■ 論 文 ■

패널교통사고자료 기반 기대교통사고건수 추정기법 평가

Assessing Estimation Methods of the Expected Crashes using Panel Traffic Crash Data

신 강 원
(경성대학교 도시공학과 전임강사)

목 차

Ⅰ. 서론
Ⅱ. 패널자료를 이용한 기대교통사고건수 추정기법
  1. 시간불변 기대교통사고건수 추정
  2. 시간변동 기대교통사고건수 추정
Ⅲ. 시간불변 기대교통사고건수 추정기법 비교
  1. 모의실험
  2. 비교결과
Ⅳ. 시간불변 기대교통사고건수 추정기법 비교
  1. 모의실험
  2. 비교결과
Ⅴ. 결론
참고문헌

Key Words : 패널교통사고자료, 평균 관측교통사고건수, 평행비교 추정치, 경험적 베이즈 추정치, 추정오차
Panel crash data, mean observed crashes, comparative parallel estimate, empirical Bayes estimate, estimation error

요 약

유사한 특성을 갖는 지점 (또는 구간)들에서 연속되고 동일한 시간 간격 동안 관측된 패널 (panel) 교통사고 자료를 시간의 흐름에 따라 비교분석하여 분석지점의 기대교통사고건수를 추정하는 과정은 교통안전 개선사업의 효과평가와 교통안전 개선사업 수행의 우선순위 결정과 같은 교통안전연구의 핵심이다. 패널 교통사고 자료를 이용한 기대교통사고건수 추정기법은 관측교통사고건수 기반 기법과 경험적 베이지안 기법으로 대별할 수 있으며, 본 연구에서는 시간의 흐름에 따른 기대교통사고건수의 변화 여부와 다양한 패널 교통사고 자료 구조에 따라 전술한 두 가지 기법의 추정오차를 모의실험을 통해 비교・분석하였다. 분석결과 시간의 흐름에 따른 기대교통사고건수의 변화 여부와 패널교통사고자료 구조의 특성과 관계없이 관측교통사고건수 기반 추정치인 평균 관측교통사고건수와 평행비교 추정치의 추정오차는 경험적 베이즈 추정치의 추정오차보다 항상 크게 나타나 향후 패널교통 사고자료를 이용한 교통안전 연구 수행 시 경험적 베이지안 추정기법의 적용이 필요하다고 판단된다. 한편 시간의 흐름에 따라 기대교통사고건수가 변화하지 않을 경우 분석기간이 늘어날수록 두 가지 기법의 추정오차는 모두 현저하게 감소하는 것으로 분석되어, 현재 국내의 교통사고 잦은 곳 선정 연구에서 기준치로 사용되고 있는 분석기간인 “1년”을 연장하여 보다 효율적으로 시간 불변 기대교통사고건수를 추정할 필요가 있다고 판단된다.

To evaluate highway safety countermeasures or identify high risk sites, the expected crashes for a site (or segment) have been estimated using the panel crash data. Past studies show that two different methods can be employed to estimate the expected crashes: observed crash based method and empirical Bayes (EB) method. This study conducts a simulation study to analyze how the estimation errors of the two estimates are affected by the different structures of the panel crash data and the presence of the change in safety over time. The results disclose that the estimation errors of the observed crash based estimates (i.e. the mean observed crash and comparative parallel estimate) are always greater than those of the EB estimates regardless of the structure of the panel crash data and the presence of the change in safety over time. Thus, it is highly recommended that the EB method be used in the study of traffic safety to obtain more reliable estimates for the expected crashes. In addition, this study corroborates that the estimation errors of the two estimates decrease as the analysis periods increase if safety does not change over time. Hence, it is also recommended that the 1-year analysis period used for identifying high risk sites in Korea be extended to produce more efficient estimates of the time-constant expected crashes.

본 논문은 2010학년도 경성대학교 학술지원연구비에 의하여 연구되었음.

I. 서론

유사한 특성을 갖는 지점 (또는 구간)들에서 연속되고 동일한 시간 간격 동안 관측된 패널 (panel) 교통사고 자료를 시간의 흐름에 따라 비교분석하여 분석지점의 기대교통사고건수를 추정하는 과정은 교통안전 개선사업의 효과평가와 교통안전 개선사업 수행의 우선순위 결정과 같은 교통안전연구의 핵심이다. 구체적으로 교통안전 개선 사업의 효과 평가 시, 사업 시행 전에 관측된 패널 교통사고 자료는 사업 미시행을 가정했을 경우 기대되는 교통사고 건수 (counterfactual expected crashes)를 추정하기 위해 사용된다. 또한 교통사고 잦은 곳을 선정할 경우에도 교통사고 이력 자료를 이용하여 각 분석지점의 기대교통사고건수 (expected crashes)를 추정하는 과정을 거치게 된다. 이처럼 여러 지점에서 매년 관측된 교통사고 자료는 다양한 교통안전연구 수행을 위해 널리 활용되고 있다.

패널교통사고 자료를 이용한 기대교통사고건수 추정을 위해 다양한 기법들이 사용되고 있는데, 이들은 "관측교통사고건수 기반 방법"과 관측교통사고건수와 분석지점과 유사한 특성을 가진 지점들 (i.e. 참조그룹)의 모수 추정치를 이용하는 "경험적 베이지안 (empirical Bayes, EB) 기법"으로 대별할 수 있다. 국외의 경우 1980년대부터 EB기법의 이론과 활용 방법에 대한 연구가 꾸준히 추진되어 현재 EB기법은 표준분석기법으로 권장되고 있으나, 국내 교통안전 연구는 관측교통사고건수 기반 방법을 표준방법으로 사용하고 있다. 따라서 두 가지 추정기법의 추정오차를 비교분석하여 국내 교통안전 연구기법의 개선 방안을 검토할 필요가 있다.

이러한 연구 배경 하에 선행연구 (신강원, 2010)는 횡단면 (cross-sectional) 교통사고 자료 분석 시 두 가지 추정기법의 추정오차를 비교분석하여 EB 기법의 도입이 필요하다는 결론을 제시한 바 있으나, 패널교통사고 자료를 이용한 기대교통사고건수 추정기법들의 추정오차의 특성, 추정오차의 차이에 대한 체계적인 비교분석은 전무한 실정이다. 따라서 본 연구에서는 패널교통사고 자료를 이용하여 교통사고 잦은 곳 선정과 다양한 교통안전 개선사업의 효과분석 수행 시 사용될 수 있는 두 가지 추정기법의 추정오차를 비교분석하여 국내 교통안전 연구기법의 개선방안을 제시하였다.

II. 패널자료를 이용한 기대교통사고 추정 기법

패널교통사고 자료를 이용한 기대교통사고건수 추정기법은 전술한 바와 같이 관측교통사고건수 기반기법과 경험적 베이지안 기법으로 분류할 수 있는데, 이 기법들은 "시간의 흐름에 따른 기대교통사고건수의 변화 가정"에 따라 각기 다른 추정량 (estimator)을 갖는다. 따라서 본 장에서는 한 지점의 기대교통사고 건수가 시간에 흐름에 따라 동일 (시간불변 기대교통사고 건수)하다는 가정과 시간의 흐름에 따라 기대교통사고건수는 변화 (시간변동 기대교통사고 건수)한다는 가정하에서 사용될 수 있는 관측교통사고건수 기반 추정량과 경험적 베이즈 추정량의 특성과 관련이론을 고찰하였다.

1. 시간불변 기대교통사고 건수 추정

시간의 흐름에 따라 기대교통사고건수가 변화하지 않을 때 분석지점의 기대교통사고건수 추정은 비교적 쉽게 수행될 수 있다. 먼저 분석기간 동안 관측된 교통사고자료만을 이용하는 기법에서는 분석지점 i의 t년도의 관측교통사고건수 y_{it}를 식(1)과 같이 시간불변 기대교통사고건수 λ_i를 갖는 포아송 확률변수로 가정한다.

$$y_{it} \sim P(\lambda_{it})=P(\lambda_i) \quad (t=1,2,\cdots,T). \tag{1}$$

이 경우 분석지점 i의 기대교통사고건수는 분석기간 T년 동안 관측된 교통사고건수의 로그우도를 최대화하여 추정할 수 있으며, 추정치는 식(2)와 같이 평균 관측교통사고건수와 같다.

$$\hat{\lambda}_i = T^{-1}\sum_{t=1}^{T} y_{it}. \tag{2}$$

이처럼 분석지점 i의 기대교통사고건수는 분석기간 동안 관측된 교통사고자료의 평균값으로 추정될 수 있으나 이 추정치는 교통사고 발생의 희귀성과 랜덤성으로 인한 "평균으로의 회귀 편의 (regression-to-the-mean bias)"를 갖기 때문에 경험적 베이지안 기법이 널리 사용되고 있다 (Abbess et al., 1981; Hauer, 1997; Hauer et al., 2002).

경험적 베이지안 이론에 대한 선행연구 결과를 요약

하면, 분석지점 i의 시간불변 (time-constant) 기대교통사고건수 i는 식(3)과 같이 포아송-감마 혼합모형의 감마 사후평균 (gamma posterior mean)과 같음을 알 수 있다.

$$\begin{aligned}\hat{\lambda}_i &= \hat{E}\left(\lambda_i \mid y_{i1}, y_{i2}, \cdots, y_{iT}\right)\\ &= w\hat{E}\left(\lambda_i\right) + (1-w)\frac{\sum_{t=1}^{T} y_{it}}{T}.\end{aligned} \tag{3}$$

식(3)에서 감마 사전분포의 평균 $E(\lambda_i)$의 추정치는 포아송-감마혼합모형의 주변분포인 음이항분포 (negative binomial distribution)의 평균과 같고 0과 1사이의 값을 갖는 베이지안 가중치 w는 식(4)와 같이 나타낼 수 있다.

$$w = \frac{1}{1 + T\hat{E}(\lambda)\hat{\phi}}, \tag{4}$$

여기서 ϕ는 음이항분포의 과분산계수 (over-dispersion parameter)로 정의되는데 이 값이 클수록 감마 사전분포의 분산은 커지는 성질을 갖는다. 즉 식(3)에 제시된 EB 추정치는 분석기간 T년 동안 관측된 교통사고건수가 y_{i1}, y_{i2},⋯, y_{iT} 일 때 분석지점 i와 유사특성을 가지는 지점들 (참조그룹)의 기대교통사고건수의 평균을 나타내며, 이 값은 항상 감마 사전분포의 평균 $E(\lambda_i)$과 식(2)의 최우추정치 (maximum likelyhood estimate)인 평균 관측교통사고건수 사이의 값을 갖는다.

2. 시간변동 기대교통사고 건수 추정

한 지점의 교통안전도는 시간의 흐름에 따른 다양한 외생변수 (e.g., 교통량, 도로 기하구조, 차량성능 등)의 변화로 인해 영향을 받기 마련이다. 따라서 한 지점의 기대교통사고건수가 분석기간 동안 동일하다는 시간불변 기대교통사고건수 가정은 적절하지 못한 경우가 많고, 이러한 변화를 적절하게 반영하기 못할 경우에는 교통사고 잦은 곳이 잘못 선정되거나 교통안전 개선사업의 효과가 오추정 (mis-estimation) 될 수 있다. Hauer (1997)는 한 지점의 시간변동 기대교통사고건수를 추정하기 위해 다양한 방법론을 제시하였다. 제시된 방법론은 주로 교통안전대책 평가를 위한 목적으로 개발되었으며, 이 기법들을 간단히 요약하면 분석지점에서 T-1년 동안 관측된 교통사고 자료와 유사지점에서 T년 동안 관측된 교통사고 자료를 이용하여 T년도 분석지점의 기대교통사고건수를 추정하는 흐름을 갖는다.

구체적으로 관측교통사고건수 기반 방법에서는 지점 i의 T년도의 기대교통사고건수 λ_{iT}의 추정치를 식(5)와 같이 제시하고 있다 (Hauer, 1997).

$$\hat{\lambda}_{iT} = \frac{N}{M}\left(1 + \frac{1}{M}\right)^{-1} \cdot \sum_{t=1}^{T-1} y_{it}, \tag{5}$$

여기서 y_{it}는 분석지점 i의 t년도 관측교통사고건수, M은 유사지점에서 T-1년 동안 관측된 교통사고건수의 합, N은 유사지점에서 T년도에 관측된 교통사고건수의 합을 나타낸다. 식(5)를 이용한 기대교통사고건수 추정방법을 도로교통공단(2009)에서는 평행 비교 방법 (comparative parallel study)으로 명명하고 있으며, 이는 시간의 흐름에 따른 유사지점의 교통사고건수 변화가 분석지점의 기대교통사고건수의 변화를 반영한다는 가정에 기반한다.

한편 Hauer (1997)는 시간에 흐름에 따라 기대교통사고건수가 변화할 경우 사용할 수 있는 경험적 베이지안 방법 (시간변동 EB 기법)도 제시하였으며, 그 결과를 요약하면 지점 i의 T년도의 기대교통사고건수 λ_{iT}의 추정치는 식(6)과 같다.

$$\begin{aligned}\hat{\lambda}_{iT} &= \hat{E}(\lambda_{iT} \mid y_{i1}, \cdots, y_{iT-1})\\ &= \hat{C}_T\left[w\hat{E}\left(\lambda_{i1}\right) + (1-w)\frac{\sum_{t=1}^{T-1} y_{it}}{\sum_{t=1}^{T-1} C_t}\right],\end{aligned} \tag{6}$$

여기서 C_t는 시간의 흐름에 따라 변화하는 해당 분석지점의 기대교통사고건수 변화를 나타내는 것으로 이는 t년도의 기대교통사고건수의 사전평균 $E(\lambda_{it})$를 분석 초기 년도의 기대교통사고건수의 사전평균 $E(\lambda_{i1})$로 나눈 값과 같다 (i.e. $C_t = E(\lambda_{it})/E(\lambda_{i1})$). 식(6)의 베이지안 가중치 w는 식(7)과 같이 표현된다.

$$w = \frac{1}{1 + \sum_{t=1}^{T-1} C_t\hat{E}\left(\lambda_{i1}\right)\phi}. \tag{7}$$

따라서 시간변동 경험적 베이지안 기법 적용의 핵심은 분석년도 별로 각기 다른 사전분포의 추정이라 할 수 있다.

Shin and Washington (2010)은 전술한 시간변동 경험적 베이지안 분석에서 분석년도 별 사전분포를 이질적 음다항 모형 (heterogeneous negative multinomial model, HNM 모형)을 이용하여 추정할 수 있음을 증명하였으며, 교통사고자료의 공간적 이질성이 유의하지 않을 경우 이 모형은 랜덤효과 포아송 모형 (random effect Poisson model)과 같다는 사실[1)]을 밝혔다. 이러한 결과는 시간변동 포아송-감마 혼합모형에 기반하는 주변분포가 닫힌 형태인 음다항 분포와 같다는 사실을 이용함으로써 시간변동 기대교통사고건수를 경험적 베이즈 방법을 통해 추정할 수 있음을 나타낸다.

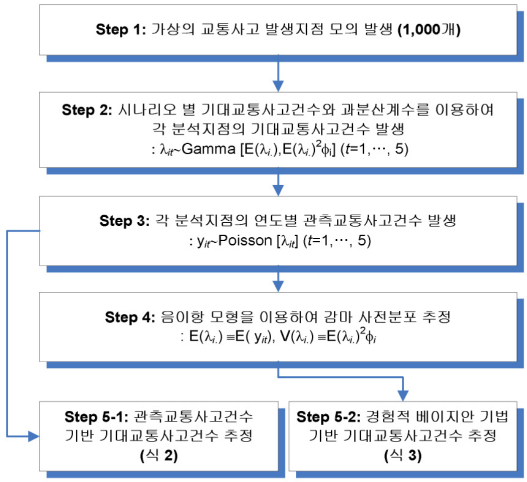
III. 시간불변 기대교통사고건수 추정기법 비교

본 장에서는 시간의 흐름에 따라 기대교통사고건수가 동일할 때 전술한 추정기법들 (i.e. 관측교통사고건수 기반방법과 베이지안 추정법 식(2)와 식(3))의 추정오차와 그 특성을 모의실험을 통해 비교분석하였다.

1. 모의실험

전술한 두 가지 추정기법의 추정오차를 비교분석하기 위해 신강원 (2010)의 모의실험 절차를 패널교통사고자료 분석에 적합하게 수정하였다. 본 연구에서는 총 100개의 모의실험 시나리오를 설정하였는데, 구체적으로 연간 기대교통사고건수의 평균과 분석년도를 각각 10개 수준 (1건, 2건, 3건, 4건, 5건, 6건, 7건, 8건, 9건, 10건)과 5개 수준 (1년, 2년, 3년, 4년, 5년)으로, 사전분포의 분산을 나타내는 과분산계수를 2개 수준 (0.1, 1.0)으로 설정하여 분석을 수행하였다.

이 모의실험 시나리오에서는 최대 5년 동안 기대교통사고건수가 변화하지 않는다는 가정 하에 분석년도의 최대값을 5년으로 설정하였으며, 사전분포의 불확실성에 따른 추정오차를 비교하기 위해 과분산계수의 값을 각각 0.1 (low heterogeneity), 1.0 (high heterogeneity)으로 설정하였다 (e.g. 사전분포의 평균이 "10건/년"이고 과분산계수가 "1.0"일 때 사전분포의 분산은 100 ($10^2 \times 1.0$)으로 사전분포의 불확실성이 매우 높음을 의미, $V(\lambda_i) = \mu^2\phi$).



〈그림 1〉 시간불변 기대교통사고건수 추정 흐름

전술한 모의실험 시나리오에 따른 두 가지 추정기법의 추정오차 비교분석절차는 〈그림 1〉과 같이 총 5단계를 거쳤으며, 분포발생에 따른 오차를 줄이기 위해 각 시나리오 별로 모의실험을 30회씩 반복하였다 (각 시나리오 별 총 모의 발생 지점 수=30,000).

마지막으로 두 가지 분석기법을 통해 추정된 기대교통사고건수 (Step 5의 결과물)와 각 분석지점의 기대교통사고건수 (Step 2의 결과물)의 차이를 식(8)에 제시한 평균제곱오차 (mean squared error, MSE)식을 이용하여 계산하였다.

$$MSE_{EB} = \frac{\sum_{i=1}^{30000} \left\{ \hat{E}(\lambda_i \mid y_{i1}, y_{i2}, \cdots, y_{iT}) - \lambda_i \right\}^2}{30000},$$
$$MSE_y = \frac{\sum_{i=1}^{30000} \left\{ \dfrac{\sum_{t=1}^{T} y_{it}}{T} - \lambda_i \right\}^2}{30000}. \tag{8}$$

한편 각 분석 시나리오 별로 계산된 MSE값은 추정치의 값이 커짐에 따라 증가 (추정치의 값이 커질수록 분산이 커지기 때문임)하기 때문에 두 가지 기법의 추정오차의 크기는 관측교통사고건수를 이용한 추정치의 MSE (MSE_y)를 경험적 베이즈 추정치의 MSE (MSE_{EB})로 나누어 계산된 상대평균제곱오차(Relative MSE, RMSE)를 이용하여 비교하였다.

1) 이질적 음다항 모형, 랜덤효과 포아송 모형에 대한 자세한 설명은 Johnson and Kotz (1969), Guo (1996), Hilbe (2007), Shin and Washington (2010) 참조.

2. 비교결과

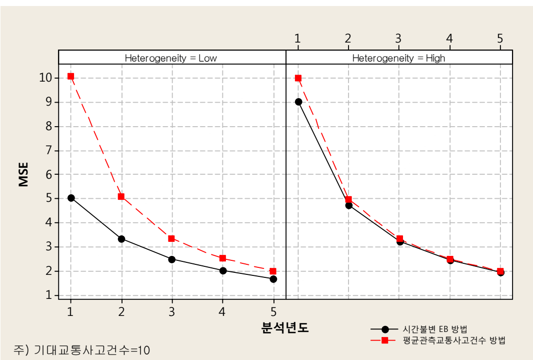
시간의 흐름에 따라 기대교통사고건수가 변화하지 않을 때 모의실험 자료를 이용하여 전술한 두 가지 추정기법의 추정오차를 비교분석한 결과를 요약하면 다음과 같다.

- 사전분포의 형태와 분석기간에 관계없이 EB 추정치의 추정오차는 기존의 관측교통사고건수 기반 추정치의 추정오차보다 항상 작다 (〈표 1〉에 제시된 시나리오 별 RMSE의 값이 모두 1보다 큼).

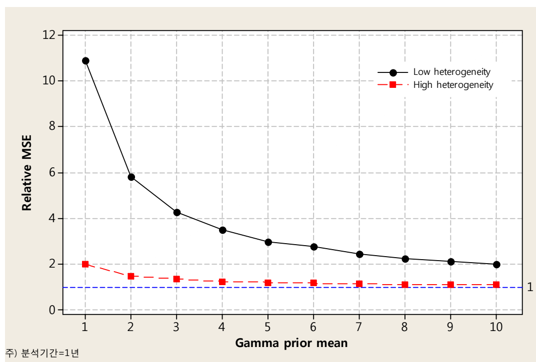
- 분석기법에 관계없이 동일한 사전분포 하에서 각 기법의 추정오차는 분석기간이 길어질수록 감소한다. 〈그림 2〉는 기대교통사고건수가 10건일 때 분석기간에 따른 두 가지 기법의 추정오차를 나타내는데 분석기간이 늘어날수록 두 가지 추정기법의 추정오차는 감소함을 알 수 있으며, 이러한 패턴은 모든 분석 시나리오에서 관측되었다. 따라서 현재 국내에서 교통사고 잦은 곳 선정 시 사용되고 있는 분석기간 (1년)의 연장이 필요하다고 판단된다.

- 감마사전분포의 불확실성이 커질수록 (i.e. 과분산계수의 값이 커지거나 감마사전평균의 값이 커질 때) RMSE의 값은 감소한다. 먼저 과분산계수가 커질 때 RMSE의 감소현상은 〈표 1〉, 〈그림 2〉, 〈그림 3〉에서 확인할 수 있다. 예를 들어 기대교통사고건수가 "10건"이고 분석기간이 "1년"일 때 과분산계수 값이 작을 경우 (low heterogeneity)에는 평균 관측교통사고건수의 추정오차는 시간불변 EB 추정치의 추정오차보다 약 2배 크지만 과분산계수 값이 클 경우(high heterogeneity)에는 상대평균제곱오차가 1.105로 감소함을 알 수 있다.

- 한편 감마사전평균의 값이 커질 때 RMSE의 감소



〈그림 2〉 분석기간에 따른 추정오차 (예)



〈그림 3〉 감마사전평균에 따른 상대평균제곱오차 (예)

〈표 1〉 분석 시나리오 별 상대 평균제곱오차 (MSE_Y/MSE_{EB})

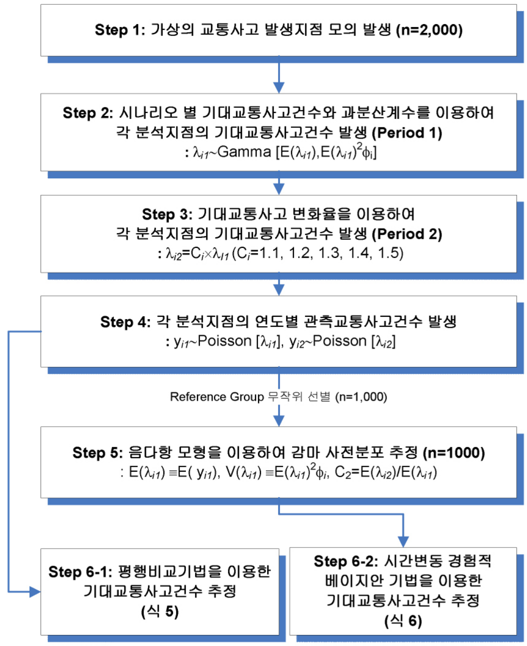
구분	사전평균 / 분석기간	1	2	3	4	5	6	7	8	9	10
Low heterogeneity (ϕ=0.1)	1	10.882	5.835	4.272	3.496	2.963	2.762	2.452	2.240	2.130	2.002
	2	5.963	3.465	2.658	2.267	1.974	1.881	1.717	1.624	1.575	1.514
	3	4.313	2.658	2.120	1.843	1.641	1.592	1.487	1.420	1.391	1.341
	4	3.462	2.236	1.845	1.631	1.480	1.440	1.368	1.320	1.293	1.254
	5	2.980	2.000	1.669	1.503	1.389	1.349	1.298	1.254	1.231	1.205
High heterogeneity (ϕ=1.0)	1	1.993	1.483	1.342	1.240	1.201	1.164	1.127	1.116	1.109	1.105
	2	1.476	1.244	1.168	1.110	1.103	1.081	1.068	1.060	1.059	1.052
	3	1.314	1.161	1.104	1.072	1.070	1.057	1.040	1.040	1.039	1.034
	4	1.241	1.111	1.078	1.056	1.052	1.044	1.034	1.030	1.027	1.021
	5	1.190	1.092	1.059	1.044	1.040	1.037	1.026	1.023	1.021	1.18

주) 각 분석 시나리오 별 상대평균제곱오차는 관측교통사고건수 기반 추정치 (식 2)의 평균제곱오차를 EB 추정치 (식 3)의 평균제곱오차로 나눈 것으로 그 값이 1을 넘으면 관측교통사고건수 기반 추정치의 추정오차가 EB 추정치의 추정오차보다 큼

현상은 〈표 1〉과 〈그림 3〉을 통해 확인할 수 있다. 예를 들어 과분산계수 값이 "0.1"이고 분석기간이 "1년"일 때 감마사전평균값이 작을 경우 (e.g. E〔λ_i〕=1) RMSE는 "10.882"지만 감마사전평균값이 클 경우 (e.g. E〔λ_i〕=10) RMSE는 "2.002"로 크게 감소함을 알 수 있다.

이처럼 모의실험 결과를 살펴본 결과 기대교통사고건수가 시간불변일 경우 현재 국내에서 사용되고 있는 평균 관측교통사고건수의 추정오차는 EB 추정치의 추정오차에 비해 항상 큼을 알 수 있는바 패널교통사고자료를 이용한 교통안전 연구 수행 시에도 EB 추정기법의 적용이 요구된다 할 수 있다.

그러나 감마사전분포의 불확실성이 증가할 때 두 기법의 RMSE는 감소할 수 있어 무조건적인 EB 기법의 적용보다는 패널교통사고자료를 이용한 음이항 모형 추정치에 대한 이론적 검증과 분석을 수행한 후 기대교통사고건수를 추정해야 할 것이다.



〈그림 4〉 시간변동 기대교통사고건수 추정 흐름

IV. 시간변동 기대교통사고건수 추정기법 비교

본 장에서는 한 지점의 기대교통사고건수가 시간의 흐름에 따라 변화한다는 가정 하에 "관측교통사고건수 기반 추정치" (i.e. 평행 비교 방법을 적용한 추정치 식(5))와 "경험적 베이즈 추정치" (i.e. 시간변동 경험적 베이즈 추정치 식(6))의 추정오차를 모의실험을 통해 비교분석하였다.

1. 모의실험

시간변동 기대교통사고건수 추정기법의 추정오차를 비교분석하기 위해 전장의 모의실험 시나리오와 동일하게 연간 기대교통사고건수의 평균을 10개 수준 (1건, 2건, 3건, 4건, 5건, 6건, 7건, 8건, 9건, 10건), 과분산계수를 2개 수준 (0.1, 1.0)으로 설정하여 분석을 수행하였다. 또한 시간의 흐름에 따른 기대교통사고건수의 변화를 반영하기 위해 5개 수준의 증가율[2] (1.1, 1.2, 1.3, 1.4, 1.5)을 설정하였다. 본 연구에서는 분석년도를 2년으로 설정하였는데 이는 현재 도로교통공단 (2009)에서 교통안전 개선사업의 효과분석 시 사용되고 있는 분석기간을 반영하기 위함이다. 따라서 본 연구에서는 전술한 20가지의 사전분포의 구조 (102)와 두 번째 분석년도 (Period 2)의 기대교통사고건수는 첫 번째 분석년도 (Period 1)의 기대교통사고건수에 비해 전술한 5개 수준의 증가율만큼 변화한다 (총 모의실험 시나리오 수=100개).

각 모의실험 시나리오에 따른 두 가지 추정기법의 추정오차 비교분석절차는 〈그림 4〉과 같이 총 6단계를 거치며, 각 시나리오 별 분포발생에 따른 오차를 줄이기 위해 모의실험을 각각 30회씩 반복하였다. 또한 기존의 교통안전 개선사업 효과 평가 흐름을 반영 (개선사업 평가 방법에 대한 자세한 설명은 Hauer (1997), 도로교통공단 (2009) 참조)하기 위해 각 시나리오 별로 발생된 총 2,000개의 교통사고 발생 지점 중 1,000개 지점을 참조그룹으로 선정하여, 기대교통사고건수 추정 시 필요한 값인 M, N, $E(\lambda_{i1})$, E(λ_{i2}), ϕ_i를 추정하였다 (각 시나리오 별 총 모의 발생 지점 수=60,000).

식(9)는 두 가지 분석기법의 추정오차 (MSE)의 계산식[3]을 나타내며, 두 가지 분석기법의 추정오차의 비

2) $C_2 = E(\lambda_2)/E(\lambda_1)$

3) 시나리오 별 총 모의발생 교통사고지점 60,000지점 중 30,000지점은 참조그룹으로 나머지 30,000지점은 분석지점으로 분류하여 분석을 수행하였음.

교는 전장과 같이 상대 평균제곱오차 (MSE_{PC}/MSE_{MEB})를 이용하였다.

$$MSE_{MEB}=\frac{\sum_{i=1}^{30000}\{\hat{E}(\lambda_{i2}\mid y_{i1},y_{i2})-\lambda_{i2}\}^2}{30000},$$

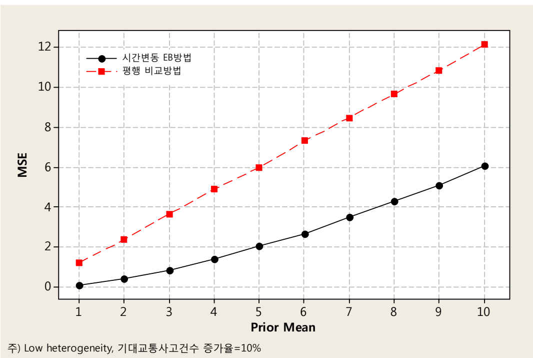
$$MSE_{PC}=\frac{\sum_{i=1}^{30000}\left[\left\{\frac{N}{M}\left(1+\frac{1}{M}\right)^{-1}\cdot y_{i1}\right\}-\lambda_{i2}\right]^2}{30000}. \qquad (9)$$

2. 비교결과

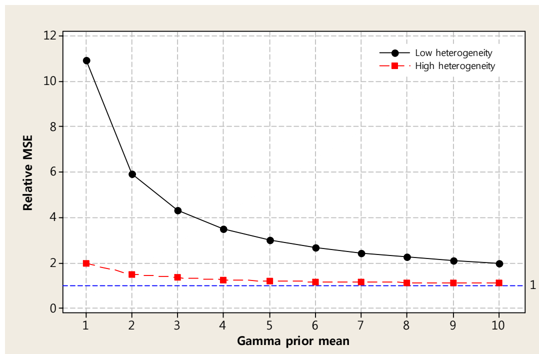
모의실험 결과를 이용하여 각 기법의 기대교통사고건수 추정오차의 특성을 살펴보면, 먼저 추정 방법론에 관계 없이 추정오차는 감마사전평균의 값이 커질수록 증가(감마사전평균의 값이 커질수록 관측교통사고건수의 분산은 커지기 때문)하게 되지만 시간변동 경험적 베이즈 추정치의 추정오차는 평행비교 추정치의 추정오차보다 항상 작음을 알 수 있다 (그림 5 참조).

전술한 결론은 모든 분석 RMSE의 값이 1보다 크다는 사실 (표 2 참조)을 통해 확증할 수 있으며, 시간의 흐름에 따라 기대교통사고건수가 변화할 때 두 가지 기법의 추정오차 비교분석 결과는 다음과 같이 요약할 수 있다.

- 사전분포의 형태와 기대교통사고 변화율에 관계 없이 기존의 평행 비교 추정치의 추정오차는 시간변동 경험적 베이즈 추정치의 추정오차보다 항상 크다. 이는 시간의 흐름에 따라 기대교통사고건수가 변화할 경우에도 경험적 베이지안 기법은 추정오차 감소측면에서 기존 관측교통사고건수 기반방법인 평행 비교방법보다 우월하다는 것을 나타낸다.
- 전장의 결과와 마찬가지로 감마사전분포의 불확실성이 커질수록 RMSE의 값은 감소한다 (그림 6[4])



〈그림 5〉 감마사전평균 증가에 따른 추정오차 변화 (예)



〈그림 6〉 분석 시나리오 별 상대 평균제곱오차 변화

〈표 2〉 분석 시나리오 별 상대 평균제곱오차 (MSE_{PC}/MSE_{MEB})

구분	사전평균 / 증가율	1	2	3	4	5	6	7	8	9	10
Low heterogeneity (ϕ=0.1)	10%	10.967	5.807	4.328	3.538	2.933	2.727	2.411	2.259	2.122	1.990
	20%	10.970	5.926	4.347	3.452	3.015	2.719	2.441	2.259	2.131	1.998
	30%	10.877	6.053	4.290	3.449	2.979	2.670	2.418	2.267	2.080	2.003
	40%	10.934	6.037	4.361	3.532	3.052	2.699	2.413	2.258	2.117	1.984
	50%	10.968	5.822	4.277	3.454	2.999	2.653	2.416	2.286	2.134	1.979
High heterogeneity (ϕ=1.0)	10%	1.938	1.519	1.348	1.257	1.180	1.170	1.140	1.137	1.124	1.093
	20%	2.030	1.514	1.339	1.251	1.200	1.165	1.158	1.132	1.105	1.100
	30%	1.975	1.449	1.369	1.236	1.195	1.187	1.146	1.134	1.123	1.088
	40%	1.943	1.505	1.344	1.247	1.196	1.165	1.155	1.129	1.123	1.100
	50%	2.017	1.518	1.358	1.259	1.187	1.167	1.159	1.128	1.111	1.100

4) 각 분석 시나리오에서 기대교통사고건수 변화율에 따른 RMSE의 변화는 미미한 것으로 나타나, 사전평균과 과분산계수에 따른 분석 시나리오 별 RMSE의 평균값을 〈그림 6〉의 RMSE 값으로 표현하였음.

참조)참조). 예를 들어 감마사전분포의 불확실성이 매우 낮을 때 (감마사전평균이 1이고 과분산계수가 0.1), 평행 비교 추정치의 추정오차는 시간변동 경험적 베이즈 추정치의 추정오차에 비해 약 11배 크게 나타날 수 있다. 그러나 감마사전분포의 불확실성이 매우 높을 때 (감마사전평균이 10이고과분산계수가 1.0), 평행 비교 추정치의 추정오차는 시간변동 경험적 베이즈 추정치의 추정오차와 유사해짐을 알 수 있다 (표 2 참조).

- 한편 시간에 흐름에 따라 달라질 수 있는 기대교통사고건수의 변화율은 두 가지 추정기법의 성능에 큰 변화를 주지 않는 것으로 나타났다. 즉 기대교통사고건수의 증가율이 변화한다 할지라도 동일한 감마사전분포의 형태하에서 RMSE는 1보다 큰 일정한 값을 유지한다 (표 2 참조). 이러한 결과는 두 가지 추정 기법 모두 시간의 흐름에 따른 기대교통사고건수의 변화를 반영할 수 있으나 평행 비교 방법은 평균으로의 회귀편의를 제거하지 못해 그 추정오차가 시간변동 경험적 베이즈 추정치의 추정오차보다 크다는 사실을 반영한다.

V. 결론 및 향후과제

본 연구에서는 패널 교통사고 자료를 이용한 기대교통사고건수 추정기법의 특징과 각 기법의 추정오차 비교·분석 결과를 시간의 흐름에 따라 기대교통사고건수가 동일할 경우와 변화하는 경우로 나누어 제시하였다. 한 지점의 기대교통사고건수는 알려지지 않은 모수라는 점에 착안하여 정교한 모의실험 시나리오를 통해 기대교통사고건수와 관측교통사고건수를 발생시킨 후 각 기법의 추정오차를 분석하였으며, 본 연구의 주요 결과를 요약하면 아래와 같다.

- 현재 국내 교통안전 연구에서 주로 사용되고 있는 관측교통사고건수 기반 기법인 평균 관측교통사고건수와 평행비교 추정치의 추정오차는 경험적 베이즈 추정치의 추정오차보다 항상 크게 나타났다. 따라서 향후 패널교통사고 자료를 이용한 교통안전 연구 수행 시 추정오차를 감소시킬 수 있는 경험적 베이즈 방법의 적용이 보다 바람직하다고 판단된다.
- 시간의 흐름에 따라 기대교통사고건수가 변화하지 않을 경우, 분석기간이 늘어날수록 두 가지 기법의 추정오차의 크기는 현저하게 감소하는 것으로 분석되었다 (이러한 결론은 표본의 크기가 커질수록 추정치의 효율성이 증가한다는 통계이론에 기반함). 따라서 시간불변 기대교통사고건수 가정이 유의할 때 현재 국내 교통사고 잦은 곳 선정 연구에서 기준치로 사용되고 있는 분석기간인 "1년"보다 더 긴 분석기간을 이용하여 추정된 "효율적인 기대교통사고건수 추정치"를 이용하여 관련분석을 수행할 필요가 있다.
- 시간의 흐름에 따른 기대교통사고건수 변화 여부와 관계없이 경험적 베이지안 기법은 교통안전연구 결과의 신뢰성을 향상시킬 수 있을 것으로 판단된다. 그러나 모의실험 결과 불확실한 사전분포를 선택했을 경우 경험적 베이즈 추정치의 추정오차는 기존 관측교통사고건수 기반 추정치와 유사해짐을 알 수 있었다. 따라서 교통안전 연구 수행 시 올바른 경험적 베이지안 기법 적용을 위해서는 불확실성이 낮은 사전분포 선택에 필요한 이질적 음이항 또는 음다항 모형 개발을 위한 체계적이고 지속적인 연구가 필요하다.

본 연구는 패널 교통사고 자료를 이용한 교통안전 연구에서 경험적 베이지안 기법의 적용이 기대교통사고건수의 추정오차를 줄일 수 있다는 이론적 사실을 모의실험을 통해 입증하였으며, 특히 다양한 조건하에서 발생하는 각 추정기법의 추정오차의 크기를 비교·분석하여 국내 교통안전연구 기법의 개선방향을 제시 했다는 점에서 선행연구와 차별성을 갖는다. 그러나 본 연구는 시간의 흐름에 따라 기대교통사고건수가 변화할 때 분석기간의 길이를 2년으로 한정 지어 두 추정치의 추정오차를 분석한바 향후 분석기간 연장에 따른 추정오차의 변화를 분석할 필요가 있다. 또한 선행연구의 결과에 기반하여 모의실험에서 가정한 포아송-감마 혼합모형의 현실적 적합도 분석에 대한 연구도 수행되어야 할 것으로 판단된다. 마지막으로 경험적 베이지안 기법 뿐 아니라 사전분포를 규정짓지 않는 전베이즈 기법 (full Bayes method)을 이용한 추정방법 (Miaou and Lord, 2003; Miranda-Moreno and Fu, 2007; Lord and Miranda-Moreno, 2008)의 적용 가능성에 대한 연구도 병행되어야 할 것이다.

참고문헌

1. 도로교통공단(2009), "2009년 교통사고 잦은 곳 기본개선계획 및 효과분석".
2. 신강원(2010), "교통사고 추정방법 비교 연구: 경험적 베이즈 추정치 vs. 관측교통사고건수", 대한토목학회 논문집, 제30권 제5D호, 대한토목학회, pp.453~459.
3. Abbess, C., Jarrett, D., and Wright, C. C. (1981), "Accidents at Blackspots: Estimating the Effectiveness of Remedial Treatment, with Special Reference to the 'regression-to-mean' Effect", Traffic Engineering and Control, Vol. 22, No. 10, pp.535~542.
4. Guo, G.(1996), "Negative Multinomial Regression Models for Clustered Event Counts", Sociological Methodology, Vol. 26, pp.113~132.
5. Hauer, E.(1997), "Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety", Pergamon, Elsevier Science Ltd..
6. Hauer, E., Harwood, D. W., Council, F. M., and Griffith, M. S.(2002), "Estimating Safety by the Empirical Bayes Method: A Tutorial", In Transportation Research Record: Journal of the Transportation Research Board, No. 1784, pp.126~131.
7. Hilbe, J. M.(2007), "Negative Binomial Regression", Cambridge University Press.
8. Johnson, N. L., and Kotz, S.(1969), "Distributions in Statistics: Discrete Distributions", Houghton Mifflin Company.
9. Lord, D., and Miranda-Moreno, L. F.(2008), "Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter of Poisson-gamma Models for Modeling Motor Vehicle Crashes: A Bayesian Perspective", Safety Science, Vol. 46, No. 5, pp.751~770.
10. Miaou, S.-P., and Lord, D.(2003), "Modeling Traffic Crash-flow Relationships for Intersections: Dispersion Parameter, Functional Form, and Bayes versus Empirical Bayes Methods", In Transportation Research Record: Journal of the Transportation Research Board, No. 1840, pp.31~40.
11. Miranda-Moreno, L. F., and Fu, L.(2007), "Traffic Safety Study: Empirical Bayes or Full Bayes?", Transportation Research Board 86th Annual Meeting, Washington, D.C.
12. Shin, K. and Washington, S.P.(2010). "Empirical Bayes Methods in the Study of Traffic Safety via Heterogeneous Negative Multinomial Model", Transportmetrica, DOI: 10.1080/18128601003680976

♧ 주 작 성 자 : 신강원
♧ 교 신 저 자 : 신강원
♧ 논문투고일 : 2010. 8. 23
♧ 논문심사일 : 2010. 11. 9 (1차)
2011. 1. 10 (2차)
2011. 1. 25 (3차)
♧ 심사판정일 : 2011. 1. 25
♧ 반론접수기한 : 2011. 6. 30
♧ 3인 익명 심사필
♧ 1인 abstract 교정필