

한국 통계이론의 과거 · 현재 · 미래¹⁾

김 우 철²⁾, 박 병 옥²⁾

1. 머리말

한국통계학회 창립 30주년을 기념하는 이번 가을 학회에서 우리나라 통계이론의 현황에 대한 주제 발표를 부탁 받고 필자는 통계이론이 과연 무엇인지 한번 생각해 보았다. 이는 필자가 학생들에게 종종 말하는 바처럼 분명 학부나 대학원에서 정규 교과목으로 가르치는 “수리통계학” 또는 “이론통계학”은 아닐 것이다. 더욱이 통계학의 어떤 특정 분야에서 다루는 내용을 뜻하는 것도 아닐 것이다. 당연한 이야기 같지만 궁극적인 목적이 응용이든 아니든 통계적 방법의 적용에 대하여 제시될 수 있는 일체의 논리적 근거를 뜻하는 것으로 해석해야 할 것이다.

우리나라 통계이론의 현주소는 어디인가라는 질문에는 세계 속에서 한국이 어느 위치에 와 있고 국제적인 흐름을 어느 정도 따라가고 있는가를 살펴봄으로써 답을 해야 할 것 같다. 따라서 앞서 백운봉(1981)교수와 우정수(1991)교수가 10주년과 20주년 기념 논문에서 밝힌 바와 같이 국제적인 현황과 추세를 알아보는 것이 선행되어야 할 것이다. 통계이론의 국제적인 현황과 추세는 학술지에 발표되는 논문들의 성향으로부터 파악할 수 있을 것이다. 그러나, 현재 발간되는 통계학 관련 학술지 모두를 대상으로 이를 조사하는 것은 불가능하다고 생각한다. 따라서, 이론적인 연구 활동을 국제적으로 선도한다고 생각되는 The Annals of Statistics(AS)와 Journal of the American Statistical Association(JASA)의 Theory and Methods section 만을 대상으로 하였다.

이 논문에서는 먼저 위 두 학술지에 발표된 논문들을 각각 주제별로 분류하여 1980년대 이후의 국제적인 흐름을 알아보았다. 여기에 사용된 주제별 분류 항목은 [표1]과 같으며 이는 AMS 2000 subject classifications에 기초한 것이다. 또한, 우리나라 통계이론의 현황을 알아보기 위하여 한국 통계학회지인 Journal of the Korean Statistical Society(JKSS)에 실린 논문들도 같은 방법으로 분류하여 AS와 JASA의 경향과 비교해 보았다. AS와 JASA는 대표적인 국제학술지로 평가받고 있다. 따라서, 여기에 국내 통계학자가 얼마나 많은 논문을 발표하느냐는 우리나라 통계이론의 질적 수준을 가늠할 수 있는 좋은 잣대가 될 수 있을 것이다. 여기에서는 1970년부터 2000년까지 국내 통계학자들에 의해 AS와 JASA에 발표된 논문 수를 일본, 중국, 대만과 비교해 보았다. 끝으로, 이와 같은 비교 분석을 바탕으로 향후 우리나라 통계이론 교육에 대한 바램을 간단히 적어 보았다.

-
- 1) 본 논문에서 학술지 논문의 주제별 분류작업은 서울대학교 통계학과 유규상, 이영경, 황준혁, 하소영, 정계식, 이명희, 전현호, 김수연, 김정연 대학원생들에 의해 수행되었음을 밝힙니다.
 - 2) 서울대학교 통계학과

[표 1] AMS 2000 Subject Classifications

분류 기호	주제	분류 기호	주제
XX	Statistics	A	Foundations
B	Sufficiency and Information	C	Decision Theory
D	Sampling Theory	E	Distribution Theory
F	Parametric Inference	G	Nonparametric Inference
H	Multivariate Analysis	J	Linear Inference
K	Design of Experiments	L	Sequential Method
M	Inference from Stochastic Processes	N	Survival and Censored Data
P	Applications	Q	Statistical Tables
R	Probability theory and Stochastic process		
S	Combinatorics, Linear Multilinear Algebra, Matrix Theory, Real Function, Ordinary Differential Equations, Partial Differential Equations, Finite Differences and Functional Equations, Approximations and Expansions, Integral Transforms, Operational Calculus, Calculus of Variations and Optimal Control, Convex and Discrete Geometry, Differential Geometry, Global Analysis, Analysis on Manifolds, Numerical Analysis		
T	Computer Science, Statistical Mechanics, Structure of Matter, Geophysics, Economics, Operations Research, Programming, Games, Biology and Other Natural Sciences, Behavioral Sciences, System Theory, Control, Information and Communication, Circuits		

(주) 위 표에서 R, S, T는 AMS 2000 subject classifications 분류기호가 아니고 필자가 편의상 다른 학문 분야 군에 부여한 것임.

2. 통계이론의 국제 현황

1990년 이전까지 AS 또는 JASA에서 발표된 논문의 주제별 분류는 창립 10, 20주년 기념 논문에서 이미 조사된 바 있다. AS의 경우 1978년부터 1990년까지 조사되었으며, JASA의 경우에는 1986년부터 1990년까지 조사되었다. 여기에서는 1991년부터 2000년까지 발표된 논문을 대상으로 주제별 분류를 먼저 해보고 이를 1980년대의 조사결과와 비교하여 통계이론에 대한 국제적인 연구동향을 살펴보기로 한다.

[표2]와 [표3]은 주제별 분류 결과이다. AS의 경우에는 저자들 자신이 분류한 것이고, JASA의 경우에는 저자들에 의한 분류가 되어 있지 않아 MathSciNet에서 제공되는 분류 서비스를 이용하였다. MathSciNet의 리스트에서 누락된 논문은 필자에 의해 분류되었다.

[표2]에서 AS에 발표된 논문의 1차 분류를 살펴보면 G(nonparametric inference)가 35%로 압도적으로 많은 비중을 차지하고 있고, 그 다음으로 F(parametric inference)의 15%, M(inference from stochastic processes)의 7.6%, 그리고 H(multivariate analysis)의 6.8%가 뒤를 잇고 있다. 2

차 분류의 경우에도 G가 여전히 압도적이거나 그 뒤를 이어 R(probability theory and stochastic process)이 큰 비중을 차지하고 있다. 2차 분류에서 R로 분류된 논문들의 대부분은 1차에서 E(distribution theory), F, G, M 등으로 분류된 것으로 1980년대의 조사결과(우정수, 1991)에서도 비슷한 현상을 발견할 수 있다.

[표 2] AS(1991~2000)에 발표된 논문의 분류현황

2차		1차																			
1차	XX	A	B	C	D	E	F	G	H	J	K	L	M	N	P	Q	R	S	T	계	
XX																		1		1	
A	8		3	1	2		1	4	6		1		1				5	1	2	35	
B	1		2					1		2										6	
C	12		3		7	3	4	7		2	1	1	1				5	1		47	
D	2		1		1		3	1	4								1			13	
E	15		2	1			4	4	2	4		1	1		2		19	4		59	
F	42		3		15		10	27	11	11	10	1	7	1	1		20	2	6	167	
G	104	1	5	5	9	2	29	13	126	15	24	2	14	3	5		19	14		390	
H	18	1			3		4	9	11	12	6		2				6	3	1	76	
J	10		2		1		3	6	6	3	3	1	3		1		5	2		46	
K	22				1		1	1		5	15	1	1		1		1	10		59	
L	8		1	1	2		4	4	1			1		14	1		5		1	43	
M	10		1				7	11	19		4		8		1		19	4	1	85	
N	1		1					1	2											5	
P	2																		1	3	
Q																					
R	18		3		1		2	2	16	2	3	2	1	3			8	3		64	
S	3						3	1	1				1						1	11	
T	1						1										1			3	
계	277	2	27	8	42	5	72	94	206	52	57	19	7	42	18	12	0	114	46	13	1,113

(주) 표에서 분류기호 다음 열에 있는 숫자는 1차 분류 만 한 논문 수임.

한편, [표3]의 JASA에 발표된 논문의 1차 분류를 살펴보면 여기에서도 G(nonparametric inference)가 24%로 가장 많은 비중을 차지하고 있음을 알 수 있다. 그 다음으로 F(parametric inference)의 18.1%, J(linear inference)의 17.5%, M(inference from stochastic processes)의 10.1%, 그리고 H(multivariate analysis)의 8.9% 순이다. AS와 비교해 볼 때 G, F, M, H 들 사이에 순위 변동은 없으나 AS에서는 4.1%에 불과한 J영역이 JASA에서는 G, F 다음으로 많은 비중을 차지

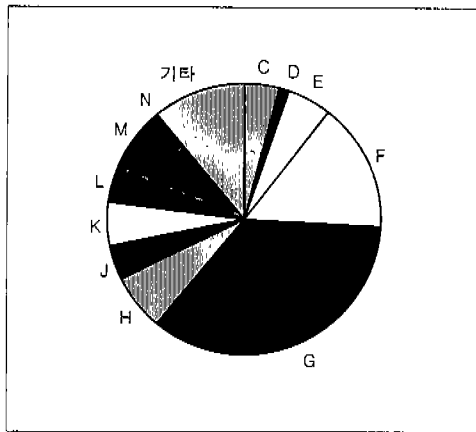
하고 있음을 알 수 있다.

[표 3] JASA (Theory and Methods, 1991~2000)에 발표된 논문의 분류현황

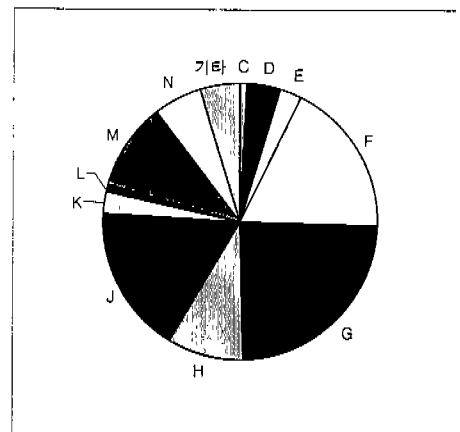
1차	2차	XX	A	B	C	D	E	F	G	H	J	K	L	M	N	P	Q	R	S	T	계	
		XX	1								1											
A	2		2		1			5						1								11
B	5							1												1		7
C	5		1					1						1								8
D	31				4			1	3		2		1									42
E	20	1						1	1	1					1			2				27
F	110	1	6	2	5	1	1	19	8	4	16	2		4	2	2		5	1	1		190
G	146	2		1	1	1	5	7	40	4	15	3	1	6	10	5		4	1			252
H	58		1		2		3	5	6	7	7			2	1			1				93
J	113	2					2	33	16	7	5	2			1	2						183
K	18				1			1			3	1								1		25
L	7							1	1		1											10
M	68		1				2	11	12	3	1			7		1						106
N	27			1			1	6	19		1		1	1	2	2		1				62
P	10										1			1	2							14
Q																						
R	1				1		1	2	1	1			1		1			1				10
S	1											1								1		3
T						1			1										1			3
계	623	6	11	4	15	3	15	94	109	27	52	9	4	23	20	12	0	14	5	2		1,048

(주) 표에서 분류기호 다음 열에 있는 숫자는 1차 분류 만 한 논문 수임.

[그림1]은 [표2]와 [표3]을 원그림(pie chart)으로 나타낸 것이다. [그림2]는 1980년대와 1990년대의 연구 동향을 1차 분류를 통하여 비교한 것이다. 이 그림에서 보듯이 대부분의 영역에서는 큰 변화가 없으나 AS의 G(nonparametric inference)영역은 24%에서 35%로 11%의 급격한 증가세를 보이고 있다. AS의 G영역은 1978년 ~1980년(백운봉, 1981)과 1981년 ~1990년(우정수, 1991)의 비교에서도 18.5%에서 24%로 증가한 바 있다.

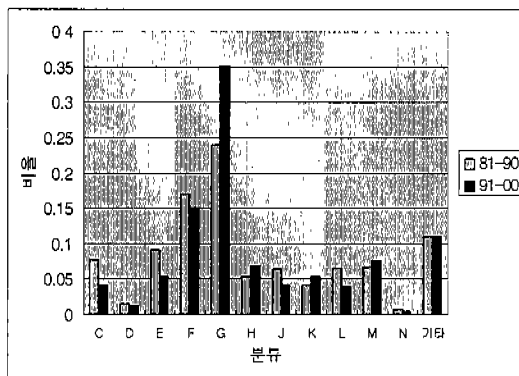


(a) AS

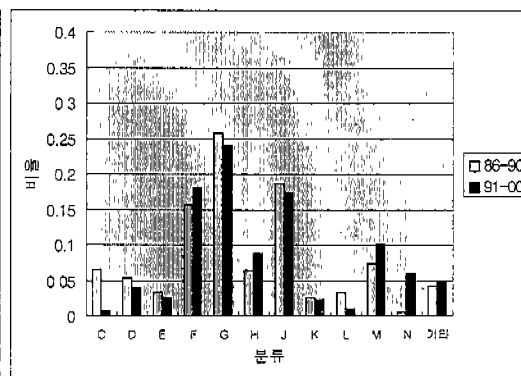


(b) JASA

[그림1] AS와 JASA (1991~2000) 논문의 1차 분류 현황



(a) AS



(b) JASA

[그림2] 1980년대와 1990년대의 비교

3. 통계이론의 국내 현황

한국 통계학회가 창립된 1970년대를 우리나라 통계학의 1세대라고 한다면, 1980년대와 1990년대를 각각 2, 3세대라고 할 수 있을 것이다. 1980년대는 1970년대에 통계학 교육을 받았던 세대가 활동을 시작했던 시기이고, 1990년대는 다시 그들로부터 교육을 받았던 세대가 국내 통계학계에 첫 발을 내 디뎠던 시기였다고 할 수 있다. 학회가 창립된 지 30년이 지난 지금, 회원 수는 학생회원을 포함하여 현재 655명이 되었고 학회를 대표하는 학술지인 JKSS에 발표되는 논문 수도 연간 약 40편에 이르는 등 양적으로 많은 팽창을 하였다고 생각한다. 우리나라 통계학의 3세대가 활동을 시작한지 10년이 지났고 이제 4세대의 활동을 눈앞에 둔 지금, 국내 통계이론의 현황과 함께 그 질적 성장에 대해 평가해 보는 것은 의미 있는 일이라 하겠다.

[표 4] JKSS (1991~2000)에 발표된 논문의 분류현황 (단, 외국인 논문은 제외)

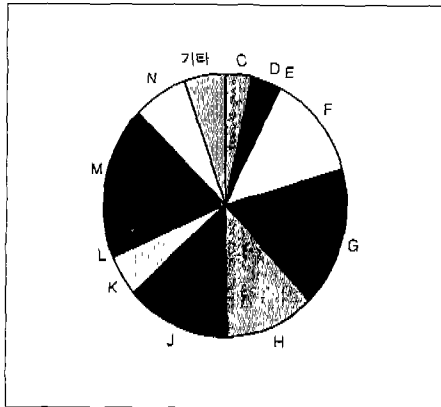
1차	2차	XX	A	B	C	D	E	F	G	H	J	K	L	M	N	P	Q	R	S	T	계	
		XX																				
A																						3
B	2			1																		3
C	6							2			1			1								10
D	10				1																	11
E																						
F	17			1	2			9	2	1	2				2							37
G	17			1	1		2	1	8	1	4			2	10			1				48
H	21			1	1			3	3	3	1											33
J	19						2	11	4		2			1								39
K	10										3									1		14
L	3				1				1					1								6
M	37			1				1	2		2			1				3				47
N	13							3	2									2				20
P											2			1								3
Q																						
R									4						2	1						7
S										1		1										2
T											1											1
계	155	0	0	5	6	0	4	30	26	6	18	1	1	9	13	0	0	6	1	0		281

(주) 표에서 분류기호 다음 열에 있는 숫자는 1차 분류 만 한 논문 수임.

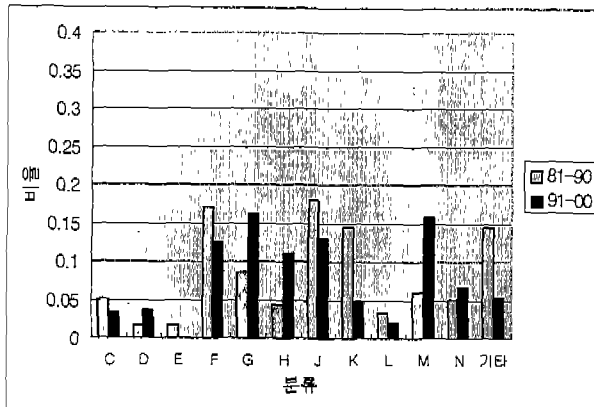
[표4]는 1991년부터 2000년까지 JKSS에 발표된 국내 통계학자들의 논문을 AS와 JASA에서와 같은 방법으로 분류한 결과이다. JKSS의 경우에도 JASA처럼 G(nonparametric inference), M(inference from stochastic processes), J(linear inference), F(parametric inference), H(multivariate analysis) 다섯 영역이 전체의 70%이상을 차지하고 있음을 알 수 있다. 그러나, G 영역이 가장 많은 17.1%를 차지하고는 있지만 AS나 JASA 만큼 뚜렷하지는 않고 오히려 M영역이 G와 거의 대등한 16.7%를 기록하고 있는 점이 특기할 만하다. 또한, AS나 JASA에서 J, M, H에 비해 높은 비중을 차지했던 F가 상대적으로 낮은 13.2%의 점유율을 나타내고 있다. 전반적으로 JKSS의 분류에서는 위의 5개 주요 영역이 AS나 JASA에 비해 고른 분포를 하고 있다는 점이 특징이다. [그림3]은 각 영역의 비중을 원그림으로 나타낸 것이다.

한편, [그림4]는 1980년대에서 1990년대로 넘어 오면서 JKSS에 발표된 논문의 성향이 어떻게 변했는지를 보여 주고 있다. 이 그림도 [그림2]와 같이 1차 분류를 기준으로 만든 것이다. 이 그림을 살펴보면 K(design of experiments)영역의 비중이 두드러지게 감소했으며, F와 J는 다소 감소 추세를 알 수 있다. 반면, G, H, M 영역의 논문은 상당히 증가하였다. 여기에서, G영역의 증가는 국제적인 수준에 근접해 가는 과정으로 해석할 수 있으나 M영역의 현저한 증가 추세는 매우

주목할 만하다. 필자의 소견으로는 M영역의 경우 상당수의 논문들이 Journal of Time Series Analysis, Stochastic Processes and Their Applications, Econometric Theory 등의 학술지에 발표되는 관계로 AS나 JASA에는 상대적으로 적은 수의 논문이 발표되는 것이 아닌가 생각된다. 따라서, JKSS에서 M영역의 급격한 증가 추세가 국제적인 경향과는 거리가 있다고 속단하는 것은 옳지 않다고 생각한다.



[그림3] JKSS (1991~2000) 논문의 1차 분류 현황



[그림4] 1980년대와 1990년대의 비교

논문의 질을 평가한다는 것은 매우 어려운 일이라 생각한다. 더욱이 자신의 전공 분야가 아닌 논문을 평가하는 것이란 더더욱 어려운 일이다. 우리나라 통계이론의 질적 수준을 가늠해보기 위하여 부득이 AS와 JASA에 발표된 국내학자의 논문 수를 조사해 보기로 하였다. 앞서 언급한 바와 같이 AS와 JASA는 대표적인 국제학술지로 인정받고 있는 만큼, 이러한 잣대가 터무니없는 것은 아니라는 데 이견이 없을 것 같다.

우리나라의 수준을 미국이나 서구라파의 나라와 비교하는 것은 아직 이르다고 생각하여 이웃나라인 일본, 중국, 대만과 비교해 보았다. [표5]와 [표6]은 1971년부터 2000년까지 10년 단위로 각 나라의 통계학자들이 발표한 논문 수를 조사한 것이다. 여기에서 국가별 분류는 논문에 나타난 저자의 소속기관에 따른 것이다. 예를 들면, 한국 사람일지라도 논문에 발표된 소속기관이 미국의 대학이나 연구소인 경우에는 한국 통계학자의 논문 수에 포함하지 않았다.

[표 5] 국가별 AS 발표 논문수

국가 \ 연도	한국	일본	중국 (홍콩제외)	대만
1971-1980	2	36	0	4
1981-1990	0	34	6	8
1991-2000	10	15	9	31

[표 6] 국가별 JASA 발표 논문수

국가 연도	한국	일본	중국 (홍콩제외)	대만
1971-1980	0	11	0	0
1981-1990	6	19	0	3
1991-2000	13	7	3	28

[표5]와 [표6]을 살펴보면 분명 우리나라의 수준이 70, 80년대에 비해 상당한 발전을 하였다고 볼 수 있다. AS와 JASA의 논문 수를 합하면 1990년대에는 23편으로 일본의 22편과 비슷하나 대만의 59편에는 많이 뒤지고 있음을 알 수 있다. 특기할 사항은 일본의 경우는 80년대에 비해 현저히 줄고 있으며 대만은 급격한 증가를 보이고 있다는 점이다.

이들 네 나라에서 AS와 JASA에 발표한 논문들의 성향을 비교해 보기 위하여 1991~2000에 발표된 논문들을 AMS 2000 subject classifications에 따라 분류해 보았다. AS에 발표된 논문들을 살펴보면, 우리나라와 대만의 경우에는 G(nonparametric inference)영역이 각각 50%와 45%로 가장 많고, 일본은 H(multivariate analysis)영역이 33%를 차지하면서 가장 비중이 높았다. 한편, JASA의 논문들을 분류해 본 결과, 우리나라는 G영역이 62%로 압도적이었으며, 대만은 G와 J(linear inference)가 각각 25%와 21%로 상당한 비중을 차지했고 일본의 경우는 M(inference from stochastic processes)영역이 43%로 가장 많았다. 여기에서도 특기할 사항은 우리나라와 대만은 정도의 차이는 있지만 AS와 JASA 모두 G영역의 논문이 가장 많이 발표된 데 반해, 일본은 G영역의 비중이 AS의 경우에는 7%, JASA의 경우에는 0%를 기록하여 상당한 대조를 이루고 있다는 점이다.

4. 맺음말

지금까지 살펴본 바와 같이 통계이론의 연구활동에서 우리나라는 지난 10년 간 상당한 발전을 하였으며, 이러한 발전 추세는 이웃나라인 대만에 비해서는 뒤떨어지지만 답보상태인 일본에 비하면 긍정적이라고 할 수 있다. 제한된 조사에 의한 결론이기는 하나, 이러한 세 나라의 추세는 각 나라에서의 통계학 교육 현실과 관련이 있다는 것이 필자의 견해이다. 대만의 경우에는 중앙연구원(Academia Sinica)에 통계연구소가 설립된 이후 여러 대학에서 대학원과정에 통계학과가 설립되고 있는 반면에 일본에는 통계학과가 거의 독립적으로 운영되고 있지 않는 교육현실과 연관이 있다고 판단된다. 한편, 우리나라의 경우에는 통계 관련 학과가 10년 동안 60여 개에서 100여 개로 증가하고 통계학회의 회원수가 300여 명에서 600여 명으로 증가(Kim and Lee, 2001)하는 양적인 발전이 있었다. 이와 같이, 통계이론의 발전을 위해서는 통계이론 연구자의 저변 확대와 정체성이 있는 통계학 교육이 필수적이라고 생각된다.

장래 우리나라 통계이론의 발전을 위해서는 국제수준의 대학원 교육이 이루어져야하고, 이를 위

해서는 학사과정에서 기초 교육이 체계적으로 수행되어 대학원 수준의 교육과 원활하게 연계되어야 할 것이다. 그러나, 최근 각 대학에서의 학부제 실시에 따라 학사과정에서의 통계학 교육은 그 정체성에 심각한 도전을 받고 있는 것이 현실이다. 따라서, 통계이론 연구의 후속세대 양성을 위해서는 통계이론 연구를 위한 통계학은 물론 기초적인 수학, 계산과학 등을 체계적으로 이수하도록 하여야 할 것이다. 이러한 후속세대의 양성과 더불어, 국제적 학술교류의 기회를 증대시키고 국내 통계이론의 연구수준과 국제수준을 비교, 평가할 수 있는 기회가 많아져야 할 것이다. 이를 위해서는 통계학회 회원의 능력을 결집하여 통계학회를 중심으로 국제적 학술활동의 기회를 늘려 나가야 할 것이다.

참고문헌

- [1] Kim, H.-J. and Lee, J.J. (2001) Milestones in the Korean History of Statistics, *Statistics in Korea*, 187-202
- [2] 백운봉 (1981), 한국통계의 현황과 장래-통계이론, 통계학연구, Vol. 10(창립 10주년 기념호), 9-13.
- [3] 우정수 (1991), 80년대 한국통계의 현황과 장래-통계이론을 중심으로, 통계학연구, Vol. 20(창립 20주년 기념호), 10-18.