

**연구노트**

**국회의원 선거에서의 주요정당 의석 수 예측**

Predicting Major Political Parties' Number of Seats in General Election:  
The Case of 2004 General Election of Korea

허명회\*

Myung-Hoe Huh

200여개의 지역구 선거가 동시에 치러지는 국회의원 선거에서 주요정당의 의석 수를 예상해야 할 필요가 있는데, 이제까지는 정당별로 당선확실 선거구 수에 경합 선거구 수를 적당히 더하는 상식적 수준의 선택에 의존하여 왔다. 그러나 선거 예측 조사 자료를 베이지 추론의 틀에 넣어 활용함으로써 정당 의석 수에 대한 합리적 점 예측과 구간 예측이 가능하다. 2004년의 제 17대 국회의원 선거에 적용하여 이 방법의 실용성을 살펴보았다.

주제어: 당선 가능성, 정당 의석 수 예측, 베이지 신뢰구간, 열세자 침묵

We calculated the predictive interval for the number of seats belonging to major political parties in the case of the 2004 General Election of Korea, using Bayesian frame of inference. Moreover, we proposed the adjustment procedure for correcting the minor group's propensity of refusal or nonresponse due to effect of the spiral of silence.

Key words: Bayesian inference, interval prediction for the number of seats, spiral of silence

\* 교신저자(corresponding author): 고려대학교 정경대학 통계학과 교수 허명회.  
E-mail: stat420@korea.ac.kr

## I. 베이즈 추론의 틀

베이즈 추론은 선거 조사자료의 해석을 보다 풍부하게 한다(Jackson 2004; 허명회 2008). 크기  $n$ 의 임의표본을 조사한 결과, 세 명의 후보 A, B, C가 경합하는 경우에서 각 후보에 대한 지지 수가  $n_A, n_B, n_C$ 로 나타났다면 ( $n_A + n_B + n_C = n$ ),

$$(n_A, n_B, n_C) \sim \text{Multinomial}(n, \theta_A, \theta_B, \theta_C)$$

로 볼 수 있다. 여기서  $\theta_A, \theta_B, \theta_C$ 는 A, B, C의 모비율로 합이 1이다. 따라서 사전분포(prior)를  $\text{Dirichlet}(\alpha_A, \alpha_B, \alpha_C)$ 로 놓으면 ( $\theta_A, \theta_B, \theta_C$ )는 사후적으로(a posteriori)

$$(\theta_A, \theta_B, \theta_C) \text{ given } (n_A, n_B, n_C) \sim \text{Dirichlet}(n_A + \alpha_A, n_B + \alpha_B, n_C + \alpha_C)$$

를 따른다. 이 연구에서는 무정보적(non-informative)인 균일분포를 ( $\theta_A, \theta_B, \theta_C$ )에 대한 사전분포로 놓기로 한다(즉,  $\alpha_A = \alpha_B = \alpha_C = 1$ ). 개별 지역구에 대한 추가적 정보가 가용한 경우에는 더 정교하게  $\alpha_A, \alpha_B, \alpha_C$  파라미터를 선정할 수 있겠지만  $\alpha_A + \alpha_B + \alpha_C \leq 10$ 이고  $n (= n_A + n_B + n_C)$ 이 1,000 정도라면  $\theta_A, \theta_B, \theta_C$ 에 대한 사후분포는  $\alpha_A, \alpha_B, \alpha_C$  파라미터 설정에 거의 영향을 받지 않을 것이다.

허명회(2008)는 A 후보가 1위일 확률, 즉 당선가능도를

$$P\{\theta_A \geq \theta_B, \theta_A \geq \theta_C \mid n_A, n_B, n_C\} (= w_A)$$

로 나타내어 선거구에서 후보자간 우열이 뚜렷하지 않은 경우에서도 조사 결과를 유용한 정보로 활용할 수 있음을 주장하였다. 예컨대  $w_A = 0.8$ ,  $w_B = 0.16$ ,  $w_C = 0.04$ 인 경우, 조사 자료에 기반하여 추정컨대 A 후보가 당선될 가능성이 (베이즈 사후확률) 80%로 다른 후보들의 당선 가능성보다

크다고 하겠다. 그러나 빈도학파(Frequentist)의 방법으로는 후보간 차이가 유의하지 않은 것으로 나타나므로 후보들이 오차 범위 내에서 접전이라고밖에 말하기 어렵다. 이상의 베이즈적 방법은 후보자 수가 2명, 또는 4명 이상인 경우로 쉽게 확장 가능하다.

$s$ 개의 의석이 걸린 국회의원 선거에서 정당 A, B, C, ...의 지역구 선거 예측 문제를 생각하기로 하자. 베이즈 방법으로 산출한 선거구  $j(=1, \dots, s)$ 에서의 정당 A 후보자의 당선가능도를  $w_A^{(j)}$ 로 표기하면 정당 A의 총 의석 수는 다음과 같이 표현된다.

$$N_A = Z_A^{(1)} + \dots + Z_A^{(s)},$$

여기서  $Z_A^{(j)}$ 는 독립적으로 Bernoulli ( $w_A^{(j)}$ ) 분포를 따르는 확률변수이다( $j=1, \dots, s$ ). 확률변수  $N_A$ 의 분포는 몬테칼로 시행을 통해 모사할 수 있겠다. 마찬가지로 방식으로, 정당 B의 총 의석 수와 C의 총 의석 수의 확률적 행태도 파악할 수 있다. 이런 해법은 서혜선·허명회(1997)에 의해 제안된 바 있지만 실제 쓰이지 않았다.

2004년 제17대 국회의원 선거에서 지역구 총 의석 수는 243개였으며 KBS와 SBS는 조사전문회사인 미디어리서치(MRI)와 TNS소프레스에 의뢰하여 선거 예측조사를 하였다. 모든 선거구에 대하여 전화조사가 실시되었고 결합으로 판단된 114개 선거구에 대하여는 출구조사가 추가되었다. 최종 지지율은 선거구별로 마지막(전화 또는 출구) 조사자료를 토대로 산출되었다.

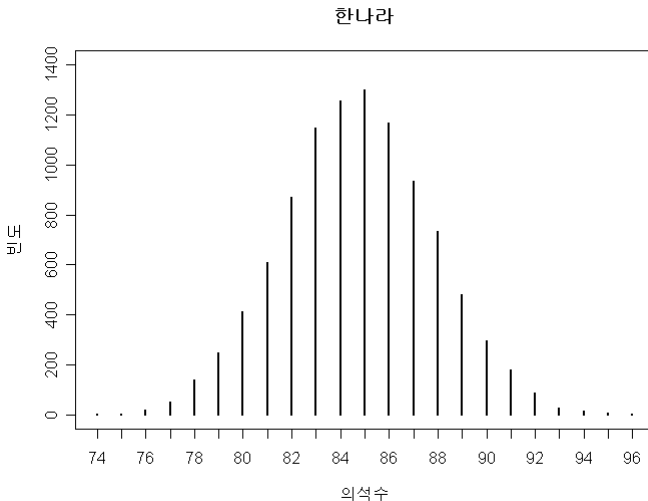
<그림 1>과 <그림 2>는 한나라당과 열린우리당의 의석 수에 대한 예측 분포를 그린 것이다.<sup>1)</sup> 신뢰수준 95%에서 한나라당 의석 수는 79석~91석으로 예측되고 열린우리당 의석 수는 138석~151석으로 예측된다(민주당은 4~7석, 자

1) 출구조사는 투표소를 추출하여 조사되므로 다단계 집락추출로 볼 수 있다. 이에 따라 설계효과(design effect)를 감안하여 유효표본크기를 보정할 필요가 있는데, 여기서는 설계효과를 2005년 지방선거에서 추정된 5로 간주하였다. 추가적으로, 20% 내외의 무응답이 있음을 감안하여 각 선거구 당선가능도 산출 시 표본크기를 20% 정도 감축하여 썼다. 정당 의석 수 분포 산출을 위해 몬테칼로 시행이 총 10,000회 반복되었다.

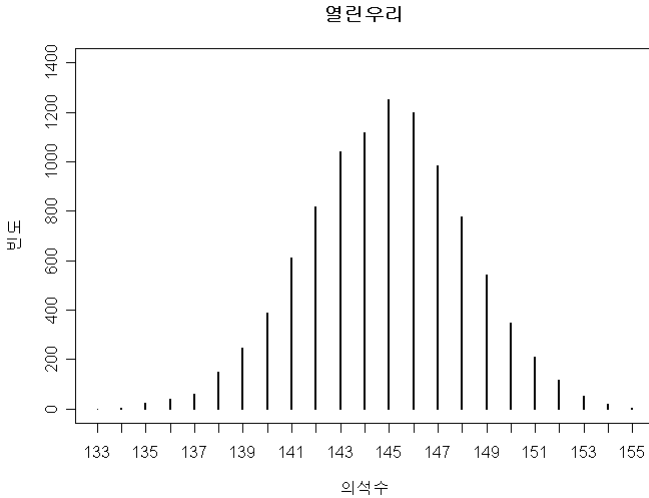
민련은 3~4석). 최빈값은 한나라당이 85석, 열린우리당이 145석이다.

상식적인 단순한 셈으로 KBS는 한나라당 지역구 의석 수를 82석으로, 열린우리당 의석 수를 148석으로 예상하였다. 구간으로는 한나라당에 대하여 최소 68석·최대 110석으로, 열린우리당에 대하여 최소 118석·최대 164석으로 예상하였다. 따라서 베이스 예상치(한나라당 85석, 열린우리당 145석)가 상식적 셈에 비해 한나라당에 대해 3석이 많고 열린우리당에 대해 3석이 적다. 그리고 구간의 길이를 보면 베이스 예상치 한나라당과 열린우리당에 대해 각각 13석과 14석이나 KBS의 상식적 셈은 33석과 47석이다. 상식적 구간은 보도하기엔 민망할 정도로 폭이 너무 넓다.

한편, 중앙선거관리위원회의 공식개표 결과 한나라당의 의석 수는 100석, 열린우리당의 의석 수는 129석이어서 결과적으로 베이스 예상치가 실제의석수에 비해 한나라당 15석 과소, 열린우리당 16석 과다하다. 또한 베이스 예측 구간도 참값을 포함하지 못한다. 2절에서 이렇게 된 원인을 살펴보고 3절에서 자료의 편향성을 보정한 예측 방법을 제안할 것이다.



〈그림 1〉 2004년 제 17대 국회의원 선거에서의 한나라당 의석 수 예측분포



〈그림 2〉 2004년 제 17대 국회의원 선거에서의 열린우리당 의석 수 예측분포

## II. 열세자 침묵 현상

2004년의 제 17대 국회의원 선거에서 지역 선거구는 총 243개였다. KBS/SBS는 공동으로 모든 지역선거구에서 조사에 근거한 예측을 하였는데 조사방법은 전화조사 또는 출구조사였다. 실사는 미디어리서치(MRI)와 TNS 소프레스가 하였다. <표 1>을 보라.

243개 지역선거구에 대한 예측결과, 19개 선거구에 대한 예측이 잘못된 것으로 나타났다. <표 2>를 보라. 19개의 오류 중 18개는 열린우리당 후보의 당

〈표 1〉 243개 지역선거구에 대한 최종 조사방법과 조사회사의 분류

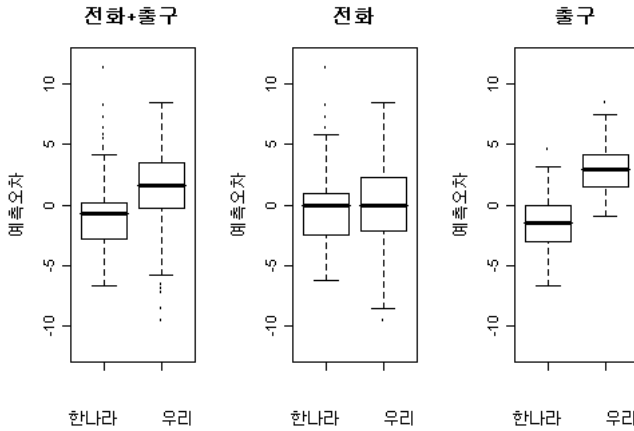
	MRI	TNS	합계
전화조사	57	72	129
출구조사	64	50	114
합계	121	122	243

〈표 2〉 243개 지역구의 당선예측 후보자 정당과 실제당선 후보자 정당

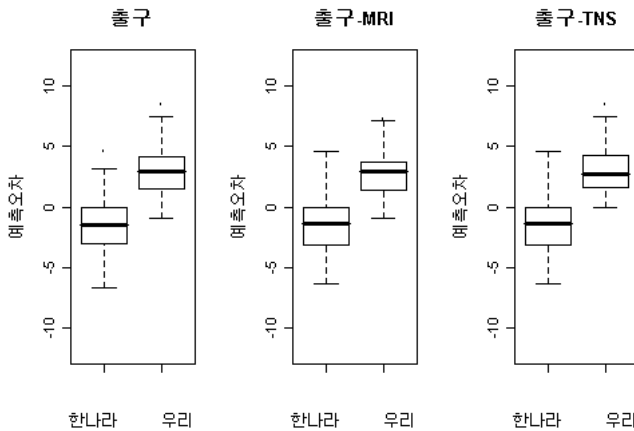
		결과					
		한나라	민주	열린우리	자민련	기타	합계
예측	한나라	82	0	0	0	0	82
	민주	0	5	0	0	0	5
	열린우리	18	0	129	1	0	148
	자민련	0	0	0	3	0	3
	기타	0	0	0	0	5	5
	합계	100	5	129	4	5	243

선이 예측되었으나 결과적으로는 한나라당 후보가 당선된 경우이고 1개는 열린우리당 후보의 당선이 예측되었으나 결과적으로는 자민련 후보가 당선된 경우이다. 2004년의 제 17대 국회의원 선거에서 야당인 한나라당이 열세인 것은 자명하였다. 예측조사에서 한나라당이 선두인 선거구 수는 82개에 불과하였다. 그러나 실제로 100개 선거구에서 당선됨으로써 예상외의 선전을 하였다. 반대로 여당인 열린우리당은 148개 선거구에서 우세하였으나 129개 선거구에서 승리하는 데 그쳤다.

예측오차를 두 주요정당인 한나라당과 열린우리당으로 나누어 살펴보기로 하자. 여기서 예측오차의 정의는 “예측득표율-실제득표율”이다. 〈그림 3〉은 조사방법별로 예측오차 분포를 상자그림으로 본 것인데, 전화조사와 출구조사를 합하는 경우 한나라당 후보들의 예측오차의 75% 정도가 음의 값을 갖는 데 반하여 열린우리당 후보들에 대한 예측오차는 75% 정도가 양의 값을 갖는 것을 볼 수 있다. 전반적으로 “열세정당 과소예측, 우세정당 과다예측” 현상을 확인할 수 있다. 그러나 전화조사에서는 이러한 열세자 과소 및 우세자 과다 현상이 나타나지 않는다. 반면 출구조사에서 이런 현상이 뚜렷하다. 즉 응답자들은 개방된 장소인 투표소 출구에서 강한 사회심리적 압박감을 느꼈음을 알 수 있다. 〈그림 4〉는 조사회사와 관련 없이 출구조사에서 편향이 편재하였



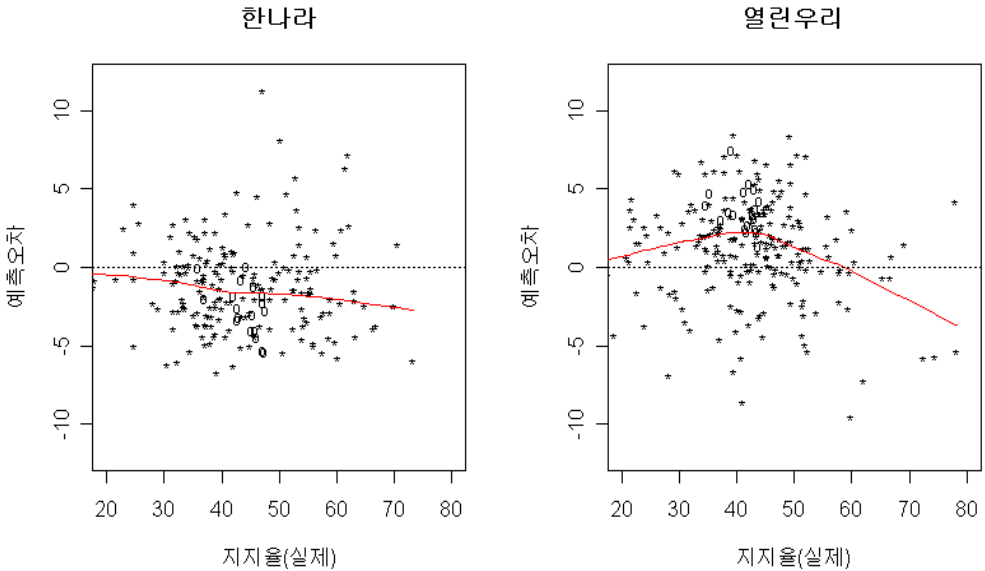
<그림 3> 조사방법별 예측오차 분포



<그림 4> 출구조사에서의 조사회사별 예측오차 분포

음을 보여준다.

노엘노이만(Noelle-Neuman 1984)은 침묵의 나선(spiral of silence) 효과가 열세자에 작용하여 조사를 회피하는 결과를 빚는 것으로 보았다. 이에 따라 열세자에 대하여는 예측오차가 음이 되고, 우세자에 대하여는 예측오차가 양이 된다.



〈그림 5〉 정당별 예측오차와 지지율의 관계  
(실선은 lowess 적합선, o점은 예측실패 사례)

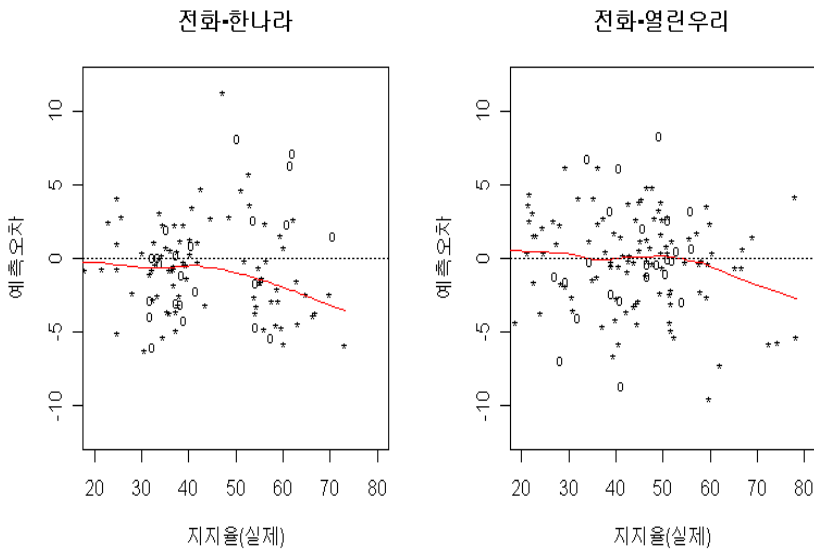
- 1) 예측오차는 지지율과 연관이 있지 않을까? 〈그림 5〉는 정당별로 예측오차와 지지율의 관계를 보여준다(실선은 lowess, 즉 국소가중치산점도평활 선이다). 한나라당에 대하여는 자당의 지지율이 20~60% 범위에서  $-1.35\%P$  (중간값) 정도의 열세자 과소예측이 있었고, 열린우리당에 대하여는 자당 지지율이 20~60% 범위에서  $1.80\%P$  (중간값) 정도의 우세자 과다예측이 있었다. 양 당 모두 자당 지지율이 60%를 넘어가면서 예측오차가 음의 방향으로 하강하는 패턴을 보인다. 선거구에서 한 당의 지지율이 60%를 넘어서는 압도적 상황에서는 응답자들이 분위기에 휩쓸리지 않는 것처럼 행동하는 것이다. 사회적 요망성 편향(social desirability bias)이 작동한 결과로 볼 수 있다. 열세자 과소예측 및 우세자 과다예측은 지역적인 현상이라기보다는 전국적인 현상임을 알 수 있다.
- 2) 〈그림 6〉과 〈그림 7〉은 조사방법 별로 나누어 예측오차와 지지율의 관계를 본 것이다. 전화조사에서 열세자 과소는  $-0.5\%P$  정도에 불과하다. 우세



자 과다는 거의 없다. 그러나 출구조사에서 열세자 과소는  $-2\%P$  정도, 우세자 과다는  $3\%P$  정도로 나타난다.

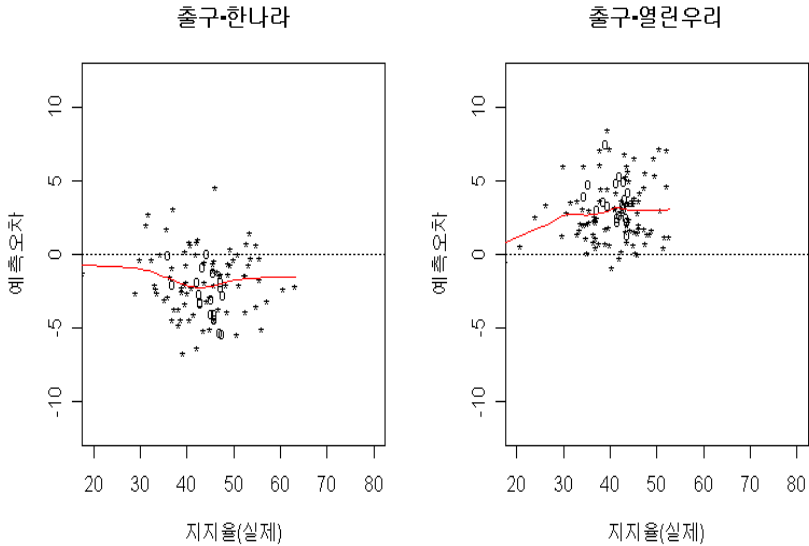
따라서 오차의 원인이 출구조사 자료가 한나라당에 대하여는 과소한 방향으로 편향되었고 열린우리당에 대하여는 과다한 방향으로 편향되었다는 데 있다고 결론지을 수 있다. 이러한 출구조사 자료의 심각한 편향은 당시 탄핵 정국에서 적지 않은 보수적 유권자들이 심리적으로 위축되어 출구면 접에서 무응답하였던 데 기인하는 것으로 보인다(이준웅 2004).

소위 침묵의 나선(spiral of silence)으로 인한 열세자 조사 기피 및 무응답 성향은 1996년의 제16대 국회의원 선거에서도 예측실패의 큰 원인이었다(이준웅 2001; 홍내리·허명희 2001).<sup>2)</sup>



〈그림 6〉 정당별 전화조사의 예측오차와 지지율의 관계

2) 나선의 침묵(소용돌이)에 대한 커뮤니케이션적 논의에 대하여는 양승찬(1998)과 이준웅(2001)을 보라. 이준웅은 조사 거절 또는 무응답 성향이 열세인 보수주의자에 작용하는 것으로 보았다.



〈그림 7〉 정당별 출구조사의 예측오차와 지지율의 관계

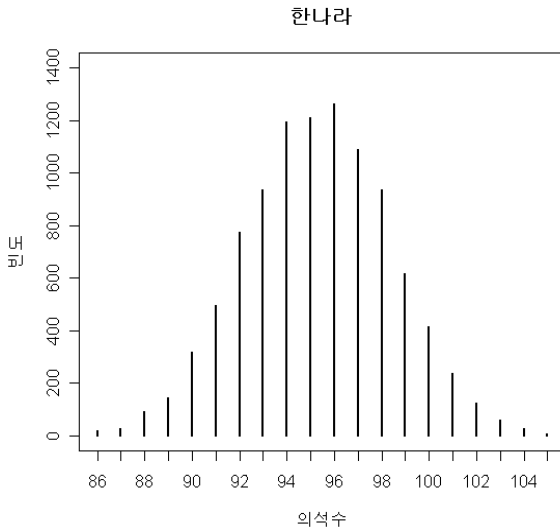
### III. 보정 베이스 예측

베이스 예측구간도 열세자 침묵으로 인한 자료 편향성으로부터는 자유롭지 않다. 이런 문제를 해결하기 위하여, 최종조사가 출구조사인 경우 우세당(열린우리당) 후보 지지율에서  $\alpha(\%P)$ 를 빼서 열세당(한나라당) 후보자들에게 그 만큼을 각자의 지지율에 비례하여 배분하는 방안을 고려하여 볼 필요가 있다. 예컨대  $\alpha=2\%P$ 라면, 보정 이전 한나라당 후보 40%, 열린우리당 후보 40%, 무소속 후보 20%가 보정 이후는 각각 41.3%, 38%, 20.7%가 된다.

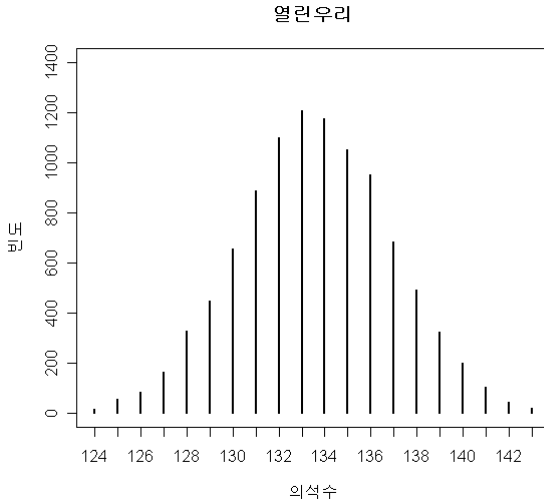
〈표 3〉은 제17대 국회의원 선거 예측조사에서 열세자 무응답 보정에 따른 예측구간 및 최빈값을 보여준다. 보정크기  $\alpha$ 를 2~3%P로 하였을 때 베이스 예측이 실제 결과에 근접하는 것을 볼 수 있다. 〈그림 8〉과 〈그림 9〉는 2%P 보정 이후 한나라당과 열린우리당의 의석 수 예측분포이다.

〈표 3〉 열세자 무응답 보정에 따른 예측 구간 및 최빈값

보정크기 alpha	한나라당			열린우리당		
	예측구간	예상치	개표결과	예측구간	예상치	개표결과
0%P	(79, 91)	85	100	(138,151)	145	129
1%P	(84, 96)	90		(133,146)	141	
2%P	(89,101)	96		(127,140)	133	
3%P	(95,107)	101		(120,134)	127	
4%P	(100,112)	106		(115,127)	120	
상식적 방법	(68,110)	82	100	(118,164)	148	129



〈그림 8〉 한나라당 의석 수 예측분포: 열세자 보정 크기 alpha=2%P



〈그림 9〉 열린우리당 의석 수 예측분포: 열세자 보정 크기  $\alpha=2\%P$

#### IV. 맺음말

베이즈 추론의 틀은 국회의원 선거에서 주요정당 당선자 수 예측에 매우 유효하다. 1절에서는, 베이즈 당선가능도를 활용하여, 2004년 제17대 국회의원 선거(243개 지역구)에서 정당 의석 수 예측구간을 산출해 보았다. 그 결과 열세당(한나라당)에 대하여는 예측이 과소한 것으로, 우세당(열린우리당)에 대하여는 예측이 과다한 것으로 드러났다. 그러나 2절에서 그 원인이 출구조사 자료가 열세자 침묵으로 인해 편향되었기 때문임을 밝힐 수 있었다. 따라서 전국적 선거판도에 명확한 우세당이 있는 경우 열세자 침묵을 어느 정도 감안한 보정이 있어야 할 것으로 생각한다. 그 일환으로 출구조사의 우세당 후보 지지율에서  $\alpha$ 를 빼서 열세당 후보자들에게 그 만큼을 각자의 지지율에 비례하여 배분하는 방안을 제안하였고 2004년 제17대 국회의원 선거의 출구조사에서 우세당과 열세당 후보자에 대하여 2~3%P 정도의 보정이 필요함을 밝혔다.

본 연구에서 제안하는 보정 베이스 예측에 있어 관건은 우세자 (또는 정당)에 대한 과다 편향인  $\alpha$  값의 설정에 있다. 이 값이 편향의 크기를 의미하므로 조사 자료에 의한 추정은 어렵다. 따라서 최근의 선거에서 사후적으로 나타난 편향 크기와 추세를 고려하는 방안이 최선이 아닐까 싶다. 2004년의 국회의원선거에서는 그 값이 3%P 정도였다. 2006년도 지방선거의 경우 서울시장 출구조사는 2.5%P였고 경기도지사 출구조사는 1.7%P였다(김영원 · 김지현 2007).

## 참고문헌

- 김영원·김지현. 2007. “2006년 지방선거 출구조사 현황 및 예측오차.” 《조사연구》 8(1): 55-79.
- 서혜선·허명희. 1997. “당선자 수에 관한 구간 예측: 4/11 총선을 중심으로.” 《한국분류학회지》 1: 57-68.
- 양승찬. 1998. “제3차 효과 가설과 침묵의 나선 이론의 연계성.” 《한국언론학보》 43: 109-141.
- 이준웅. 2001. “여론환경에 대한 인식과 정치적 의견 표명: 제16대 총선 예측조사 실패에 대한 정치 커뮤니케이션적 논의.” 《한국방송학회》 15(1): 199-236.
- 이준웅. 2004. “제17대 총선 예측조사의 문제: 조사거절자와 응답불성실자 편향을 중심으로.” 《언론정보연구》 41(1): 110-136.
- 허명희. 2008. “선거여론조사결과의 베이즈적 해석: 저널리즘과 아카데미즘의 조화.” 미발표논문.
- 홍내리·허명희. 1991. “제16대 국회의원 선거의 예측조사에 대한 사후적 검증.” 《조사연구》 2(1): 1-35.
- Jackson, S. 2004. “Bayesian analysis for political research.” *Annual Review of Political Science* 7: 483-505.
- Noelle-Neuman, E. (1974). “The spiral of silence: a theory of public opinion.” *Journal of Communication*, 24(2), 43-51.