

토론: 통계계산분야의 반성과 대비

김성수¹⁾

1. 통계계산 분야에 대한 우리의 반성

우리 학계에서 통계계산분야의 중요성은 통계학회 10주년, 20주년 기념호 등에 언급이 되어 있는 바대로 학계의 끊임없는 관심을 받아왔고, 1988년 창립된 통계계산 연구회 등을 중심으로 통계계산 분야에 대한 지속적인 관심과 주기적인 발표, 교육 등을 통하여 많은 발전을 이루어 왔다. 또한 통계계산 분야의 체계적인 발전을 위해서 교육과정의 중요함을 인식하고 통계교육 상담연구회 등을 중심으로 "정보관련 통계학과의 교과과정(1999, 조신섭 외 3인)" 등에서 통계학과에서 가르쳐야 할 통계계산 분야 교과목 등에 대한 논의를 진지하게 한 바 있다. 여기서 통계계산 분야는 "statistical computing", "computational statistics"를 통틀어 의미하는 것으로 한다.

통계계산 연구회나 통계교육 상담연구회 등을 중심으로 통계계산 분야에 대한 활발한 논의와 교육을 통하여 많은 발전을 이룬 것은 사실이지만, 실질적인 발전면에서는 아직도 더욱 노력해야 할 점들이 많다고 할 수 있다. 왜냐하면 통계계산 분야의 발전은 컴퓨터분야, 수리·통계알고리즘 분야 등과 더불어 발전이 이루어지며 따라서 조금 생소한 분야로의 노력을 과다히 요구하기 때문이다. 하지만 실제적으로는 단순한 언어 교육, 통계소프트웨어의 교육, 웹의 활용, 웹 프로그래밍 기법 등의 교육이 주로 이루어지고 있는 현실에서는 통계계산의 발전이 단순히 전산분야의 발전을 통계분야와 접목하는 방향으로 가고 있는 게 현실이라고 할 수 있다.

통계계산 분야의 발전은 김철웅 등이 이야기한 것처럼 전산학 또는 컴퓨터의 발전에 힘입은 바 크지만, 실제로 여기서 더욱 중요한 것은 수리적인 통계방법론 및 통계알고리즘이라고 할 수 있다. 왜냐하면 컴퓨터가 아무리 발전한다고 할 지라도 수리적인 방법론의 발전, 적합한 알고리즘의 발전이 없이는 통계계산 분야의 발전을 이룰 수 없기 때문이다. 예를 들어 김철웅 등이 예로 든 통계그래픽 분야의 "representation of high dimensional data"를 위해서는 projection pursuit 알고리즘이나 grand tour 방법 등이 필요하고, MCMC(Monte Carlo Markov Chain)를 위해서는 Metropolis-Hastings 알고리즘 등에 대한 연구가 반드시 필요하기 때문이다. 따라서 김철웅 등이 이야기 한 것처럼 "전산학 또는 컴퓨터의 발전이 새로운 통계방법을 탄생시키고, 통계계산 분야의 발전이 모두 컴퓨터의 발전에 기초하였다"고 강조하는 것보다는 컴퓨터의 발전이 통계계산 분야 발전의 한 계기로서만 인식이 되면 좋겠다는 생각이다.

이러한 의미에서 우리 학계에서 주로 이루어진 통계계산의 발전은 수리적 통계방법론, 통계알고리즘의 연구라는 면보다는 통계에서 필요한 계산의 구현 등에 치중한 면이 많았다고 할 수 있다. 이러한 면에서 조신섭 등(1999)이 "정보관련 통계학과의 교과과정"에서 소개한 통계알고리즘의 교육사례는 학부과정에서 이루어질 수 있는 한 좋은 예라고 할 수 있고, 학문의 길로 접어드는 대학원의 경우엔 고급적인 통계계산 분야의 알고리즘도 연구가 되어야 할 것이다.

1) (110-791) 서울시 종로구 농성동 한국방송통신대학교 정보통계학과, 부교수
E-mail : sskim@knou.ac.kr

2. 통계계산분야의 대비

통계계산의 현황과 미래는 김철웅 등이 자세히 소개하고 있는 바와 같이 새로운 통계적 방법론의 출현, 컴퓨터의 발전 등으로 인하여 갈수록 그 영역이 넓혀질 것으로 인식되고 있다. 그럼, 우리는 앞으로 어떻게 준비하고 대비하여야 할 것인가. 더욱 더 경쟁이 치열해지고, 학문이 다양화해 가는 소용돌이 속에서 통계계산 분야를 세계적으로 우뚝 설 수 있게 하기 위해서는(또는 최소한 다른 선진 나라와 어깨를 나란히 할 수 있게 위해서는) 어떠한 준비가 필요할 것인가. 이를 위해서 우리 학계, 정부, 기업체 등에서는 어떠한 준비가 필요할 것인가 등에 대하여 간략히 언급하고자 한다.

1) 학계에서의 통계계산분야의 대비

통계계산 분야의 발전은 하루 아침에 이루어지는 것이 아니고, 끊임없이 쌓여 가는 지식의 경험, 지식의 창고 속에서 발전이 이루어지는 법이다. 물론 모든 학문의 발전도 마찬가지겠지만 특히 통계계산 분야의 발전은 과거의 토대 위에서 더욱 자리를 든든히 잡아나갈 수 있는 분야이다. 우리 학계는 통계계산연구회를 중심으로 꾸준한 활동을 펼쳐왔고, 또한 앞으로도 그 활동 범위를 넓혀 갈 것이다. 하지만 우리가 지닌 문제점들을 먼저 해결하지 않고서는 그 발전속도와 방향에서 문제점이 생길 수 있다.

먼저 교육 문제의 경우는 김철웅 등이 “누적되어온 문제점”에서 지적한 바와 같이 실질적인 인식과 교육 방향의 연계가 잘 이루어지지 않고 있는 모습이다. 통계계산 분야는 단순히 통계 문제를 컴퓨터를 이용하여 푼다는 것이 아니라 통계적 알고리즘과 컴퓨터를 이용한 수치적 계산을 이용하여 통계적 문제를 해결해 나가는 중요한 분야이다. 따라서 김철웅 등이 지적한 사항들을 잘 인식하여 통계계산 분야의 전문인력 양성에 더욱 더 힘을 기울여야 할 것이다.

통계계산 분야가 더욱 발전을 이루기 위해서는 사회의 필요성을 충분히 제 때에 해결할 수 있을 때에 더욱 더 빛을 발할 수 있다. 따라서 김철웅 등이 “통계계산의 미래”에서 이야기 한 것처럼 회사에서 CRM, Financial Engineering, Data Mining 등의 분야의 사람들이 필요할 때 이러한 요구에 부응할 수 있는 교과목도 개설하여 이론 뿐만 아니라 실질적인 전산 능력도 겸비한 인력을 배출할 수 있도록 하여야 할 것이다. 다만, 우리가 명심해야 할 것은 단순히 소프트웨어의 활용, 해석, 방법론 등에만 국한시킬게 아니라 반드시 알고리즘적인 접근을 할 수 있도록 해야 할 것이다. 이러한 면에서 영국 캠브리지 대학의 Statistical Laboratory나(www.statslab.cam.ac.uk)이나 미국 George Mason 대학의 Computational Statistics part(<http://science.gmu.edu>)를 참조하면 좋을 것이다.

2) 정부, 기업에서의 역할

통계계산 분야의 발전을 이루기 위해서는 다양한 분야의 전공자들이 공동 작업을 할 수 있어야 한다. 왜냐하면 통계계산 분야는 이론적인 내용, 알고리즘적인 방법론, 시스템의 구현 등이 종합적으로 이루어져야 하기 때문이며, 또한 이를 위해서는 많은 경비가 소요된다.

현재 우리나라에서 기초분야에 대한 연구는 거의 도외시된 채 단순히 기술적인 구현 문제에 너무 많은 비중을 두고 있다. 이러한 문제는 통계계산 분야에 있어서도 어김없이 적용되고 있다. 예

를 들어 통계적 문제 구현, 그래픽 구현 등에 있어서 필수적으로 이용되는 행렬 라이브러리, 그래픽 라이브러리 등에 대한 기초연구 개발지원이 이루어진 적이 있는가(내가 아는 한 없는 것으로 알고 있다). 물론 이러한 문제는 각자 연구자들의 책임이라고 할 수 있겠지만 이러한 작업들은 한 개인의 힘으로 이루어질 수 있는 것이 아니라 관련 연구자들의 공통 작업이 있을 때에야 만이 이루어질 수 있는 방대한 작업이다.

현재 미국, 유럽뿐만이 아니라 일본, 대만, 중국 등에서는 통계분야의 종합적인 연구, 지원 등을 위하여 통계연구소를 두고 있을 뿐만 아니라, 통계계산 분야와 관련된 다양한 연구소를 두고 있다. 우리나라도 체계적인 통계 문제의 연구, 통계의 활용, 타분야와의 연계, 통계계산 분야의 국가적인 지원을 위해서는 반드시 통계연구소를 두고, 산하에 통계계산 분과를 두어 체계적인 연구가 이루어질 수 있도록 노력해야 할 것이다.

참고로 유럽연합에서의 통계계산 관련 분야의 지원의 예로서는 Spatial and Computational Network(<http://www.maths.lancs.ac.uk/~breyer/essn/index.html>)를 참조할 수 있겠다.

통계계산 분야는 전체적으로 볼 때 각 기업에서 갈수록 수요가 늘어나고 있는 입장이다. 특히 최근 들어 CRM, Data Mining, ERP 연계 웹 분석시스템 등의 다양한 수요가 일어나면서 이 분야의 전공자들에 대한 수요가 급속히 늘어나고 있는 실정이다. 이러한 면에서 각 기업체에서도 필요한 시스템을 구축하고, 효과적인 분석시스템을 가동하기 위해서는 통계 이론 뿐만 아니라 통계계산 분야의 지식을 소유한 사람이 필요하게 된다.

기업은 기본적으로 당장 필요하지 않은 일에는 돈을 지출하지 않는 경향이 있으므로 통계계산 분야에 있어 기본적인 이론적인 통계적 방법론, 통계알고리즘 분야의 개발은 국가 지원 통계연구소 등에서 담당하고 기업은 산학과 연계하여 응용분야 연구 지원을 할 수 있도록 유도하면 좋을 것이다.

3. 우리의 자세

통계계산 분야는 컴퓨터의 발달, 통계적 방법론의 발달, 통계알고리즘의 발달과 더불어 통계에서 차지하는 역할이 갈수록 커지고 있다. 하지만 현실적으로 학계에서의 통계계산 분야의 교육은 단순한 프로그램의 교육, 통계패키지의 활용법 등에 그치고 있다. 김철웅 등이 통계계산의 현황에서 밝히고 있는 것처럼 현시대에는 다양한 통계적 지식 및 통계계산 분야의 지식이 요구되고 있고, 또한 우리의 할 일들이 그만큼 커나가고 있다. 미래에 있어서의 발전 속도와 응용범위의 증가는 또 어떠한 것인가.

통계분야에서 발생하는 문제뿐만이 아니라 타 분야에서 발생하는 문제들도 통계적으로 재조명하여 해석하고 통계학의 한 분야로 만들기 위해서는 반드시 통계계산 분야의 지식이 필요하다. 즉, 통계적 방법론의 해석, 통계적 알고리즘의 연구 및 구현, 새로운 방법의 창조 등이 요구되게 된다. 따라서 우리는 지금까지의 교육에서 한 단계 더 올라가 종합적인 통계계산 분야의 지식교육이 될 수 있도록 노력하고, 또한 사회의 관심을 이끌어 낼 수 있도록 더욱 더 매진해야 할 것이다.

참고문헌

- [1] 김병천(1991), “통계계산분야의 현재와 미래”, 20권, 창립 20주년 기념호, 105-110
- [2] 조신섭, 신봉섭, 이상복, 한경수(1999), “정보관련 통계학과의 교과과정”, 응용통계연구,
- [3] 허문열(1981), “한국통계의 현황과 장래-통계와 전산”, 10권, 창립 10주년 기념호, 77-80 제12권 2호, 683-703.
- [4] Friedman, J.H.(1997), “Data Mining and Statistics: What’s the connection?”, 29th Symposium on the Interface: Computing Science and Statistics,