

베이지안 사회과학 방법론이란 무엇인가?*

박종희 | 서울대학교

베이지안 방법이란 자료분석에서 확률이라는 수학적 언어를 이용해 연구자가 가진 불확실성을 체계적으로 기술하고자 하는 경험적 연구방법이다. 본 논문은 베이지안 통계를 간략히 소개하고 베이지안 통계가 경험적 사회과학 방법론으로서 갖는 의미를 논한다. 구체적으로, 베이지안 방법론은 빈도주의에 기반한 전통적 회귀모형이 갖는 제한성을 비표본자료 분석, 가설검정, 다층모형과 비선형모형 추정, 그리고 시간적·공간적 맥락의 존성의 측정 등의 측면에서 극복할 수 있는 중요한 대안적 연구방법이다. 이러한 논의를 뒷받침하기 위해 본 논문은 정치학 안에서 진행된 베이지안 연구의 대표적인 성과를 소개한다.

주제어: 베이지안 통계, 사전분포, 사후분포, 가설검정, 마르코프체인 몬테카를로 방법

* 이 논문은 2013년도 정부재원(교육과학기술부 인문사회연구역량강화사업비)으로 한국연구재단의 지원을 받아 연구되었음(2013S1A3A2053683).

I. 들어가며

존 스튜어트 밀(John Stuart Mill)은 『논리학 체계』(A System of Logic)에서 비교연구를 사회과학의 핵심방법론으로 소개한 바 있다. 밀의 비교방법론은 이후 차이를 통한 비교방법론(Method of Difference)과 유사성을 통한 비교방법론(Method of Agreement)으로 크게 양분되어 정치학을 비롯한 제반 사회과학 방법론의 발전에 지대한 영향을 미쳤다(예: Weber 2010, Moore 1966, Skocpol 1979, Katzenstein 1985, Rogowski 1989, Collier and Collier 2002).

그러나 제한된 수의 사례를 연구자의 직관과 관찰에 의존하여 비교하고 그 결과를 이야기(narrative)로 풀어내는 작업은 연구자의 편향이 개입하기 쉽고, 사례선택의 객관성이 담보되기 어려우며, 반사실적 추론이 쉽지 않고, 그 결과가 일반화되기 어렵다는 약점이 있다. 뿐만 아니라 언어적 서술구조의 특징상 사회현상에서 본질적인 부분과 비본질적인 부분을 체계적으로 구분하는 것과, 본질적인 부분의 발현을 비결정론적인 방식으로 설명하는 것이 쉽지 않았다.

사회과학의 정량적 연구방법(quantitative methods)은 이러한 정성적 비교연구(qualitative comparative methods)의 약점을 보완할 수 있는 중요한 대안으로 제시되었다. 통계기법과 확률모형을 사용하는 정량적 연구방법은 많은 양의 사례를 체계적으로 비교하는 것이 용이하고 연구자가 주목하고자 하는 사회현상의 본질적인 부분을 비본질적인 부분과 체계적으로 구분하는 것이 가능하다는 장점으로 인해 그 수와 적용범위가 폭발적으로 증가하고 있다. 급기야 킹, 코헤인, 그리고 버바(King, Keohane, and Verba 1994, 이하

KKV)는 Designing Social Inquiry라는 저서를 통해 정량적 연구방법을 사회과학 연구의 표본으로 제시한 바 있다.

여기서 정치학을 비롯한 사회과학 전반의 정량적 연구는 대부분 20세기 중반 이후에 통계학의 주류로 등장한 빈도주의 통계학(frequentist statistics)이라는 사실에 주목할 필요가 있다. 빈도주의 통계학은 모든 자료를 모집단(population)으로부터 추출된 표본(sample)으로 간주하고 표본에서 관찰된 자료의 특성-표본통계(sample statistics)라고 불리는-을 통해 모집단에 실제로 존재할 참값(true parameters)을 추정하는 것을 목표로 한다. 빈도주의 통계는 표본에서 관찰된 자료의 특성이 얼마나 모집단의 참값과 가까운 것인지를 보기 위해 거의 모든 경우에 영가설검정(Null Hypothesis Test)이라는 고정된 분석틀을 사용한다. 간단히 요약하면, 영가설검정은 영가설을 참값으로 갖는 가상의 모집단으로부터 표본추출을 무한 반복할 때 연구자가 현재 가진 자료보다 더 극단적인 자료를 얻게 될 빈도(frequency)를 확인하는 것이다. 그 가상의 빈도가 $1/20(0.05)$ 보다 작다면 현재 가진 자료는 영가설을 참값으로 갖는 가상의 모집단으로부터 추출된 표본으로 볼 수 없으며 따라서 영가설은 기각된다. KKV가 질적 연구의 표본으로 제시한 사회과학 연구의 원칙들은 사실상 빈도주의 통계학의 분석틀을 질적 연구에 투사한 것이다.¹⁾

그러나 영가설검정과 모집단-표본의 구분을 핵심으로 하는 빈도주의 통계학의 패러다임은 사회과학 자료조사에 몇 가지 중요한 한계를 가지고 있다. 첫째, 사회과학에서 얻게 되는 대부분의 자료는 무한 표본추출이 불가능한 시공간 자료이다. 예를 들어 비교정치나

1) Brady and Collier(2004)는 KKV의 사회과학 연구방법에 대한 지침이 지나치게 빈도주의 통계로 경도되었음을 비판하고 있다.

국제정치에서 사용하는 국가간시계열자료(time-series cross-sectional data)에서 국가의 수가 무한으로 증가하는 모집단을 상상하기는 어렵다. 따라서 실험이나 설문조사와 같이 반복된 표본추출이 가능한 극히 예외적인 경우를 제외한다면 영가설검정과 같은 빈도주의 패러다임을 무한 표본추출이 불가능한 시공간자료의 분석틀로 삼는 것은 심각한 인지적 불일치를 야기한다. 빈도주의 통계의 핵심개념인 p값(p-value)을 정확하게 이해하고 있는 자료분석가들이 많지 않다는 점과 p값만을 기준으로 가설의 타당성을 판단하는 것이 위험하다는 최근의 연구가 그 대표적인 예이다(Gigerenzer 2004. Ioannidis 2005).

둘째, 빈도주의 통계를 통해서 사회과학의 통상적인 연구 질문에 대한 확률론적 답변이 어렵다. 예를 들어 “북한에서 자스민혁명과 같은 내부로부터의 변화가 발생할 가능성은 어느 정도인가?”라거나 “다음 대통령 선거에서 현재의 야당이 승리할 가능성은 얼마인가?”와 같은 질문은 모두 반복될 수 없는 사건의 발생가능성에 대한 확률론적 질문이다. 이러한 질문은 사회과학, 특히 정치학 연구에서 매우 일반적이며 연구자들은 경험자료를 수집하여 그에 대한 확률론적 답변을 찾고자 한다. 그러나 빈도주의 통계에서는 반복불가능한 사건에 대한 확률론적 접근이 어려울 뿐 아니라 통계분석의 결과를 확률론적으로 표현하는 것이 곤란하다. 빈도주의 통계학은 표본의 특성을 정리하고 이를 추론하기 위해서 확률모형을 사용하지만 연구의 최종결과라고 볼 수 있는 가설검정의 결과나 관심값(quantity of interest)을 확률론적으로 표현하지 못하는 한계를 가지고 있다.

이러한 빈도주의 통계의 약점을 비판하면서 1990년대 초반 이후 통계학 내에서 베이지안 통계(Bayesian statistics)가 다시 주목받기

시작했으나, 사실상 20세기 이전까지 확률이론과 통계학의 발전은 베이지안 패러다임 내에서 진행되었다. 20세기 초에 확률을 객관적으로 실재하는 빈도로 보는 빈도주의학파가 등장하여 베이지안 패러다임을 비판하면서 확률을 주관적 관념으로 보는 베이지안은 수적으로 고립되었고 대부분의 통계학 교과서의 목차에서 밀려났다.²⁾ 확률에 대한 철학적 견해 차이와 사전확률분포의 사용에 대한 논쟁에서 시작된 빈도주의와 베이지안 통계의 경쟁은 20세기 전체에 걸쳐 평행선을 그리며 지속되었다.

빈도주의와의 경쟁에서 베이지안 연구자들이 수세에 몰린 가장 중요한 이유는 베이지안 추정방법이 고차원 적분을 포함하는 계산적 복잡성을 가졌기 때문이다. 우도함수를 미분하여 최대값을 찾는 빈도주의 통계학과는 달리 베이지안 통계는 확률분포를 적분하여 모수의 사후분포를 찾아내야 하므로 고차원적분의 해가 분석적으로 확보될 수 없는 경우에는 그 해를 찾을 수 있는 방법이 마땅치 않았다.

그러나 20세기 후반에 이르러 고차원적분의 해를 컴퓨터 시뮬레이션을 통해 구할 수 있는 마르코프 체인 몬테카를로 방법(Markov chain Monte Carlo Methods, 이하 MCMC)이 개발되면서 베이지안 통계는 ‘혁명’이라고 일컬을 수 있는 중요한 도약을 한다.³⁾ 그동안 이론적 가능성에 머물렀던 베이지안 통계방법은 MCMC 방법을 통해 주류 통계학인 빈도주의에 상응하는 대안적인 통계분석 패러다임으로 자리잡게 되었다.

2) 빈도주의 학파의 대표인물은 20세기 통계학의 발전에 가장 중요한 공로를 한 피셔(R.A. Fisher), 니만(Jerzy Neyman), 그리고 피슨(Karl Pearson)이라고 할 수 있다.

3) MCMC 방법에 대한 가장 권위있는 저술로는 Robert and Casella(2004)가 있다.

1990년대 이후 일군의 정치학자들은 베이지안 통계를 통해 사회과학 정량적 연구의 한계를 극복하려는 다각적인 시도를 전개해왔다. 정치학 내에서 진행된 베이지안 분석은 정치인의 이념형 연구(Clinton, Jackman and Riverse 2000, Martin and Quinn 2002, Park 2011b), 다층모형 연구(Gelman and Hill 2007), 전문가 지식을 이용한 자료연구(Gill and Freeman 2013), 다변량 시계열 자료 분석(Brandt and Freeman 2009), 전환점 분석(Park 2009, Park 2011a, Park 2012) 등의 영역에서 기존 빈도주의 연구방법을 뛰어넘는 다양한 성과를 보여주었다.

본 논문은 베이지안 방법론을 간략히 소개하고 정치학 내에서 진행된 베이지안 분석의 주된 연구성과를 살펴보는 것을 목적으로 한다. 다만 베이지안 연구의 지평이 매우 넓고 중요한 연구와 그렇지 않은 연구를 나누는 기준이 자의적일 수 있다는 이유, 그리고 한정된 지면으로 인해 다양한 연구내용을 정확하게 소개하기 어렵다는 점을 고려하여, 부득불 필자의 연구를 중심으로 소개하게 되었다는 점에 대해 미리 독자의 양해를 구한다.

2. 베이지안 통계란 무엇인가?

2.1. 베이지안 분석의 3요소: 자료의 확률분포, 사전확률분포, 그리고 사후확률분포

베이지안 통계는 “확률이라는 수학적 언어를 이용해 인식론적 불확실성을 체계적으로 기술”(a system for describing epistemological uncertainty using the mathematical language of probability)하는 과학적 접근법으로 정의될 수 있다.⁴⁾ 자료를 모집

www.kci.go.kr

단으로부터 추출된 무작위적인 것으로 보고 고정된 모수값을 찾는 빈도주의와는 달리 베이지안 통계는 모수를 확률분포를 갖는 확률변수(random variable)로 보고 자료를 통해 모수의 사후분포를 찾는 것을 목적으로 한다. 다시 말하면 빈도주의는 실재론(realism)적 입장에서 모수의 참값은 객관적으로 존재하는 고정된 것이라 보는 반면, 베이지안 통계는 모수는 주관적 세계에만 존재하는 것이며 객관적 세계를 이해하는 것을 돕는 개념적 도구라고 본다.⁵⁾

모든 베이지안 자료분석은 세 종류의 확률분포(probability distribution)를 사용한다. 먼저 자료의 확률분포(the probability distribution of data)이다. 자료의 확률분포는 특정한 모형을 가정했을 때 현재 연구자가 가진 자료가 그 모형을 통해 생성될 수 있는 개연성을 확률분포의 형태로 보여준다. 자료의 확률분포는 빈도주의에서 말하는 우도(likelihood)와 그 형태는 같으나 해석은 같지 않다. 자료의 확률분포는 그 총합이 1이며 0과 1사이에 모든 값이 위치하는 적정확률분포인 반면 빈도주의에서 말하는 우도는 하나의 상수값에 불과하다. 자료를 모집단의 무작위표본으로 보고 모수를 고정된 것으로 보는 빈도주의 패러다임에서는 현재 연구자가 가진 자료에 대한 확률분포를 표현하는 것이 불가능하다.

베이지안 자료분석의 두 번째 요소는 사전확률분포(the prior probability distribution)이다. 사전확률분포는 주로 모수에 대해 설

4) http://www.scholarpedia.org/article/Bayesian_statistics

5) 이러한 견해 차이는 사실 확률의 존재론적 특성에 대한 오래된 논쟁에 그 기원을 두고 있다. 빈도주의 학파는 확률은 객관적으로 존재하는 실체라고 주장한 반면 베이지안 학파는 객관적 확률이란 존재하지 않으며 확률은 주관적인 것이라고 본다. 오늘날에는 확률에 대한 존재론적 대립은 자취를 거의 감추었다. 존재론적 논쟁이 과학철학적 입장에서는 나름의 의미를 가질 것이나, 객관적 확률과 주관적 확률이 동일한 수학적 공리에 기반 한 이상 그 존재론적 차이를 두고 다투는 것은 자료분석의 관점에서는 큰 의미가 없어졌다고 볼 수 있다.

정하며 연구자가 가지고 있는 사전정보를 반영하는 것이다. 언뜻 생각하면 자료이외의 정보를 자료분석에서 사용한다는 것이 객관성을 중시하는 과학적인 연구의 원칙에 어긋나는 것으로 생각할 수 있다. 실제로 사전확률분포의 사용에 대한 논쟁은 베이지안 통계에 대한 가장 오래되고 중요한 논쟁 중의 하나이다. 이에 대한 베이지안들의 입장을 간략하게 요약하면 베이지안은 확률이론을 체계적으로 적용하는 모든 자료분석은 사전확률분포를 가정하는 것이 필수적이며 사전확률분포의 자의성을 통제할 수 있는 다양한 방법이 있음을 강조한다. 이에 대해서는 다음 장에서 좀 더 구체적으로 살펴보기로 한다.

베이지안 자료분석의 세 번째 요소는 사후확률분포(the posterior distribution)이다. 사후확률분포란 연구자가 가진 사전확률분포를 자료의 확률분포를 통해 업데이트한 결과로 베이지안 분석의 결과라고 볼 수 있다. 연구자가 사전에 가지고 있었던 모수에 대한 정보를, 자료에서 얻게 된 새로운 정보를 이용하여 수정한 것이 사후확률분포로 나타난다. 베이지안 자료분석의 중요한 장점은 사후확률분포를 통해 가설검정, 반사실적 시뮬레이션(counterfactual simulation), 혹은 미래값 예측(forecast)과 같은 다양한 추가분석을 할 수 있다는 점이다.

2.2. 베이즈의 정리(The Bayes' Theorem)

베이지안 자료분석에서 사전확률분포가 자료의 확률분포를 이용하여 사후확률분포로 전환될 수 있는 원리는 베이즈의 정리를 통해서이다. 베이즈의 정리는 조건부확률들 간의 등가성에 대한 정의이다. 확률에 대한 정의를 따른다면 베이즈의 정리는 수학적으로 항상

www.kci.go.kr

참이다. 이렇게 간단한 정의가 통계학의 역사를 바꿀 만큼 중요성을 가지게 된 이유를 이해하기 위해서는 통계학의 역사를 거슬러 올라가서 이야기를 시작해야 한다.

18세기 수학에 남다른 재능을 가졌던 베이즈주교(Reverend Thomas Bayes)는 당시 과학자들을 괴롭혔던 문제에 대해 독창적인 접근을 제시한다. 훗날 베이즈의 정리라고 알려지게 될 이 정리는 우리가 알고 있는 정보로부터 모르는 정보를 업데이트하기 위한 가장 체계적인 접근법이다. 베이즈는 우리가 모르는 정보를, 우리가 기존에 알고 있는 정보를 이용하여 구할 수 있음을 조건부확률의 특성을 이용해 찾아낸다. 베이즈의 정리는 다음과 같이 쓸 수 있다:

$$\Pr(A_i|D) = \frac{\Pr(A_i)\Pr(D|A_i)}{\Pr(D)} = \frac{\Pr(A_i)\Pr(D|A_i)}{\sum_{i=1}^N \Pr(A_i)\Pr(D|A_i)}$$

<수식 1> 베이즈의 정리

주사위를 던져 6이 나올 확률을 추정하는 도박사의 문제를 예로 들면, $\Pr(A_i|D)$ 는 자료에 기반하여 주사위를 한 번 던져 6이 나올 확률을 업데이트한 사후확률분포이다. $\Pr(A_i)$ 는 6이 나올 확률에 대한 사전확률분포로 주사위의 대칭성을 고려한 이론적 추측이나 연구자가 옳다고 믿는 사전정보를 반영한다. $\Pr(D|A_i)$ 는 연구자가 가진 자료의 분포로 주사위를 여러 번 던져 나온 수의 기록들이 연구자가 가정한 확률모형을 따라 어떻게 분포하는지를 반영한다. 마지막으로 $\Pr(D)$ 는 자료의 한계확률분포(the marginal density of data)로 주사위를 한 번 던져 6이 나올 확률에 대한 모든 이론적 가능성을 고려할 때, 현재 연구자가 가진 자료가 관측될 가능성을

나타낸다. 자료의 한계확률분포($Pr(D)$)는 관심값(A_i)에 의존하지 않는 상수이므로 베이저안 추정에서 계산하지 않지만 베이저안 모형비교에서는 매우 중요한 역할을 한다.

가장 최근 한국에서 벌어진 논쟁의 예를 들어 베이지의 정리를 이해해 보자. 2008년 한국 산업안전공단은 삼성전자 백혈병 발생과 관련하여 대대적인 통계조사와 역학조사를 진행했다. 고용보험자료로부터 217,131명의 자료와 반도체제조 기업들의 인사자료 139,763명의 기록을 통계조사한 결과, 산업안전보건연구원은 반도체 노동자의 백혈병 위험도에 대해 남성의 경우 “백혈병 발생은 [일반인구와] 비슷한 수준”이었으며 여성의 경우 “인사자료 코호트에서 표준화사망비는 1.48(95% 신뢰구간 0.54 - 3.22), 표준화 암등록비는 1.31(95% 신뢰구간 0.57 - 2.59)으로 일반인구집단에 비해 약간 높았으나 통계적으로 유의하지 않았다”라고 결론지었다 (한국 산업안전공단 2008, 34쪽).⁶⁾

여기서 표준화사망비(Standardized Mortality Ratio)란 관찰사망수를 기대사망수로 나누어 100을 곱한 값으로 관찰사망수가 인구평균 기대사망수에 비해 얼마나 이례적인가를 표시하는 통계이다. 표준화사망비는 혈액암발병률을 인구표준에 대한 비율(ratio)로 표현하는 것이다. 그러나 정작 중요한 것은 임의의 노동자를 반도체 제조공정에 배치했을 때 이 노동자로부터 백혈병이 발생할 확률, 즉 조건부확률이다. 임의의 노동자가 반도체 제조공정에 배치되는지의 여부를 나타내는 임의변수를 X 라고 하고 반도체 제조공정에 배치되

6) 이후 서울대학교 보건대학원의 백도명팀(서울대 산학협력단 2009)과 연세대학교 보건대학원의 김인아 등(Kim et al. 2012)은 역학조사를 통해 반도체제조공정이 백혈병과 관련 혈액암 발병에 유의미한 영향을 미친다고 결론을 내려 백혈병과 관련 혈액암이 산업재해로 법원에 의해 판정되는 데에 중요한 기여를 했다.

는 경우를 $X=1$ 로, 백혈병의 발병여부를 Y 라고 놓고 백혈병 발병한 경우를 $Y=1$ 로 놓자.

우리의 관심은 임의의 노동자가 반도체 제조공정에 배치되었을 때 그로부터 백혈병 발병을 관찰하게 될 조건부확률, 즉 $\Pr(Y=1|X=1)$ 이 비반도체노동자에게서 백혈병을 관찰하게 될 조건부확률, 즉 $\Pr(Y=1|X=0)$ 보다 큰 가하는 것이다. 산업안전공단의 원자료가 공개되지 않은 관계로 발표된 통계치를 기반으로 위의 조건부확률들을 계산해보자. 먼저 인사자료코호트의 총인원은 139,763명이며 이 중에서 백혈병 암등록자는 총 16명(남자 8명, 여자 8명)이다(산업안전보건연구원 2008, 28쪽). 따라서

$$\Pr(Y=1 \text{ and } X=1) = \frac{16}{139763} = 0.0001144795$$

우리가 임의의 관찰대상을 전체 인구구성에서 뽑았을 때, 그 임의의 관찰대상이 반도체 근무경력을 가진 사람일 확률은 전체 인구구성(2002년 기준) 대비 인사자료코호트의 총인원의 비율로 근사값을 구할 수 있다.

$$\Pr(X=1) = \frac{139763}{47622179} = 0.00293483$$

따라서,

$$\Pr(Y=1|X=1) = \frac{\Pr(Y=1 \text{ and } X=1)}{\Pr(X=1)} = \frac{0.0001144795}{0.00293483} = 0.0390072$$

즉 반도체 제조공정에서 근무하는 노동자가 있을 때 그들에게서 백혈병암등록자가 관찰될 조건부확률은 약 0.04라고 볼 수 있다.

이제 일반인의 백혈병 발병률을 계산해보도록 하자. WHO (2010)에 따르면 한국의 백혈병 발병률은 10만명 당 2.9명, 즉

0.000029이다(산업안전보건연구원 2008, 28쪽). 반도체 노동자를 뺀 나머지 일반인의 발병률을 구하는 것이 정확하나 그에 대한 통계를 구할 수 없는 관계로 백혈병의 표준화 암등록비인 1.05배를 이용하여 반도체제조공정의 노동자를 뺀 나머지 일반인의 백혈병 발병률을 구하면 약 0.00002899575가 나온다.⁷⁾ 즉,

$$\begin{aligned} \Pr(Y=1 \text{ and } X=0) &= 0.00002899575 \\ \Pr(X=0) &= 1 - \Pr(X=1) = 0.9970652 \end{aligned}$$

베이즈정리를 이용하여 반도체 제조공정에 노출되지 않은 일반인이 백혈병에 걸릴 확률을 계산해 보면,

$$\begin{aligned} \Pr(Y=1|X=0) &= \frac{\Pr(Y=1 \text{ and } X=0)}{\Pr(X=0)} = \frac{0.00002899575}{0.9970652} = 0.0000290811 \\ \frac{\Pr(Y=1|X=1)}{\Pr(Y=1|X=0)} &= \frac{0.0390072}{0.0000290811} = 1341.325 \end{aligned}$$

즉 우리가 임의의 반도체 노동자 1명을 뽑았을 때 이 노동자가 백혈병에 걸릴 확률은 비반도체 종사자 보다 약 1341배나 더 높다고 볼 수 있다.⁸⁾

7) 인사기록 코호트의 반도체 노동자들의 수를 X2라하고 이를 제외한 일반국민을 X1이라고 하자. 그리고 해당 집단의 백혈병 환자수를 각각 x2, x1이라고 놓자. 그러면 $(x1 + x2)/(X1 + X2) = 0.000029$ 이며 $1.05 \cdot x1/X1 = x2/X2$ 이다. 이 두 방정식을 풀기 위해 $X1 = 47622179 - 139673$ 이고 $X2 = 139673$ 이라고 놓으면, $x1 = 1376.788$ 그리고 $x2 = 4.25516$ 임을 알 수 있다. 따라서 $x1/X1 = 0.00002899575$ 이다.

8) 베이지안 접근법을 통해 구한 관심값의 분포(불확실성 구간)를 알고자 한다면 각 확률들의 확률분포를 가정해서 구할 수 있으나 예의 단순함을 위해 본 지면에서는 생략하였다. 관심있는 독자를 위해 그 스케치만을 설명한다면, 예를 들어 백혈병암 발생의 자료분포가 반도체 노동자와 일반인 집단에서 각각 서로 다른 프와송분포(Poisson distribution)를 따른다고 가정하고 평균의 사전분포를 감마분포(Gamma distribution)로 놓으면 두 집단의 평균의 차이가 갖는 분포를 사후분포를 대조하여 쉽게 계산할 수 있다.

2.3. 현대 베이지안 추정방법: MCMC 혁명

베이지안 분석에서 추정(estimation)이란 사전분포와 자료분포로부터 모수의 사후분포를 찾아내는 것이다. 그러나 하나 이상의 모수를 가진 모형의 경우 베이지안 추정은 다차원적분을 포함하기 때문에 분석적인 해를 구하는 것이 쉽지 않다. 이러한 계산적 어려움이 오랫동안 베이지안 통계의 발전가능성을 가로막은 중요한 제약이었다.

이러한 계산적 어려움은 새로운 추정방법의 출현을 통해 중요한 돌파구가 마련되었다. 2차 대전당시 미국에서 맨하탄 프로젝트에 참여했던 메트로폴리스를 포함한 일군의 과학자들(Metropolis et al. 1953)은 컴퓨터 시뮬레이션을 통해 다차원적분의 해를 구하는 방법을 개발하였다. 메트로폴리스 등의 방법은 통계학자들에게 한동안 주목받지 못하다가 1970년 해이스팅스(Hastings 1970)에 의해 재발견되었다. 1980년대 말에 이르면 베이지안 통계학자들이 메트로폴리스와 해이스팅스의 시뮬레이션 방법의 통계적 적용가능성을 다양한 통계분야에서 확인하고 그 방법을 마르코프 체인 몬테카를로 방법(Markov chain Monte Carlo methods, 이하 MCMC)이라 명명하였다.

MCMC의 원리를 가장 쉽게 이해하는 방법은 분절적 마르코프 체인(discrete Markov chain)을 통해서이다. 마르코프 체인은 인접한 상태에 대해서만 의존성을 갖는 시계열 확률변수이다. 예를 들어 민주주의 이행이 다음과 같은 마르코프 체인을 따라 이루어진다고 가정해 보자.

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.1 & 0.7 \\ 0.3 & 0.2 & 0.5 \\ 0.1 & 0.1 & 0.8 \end{pmatrix}$$

a_{ij} 는 현재 i 상태에 있는 국가가 다음 시기에 j 상태로 이행할 확률을 나타낸다. 권위주의 체제는 1, 민주주의 이행기는 2, 그리고 3은 민주주의 공고화 상태를 나타낸다고 하자. 예를 들어 a_{12} 는 현재 권위주의 체제에 있는 국가가 다음 기에 민주주의 이행기로 이동할 가능성이며 A에서는 0.2로 가정되어 있다. 세 번째 열이 가장 높은 확률을 나타내는 것은 일단 민주주의 공고화 상태로 이전되면 민주주의가 지속될 가능성이 높다는 의미이다.⁹⁾

전이행렬 A는 몇 시기만 지나면 출발점에 상관없이 곧 안정된 상태로 이행하는 특성을 지니고 있다. 다른 말로 하면 임의로 관측된 국가가 전체주의, 이행기, 공고화된 민주주의일 확률이 무엇인지 위 전이행렬을 통해 알 수 있다는 것이다. 이를 살펴보기 위해 서로 다른 초기분포가 20기 뒤에 어떤 확률분포로 변하는지 계산해 보자.

9) 민주주의 이행이 제 1차 마르코프 체인(first order Markov chain)을 따른다고 가정하는 것은 한 국가의 체제이행 여부는 현재 그 국가가 어떤 레짐에 있느냐에 전적으로 좌우되며 현재 레짐 이전의 체제이행 과정이 어떠한지의 여부는 중요하지 않다는 것을 의미한다. 만약 민주주의 이행이 제 1차 마르코프 체인이 가정하는 것보다 더 오랜 기억을 가진 것이라면 고차원의 마르코프 체인을 사용하거나 역사적 지속성을 가정하는 시계열 모형 혹은 사건역사모형(event history model)을 사용하는 것이 바람직할 것이다.

$$\begin{aligned}
A^2 &= A * A = \begin{pmatrix} 0.14 & 0.11 & 0.75 \\ 0.17 & 0.12 & 0.71 \\ 0.13 & 0.11 & 0.76 \end{pmatrix} \\
A^3 &= A^2 * A = \begin{pmatrix} 0.136 & 0.111 & 0.753 \\ 0.141 & 0.112 & 0.747 \\ 0.135 & 0.111 & 0.754 \end{pmatrix} \\
&\vdots \\
A^{20} &= A^{19} * A = \begin{pmatrix} 0.1358025 & 0.1111111 & 0.7530864 \\ 0.1358025 & 0.1111111 & 0.7530864 \\ 0.1358025 & 0.1111111 & 0.7530864 \end{pmatrix}
\end{aligned}$$

$A * A$ 는 A 의 행렬곱으로, 이를 통해 2기 이후의 변화를 보여주는 전이행렬을 구할 수 있다. 같은 방식으로 20기 이후의 변화를 보여주는 전이행렬은 A^{20} 으로 구할 수 있다. 여기에 상이한 초기분포값을 곱해 보면 다음과 같다:

$$\begin{aligned}
p_1 &= \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \\
p_2 &= \begin{pmatrix} \frac{1}{10} & \frac{1}{10} & \frac{8}{10} \end{pmatrix} \\
p_3 &= \begin{pmatrix} \frac{8}{10} & \frac{1}{10} & \frac{1}{10} \end{pmatrix} \\
p_4 &= (1 \ 0 \ 0) \\
p_1 \times A^{20} &= p_2 \times A^{20} = p_3 \times A^{20} = p_4 \times A^{20} = (0.1358025 \ 0.1111111 \ 0.7530864)
\end{aligned}$$

즉, 상이한 초기분포에도 불구하고 20기 이후에 관측된 확률분포는 소수점 7자리까지 동일함을 확인할 수 있다. 초기분포에 상관 없이 임의의 한 국가를 국제체제로부터 관측했을 때, 그 국가가 공고화된 민주주의를 가진 국가일 확률은 0.75이며 독재국가일 확률은 약 0.14, 그리고 이행기에 있을 확률은 약 0.11인 것이다. 마르코프 체인의 시작점이 달라도 몇 차례의 전이를 겪으면서 초기분포의 영향이 사라지고 안정된 확률분포로 수렴하고 있음을 확인할 수 있다.

MCMC는 마르코프 체인의 이러한 특성을 이용하여 사후분포로

수렴하는 마르코프 체인을 찾아 일정수준까지 시뮬레이션을 수행하는 기법이다. 마르코프 체인의 몇 가지 중요한 속성상, 처음 몇 번의 시뮬레이션을 버리고 순차적으로 추출된 나머지 시뮬레이션 값을 모으면 사후분포로부터 직접 추출한 값과 거의 유사하게 된다. 기존 시뮬레이션 방법이 초기값이나 시뮬레이션 순서에 민감하고 균형점에 도달하는지의 여부가 불확실했던 반면 MCMC는 초기값이나 시뮬레이션 순서로부터 독립적이고 균형점 도달이 이론적으로 증명된 시뮬레이션 기법이라는 중요한 차이를 갖는다.

가장 간단하면서도 유용한 MCMC 기법은 깁스 추출법(Gibbs sampling)이다 (Geman and Geman 1984). 깁스 추출법은 전통적인 통계학에서 두 개의 조건부 정규분포(conditional normal distribution)로부터 결합정규분포(joint normal distribution)를 추출하는 방법과 유사하다. 신호처리학(signal processing studies)에서 오랫동안 사용되었던 칼만필터링(Kalman filtering) 역시 깁스 추출법과 사실상 동일한 방법이라고 볼 수 있다. 다만 깁스 추출법은 정규분포 이외의 분포로 확장될 수 있다는 장점이 있으며, 보다 일반적인 MCMC기법인 메트로폴리스-헤이스팅스 기법(Metropolis-Hastings methods)은 적절한 제안분포(proposal distribution)를 찾을 수만 있다면 모든 확률분포에 적용할 수 있다는 매우 큰 장점이 있다(Chib 2001, Gelfand and Smith 1990, Gelman et al. 2004 참조).

3. 베이저안 사회과학 방법론

지금까지 통계학에서 개발되고 발전되어온 베이저안 통계방법에 대해 살펴보았다. 본 장에서는 보다 구체적으로 베이저안 방법론이

하나의 사회과학 방법론으로서 가진 의미를 세 가지 측면에서 살펴볼 것이다.

3.1. 사전분포와 디 피네티의 정리

경험자료 분석에 대한 하나의 대안적 사회과학 방법론으로서 베이저안 방법론이 가지는 가장 중요한 의의는 사회과학에서 개발된 이론적 모형 혹은 가설이 경험적 자료와 맺는 관계를 확률이론(probability theory)을 이용하여 체계적으로 기술할 수 있다는 점이다.

사전분포와 자료분포로부터 사후분포를 추출하는 베이저안 통계는 오랫동안 사전분포의 사용이라는 점 때문에 집중적인 공격을 받아왔다. 과학적 연구기법이라면 연구자의 사전적 지식과 의견을 배제하고 자료에 있는 정보에만 의존해야 한다는 피셔(R. A. Fisher)의 주장(Aldrich 2008)은 많은 학자들의 공감대를 불러왔다. 이에 대한 베이저안들의 반응은 크게 두 가지이다. 하나는 소극적 방어로서, 정보가 거의 없는 사전분포(noninformative prior)를 사용하여 사전분포의 영향을 분석에서 사실상 배제할 수 있으며, 사전분포의 영향은 자료가 증가하면서 거의 사라지게 되므로 자료의 양이 극단적으로 적은 경우가 아니라면 전통적 통계분석과 크게 다르지 않다는 점이다. 그러나 이러한 주장은 사전분포의 영향이 단순히 분포를 넓게 잡음으로써 쉽게 배제되지 않는다는 점을 경시한 것이며,¹⁰⁾ 자료의 양이 많고 적음은 모형의 복잡성 정도에 대해 상대적일 수밖에 없다는 점에서 매우 제한적인 방어이다. 바로 이러한 점 때문에 사전분포가 분석결과에 미치는 영향을 최소화 또는 탈각하는 방법을 연구하는 객관적 베이저안(objective Bayesians)들이 등장했다

10) 대표적인 예로 위계모형에서 에러분포에 대한 가정이 분석결과에 미치는 영향에 대한 겔만(Gelman 2006)의 연구를 보라.

(Berger 2006).

베이저안들은 사전분포의 사용에 대한 보다 적극적인 해석으로 디 피네티의 정리(de Finetti's theorem)가 경험적 자료분석에서 가진 함의를 강조한다(예: Gelman et al. 2004, 121-125). 디 피네티의 정리는 사전분포의 사용이 확률이론을 이용한 자료분석의 필수적 요소라는 점을 보여주었다(de Finetti 1931). 디 피네티의 정리에 따르면, 우리가 관측한 자료들이 서로 교환가능한(exchangeable) 관측치들이라고 가정한다면,¹¹⁾ 모든 관측자료는 적절한 사전분포와 조건부 자료분포의 곱으로 표현될 수 있으며 이는 곧 사전분포를 포함시키지 않는다면 관측자료를 교환가능한 자료로 보기 매우 어렵다는 점을 의미한다. 따라서 자료분석에서 확률론적인 분석을 일관되고 체계적으로 진행하고자 한다면 연구자들은 자신이 사전에 가진 지식-그 정보의 양이 많은 적든-을 사전분포의 형태로 분석에 포함하는 것이 필수적이다.¹²⁾

사전분포의 사용이 자료분석에서 확률이론의 일관되고 체계적인

11) 교환가능성(exchangeability)란 쉽게 말해, 관측치들이 각각 고유한 것이 아니라 하나의 공통된 자료생성과정(data generating process)을 통해 생성되었음을 의미한다. 따라서 하나의 분포로 표현될 수 있다는 것이다. 전통적 통계학에서 말하는 i.i.d.(independently, and identically distributed)와 유사한 개념이다. 다만 i.i.d.는 교환가능성을 함의하지만 그 역은 성립하지 않는다.

12) 사전분포의 역할은 단순히 자료분석의 필수요소라는 수동적 기능에 국한되지 않는다. 예를 들어 퍼스(Firth 1993)는 사전분포에 관한 한 가지 재미있는 연구결과를 발표한 바 있다. 로짓모형을 대표적인 빈도주의 추정방법인 최대우도추정(maximum likelihood estimation)방법으로 추정하면 계수들이 상당한 편향을 가지고 있으며 그 편향을 제거하기 위해서는 스코어함수를 변형해야 함을 증명하였다. 그런데 퍼스는 이렇게 편향을 제거한 우도함수의 최종형태가 제프리 사전분포(Jeffrey's prior)를 사용한 베이저안 로짓모형과 일치한다고 밝혔다. 즉 베이저안의 사전분포가 최대우도추정의 모수값을 0에 가깝도록 수축(shrinkage)하여 우도만을 기반으로한 추정에서 나타날 수 있는 편향을 제거하는 역할을 하고 있다는 것이다.

적용을 가능케 한다면, 확률이론의 일관된 적용이 갖는 이점은 무엇인가? 이에 대한 대답은 매우 간단하다. 많은 사회과학 연구자들은 인식론적으로나 존재론적으로 결정론적인 세계관을 배격한다. 인간이 가진 인지적 한계 때문이거나, 측정의 제한성 때문이거나, 혹은 물리적 세계 그 자체의 속성 때문으로 인해 사회과학자의 자료분석은 불확실성의 영역을 고려해야 한다는 점에 대해서는 대부분의 사회과학자들이 합의할 것이다. 바로 이러한 사회과학자들의 합의를 고려한다면, 불확실성을 가장 체계적으로 표현하는 과학적 언어인 확률이론을 자료분석에서 일관되게 사용하자는 베이지안 방법론의 주장은 설득력이 높다.

3.2. 위계모형

베이지안 통계는 사전분포를 포함한다는 측면에서 사실상 항상 위계모형을 사용한다. 따라서 한 단계 이상의 사전분포를 사용하여 다양한 다층모형분석(multilevel model analysis)을 별도의 가정 없이 수행할 수 있다는 점은 베이지안 통계의 중요한 장점이라고 볼 수 있다. 위계모형을 이용한 다층모형분석이 사회과학 자료분석에서 갖는 중요성은 매우 크다. 사회과학 분야의 많은 관측자료는 군집구조(clustering)나 복수의 분석수준(multiple levels of analysis)을 가지고 있다. 이러한 군집구조나 상이한 분석수준을 자료분석에서 적절하게 반영하지 못할 경우, 심슨의 역설(Simpson's paradox)과 같은 오류에 빠지거나 정확하지 못한 추정값을 찾을 수 있다.

농업보조금의 예를 들어 위계모형을 통한 상이한 분석수준과 군집구조의 반영을 이해해 보자.¹³⁾ 농업보조금 지급에 관한 국가간

13) 이하의 논의는 Park and Jensen(2007)에서 참조하였다.

그리고 시계열간 연구를 진행한다고 가정하자. 이론적 모형을 통해 농업보조금 결정에 중요한 영향을 미치는 국가간 변수를 발견하고 이를 경험자료를 통해 분석해 보고자 한다. 수집된 농업보조금 자료는 국가, 연도, 그리고 작물별로 관측되어 있다. 이에 기초하여 농업보조금의 결정에 유의미한 영향을 미치는 변수를 국가, 시간, 그리고 작물로 크게 구분할 수 있다고 가정하면 다음과 같은 세 가지 모형을 생각해 볼 수 있다.

모형 1: $p(\text{농업보조금}) = f(\text{국가변수}) + \text{에러}$

모형 2: $p(\text{농업보조금}) = g(\text{국가변수}, \text{시간변수}) + \text{에러}$

모형 3: $p(\text{농업보조금}) = h(\text{국가변수}, \text{시간변수}, \text{작물변수}) + \text{에러}$.

모형 1은 농업보조금의 결정이 전적으로 국가변수에 의해서만 이루어지고 나머지는 모두 정규분포를 띠는 노이즈라고 가정하는 것이다. 모형 1을 사용하는 것은 연구자가 암묵적으로 시간변수나 작물변수의 평균적 영향은 0에 가깝다는 가정을 하는 것이며 만약 이러한 가정이 틀렸을 경우, 국가변수의 영향을 정확히 측정하는 것은 매우 어렵다. 마찬가지로 모형 2의 경우, 작물변수의 평균적 영향을 0으로 봄으로써 모형의 단순화를 가져왔으나 실제 작물변수의 평균적 영향을 알 수 없으므로 타당성을 확인하기 어렵다.

반면 모형 3은 농업보조금이 국가, 시간, 그리고 작물변수 모두에 의해 복합적으로 결정된다고 본다. 모형 3이 모형 1과 모형 2에 비해 더 적절한 모형이라는 점에는 쉽게 동의할 수 있지만, 문제는 모형 3을 자료분석에 사용하는 것이 쉽지 않다는 점이다. 세 가지 문제가 제기된다. 먼저, 국가, 시간, 그리고 작물이라는 세 가지 서로 다른 분석수준에서 관측된 자료들을 어떻게 하나의 모형에 온전히 담을 수 있을 것인가? 둘째, 국가, 시간, 그리고 작물이라는 세 가지 분석수준에서 온전히 관측되지 않은 변수들의 영향을 어떻게

통제할 것인가? 즉 동일한 국가의 보조금은 보이지 않는 공통요인을 공유할 것이고 동일한 작물의 보조금과, 동일한 시점의 보조금 역시 보이지 않는 공통요인을 공유할 수 있다. 마지막으로, 국가 변수의 영향이 시간과 공간, 혹은 작물에 따라 상이하게 나타날 수 있는 가능성-소위 맥락효과(contextual effects)-을 어떻게 분석에 반영할 것인가 라는 문제가 있다.

전통적 통계학의 고정효과모형(fixed effects model)은 국가, 시간, 혹은 작물변수의 영향이 각각 고정값을 가진 상수라고 보고 이들이 농업보조금에 미치는 영향을 산술적으로 소거하는 방식을 택한다. 또는 확률효과모형(random effects model)으로 국가, 시간, 작물변수가 고유의 오차분포를 갖는다고 본다. 반면 베이지안 위계모형은 국가, 시간, 작물변수가 각각 독자적인 확률모형을 가지고 있으며 상이한 분석수준들이 개별적인 사전분포의 형태로 농업보조금의 결정에 영향을 준다고 간주한다. 숨은 군집구조와 상이한 분석수준으로 인해 농업보조금의 한계자료분포($p(Y|\theta)$)를 이용하여 모수의 사후분포를 직접 찾기는 매우 어렵다. 그러나 베이지안 위계모형을 통해서 각각의 분석수준에 대한 분포를 설정하고 그 분포에 의존하는 농업보조금의 조건부분포($p(Y|\theta, \text{국가}, \text{시간}, \text{작물})$)를 찾는 것이 가능하고 이로부터 모수의 사후분포를 추출하는 것이 비교적 용이하다. 이 경우 국가, 시간, 작물의 군집구조와 상이한 분석수준이 사전에 통제되었기 때문에 조건부분포를 이용하여 추출된 모수의 사후분포는 애초 우리가 얻고자 했던 사후분포 즉, $p(\theta|Y)$ 와 동일한 것이 된다. 이를 대략적으로 수식화하면,

$$\begin{aligned}
p(\Theta|Y) &\propto p(Y|\Theta)p(\Theta) \\
p(Y|\Theta) &= \int p(Y|\Theta, \text{국가}, \text{시간}, \text{작물})p(\text{국가})p(\text{시간})p(\text{작물})d\text{국가}d\text{시간}d\text{작물} \\
p(\text{국가}) &= N(\alpha_0 + \sum_{i=1}^K \alpha_i \text{국가변수}_i, \sigma_{\text{국가}}^2) \\
p(\text{시간}) &= N(\beta_0 + \sum_{i=1}^L \beta_i \text{시간변수}_i, \sigma_{\text{시간}}^2) \\
p(\text{작물}) &= N(\gamma_0 + \sum_{i=1}^M \gamma_i \text{작물변수}_i, \sigma_{\text{작물}}^2).
\end{aligned}$$

이와 같은 방법으로 베이저안 위계모형은 다양한 모형과 자료에 확장되고 변용되어 적용될 수 있다. 위계모형을 통한 모형의 확장과 변용은 베이저안 자료분석의 가장 중요한 실용적 장점이라고 볼 수 있다.

지면관계상 본 논문에서 자세하게 언급할 수는 없지만 위계모형은 증대변수(augmented variable)를 도입하는 방식으로 모형 외에 존재하는 가상의 모수를 도입하여 확장될 수 있다. 예를 들어 프로빗 이분모형의 추정치는 모형의 비정규성으로 인해 추정방법이 매우 까다롭지만 프로빗 모형의 모수 외에 이분자료의 상태를 결정하는 $y_i^* = X_i\beta + \epsilon_i$ 라는 연속변수를 가정하면 이 연속변수에 대해 정규분포를 취하는 β 를 쉽게 추출할 수 있다 (Albert and Chib 1993). 이와 같이 위계모형 안에서는 변수의 증대가 자유롭게 될 수 있기 때문에 비선형모형과 같은 비정규분포에서의 모수추정에 매우 효과적이다.

3.3. 비표본자료의 분석과 가설검정

웨스턴과 잭만(Western and Jackman 1994)은 비교연구에서 베이저안 방법론의 도입을 촉구하는 글에서, 빈도주의 통계방법은 비표본자료의 분석에 “적용하는 것이 불가능(inapplicable)”함에도 불

구하고 무비판적으로 광범위하게 사용되고 있다고 비판한다 (Western and Jackman 1994, 415쪽). 여론조사나 실험을 통해 얻게 된 자료를 제외하면 대부분의 사회과학 자료는 반복적 표본추출이 가능하지 않는 비표본자료 또는 비확률자료(nonsampling or nonstochastic data)이다. 특히 비교정치나 국제정치, 그리고 지역연구와 같이 고정된 분석단위를 이용한 분석의 경우 자료의 한계가 물리적으로 설정되어 있어서 수집된 자료가 곧 모집단을 구성하는 경우가 대부분이다.

비표본자료의 분석에 빈도주의 통계방법을 사용할 경우 가장 심각한 인지적 불일치를 일으킬 수 있는 부분이 바로 영가설검정이다. 영가설검정은 표본에서 관찰된 표본통계값(예: t-statistic)이 영가설을 참으로 가정하는 모집단의 관점에서 볼 때 얼마나 극단적인가를 통해 영가설을 기각 또는 기각하지 않는 방식으로 이루어지며 이때 표본자료의 극단성을 보여주는 지표로 p값(p-value)을 사용한다. 이러한 영가설검정이 비표본자료에 적용될 때 나타나는 가장 큰 문제점은 바로 p값에 대한 해석이 매우 어렵거나 혹은 잘못되기 쉽다는 것이다.

예를 들어 민주주의 국가들은 전쟁을 잘 하지 않는다는 민주평화론 가설을 검증하기 위해 국가가 시계열 자료를 모아서 분석했고 그 결과 민주주의 국가쌍(a pair of democracy) 더미변수의 p값이 양측 검정(two-tailed test)에서 0.049가 나왔다고 하자. 이 경우 연구자는 민주주의 국가쌍이 다른 국가쌍과 전쟁발생의 가능성에서 차이가 없다는 영가설을 95% 신뢰수준에서 기각할 것이며 민주평화론의 주장은 경험자료와 일치한다고 주장할 것이다. 또 다른 연구자가 같은 자료에 대해 다른 통제변수를 사용해서 분석한 결과 양측 검정 p값이 0.051이 나왔다고 가정하자. 이제 이 연구자는 앞선

연구자의 결론을 비판하면서 민주주의 국가쌍의 전쟁발생 가능성이 다른 국가쌍의 그것과 차이가 없다는 영가설의 주장은 95% 신뢰수준에서 기각될 수 없다고 결정할 것이며 따라서 민주평화론은 잘못된 이론이라고 결론지을 것이다. 이 두 연구에서 발견된 p값의 차이는 불과 0.002에 불과하지만 이 작은 차이가 만들어내는 결론의 간극은 매우 크다.

위의 예에서 확인할 수 있는 비표본자료에서 영가설검정의 문제점은 크게 두 가지이다. 첫째, p값은 자료로부터 확인된 영가설의 개연성을 보여주는 지표가 아니다. 만약 p값이 자료로부터 확인된 영가설의 개연성이라면 $p\text{값} = p(\text{영가설}|\text{자료})$ 라고 표현되어야 할 것이다. 그러나 사실 p값은 영가설을 참이라고 가정한 상태에서 자료의 관측가능성을 의미하는 $p(\text{자료}|\text{영가설})$ 에 더 가깝다.¹⁴⁾ 그럼에도 불구하고 p값을 자료를 기반으로 한 영가설의 개연성, 즉 $p\text{값} = p(\text{영가설}|\text{자료})$ 으로 해석하는 경우가 매우 많다. 이 둘이 같기 위해서는 아래의 식에서 $p(\text{자료}) = p(\text{영가설})$ 이어야 한다:

$$p(\text{자료}|\text{영가설}) = \frac{p(\text{영가설}|\text{자료})p(\text{자료})}{p(\text{영가설})}$$

그러나 이 두 확률분포의 등가성을 전제할 아무런 이유가 없기 때문에 p값을 영가설의 개연성으로 해석하는 것은 잘못된 것이다.¹⁵⁾ 이러한 혼돈이 일어나게 된 가장 중요한 이유는 바로 영가설 검정과 p값이 일관된 이론적 기반 위에 있지 않기 때문이다. 기거렌저(Gigerenzer 2004)는 빈도주의 통계에 의해 사용되는 영가설검정

14) 정확하게 말하면 p값은 영가설이 참일 때, 반복되는 표본추출로부터 계산된 표본 통계값이 현재의 통계값보다 더 극단적인 값이 될 확률이라고 볼 수 있다. 즉, $p\text{값} = 2(1 - \Phi(\text{표본통계값}))$.

15) 이에 대한 자세한 논의는 Carlin and Louis(2000, 6-10)와 Gill(1999)를 참조하라.

의 틀은 피셔(R.A. Fisher)에 의해 제시된 p값과 니만과 퍼슨(Jerzy Neyman and Egon S. Pearson)의 가설개연성에 대한 결정이론을 뒤섞어 놓은 하나의 ‘관행(the null ritual)’에 불과할 뿐, 일관성이나 이론적 토대가 없다고 지적한다. 이러한 영가설검정의 비밀관성과 논리적 취약성은 비표본자료에 대한 분석에서 더욱 심각한 문제로 대두된다. 반복되는 표본추출을 전제할 때만 의미를 갖는 p값을 반복될 수 없는 비표본자료의 안에서 해석하는 것은 존재하지 않는 물질의 질량을 재려는 시도와 유사하다고 볼 수 있다.

영가설검정이 비표본자료의 분석에서 갖는 두 번째 문제점은 통계적 유의기준-95% 신뢰구간 혹은 양측검정 p값 0.05-의 자의성이다. 흔히 0.05라는 기준은 20번의 반복추출 중에서 많아야 한 번 관측될 만한 자료라는 의미로 해석된다. 그러나 위의 예에서 드러난 바와 같이 왜 0.047과 0.049의 차이는 무시되어야 하고 0.049와 0.051의 차이는 주목되어야 하는가에 대한 이론적 근거로는 지극히 취약하다. 특히 영가설검정이 비표본자료에 적용될 때, 20번의 표본추출 중에서 한 번(1/20)이라는 빈도주의적 기준은 더더욱 그 설득력이 약화된다.

베이저안 자료분석은 가설의 타당성을 수치화하여 정확하게 표현할 수 있기 때문에, 영가설을 항상 가정할 필요가 없으며 다른 가설을 기각함으로써 연구자 가설의 타당성을 입증하는 전략을 택할 필요도 없다. 또한 두 가지 이상의 가설이나 변수의 조합이 서로 완전히 다른 가설들을 비교할 수 있다.

마지막으로 가설의 개연성이 확률의 형태로 표현되기 때문에 분석결과의 해석을 위해 자의적인 유의성 기준에 절대적으로 의존할 필요가 없다. 한마디로 요약하면 베이저안 자료분석의 목표는 $p(\text{연구자가설}|\text{자료}) > p(\text{대항가설}|\text{자료})$ 를 다양한 대항가설에 대

해 확인하는 것이다.

3.4. 시간적·공간적 특수성의 고려

사회과학과 역사학의 교류에 평생을 바쳐온 소웰(William H. Sewell Jr.)은 사회과학이 명심해야 할 역사학의 제 1명제로 이질성(heterogeneity)을 들고 있다. 역사는 사회과학자의 단순화된 틀을 거부하는 불연속과 이질성으로 가득 차 있다는 것이다. 그러나 동시에 그 이질성을 이유로 일반화(generalization)의 과업을 포기하는 것이 바로 역사학자들의 한계라고 소웰은 지적한다 (Sewell 2005).

소웰의 문제제기는 비단 시간적 차원만이 아니라 공간적 차원에 서도 이해될 수 있다. 구체성과 세세한 묘사를 특징으로 하는 개별 사례연구의 경우 그 목적이 일반이론의 수립보다는 사례에 대한 이해를 높이는 것에 있는 경우가 많다. 따라서 일반화된 이론에서 무시될 수 있는 맥락의존성과 인과성의 전개과정, 그리고 인과관계의 복잡성이 개별사례연구의 중요한 연구주제이다. 반면 다수사례연구는 일반이론을 통해 추출된 가설이 다수의 사례들을 통해서 공통적으로 관측되는 패턴인지를 확인하는 것에 초점을 둔다. 따라서 맥락의존성이나 인과관계의 복잡성과 같은 주제들보다는 단순인과성을 전제하고 일반화가가능성을 높이는 데에 집중한다. 그렇다면 공간적 이질성 혹은 맥락의존성을 인정하면서도 변인들 간의 일반화를 추구할 수 있는 방법론적 틀은 무엇인가?

베이지안 사회과학 방법론은 맥락의존성 혹은 이질성과 일반화라는 사회과학 연구의 두 가지 대립되는 연구경향을 정량적 연구방법의 틀 안에서 종합할 수 있는 유연하고 종합적이며, 효과적인 방법론적 틀이다. 시간적 일시성(temporality), 사건의 역사성, 그리고

인과관계의 맥락의존성(contextuality)을 적극적으로 고려하는 것은 베이지안 방법을 통한 거시역사연구의 중요한 장점이다(Park 2012, Western and Kleykamp 2004).

그 대표적인 예가 최근 베이지안 방법의 중요한 연구주제로 부각되고 있는 다층모형(multilevel model)에 대한 연구이다(Gelman and Hill 2007). 다층모형은 전통적인 회귀분석모형 안에서 쉽게 측정할 수 없었던 맥락효과(contextual effect)를 반영하고 추정할 수 있는 효과적인 분석틀로 주목받고 있다.

시간적 차원에서 맥락효과에 대한 고려는 베이지안 시계열분석 방법을 통해 개발되고 있다. 대표적으로 모수의 값이 시간적으로 변화하는 동적선형모형(dynamic linear model), 모수의 값이 특정 시점을 전후로 변화한다고 가정하는 전환점모형(changepoint model) (Chib 1998, Park 2009), 다변량 시계열변수의 동적 변화를 추적하는 베이지안 벡터 자기회귀모형(vector autoregression model) (Sims and Zha 1998, Brandt and Freeman 2009), 그리고 두 개 이상의 숨은 레짐이 반복되면서 하나 이상의 모수값이 출몰하는 것을 연구하는 마르코프전환모형(Markov switching model) 등이 있다(Kim and Nelson 1999). 최근에는 베이지안 방법을 이용하여 맥락효과를 시간과 공간 차원에서 동시에 확인하고자 하는 시도가 진행되고 있다(Park 2012).

4. 베이지안 분석의 예

4.1. 이념점 연구

정치인들의 선호를 일차원 혹은 이차원의 공간을 통해 표현하는

www.kci.go.kr

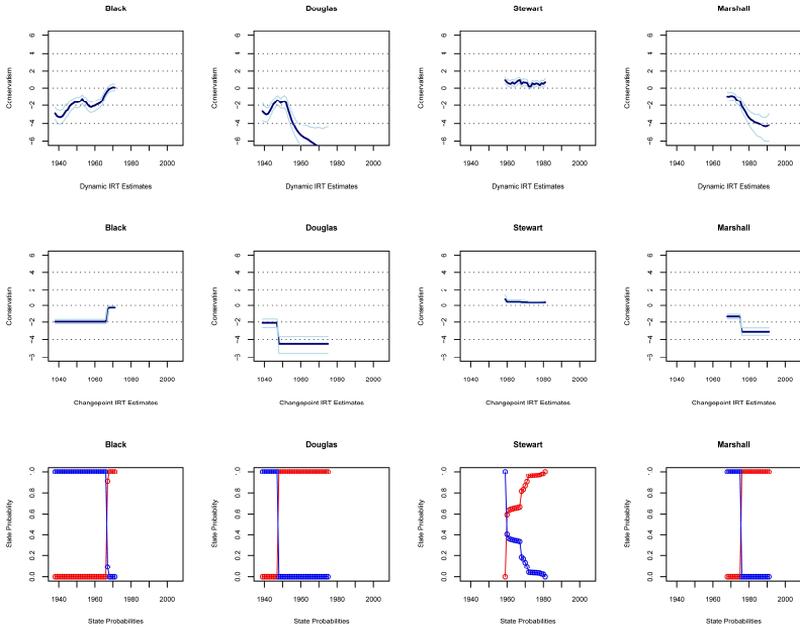
이념점(ideal point)은 의회에서의 투표와 같이 정치인들의 행위를 설명하는 핵심적인 개념이다. 미국정치에서 정치인들의 이념형 좌표는 보통 1차적으로 보수적(conservative) 이념과 자유주의적(liberal) 이념 중 어느 한 군데에 위치해 있는 것으로 추정된다. 1960년대 이래 많은 미국정치학자들은 국회의원들의 투표결과를 통해 이들의 이념점을 추정하려고 시도해왔다. 추상적인 이론적 개념으로만 치부되어오던 이념점을 객관적 경험자료를 통해 확인한다는 사실은 많은 정치학자들을 매료시켰고 다양한 방법들이 이 작업에 시도되었다.

가장 먼저 나타난 방법은 고차원 척도법(multidimensional scaling)으로서, 고유벡터분해와 같은 선형대수적 차원단순화 방법으로 투표자료에서 나타난 정치인들의 상관성행렬을 낮은 차원의 벡터로 표현하는 방법이었다. 이 방법은 그 단순성에도 불구하고 불확실성을 표현하지 못하는 기계적인 접근이었으며 또 투표행위에 대한 이론적 모형과는 유리된 접근법이었다. 이어서 등장한 것이 풀과로젠셀(Poole and Rosenthal 1985)이 개발한 노미네이트(NOMINATE, Nominal Three-Step Estimation)방법이었다. 노미네이트는 공간투표모형(spatial voting model)을 이항로짓모형(binary logit model)으로 변형한 뒤, 주요 모수들을 최우추정법(maximum likelihood method)으로 추정하는 것이었다. 그러나 노미네이트의 최우추정치는 통계적 일관성(consistency)이 취약한 추정치라는 비판이 제기되었다. 그 중요한 이유는 공간투표모형으로부터 변형된 이항로짓모형이 그 복잡성으로 인해 모수가 유일값을 갖지 못하는 모형의 확인불가능성(non-identifiability)때문이었다. 다른 말로 표현하면 모수의 수가 지나치게 많은 반면 모수의 유일값을 보여줄 수 있는 정보의 양은 제한적이라는 사실이 문제였다.

베이지안 이념점 추정법(Bayesian ideal point estimation methods)은 이러한 빈도주의 추정방법에 대한 대안적 접근법으로 등장하였다. 가장 먼저 클린턴, 잭만, 그리고 리버스(Clinton, Jackman, and Rivers 2004)는 교육통계학과 심리학에서 자주 사용되는 2모수 문항반응모형(two parameter item response model)을 통해 공간투표모형을 경험적 자료로 일관되게 표현하였다. 베이지안 추정방법은 사전확률분포의 사용을 통해 최우추정법에서 나타난 모형의 확인불가성 문제를 상당부분 피할 수 있었고 보수-자유주의의 일차원 이념좌표 뿐만 아니라 2차원 이상의 이념좌표에 대한 추정도 가능했다. 여기서 더 나아가 베이지안 방법에 기반하여 이념점이 시간적으로 변화하는 양태를 추적하는 동적 이념형 모형을 추정하는 것이 가능하다(Martin and Quinn 2000, Park 2011).

[그림 1]은 1937년부터 2006년까지의 미국대법원 판결자료를 토대로 미국 대법원 판사의 이념점을 베이지안 방법을 이용하여 추적한 것이다. 2년 이하의 임기를 가진 6명의 판사를 제외하면, 이 시기 동안 43명의 대법원 판사가 4868개의 사례에 대해 판결을 했으며 판사 일인당 약 113개의 사례를 심의했다. 마틴과 퀸(Martin and Quinn 2002)은 베이지안 이념점 분석모형-프로빗 링크를 가진 베이지안 2모수 문항반응모형-에 동적선형모형을 가미하여 대법원 판사의 이념점이 시간적으로 어떠한 변화를 보였는지 분석한 바 있다. 주목할 점은 이 분석에서 사용된 유일한 자료는 종속변수인 판결결과 밖에 없으며 판결결과 또한 승인/기각의 이분자료(binary data)라는 점이다. 따라서 이렇게 제한된 자료를 통해 개별 사례의 특성을 통제하고 이념점의 동학을 추정하는 것은 계산적 부담이 매우 큰 작업이며 최우추정법과 같은 빈도주의 추정방법을 통해서서는 안정된 추정값을 찾는 것이 매우 어렵다.

[그림 1] 미국 대법원 판사의 이념점에 대한 전환점 분석 결과



* 출처: Park(2011b)

[그림 1]의 첫 번째 줄은 4명의 판사에 대한 마틴과 퀴의 분석 결과를 보여주고 있다. 음수는 진보적인 이념을 나타내고 양수는 보수적인 이념을 나타낸다. 스튜어트 판사(Justice Potter Stewart, 1958-1981)를 제외한 판사들의 이념점이 시간적으로 변화하고 있음을 알 수 있다. 만약 이념점은 시간적 변화가 없다는 모형을 사용해 분석하게 되면 이러한 시간적 변화를 시야에서 완전히 놓치게 되어 잘못된 결론에 도달할 수 있다.

[그림 1]의 두 번째와 세 번째 줄은 박종희(Park 2011b)의 이념점에 대한 숨은 마르코프 모형(hidden Markov model of ideal point estimation model) 분석의 결과이다. 모든 판사들이 한 번의

이념점 변화를 겪었을 수 있다는 가정을 통해 이념점의 가장 급격한 변화가 일어난 시점을 찾아 이념점의 숨은 레짐을 찾아낸 것이다. 마틴과 퀴의 연구와 유사하게 스튜어트 판사를 제외한 나머지 세 명의 판사들의 이념점에서 유의미한 시간적 변화를 관측할 수 있다. 마틴과 퀴의 연구결과와 숨은 마르코프 모형 분석의 결과가 가진 중요한 차이점은 마틴과 퀴는 이념점의 변화를 임의보행과 같은 부드러운 진화과정으로 본 반면, 숨은 마르코프 모형은 이념점의 변화가 특정 시점을 전후로 한 단절적 변화로 본 것이다. 따라서 숨은 마르코프 모형은 이념점의 단절점(turning point)를 명시적으로 찾을 수 있고 단절점 이전과 이후의 차이를 비교하기 용이하다는 장점이 있다.

4.2. 선거예측 연구

베이저안 분석의 중요한 장점 중의 하나는 바로 예측이 용이하다는 점이다. 조건부 확률모형을 이용하는 베이저안 분석의 특성상 미래값이나 반사실적 상황에 대한 예측은 베이저안 분석결과를 통해 쉽게 계산가능하다. 수식으로 표현하자면 베이저안 예측이란,

$$\begin{aligned} p(\text{예측}|\text{자료}) &= \int p(\text{예측}, \theta|\text{자료})d\theta = \int p(\text{예측}|\theta, \text{자료})p(\theta|\text{자료})d\theta \\ &= \int p(\text{예측}|\theta)p(\theta|\text{자료})d\theta \end{aligned}$$

이다. 이 때 $p(\theta|\text{자료})$ 란 사후분포를 말하며 $p(\text{예측}|\theta)$ 란 사후분포를 통해 추출된 모수값을 이용해서 예측값을 구하는 것을 말한다. 따라서 예측을 위한 별도의 부가적인 가정이나 논리의 비약 없이 확률이론에 의해 뒷받침되는 과학적인 예측이 가능하다.

이러한 베이저안 분석의 예측적 장점은 선거예측조사에서 이미

검증된 바 있다. 2008년과 2012년 미국 대통령 선거예측에서 매우 정확한 선거예측을 하여 큰 화제를 몰고 온 야구통계학자 실버(Nate Silver)는 자신의 선거예측 방법이 베이지안 방법론에 의존하고 있다고 말한 바 있다(Silver 2012). 실버의 베이지안 모형이 아직 투명하게 공개되지는 않았지만 2012년 미국 대통령 선거 예측에서 실버보다 더 정확한 예측을 한 것으로 주목받는 린저(Drew Linzer)는 자신의 베이지안 방법과 모형을 공개하였다(Linzer 2013).

[그림 2]는 린저가 베이지안 예측모형을 통해 각 주에서 2012년 미국 대통령 선거결과를 예측한 것을 보여준다. 린저의 방법은 주(州)별로 진행된 여론조사 결과를 베이지안 위계모형을 통해 종합하고 동적선형모형을 통해 그 시간적 추이를 추적하여 여론의 흐름을 예측하는 것이다. 주별 선거결과를 통해 대통령 선거결과가 결정되는 미국 대통령선거의 특징을 고려할 때, 린저의 모형은 매우 적절하고 현실적인 모형이라고 할 수 있다. [그림 2]에서 확인할 수 있는 바와 같이 린저의 베이지안 위계모형은 미국 모든 주의 선거결과를 정확하게 예측하였다.

박중희(2013)는 린저의 방법과 유사한 베이지안 위계-동적선형모형을 이용하여 2012년 한국 대통령선거 여론조사에서 나타난 조사기관 편향을 조사하였다. 박중희는 조사기관 편향(house bias)을 조사기관 고정값(house effects) 중에서 95% 확률구간이 0을 포함하지 않는 극단적인 고정값이라고 정의하고 이를 통해 중앙일보와 엠비존의 조사결과가 편향이라고 부를 수 있는 극단적인 고정값을 보였다고 분석하였다.

[그림 3]의 왼쪽 그래프는 조사기관 고정값에 대한 베이지안 분석 결과를 보여주며 오른쪽 그래프는 빈도주의 방법인 고정효과 모

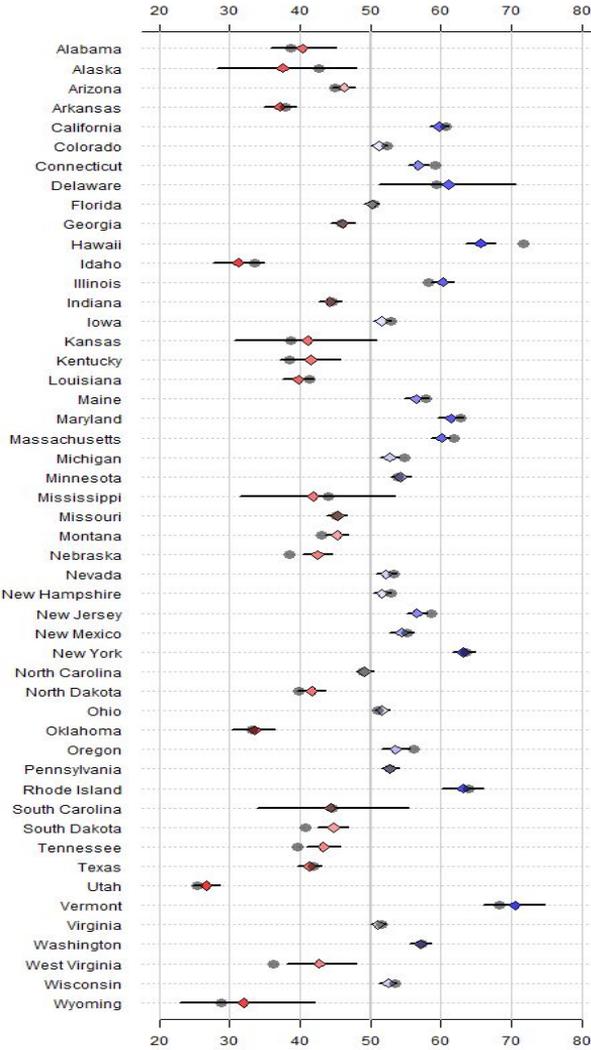
www.kci.go.kr

형을 이용한 분석결과이다. 고정효과 모형 분석을 위해서 불가피하게 한 조사기관(한국리서치)을 기준범주(base category)로 사용해야 했다. 고정효과 모형은 조사횟수에 따라 신뢰구간의 폭이 크게 변화하는 반면 베이지안 분석결과는 그 차이가 비교적 적음을 알 수 있다. 이러한 차이는 개별 조사기관들만의 정보를 사용하는 고정효과모형과는 달리 베이지안 위계모형은 정보의 양에 따라 다른 조사기관들의 정보로부터 설명력을 빌려오는 방식(borrowing strength)을 취하기 때문이다. 이를 통해 과다추정(overfitting) 또는 과소추정(underfitting)의 위험을 피할 수 있다는 것이 베이지안 위계모형이 고정값 모형에 대해 갖는 중요한 이점이다(Gelman and Hill 2007).

4.3. 전환점 연구

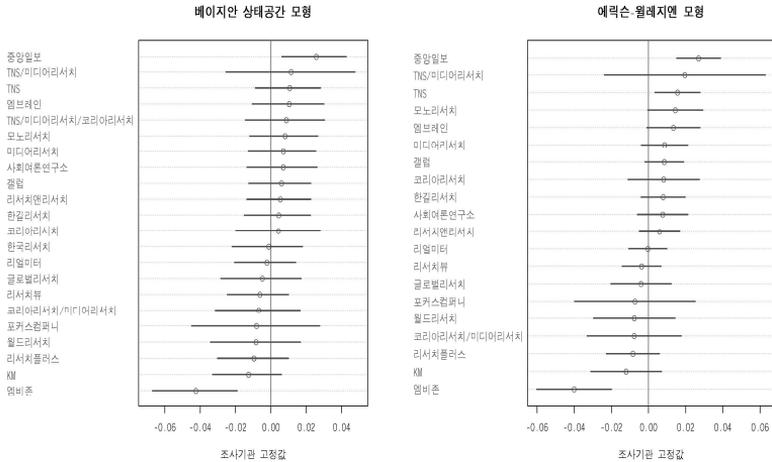
전환점모형은 인과관계의 시간적 맥락의존성을 측정할 수 있는 중요한 분석도구로 주목받고 있다. 특히 숨은 마르코프 모형을 통한 베이지안 전환점분석은 전환점의 수에 대한 제한이 없고 비선형모형에 대한 전환점추정이 가능하며 모수의 변화가능성에 대한 제약이 불필요하다는 점 등에서 장점을 지니고 있다. 대표적으로 칩(Chib 1998)에 의해 개발된 숨은 마르코프 모형을 이용한 전환점분석은 비회귀적인 특성을 지닌 거시역사과정을 분석함에 있어서 매우 유용하다.

[그림 2] 베이지안 예측모형을 통한 린저의 2012년 미국 대통령 선거
예측(마름모와 선)과 결과(회색 원)



* 출처: Votamatic(<http://votamatic.org/forecast-detail/>)

[그림 3] 제 18대 대통령 선거 여론조사에서 나타난 조사기관 고정값



- * 왼쪽 그래프는 베이지안 모형으로 추정된 조사기관 고정값이며 오른쪽은 OLS-고정효과 모형으로 추정된 조사기관 고정값이다. OLS-고정효과 모형은 한국리서치를 가변수의 기준범주로 사용하였다. (-)는 친문재인 고정값을, (+)는 친박근혜 고정값을 나타낸다.
- * 출처: 박종희(2013).

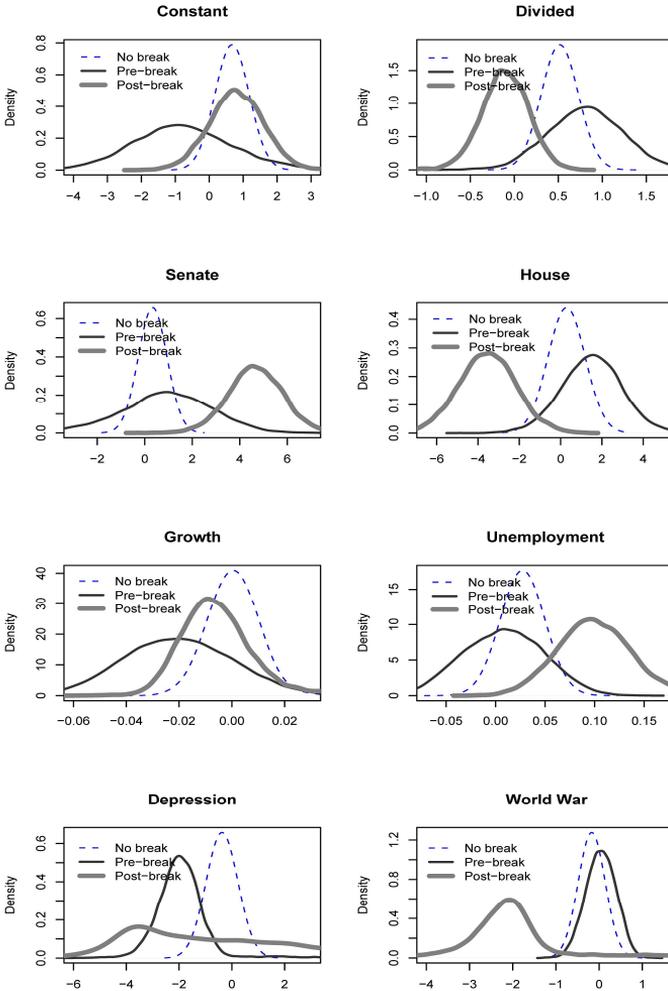
박종희(Park 2009)는 베이지안 프와송 전환점회귀모형 (Bayesian Poisson regression changepoint model)을 개발하여 미국 대통령 무력사용결정의 역사적 변화과정을 19세기 말부터 20세기 말까지 추적하였다. [그림 4]는 전환점 모형의 회귀분석 계수 (pre-break와 post-break)를 전환점을 고려하지 않는 회귀변속 계수(no break)와 비교하고 있다. 먼저 자료의 한계분포를 이용한 베이지안 모형비교를 통해 전환점을 고려하지 않은 모형은 전환점을 고려한 모형에 비해 자료에 대한 설명력이 매우 뒤떨어진다는 점을 확인하였다. 따라서 전환점을 무시한 기존 경험연구들의 결론은 그 타당성의 근거가 매우 취약하다고 볼 수 있다.

보다 구체적으로 전환점분석 결과를 살펴보면, 전환점을 고려하

지 않은 모형에서는 분점정부(Divided Government)가 무력사용의 횟수를 증가시키는 경향이 19세기 말부터 있어 온 것으로 보여주는 반면, 전환점 분석결과는 1943년 이전에는 분점정부의 영향이 거의 없었으며 분점정부와 무력사용 빈도 사이의 양의 상관성은 1943년 이후, 즉 냉전시기의 대통령들 사이에서만 관측되고 있음을 보여준다. 마찬가지로 실업률(Unemployment)이나 경제성장(Growth), 그리고 의회에서의 대통령당의 의석수(Senate과 House)의 효과도 냉전 이후와 이전에 중대한 차이를 보여주고 있음을 확인할 수 있다. 이러한 결과는 대공황(Depression)과 1·2차 대전(Great War)이 무력사용 빈도에 미친 직접적 영향을 통제한 뒤에 얻은 것이다. 결론적으로 무력사용 결정요인의 효과는 1943년을 기점으로 큰 변화를 보여주고 있는데, 이는 진주만습격으로 인해 미국 외교정책이 근본적인 변화를 겪었다는 냉전사학자 개디스(John Lewis Gaddis)의 역사적 연구결과와 일치하는 것이다 (Gaddis 1997).

베이지안 전환점분석 방법은 국가간 시계열자료의 분석에도 적용될 수 있다. 박중희(Park 2012)는 국제무역기구의 양자무역에 대한 효과를 고정값모형을 통해 분석한 골드스타인 등(Goldstein et al. 2007)의 연구를 패널전환점 분석방법(hidden Markov panel models)을 이용하여 재분석하였다. 이를 통해 양자무역에 영향을 주는 국가간 숨은 변수가 고정효과모형의 가정과는 달리 중대한 시간적 변화를 겪었음을 확인하고 그 영향을 통제한 뒤에 모형을 재분석한 결과, 골드스타인 등의 주장이 일정 부분 과장된 면이 있음을 확인하였다.

[그림 4] 미국 대통령 무력사용 결정요인의 역사적 변화



* 출처: Park(2009).

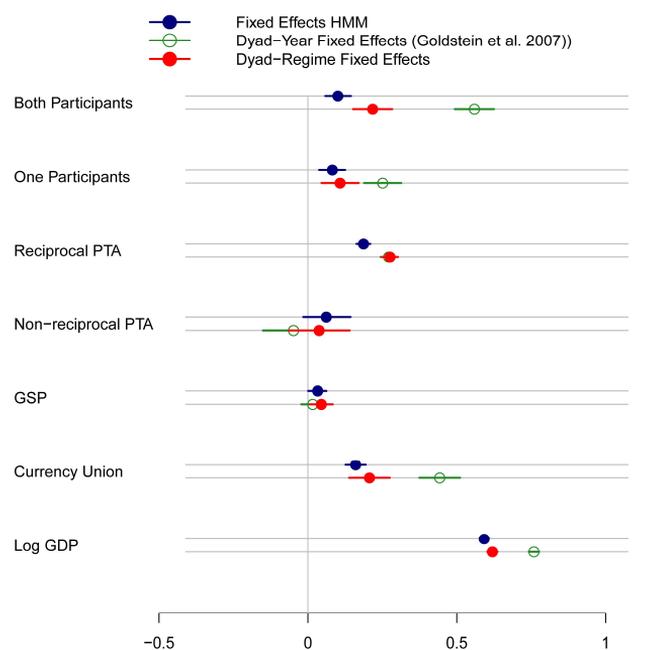
구체적으로, 골드스타인 등은 국가들이 모두 국제무역기구(the GATT/WTO)의 회원국일 때에는 그렇지 않은 경우보다 그들간의 양자무역이 70%이상 증가하였다고 보고하였다. 그러나 국가간 숨은 변수의 시간적 변화를 고려한 패널전환점모형은 쌍방 무역기구 회원국들 사이에서 양자무역의 증가가 10.6%증가에 그친 것으로 추정하였다. 또한 국제무역기구의 회원지위가 배타적인 무역협정인 특혜 무역협정(PTA)보다 양자무역 증가에 더 큰 영향을 주었다는 골드스타인 등의 주장 또한 패널전환점 분석결과에서는 오류인 것으로 밝혀졌다(Park 2012, 1048).

5. 결어

본 논문에서는 베이지안 통계를 이용한 베이지안 사회과학 방법론의 의의와 가능성, 그리고 연구경향을 개괄적으로 살펴보았다. 본문에서 구체적으로 살펴보았듯이 베이지안 통계를 이용한 연구는 빈도주의 패러다임의 약점을 보완하면서 통계연구의 중요한 대안적 패러다임으로 부상하고 있다. 본 논문은 이러한 베이지안 통계방법이 단순히 하나의 통계기법으로 머무는 것이 아니라 사회과학 연구의 독자적인 경험적 방법론(methodology)으로 정립될 수 있는 가를 살펴보았다. 본 논문의 논의를 정리하면, 베이지안 방법을 사전분포를 통한 현상의 확률적 재현, 위계모형을 통한 모형의 확장가능성, 비표본자료에서의 가설 및 이론 검증, 그리고 시공간적 맥락특수성을 고려하는 모형의 유연성 등에서 하나의 독자적 패러다임으로 정

립될 수 있는 가능성을 가지고 있다고 볼 수 있다. 바로 이와 같은 강점으로 인해 베이지안 연구는 사회과학을 포함한 과학연구 일반에서 큰 주목을 받고 있는 것이다(Eddy 2004, McGrayne 2012).

[그림 5] 국제무역기구의 양자무역에 대한 효과



* 분석단위는 방향성을 가진 국가쌍-연도(the directed dyad-year)이며 종속변수는 자연로그로 치환된 수입(1967년 미국 달러로 표준화)이다. GSP는 일반특혜관세(the Generalized System of Preferences)이며 PTA는 특별혜무역협정(Preferential Trade Agreements)이다.

* 출처: Park(2012).

그러나 모든 과학적 연구방법과 마찬가지로 베이지안 방법론 역시 약점이 없는 무오류의 연구방법이 아니다. 베이지안 방법론 역시

몇 가지 중요한 부분에서 취약성을 가지고 있다. 첫째, 베이지안 연구방법은 컴퓨터 시뮬레이션을 통한 추정방법을 사용하기 때문에, 연구방법의 교육에 있어서 프로그래밍의 중요성이 매우 크며 추정과정에서도 계산적 부담이 큰 경우가 많다. 이런 이유로 기본적인 모형을 제외하면, 베이지안 분석은 프로그래밍에 능한 소수의 전유물처럼 비춰지기도 한다. 그러나 최근 BUGS project를 통해 WinBUGS와 JAGS, 그리고 OpenBUGS와 같이 사용이 용이한 무료 베이지안 소프트웨어가 광범위하게 보급되어 베이지안 분석을 훨씬 용이하게 하고 있다. 또한 베이지안 분석에 대한 프로그램을 제공하는 다양한 패키지가 R이나 Python, 그리고 Matlab 사용자들에게 무료로 제공되고 있어서 복잡한 프로그래밍 없이도 베이지안 분석을 할 수 있는 길이 열려 있다.¹⁶⁾ 그 대표적인 예가 필자가 개발에 참여하고 있는 MCMCpack (Martin et al. 2011)이다.

베이지안 연구의 두 번째 약점은 자의적인 사전확률분포의 사용 가능성이다. 사전확률분포의 사용이 불가피하고 필수적이라는 것이 베이지안 연구자들의 주장이나 사전확률분포의 선택이 분석결과에 어떤 영향을 줄 것인가는 자료와 모형, 그리고 사전확률분포의 종류에 따라 매우 다르다. 사전확률분포의 자의적 영향을 줄일 수 있는 다양한 방법이 개발 중이지만 아직 이에 대한 통일된 틀이 존재하지 않는 것은 베이지안 연구의 확장과 발전을 막는 중요한 약점이라고 할 수 있다. 사전확률분포의 영향을 정확하게 보여주고 연구자가 그 영향을 숙지하고 사전확률분포를 선택할 수 있도록 도와주는 통일된 틀이 마련될 필요가 있다. 또한 선택된 사전확률분포에 대해 그 선택의 민감성을 확인하는 작업이 필수적이다.

16) R에서 제공되는 베이지안 패키지의 종류는 베이지안 태스크뷰(Bayesian taskview)에 잘 정리되어 있다.

(<http://cran.r-project.org/web/views/Bayesian.html>).

이외에도 마르코프 체인의 수렴여부에 대한 확증적 검증의 어려움이나 확률분포적 가정에 의존하지 않는 자료분석의 가능성, 그리고 혼합모형의 추정에서 라벨전환의 문제(label-switching problem) 등, 베이저안 통계 안에서 제기되는 다양한 문제들이 베이저안 방법론 개발자들에게 도전과 기회를 제공하고 있다.

마지막으로 베이저안 연구의 두 가지 중요한 최근 동향에 대해 소개함으로써 본 논문을 맺고자 한다. 먼저 최근 베이저안 통계방법에서 자료증가와 모수확장(data augmentation and parameter expansion) 방법을 통해 깃스 추출법과 같은 손쉬운 시뮬레이션 방법을 이용하여 복잡한 모형을 추정하는 것이 더욱 가능해지고 있다(Liu and Wu 1999). 앞서 언급한 바와 같이 이분모형과 같은 비선형모형의 경우 비정규분포를 사용하기 때문에 시뮬레이션이 매우 복잡할 수 있다. 이 때 자료증가와 모수확장 방법을 사용하면 숨은 변수를 사용하여 연구자의 관심변수를 손쉽게 추출할 수 있다. 자료증가와 모수확장은 그 기본 원리가 비교적 단순하고 다양한 모형에 적용이 가능하기 때문에 베이저안 분석의 영역을 확장함에 있어서 매우 중요한 기여를 할 것으로 기대된다.

두 번째로 주목할 만한 베이저안 연구의 최근 성과는 초(超)차원(transdimensional) MCMC 방법이다. 그린(Green 1995)에 의해 그 적용 가능성이 본격적으로 현실화된 초차원 MCMC 방법은 모수의 차원이 아니라 모형선택의 차원에 대해서도 마르코프 체인을 움직이게 하여 보다 현실적인 모형을 마르코프 체인이 더 자주 방문하도록 하는 기법이다. 이를 통해 모형의 개연성에 대한 사후분포, 즉 $p(\text{모형}|\text{자료})$ 을 정확하게 측정할 수 있다. 따라서 모수의 종류와 수가 서로 다른 모형들에 대해서도 초차원 MCMC 방법을 이용해서 모수추정과 모형선택을 동시에 수행할 수 있는 매우 효과적인

방법이다.

본 논문은 논의의 선명성을 위해 빈도주의를 베이지안 방법에 대립하는 분석틀로 제시하고 그 차이와 논쟁의 역사를 소개하였다. 그러나 앞서 언급한 바와 같이 베이지안 통계와 빈도주의는 확률이론의 수학적 공리(axiom)를 공유하며 모두 확률모형을 통해 사회현상을 설명하고자 한다는 측면에서 공통점을 갖는다. 실제로 오늘날에는 더 이상 베이지안 연구자와 비베이지안 연구자를 나누는 것이 어려울 만큼, 베이지안 방법은 연구자들이 철학적 견해 차이와 무관하게 자유롭게 택할 수 있는 하나의 방법론적 메뉴로서 통계교육과 자료분석의 영역에서 확고하게 제자리를 잡았다. 중요한 것은 연구자가 가진 이론의 검증과 자료의 분석에 무엇이 더 적합한가에 대한 고민과 판단이며, 이러한 측면에서 베이지안 방법과 빈도주의는 연구자들에게 보다 매력있는 메뉴를 제공하기 위해 끊임없이 경쟁하지만 서로 보완적인 연구방법 패러다임이라고 보아야 할 것이다.

참고문헌

- 박종희. 2013. “제 18대 대선 여론조사에서 나타난 조사기관 편향” 『조사연구』14권 1호. 1-30쪽.
- 산업안전보건연구원. 2008. 『반도체 제조공정 근로자 건강실태 역학조사 보고서 요약본』 한국산업안전공단.
- 서울대 산학협력단. 2010. 『삼성 반도체 사업장 위험성 평가 자문보고서』 서울대 산학협력단.
- Albert, James H. and Siddhartha Chib. 1993. “Bayesian Analysis of Binary and Polychotomous Response Data.” *Journal of the American Statistical Association*. Vol.88, No.422. pp. 669-679
- Aldrich, John. 2008. “R. A. Fisher on Bayes and Bayes' Theorem” *Bayesian Analysis* Vol.3. No.1. pp.161-170.
- Berger, James. 2006. “The Case for Objective Bayesian Analysis” *Bayesian Analysis*. Vol.1. No.3. pp.385-402.
- Brandt, Patrick T. and John R. Freeman. 2009. “Modeling Macro Political Dynamics” *Political Analysis*. Vol.17. No.2 pp. 113-142.
- Carlin, Bradley P. and Thomas A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. 2nd Edition. New York: Chapman & Hall/CRC.
- Chib, Siddhartha. 1998. “Estimation and comparison of multiple change-point models.” *Journal of Econometrics*, Vol.86. No.2. pp.221-241.
- Chib, Siddhartha. 2001. “Markov chain Monte Carlo methods: computation and inference,” in J.J. Heckman & E.E. Leamer (ed.), *Handbook of Econometrics*, eds. Vol.5. Ch.57. pp. 3569-3649.

- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2000. "The Statistical Analysis of Legislative Behavior: A Unified Approach." *American Political Science Review*. Vol.98. pp. 355-370.
- Collier, Ruth Berins and David Collier. 2002. *Shaping the Political Arena Critical Junctures, the Labor Movement, and Regime Dynamics in Latin America*. University of Notre Dame Press.
- de Finetti, Bruno. 1931. "Funzione caratteristica di un fenomeno aleatorio." *Atti della Reale Accademia Nazionale dei Lincei*. Vol.4. pp. 251-299.
- Eddy, Sean R. 2004. "What is Bayesian Statistics?" *Nature Biotechnology* Vol.22. pp.1177-8.
- Firth, David. 1993. "Bias Reduction of Maximum Likelihood Estimates" *Biometrika*. Vol.80. No.1. pp.27-38.
- Gaddis, John Lewis. 1997. *We Now Know: Rethinking Cold War History*. New York: Oxford University Press.
- Gelfand, Alan E. and Adrian F. M. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association*. Vol.85. pp.398-409.
- Gelman, Andrew. 2006. "Prior distributions for variance parameters in hierarchical models" *Bayesian Analysis* Vol.1. No.3. pp. 515-533.
- Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2004. *Bayesian Data Analysis*. Boca Raton, London, New York, and Washington D.C.: Chapman & Hall/CRC.
- Geman, Stuart and Donald Geman. 1984. "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images" *IEEE Transactions on Pattern Analysis and Machine*

- Intelligence* Vol.6. pp.721-741.
- Gigerenzer, Gerd. 2004. "Mindless Statistics" *Journal of Socio-Economics*. Vol.33. pp.587-606.
- Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly*. Vol.52. No.3. pp. 647-674.
- Gill, Jeff and John Freeman. 2013. "Dynamic Elicited Priors for Updating Covert Networks." *Network Science*. Vol.1. No.1. pp.68-94.
- Goldstein, Judith, Douglas Rivers, and Michael Tomz. 2007. "Institutions in International Relations: Understanding the Effects of the GATT and the WTO on World Trade." *International Organization* Vol.61. pp.37-67.
- Green, Peter J. 1995. "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination." *Biometrika* Vol.82. No.4. pp.711-732.
- Hastings, W. K. 1970. "Monte Carlo Sampling Methods using Markov Chains and Their Applications." *Biometrika*. Vol.57. pp.97-109.
- Holland. P.W. 1988. "Statistics and Causal Inference" *Journal of the American Statistical Association* Vol.81. pp.945-960.
- Ioannidis, John P.A. 2005. "Why Most Published Research Findings are False" *PLoS Medicine*. Vol.2. No.8. pp.696-701.
- Katzenstein, Peter J. 1985. *Small States in World Markets: Industrial Policy in Europe* Ithaca, Cornell: Cornell University Press.
- Kim, Chang-Jin and Charles R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. Cambridge, MA: The MIT Press.
- Kim, Inah, Hyun J. Kim, Sin Y. Lim, Jungok Kongyoo. 2012.

- “Leukemia and non-Hodgkin lymphoma in semiconductor industry workers in Korea.” *International Journal of Occupational and Environmental Health*. Vol.18. No.2. pp. 147-153.
- McGrayne, Sharon Bertsch. 2012. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. New Haven: Yale University Press
- Linzer, Drew. 2013. “Dynamic Bayesian Forecasting of Presidential Elections in the Senate” *Journal of the American Statistical Association*. Vol.108. No.501. pp.124-134.
- Liu, Jun S. and Ying Nian Wu. 1999. “Parameter Expansion for Data Augmentation.” *Journal of the American Statistical Association*. Vol.94, No.448. pp.1264-1274.
- Martin, Andrew D., Kevin M. Quinn 2002. “Dynamic ideal point estimation via Markov chain Monte Carlo for the U.S. Supreme Court, 1953-1999.” *Political Analysis*. Vol.10. No.2. pp. 134-153.
- Martin, Andrew D., Kevin M. Quinn and Jong Hee Park. 2011. “MCMCpack: Markov chain Monte Carlo in R.” *Journal of Statistical Software*. Vol.42, No.9. pp.2-21.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller E. 1953. “Equation of State Calculations by Fast Computing Machines.” *Journal of Chemical Physics*. Vol.21. pp.1087-1092.
- Moore, Barrington Jr. 1966. *Social Origins of Dictatorship and Democracy: Lord and Peasant in the Making of the Modern World*. Boston: Beacon Press.
- Park, Jong Hee. 2010. “Structural Change in the U.S. Presidents’ Use of Force Abroad.” *American Journal of Political Science*. Vol.54, No.3. pp.766-782.

- Park, Jong Hee. 2011a. "Changepoint Analysis of Binary and Ordinal Probit Models: An Application to Bank Rate Policy Under the Interwar Gold Standard." *Political Analysis*. Vol.19, No.2. pp.188–204.
- Park, Jong Hee. 2011b. "Modeling Preference Changes via a Hidden Markov Item Response Theory Model" in Steve Brooks, Andrew Gelman, Galin Jones and Xiao–Li Meng, eds., *Handbook of Markov Chain Monte Carlo Methods*. Chapman & Hall/CRC Press.
- Park, Jong Hee. 2012. "A Unified Method for Dynamic and Cross–Sectional Heterogeneity: Introducing Hidden Markov Panel Models" *American Journal of Political Science*. Vol.56. No.4. pp.1040–1054.
- Park, Jong Hee and Nathan M. Jensen. 2007. "Electoral Competition and Agricultural Support in OECD Countries" *American Journal of Political Science* Vol.51, Issue 2, pp.314–329.
- Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Poole, Keith. T. and Howard Rosenthal. 1985. "A Spatial Model for Legislative Roll Call Analysis." *American Journal of Political Science*. Vol.29. pp.357–384.
- Poole, Keith. T. and Howard Rosenthal. 1997. *Congress: A Political–Economic History of Roll–Call Voting*. Oxford: Oxford University Press.
- Robert, Christian P. and George Casella. 2004. *Monte Carlo Statistical Methods*. Springer.
- Rogowski, Ronald. 1989. *Commerce and Coalitions*. Princeton: Princeton University Press.
- Sewell, William H. 2005. *Logics of History: Social Theory and Social Transformation*. Chicago: University of Chicago Press.
- Skocpol, Theda. 1979. *States and Social Revolutions: A*

- Comparative Analysis of France, Russia, and China.*
Cambridge: Cambridge University Press.
- Silver, Nate. 2012. *The Signal and the Noise: Why So Many Predictions Fail — but Some Don't.* Penguin Press HC.
- Sims, C.A. and Tao Zha. 1998. “Bayesian Methods for Dynamic Multivariate Models.” *International Economic Review*. Vol.39. No.4. pp.949–968.
- Weber, Max. 2010. 『프로테스탄티즘의 윤리와 자본주의 정신』(Die Protestantische Ethik und der Geist des Kapitalismus). 김덕영 옮김. 도서출판 길.
- Western, Bruce. 2001. “Bayesian Thinking about Macrosociology.” *American Journal of Sociology*. Vol.107. No.2. pp.353–378.
- Western, Bruce and Simon Jackman. 1994. “Bayesian Inference for Comparative Research.” *American Political Science Review* Vol.88. pp.412–423.
- Western, Bruce, and Meredith Kleykamp. 2004. “A Bayesian Change Point Model for Historical Time Series Analysis.” *Political Analysis*. Vol.12. pp.354–74.

What is Bayesian Methodology?

Jong-Hee Park(Seoul National University)

Bayesian methods are a group of empirical methods that allow researchers to systematically describe various types of uncertainty involved in the process of data analysis using the mathematical language of probability. This paper introduces Bayesian statistics and discusses its implications as an empirical methodology for social scientists. I argue that the Bayesian methodology in social sciences is an effective empirical framework for the analysis of nonstochastic, multilevel, or historical data. Also, the Bayesian methodology can provide a more plausible framework for theory/hypothesis testing than the conventional null hypothesis test in a frequentist tradition. To illustrate the utility of the Bayesian methodology in political science, I discuss several important achievements in the application of the Bayesian methodology to questions in political science.

Keywords: Bayesian statistics, prior distribution, posterior distribution, hypothesis test, Markov chain Monte Carlo methods