라스가 죽은 1827년을 강조했는데 사실 「확률의 해석 이론」(*Théorie analytique des probabilités*, 1812)을 낸 이후 라플라스는 그의 평생에 걸친 역작으로 평가되는 「천체역학」(*Traité de mécanique céleste*; 1799-1805년 사이에 네 권으로 발행되고 1825년 한 권의 부록이 추가로 발행되었다)의 부록을 마무리하느라(Stigler, 1986, p. 162) 확률에 대해서는 거의 연구하지 않았기 때문에 Stigler가 1827년을 강조한 이유는 1827년이 확률과 통계학의 역사에서 중요한 연구가 있었던 해라기보다는 단지 라플라스에 대한 존경의 뜻을 나타내기 위해서라고 볼 수 있다. 한편 1820년대는 Porter(1986)의 책이 시작되는 시기이다. 놀랍게도 라플라스를 아예 빼고 시작하는데서 볼 수 있듯 수리통계학은 Porter의 관심사가 아니다. 대신 그의 연구에서는 사회통계가 핵심적인 부분을 차지하는데 앞서 밝혔듯 사회통계의 초기 역사에서 주인공은 케틀레이다. 확률과 통계학에 대해 아무 것도 알지 못했던 케틀레가 파리에 가서 프랑스 수학자들로부터 확률과 표본을 이용한 인구조사에 대해(그것도 겨우 몇 달간) 배우고 온 해가 1823년이었는데 통계에 매료된 케틀레는 바로 그때부터 유럽 최고의 통계학전문가이자 통계사업가 역할을 하게 된다. 간단히 말해서 1820년대는 통계학사에서 라플라스라는 대이론가가 사라지고(즉 앞에서 ‘오차이론의 시대’라 부른 시대가 막을 내리고), 케틀레라는 대사업가가 등장한 시기라고 할 수 있다. 케틀레는 라플라스가 천문학을 대상으로 주도적으로 연구했던 오차이론을 사회 데이터에 적용하려 하였는데, 창조적인 시도가 흔히 그러하듯 그 결과는 성공적이지 못했다. 한편 케틀레의 평균적인 사람이 태어난 해가 1835년이었고 이 가상의 ‘사람’으로 인해 통계적인 사고가 널리 퍼지기 시작하였으므로 1830년대를 19세기 통계학에서 중요한 시기로 잡는 경우도 종종 볼 수 있다.

1840년대: 이 시기는 과학사가인 Lorrain Daston이 그녀의 책에서 ‘고전적 확률의 시대’의 마지막 막으로 명시한 시기로서, 그녀가 고전적 확률의 시대라고 부른 시기는 1650년대부터 1840년대까지 이르는 약 200년 동안이다(Daston, 1988). 그녀는 이 시기의 확률을 한 마디로 ‘계산으로 나타낸 상식’(good sense reduced to calculus) 또는 ‘합리적인 계산’(the reasonable calculus)이라고 일컬었는데 그녀가 1840년을 고전적 확률의 마지막 시기로 정한 이유는 다음과 같다. Daston에 의하면 고전적 확률 시대의 마지막을 대표하는 사람은 포아송(Siméon-Denis Poisson)인데 1837년 포아송이 재판의 유죄결정과정에 확률을 활용하는 문제에 대한 글을 발표했을 때 밀(John Stuart Mill)을 비롯한 여러 철학자, 수학자로부터 그러한 시도는 상식에 반하는 것으로 ‘수학의 진정한 불명예’(the real opprobrium)이자 ‘지성의 일탈’에 해당하는 잘못된 것이라는 비판이 일어났고(Gigerenzer, et al. 1989, pp. 34-35), 이에 따라 확률과 그 활용 사이에 간격이 생기면서 결국 확률 자체가 주관적 확률과 객관적 확률로 구분되어 합리적인 상식을 표현하는 고전적 확률의 시대가 끝났다는 것이다(Daston, 1988, pp. 296-386). 우리는 17세기 중반 이후 줄곧 확률의 적용 범위가 늘어나기만 했다고 생각할 수도 있는데 이 시기부터는 결과적으로 점점 증거, 판단, 원인 등에 대해서는 확률을 적용하지 않게 되었으므로 어떤 면에서는 확률의 활용 폭이 줄어든 셈이다(Gigerenzer, et al. 1989, p. 36). 1840년에는 그 포아송마저 ‘고전적 확률과 더불어’ 사망했다.

2.4 통계학의 시대 (1880년대 이후)

사실 19세기 중엽까지의 통계학사는 통계학을 전공으로 공부하는 사람들에게조차 내용과 목적, 실제 활용 분야, 그리고 주요 인물 등이 모두 낯설게 느껴질 정도인데 통계학적 추론의 측면에서 본다면 지금까지 살펴본 역사는 19세기말부터 20세기 초에 영국을 중심으로 크게 발달하게 되는 현대통계학을 위한 준비과정에 지나지 않았던 것처럼 보인다. 앞에서 살펴본 시기까지는 이론과 활용 양쪽에서 유럽에서 이루어진 통계학의 발달사에서 영국이라는 나라는 거의 아무런 역할도 하지 못했었지만 19세기 말에 이르면 상황은 전혀 달라진다. 골턴(Francis Galton), 에지워스(F. Y. Edgeworth), 윌(G. U. Yule), 피어슨(K. Pearson), 고세트(W. Gosset) 그리고 피셔(R. A. Fisher)등이 19세기말부터 현대통계학의 기초를 마련했다는 점에 대해서는 이론이 없기 때문이다. 나아가 대학(University College London)에 통계학과가 처음 생긴 것도 바로 이 시기(1911년) 영국에서의 일이었으니 통계학은 이 시기 영국에서 비로소 하나의 독립된 학문으로서 탄생한 셈이다. 이와 같이 19세기말부터 시작된 통계학의 발달과정은 매우 다양하고 복잡한 모습을 띠게 되었으므로 당연하게도 20세기 통계학의 역사를 총체적으로 연구한 결과물이 나오기는 매우 어렵게 되었다. 따라서 대개의 통계학사들은 1900년 부근(Stigler, 1986; Porter, 1986; Farebrother, 1999), 또는 1930년대 무렵(Hald, 1998; Krüger et al., 1987a, 198b, Klein, 1997)에서 논의를 끝내고 있다.

먼저 19세기말 영국 통계학이 발달하게 된 시초를 찾는 문제는 어렵지 않다. 여기서 중요한 것은 ‘누가 선구자인가’라는 것이 아니고 골턴을 선두로 한 일군의 연구자들이 당시 ‘무엇을 왜 연구했는가?’라는 문제이다. 우선 이 글에서 1880년대부터 ‘통계학의 시대’가 시작되었다고 하는 이유를 몇 가지 들어보자. 골턴은 그에 대한 평전을 쓴 D. W. Forrest가 표현했듯 ‘영국 빅토리아 시대의 천재’(Forrest, 1974)라는 이름을 붙이기에 손색이 없을 정도로 다양한 분야에서 새롭고 획기적인 아이디어를 많이 낸 인물인데 통계학사에서 우선 그를 주목하는 이유는 무엇보다 상관과 회

기에 관한 그의 연구 때문이다. 그 연구가 발표되었던 해가 바로 1885년이었는데(Stigler, 1986, pp. 281-299) 그의 연구는 곧바로 에지워스, 칼 피어슨, 그리고 울의 연구로 이어지면서 케틀레의 한계를 넘어 통계학의 새로운 지평을 크게 열게 된다. 이와 같은 골턴의 연구가 케틀레와 가장 크게 달랐던 점은 무엇보다 골턴의 경우에는 고정된 참값과 오차라는 결정론적 시각에서 벗어나 데이터 값 그 자체를 주목하게 된 점이라 할 수 있다. 사실 골턴은 거기서 더 나아가 케틀레처럼 전체를 대표할 수 있는 평범한 평균을 강조하는 대신 천재와 같이 특출한 경우, 즉 분포의 끝 쪽 부분에 더 주목하고 그러한 천재적인 능력이 어떻게 유전되는가에 더 큰 관심을 갖고 연구하였다. 이와 같은 골턴의 연구는 그가 우생학(eugenics)이라는 말을 새로 만들고 '인류의 품종개량'에 관심을 가졌다는 사실과도 그대로 부합한다(MacKenzie, 1981). 그런데 결정론이라는 오랜 사고틀을 벗어난 것은 골턴 혼자만은 아니었다. 이미 시대는 물리학자 맥스웰(James Clerk Maxwell)의 열역학 연구로부터 통계역학(statistical physics)이 1870년대에 등장하고, 어떤 궁극적인 원인이 아니라 독자적으로 작용하는 통계적인 법칙의 존재를 주장하는 철학자 피어스(C. S. Peirce)의 비결정론이 1890년대에 등장하는 시대가 된 것이다(Porter, 1986, pp. 193-227). 통계학사를 공부하는 과정에서 우리는 '왜 1900년 무렵이 책제목 같은 데서 자꾸 등장하는가'라는 물음을 당연히 갖게 되는데, 이상한 점은 Porter(1986), Stigler(1986), Farebrother(1999)과 같은 저자들이 1900년 부근에서 책을 끝내는 이유에 대해서 거의 아무런 언급도 하지 않고 있다는 점이다. 통계학의 내용이라는 측면에서 본다면 1900년 무렵을 경계로 19세기 이전까지의 '근대통계학'과 20세기 이후의 '현대통계학' 정도로 구분하는 편이 가장 설득력이 있을 것 같은데 이 경우 피어슨(Karl Pearson)이 그 두 시기를 잇는 역할을 맡는 셈이다. 대부분의 통계학사 연구들이 1900년을 전후한 시기를 책의 끝으로 삼았다는 사실 자체가 짐작컨대 연구자들이 명시적이지는 않더라도 근대와 현대 통계학의 상이한 면모를 염두에 둔 탓일 것이다.

한편 19세기말에 접어들면 확률과 통계학은 더 이상 같은 연구 안에서 한꺼번에 다루지기 어려운 상태에 이르게 된다는 점도 주목할 만하다. 이 시기 이후 두 분야가 모두 크게 발달하므로 각각의 깊이와 폭을 고려할 때 둘을 한꺼번에 다루기 어려워진 것은 당연한 노릇으로서 예컨대 Hald(1990, 1998)가 붙인 두 책의 제목에서도 그러한 사정은 잘 드러난다. 대단한 박학다식자인 Anders Hald조차 그가 쓴 두 책 가운데 첫 번째 책의 제목을 *A History of Probability and Statistics and Their Applications before 1750*라고 포괄적으로 붙였다가 나중에 나온 것에는 *A History of Mathematical Statistics From 1750 to 1930*와 같이 붙일 수밖에 없었던 것이다.

마지막으로 1930년대에 대해 간단히 언급하자. 이 시기는 Krüger et al. (1987a, 1987b)에서 제시한 확률혁명의 시기 1800-1930(또는 Ian Hacking의 1823-1936)의 마지막 시기이다. Hacking은 이 시기에 이르면 결정론적 세계관은 쇠퇴하고 우연 또는 우연의 법칙이 그 자리를 차지한다고 주장하면서 확률과 통계학에 의한 이와 같은 세계관의 변혁이야말로 역사상 그 어떤 변화보다 혁명적인 것이라고 주장했다(Hacking, 1987, p. 54). 한편 Hald(1998) 역시 통계학적 추정방법에 대한 R. A. Fisher의 1930년대까지의 연구를 소개하면서 1400쪽에 가까운 두 책을 끝낸 바 있다. 마지막으로 시계열분석의 역사를 쓰면서 그 연구 대상 시기를 명시한 Klein(1997)이 1938년에서 연구를 끝낸 이유는 그 이전까지의 확률 이론과 경험적 연구를 종합한 안정적 확률 과정(stationary stochastic process)에 대한 Herman Wold의 연구가 바로 그 해에 나왔기 때문이다(Klein, 1997, pp. 282-287).

3. 결론

앞서 서론에서도 밝혔듯 이미 적지 않은 통계학사 연구들이 나와 있고 그 연구들을 바탕으로 더욱 다양하면서도 매우 정밀한 연구들도 이미 나와 있다. 그 가운데 책으로 출판된 것만 예로 들더라도 통계학사에 등장하는 인물들에 대한 인명사전(Johnson and Kotz(1997), Heyde and Seneta(2001)), 특정한 인물에 대한 평전(P. S. Laplace에 대한 Gillispie(1997); I. J. Bienaymé에 대한 Heyde and Seneta(1977); F. Edgeworth에 대한 Mirowski(1994); F. Galton에 대한 Forrest(1974)와 Gillham(2001); R. A. Fisher에 대한 Fienberg and Hinkley(1980)), 특정 국가에서의 역사(영국에 대한 Cullen(1975), MacKenzie(1981); 이탈리아에 대한 Patriarca(1996)), 통계학사에서 중요한 문헌을 재수록한 책(Kotz and Johnson(1992a, 1992b, 1997), David and Edwards(2001)), 통계학 전반이 아닌 통계학의 특정 분야만을 대상으로 한 역사(역학률의 역사를 연구한 Dale(1999); 선형관계 적합의 역사를 연구한 Farebrother(1999))등을 들 수 있다. 이와 같은 세밀한 연구들은 그만큼 통계학사 연구가 깊이를 갖추어 간다는 증명일 터이다. 그런데 이 글은 그러한 정밀한 각론 연구가 아니라 도리어 총론적으로 통계학사 전체를 훑어보고 시대를 나누어 정리하려 시도한 글이다. 한 권의 책으로도 충분하지 못할 거창한 작업을 짧은 글에서 시도한 이유는 앞서 서론에서도 강조했듯 통계학사 연구가 전문화되면서 자칫 통계학 전공자를 소외시키는 듯 보이기도 했기 때문이다. 수입학문을 공부하는 입장에서 그 역사를 덮어둔 채 현재와 미래만 보고 갈 수는 없다는 당위는 부정하기 어렵지만 그래도 어쩌면 이러한 시도 자체가 시대착오적인 것 아닌가 싶기도 하다.

무엇보다 이 글에서 시도한 시대구분은 어떠한 엄밀한 기준에 따른 것은 아니라는 점을 강조해야겠다. 사실 네 개로 시기를 구분한 원래 의도는 통계학 내부의 이론이 발달해온 역사와 그 학문이 세상 속에서 했던 역할, 그 학문이 성장했던 사회적인 배경과 조건들을 함께 아울러 드러내는 틀을 만들고 싶다는 것이었다. 아쉽게도 결과는 의욕에 비해 미진하고 허술한 구석이 승승 보이는 글이 되고 말았다. 하지만 통계학이 하나의 독립된 학문 분야로 자리를 잡은 지 일 세기가 가깝게 지난 지금 시점에서 인문, 사회, 자연과학 등으로 분할이 이루어지기 이전의 통계학의 지난 역사를 되돌아보는 과정은 한편으로 통계학만이 가진 남달리 유연한 특성을 확인하는 계기가 될 수 있을 법도 하다. 때문에 장차 정치한 통계학사의 시대구분을 위해서는 다음과 같은 질문을 제기해 볼 수도 있겠다: 통계학사 연구에서의 시대구분은 다른 분야에서의 시대구분, 가령 사회과학이 출현하는 역사를 연구한 Richard Olson의 1642년부터 1792년까지의 시기와 어떤 관련은 없을까(Olson, 1993)? 또는 수학사를 수학의 주요 분야에 따라 나눈 Grattan-Guinness의 연구에서 제시한 시대구분과는 어떤 관련이 없을까(Grattan-Guinness, 1997)? 나아가 한 분야의 역사가 아니고 포괄적인 서양 근, 현대사에서의 시대구분, 예컨대 흄스봄의 혁명의 시대(1789-1848), 일본의 시대(1848-1875), 제국의 시대(1875-1914) 등과도 어떤 연관성을 가질까(Hobsbawm, 1962, 1975, 1987)?

끝으로 이 글의 주제와 가장 근접한 글들을 몇 편 소개하자. Krüger et al. (1987a)에 실린 Thomas Kuhn(Kuhn, 1987), I. Bernard Cohen(Cohen, 1987), Ian Hacking(Hacking, 1987), 그리고 Lorenz Krüger(Krüger, 1987)의 글이 그것들이다. 이 글들은 과학사에서의 '과학혁명'(scientific revolution)에 비견될만한 것으로서, 과연 '확률혁명'(probabilistic revolution)이라고 부를 만한 것

이 있었는가? 있었다면 그 시기와 성격은 어떠했는가? 그리고 그 '혁명'을 시기에 따라 몇 개로 구분할 수 있는가?' 등의 주제를 과학사 분야의 대가급 학자들이 다룬 것들이다. 또한 Stigler(1986)는 그의 통계학사 책을 세 부분으로 나누었는데 이 글에서 오차의 시대, 통계의 시대, 통계학의 시대라고 부른 두 번째, 세 번째, 네 번째 시기는 그 책에서의 1, 2, 3부와 각각 그 기간이 비슷하다. 하지만 이 글은 통계학 내부의 역사에 치중한 Stigler(1986)에 비해 사회적인 배경을 좀 더 강조하려 시도한 한 점, 그리고 Stigler(1986)가 다루는 1750-1900 사이의 시기보다 조금 더 넓은 시기를 살펴보려 한 점 등에서 그 책과 구별된다 할 수 있겠다.

참고문헌

- [1] 김영식 (1984). 「과학혁명: 근대과학의 출현과 그 배경」, 민음사, 서울.
- [2] 허명희 (1994). 14면 주사위의 확률, 「응용통계연구」, 제 7권 1호, 113-119.
- [3] Cohen, I. B. (1987). Scientific Revolutions, Revolutions in Science, and a Probabilistic Revolution 1800-1930. Krüger, L., Daston, L. J. and Heidelberger, M. (editors) *The Probabilistic Revolution, vol. 1. Ideas in History*, MIT Press, Cambridge, Massachusetts, 23-44.
- [4] Dale, A. I. (1999). *A History of Inverse Probability: From Thomas Bayes to Karl Pearson*. Springer, New York.
- [5] Daston, L. (1988). *Classical Probability in the Enlightenment*, Princeton University Press, Princeton, NJ.
- [6] David, F. N. (1962). *Games, Gods and Gambling: A History of Probability and Statistical Ideas*, Dover, New York.
- [7] David, H. A. and Edwards, A. W. F. (2001). *Annotated Readings in the History of Statistics*, Springer, New York.
- [8] De Moivre, A. (1756). *Doctrine of Chance*, third edition, Millar, London.
- [9] Desrosières, A. (1998). *The Politics of Large Numbers: A Study of Statistical Reasoning*, Harvard University Press, Cambridge, Massachusetts, *La Politique des grands nombres: Histoire de la Raison Statistique* translated by Camille Naish.
- [10] Farebrother, R. W. (1999). *Fitting Linear Relationships: A History of the Calculus of Observations 1750-1900*. Springer, New York.
- [11] Fienberg, S. E. and Hinkley, D. V. (editors) (1980). *R. A. Fisher: An Appreciation*, Springer, New York.
- [12] Forrest, D. W. (1974). *Francis Galton: The Life and Work of a Victorian Genius*, Paul Elek, London.
- [13] Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J. and Krüger, L. (1989). *The Empire of Chance: How Probability Changed Science and Everyday Life*, Cambridge University

Press, Cambridge.

- [14] Gillham, N. W. (2001). *Sir Francis Galton: From African Exploration to the Birth of Eugenics*, Oxford University Press, Oxford.
- [15] Gillispie, C. C. (1997). *Pierre-Simon Laplace 1749-1827: A Life in Exact Science*, Princeton University Press, Princeton.
- [16] Hacking, I. (1975). *The Emergence of Probability*, Cambridge University Press, Cambridge.
- [17] Hacking, I. (1987). Was There a Probabilistic Revolution 1800-1930? Krüger, L., Daston, L. J. and Heidelberger, M. (editors) *The Probabilistic Revolution, vol. 1. Ideas in History*, MIT Press, Cambridge, Massachusetts, 45-55.
- [18] Hacking, I. (1990). *The Taming of Chance*, Cambridge University Press, Cambridge.
- [19] Hald, A. (1990). *A History of Probability and Statistics and Their Applications Before 1750*, Wiley, New York.
- [20] Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*, Wiley, New York.
- [21] Heyde C. C. and Seneta, E. (1977). *I. J. Bienaymé: Statistical Theory Anticipated*, Springer, New York.
- [22] Heyde C. C. and Seneta, E. (editors) (2001). *Statisticians of the Centuries*, Springer, New York.
- [23] Hilts, V. L. (1973). Statistics and Social Science, Giere, R. and Westfall, R. S. (editors) *Foundations of Scientific Method: The Nineteenth Century*, Indiana University Press, Bloomington, 206-233.
- [24] Hobsbawm, E. J. (1962). *The Age of Revolution*, Vintage Books, New York, 「혁명의 시대」 정도영, 차명수 옮김, 1998, 한길사, 서울.
- [25] Hobsbawm, E. J. (1975). *The Age of Capital*, Vintage Books, New York, 「자본의 시대」 정도영 옮김, 1998, 한길사, 서울.
- [26] Hobsbawm, E. J. (1987). *The Age of Empire*, Vintage Books, New York, 「제국의 시대」 김동택 옮김, 1998, 한길사, 서울.
- [27] Johnson, N. L. and Kotz, S. (editors) (1997). *Leading Personalities in Statistical Sciences*, Wiley, New York.
- [28] Klein, J. L. (1997). *Statistical Visions in Time: A History of Time Series Analysis, 1662-1938*, Cambridge University Press, Cambridge.
- [29] Kotz, S. and Johnson, N. L. (editors) (1992a). *Breakthroughs in Statistics, vol. I, Foundations and Basic Theory*, Springer, New York.
- [30] Kotz, S. and Johnson, N. L. (editors) (1992b). *Breakthroughs in Statistics, vol. II, Methodology and Distribution*, Springer, New York.
- [31] Kotz, S. and Johnson, N. L. (editors) (1997). *Breakthroughs in Statistics, vol. III*, Springer, New York.

- [32] Krüger, L. (1987). The Slow Rise of Probabilism: Philosophical Arguments in the Nineteenth Centuries, Krüger, L., Daston, L. J. and Heidelberger, M. (editors), *The Probabilistic Revolution, vol. 1. Ideas in History*, MIT Press, Cambridge, Massachusetts, 59-89.
- [33] Krüger, L., Daston, L. J. and Heidelberger, M. (editors) (1987a). *The Probabilistic Revolution, vol. 1. Ideas in History*, MIT Press, Cambridge, Massachusetts.
- [34] Krüger, L., Gigerenzer, G. and Morgan, M. S. (editors) (1987b). *The Probabilistic Revolution, vol. 2. Ideas in the Sciences*, MIT Press, Cambridge, Massachusetts.
- [35] Kuhn, T. (1987). What are Scientific Revolutions? Krüger, L., Daston, L. J. and Heidelberger, M. (editors) *The Probabilistic Revolution, vol. 1. Ideas in History*, MIT Press, Cambridge, Massachusetts, 7-22.
- [36] MacKenzie, D. (1981). *Statistics in Britain, 1865-1930: The Social Construction of Scientific Knowledge*, Edinburgh University Press, Edinburgh.
- [37] Mirowski, P. (editor) (1994). *Edgeworth on Chance, Economic Hazard, and Statistics*, Rowman & Littlefield, Lanham, MD.
- [38] Olson, R. (1993). *The Emergence of the Social Sciences 1642-1792*, Twayne, New York.
- [39] Patriarca, S. (1996). *Numbers and Nationhood: Writing Statistics in Nineteenth-Century Italy*, Cambridge University Press, Cambridge.
- [40] Plato, J. (1994). *Creating Modern Probability*, Cambridge University Press, Cambridge.
- [41] Porter, T. M. (1986). *The Rise of Statistical Thinking 1820-1900*, Princeton University Press, Princeton.
- [42] Salsburg, D. (2001). *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century*, Freeman, New York, 「천재들의 주사위」, 최정규 옮김, 2003, 뿌리와이파리, 서울.
- [43] Stigler, S. M. (1986). *The History of Statistics: The Measurement of Uncertainty before 1900*, Belknap Press of Harvard University Press, Cambridge, Massachusetts.
- [44] Todhunter, I. (1865). *A History of the Mathematical Theory of Probability: From the Time of Pascal to That of Laplace*, Macmillan, London.
- [45] Westergaard, H. L. (1932). *Contributions to the History of Statistics*, P. S. King, London.

[2003년 9월 접수, 2003년 12월 채택]