

교통사고 추정방법 비교 연구: 경험적 베イズ 추정치 vs. 관측교통사고건수

Comparative Study on the Estimation Methods of Traffic Crashes: Empirical Bayes Estimate vs. Observed Crash

신강원*

Shin, Kangwon

Abstract

In the study of traffic safety, it is utmost important to obtain more reliable estimates of the expected crashes for a site (or a segment). The observed crashes have been mainly used as the estimate of the expected crashes in Korea, while the empirical Bayes (EB) estimates based on the Poisson-gamma mixture model have been used in the USA and several European countries. Although numerous studies have used the EB method for estimating the expected crashes and/or the effectiveness of the safety countermeasures, no past studies examine the difference in the estimation errors between the two estimates. Thus, this study compares the estimation errors of the two estimates using a Monte Carlo simulation study. By analyzing the crash dataset at 3,000,000 simulated sites, this study reveals that the estimation errors of the EB estimates are always less than those of the observed crashes. Hence, it is imperative to incorporate the EB method into the traffic safety research guideline in Korea. However, the results show that the differences in the estimation errors between the two estimates decrease as the uncertainty of the *prior* distribution increases. Consequently, it is recommended that the EB method be used with reliable hyper-parameter estimates after conducting a comprehensive examination on the estimated negative binomial model.

Keywords : *traffic crashes, empirical Bayes estimate, observed crashes, estimation error*

요 지

교통안전연구에서 한 지점 (또는 구간)의 기대교통사고건수를 신뢰성 높게 추정하는 것은 매우 중요하다. 기대교통사고건수를 추정하기 위해 현재 국내에서는 주로 관측교통사고건수를 사용하고 있으나 국외에서는 포아송-감마 혼합모형에 기반한 경험적 베イズ 추정치를 활용하고 있는 추세이다. 그간 많은 연구들이 경험적 베イズ 추정치를 이용해 기대교통사고건수 추정 및 교통안전개선사업의 평가를 시도 하였으나 전술한 2가지 추정방법의 추정오차를 비교연구한 문헌은 없는 실정인 바 본 연구는 2가지 추정방법의 추정오차를 모의실험을 통해 비교·분석하여 제시하였다. 모의 발생된 총 3,000,000개 지점의 교통사고 자료를 분석한 결과 기대교통사고건수를 경험적 베イズ 추정방법을 이용해 추정했을 경우 그 추정오차는 관측교통사고건수만을 이용했을 때 발생하는 추정오차에 비해 항상 작은 것으로 나타나 국내의 교통안전연구 가이드라인에 경험적 베이지안 추정방법의 도입이 필요할 것으로 판단된다. 그러나 사전분포의 불확실성이 높아질수록 두 가지 추정방법의 추정오차의 차이는 감소하는 것으로 나타나, 기대교통사고건수 추정 시 추정된 음이항 모형에 대한 종합적인 검증을 수행한 후 신뢰성 있는 초모수의 추정치를 이용해 경험적 베이지안 방법을 적용하는 것이 바람직하다고 판단된다.

핵심용어 : 교통사고, 경험적 베イズ 추정치, 관측교통사고건수, 추정오차

1. 연구배경 및 목적

교통안전연구에서 한 지점(또는 구간)의 기대교통사고건수 (expected crashes)를 신뢰성 높게 추정하는 것은 도로교통안전을 향상시키기 위한 교통사고 잦은 곳의 선정, 교통안전 개선사업의 수행계획 수립, 교통안전 개선사업의 평가 등과 같은 다양한 활동을 위해 매우 중요하다. 기대교통사고건수를 추정하기 위해 현재 국내에서는 주로 관측교통사고건수

(observed crashes)를 사용하고 있으나 국외에서는 포아송-감마 혼합모형(Poisson-gamma mixture model)에 기반한 경험적 베イズ 추정치(empirical Bayes estimate)를 활용하고 있는 추세이다. 물론 국내외 많은 연구(Abbesse et al., 1981; Hauer and Persaud, 1987; Elvik, 1997; 박민호 등, 2006; 박규영 등, 2006; Shin and Washington, 2007; Shin et al., 2009)들이 경험적 베イズ 추정치를 이용해 기대교통사고건수 추정 및 교통안전개선사업의 평가를 시도 하였으나 전

*정회원 · 경성대학교 도시공학과 전임강사 · 경성대학교 공학기술연구소 겸임연구원 (E-mail : kangwon@ks.ac.kr)

술한 2가지 추정방법의 추정오차를 비교연구한 문헌은 없는 실정이다. 특히 전술한 선행연구는 실제 교통사고자료를 이용하여 베이지안 기법의 적용 당위성을 언급하였으나 각 지점의 기대교통사고건수는 알려지지 않은 모수(parameter)인 바 실증 데이터를 이용하여 2가지 추정방법의 오차를 비교·분석 하는 것은 한계가 있다고 할 수 있다.

따라서 본 연구는 시뮬레이션을 통해 각 지점의 기대교통사고건수를 정의하고 이를 추정하기 위해 사용될 수 있는 관측교통사고건수 기반 추정 방법과 경험적 베이지안 추정 방법의 추정오차를 비교·분석하여 제시하였다. 이를 위해 임의지점(1,000개)의 기대교통사고건수와 관측교통사고건수를 포아송-감마 혼합모형(Poisson-gamma mixture model)을 이용하여 모의발생 시켰다. 모의 발생된 교통사고자료에 음이항 모형(negative binomial model)을 적용하여 감마사전분포(gamma prior distribution)의 초모수(hyper-parameter)를 추정한 후 이를 경험적 베이스 추정치를 얻는데 사용했으며, 추정된 경험적 베이스 추정치와 모의 발생시킨 관측교통사고건수의 기대교통사고건수에 대한 추정오차를 비교하여 2가지 추정방법의 적용 적합성을 분석하였다. 궁극적으로 본 연구는 관측교통사고건수를 이용한 기대교통사고건수 추정방법과 경험적 베이지안 추정치를 이용한 기대교통사고건수 추정방법의 추정오차를 비교·분석하여 현재 국내에서 사용되고 있는 교통안전연구 방법론의 개선방향을 제시하기 위해 수행되었다.

2. 기대교통사고건수 추정방법

본 장에서는 한 분석 대상지점(또는 구간)의 기대교통사고건수를 추정하기 위해 사용되고 있는 횡단면적(cross-sectional) 관측교통사고건수 기반 추정방법과 경험적 베이지안 방법을 제시하였다.

2.1 관측교통사고건수 기반 추정방법

기대교통사고건수 추정을 위해 가장 널리 사용되는 방법은 관측교통사고건수 기반 추정방법이라 할 수 있다. 관측교통사고건수 기반 추정방법은 분석 지점에서 관측된 교통사고건수(observed crashes)를 이용하여 기대교통사고건수를 추정하는 간단한 방법으로, 분석 지점 i 의 교통사고건수가 기대교통사고건수 λ_i 인 포아송 분포를 따를 때 관측교통사고건수의 로그우도함수(log-likelihood function)은 식 (1)과 같이 표현 될 수 있다.

$$\ln L(y_i|\lambda_i) = -\lambda_i + y_i \ln \lambda_i - \ln(y_i!) \quad (1)$$

여기서 y_i 는 분석지점 i 의 관측교통사고건수를 나타내며, 위 로그 우도함수를 최대화하는 기대교통사고건수의 추정치는 식 (2)와 같이 관측교통사고건수와 동일함을 알 수 있다.

$$\hat{\lambda}_i = y_i \quad (2)$$

이처럼 관측교통사고건수를 이용하여 분석지점의 기대교통사고건수인 포아송 모수를 손쉽게 추정할 수 있으며, 그 적용의 용이성 때문에 관측교통사고건수 기반 추정방법은 교통안전연구에서 널리 사용되고 있다(McMillen, 1999; 도로

교통공단, 2009).

2.2 경험적 베이지안 기법

한 지점의 기대교통사고건수를 추정하기 위해서는 해당 지점의 관측교통사고건수를 사용할 수 있으나 관측교통사고건수만을 사용하여 기대교통사고건수를 추정했을 경우 “평균으로의 회귀(RTM: regression-to-the-mean) 현상”으로 인한 기대교통사고건수에 대한 오추정(mis-estimation)의 문제점이 제기되어 왔다(Hauer, 1980a, 1980b, 1997; Hauer and Persaud, 1983; Hauer et al., 2002). 전술한 RTM 오차를 제거하기 위해 국외에서는 1980대 이후 베이지안 방법을 적용한 기대교통사고건수 추정방법론이 적용되어 왔으며, 이 추정방법론은 분석지점에서 발생하는 기대교통사고건수의 사전분포(prior distribution)와 우도(likelihood)를 각각 감마분포(gamma distribution)와 포아송 분포(Poisson distribution)로 가정하는 포아송-감마 혼합모형으로 설명될 수 있다.

구체적으로 교통안전연구에서 포아송-감마 혼합모형은 분석 지점 i 의 기대교통사고건수인 포아송 모수(Poisson parameter: λ_i)를 추정할 때 발생하는 불확실성을 고려하기 위해 포아송 모수 λ_i 를 고정된 상수가 아닌 초모수 α, β 를 가지는 감마 확률변수로 가정한다. 따라서 포아송-감마혼합모형에서 분석 지점 i 의 관측교통사고건수를 y_i 라 할 때 포아송 모수 λ_i 의 사후분포 $f(\lambda_i|y_i)$ 는 식 (3)과 같다(Hauer, 1997).

$$f(\lambda_i|y_i) = \frac{f(y_i|\lambda_i) \cdot f(\lambda_i)}{f(y_i)} \quad (3)$$

여기서 $f(y_i|\lambda_i)$ 는 관측교통사고건수 y_i 의 포아송 우도, $f(\lambda_i)$ 는 초모수 α, β 를 가지는 포아송 모수 λ_i 의 감마사전분포, $f(y_i)$ 는 포아송-감마 혼합모형의 주변확률분포이다.

한편 포아송-감마 혼합모형의 주변확률분포 $f(y_i)$ 는 식 (4)에서 제시된 바와 같이 음이항 분포로 표현할 수 있다.

$$f(y_i) = \frac{\Gamma(y_i + \alpha) \left(\frac{1}{1 + \beta}\right)^\alpha \left(\frac{\beta}{1 + \beta}\right)^{y_i}}{y_i! \Gamma(\alpha)} \quad (4)$$

여기서 $\Gamma(\cdot)$ 는 감마함수, α, β 는 감마사전분포의 초모수, 그 외 변수는 식 (3)에서 정의된 바와 같다.

포아송-감마 혼합모형에서 전술한 감마사전분포, 음이항 주변확률분포, 포아송 우도를 이용했을 때 식 (3)에 제시된 사후분포(posterior distribution)은 식 (5)에 제시한 바와 같이 모수 $(\alpha + y)$ 와 $\beta(\beta + 1)^{-1}$ 을 갖는 감마분포와 같다.

$$f(\lambda_i|y_i) = \frac{\left(1 + \frac{1}{\beta}\right)^{y_i + \alpha} \lambda_i^{y_i + \alpha - 1} e^{-\lambda_i \left(1 + \frac{1}{\beta}\right)}}{\Gamma(\alpha + y_i)} \quad (5)$$

이러한 결과는 감마사전분포가 포아송우도의 공액사전분포(conjugate prior distribution)이기 때문이다(Hauer, 1997).

추정된 감마 사후분포는 분석지점 i 에 y_i 건의 교통사고가 관측되었을 때 참조그룹(i.e. 분석지점과 유사한 외부조건을 가진 지점의 그룹)의 기대교통사고건수의 분포를 나타내는 것으로 교통안전연구에서는 주로 사후분포의 평균(posterior mean)이 분석지점의 기대교통사고건수의 추정치로 사용된다. 감마사후분포의 평균은 형상모수(shape parameter)인 $(\alpha + y_i)$

와 척도모수(scale parameter)인 $\beta(\beta+1)^{-1}$ 의 곱과 같으며 이는 식 (6)과 같이 정리할 수 있다. 식 (6)에서 베이저안 가중치(Bayesian weight) w 는 포아송모수 λ_i 의 사전평균 $E(\lambda_i)=\alpha\beta=\phi^{-1}\beta$ 와 사전분산 $V(\lambda_i)=E(\lambda_i)^2\phi$ 의 함수로 0과 1 사이의 값을 가지며, ϕ 는 감마 사전분포의 초모수인 α 의 역수로 과분산계수(overdispersion parameter)를 의미한다. 여기서 과분산계수는 사전분포의 분산을 결정짓는 모수 중 하나로 사전평균값이 일정하다면 ϕ 값이 커질수록 사전분포의 분산은 커지는 특성을 갖는다. 따라서 분석지점 예 y_i 건의 교통사고가 관측되었을 때 기대교통사고건수 λ_i 의 사후평균은 사전평균인 $E(\lambda_i)$, 관측교통사고건수 y_i , 베이저안 가중치 w 를 이용하여 구할 수 있고, 이 사후평균은 관측사고건수 y_i 와 사전분포의 평균인 $E(\lambda_i)$ 값 사이의 값으로 나타난다.

$$\begin{aligned} E(\lambda_i|y_i) &= \frac{(y_i + \alpha)\beta}{(\beta+1)} \\ &= \left\{ \frac{E(\lambda_i)}{E(\lambda_i) + V(\lambda_i)} \right\} E(\lambda_i) + \left\{ \frac{V(\lambda_i)}{E(\lambda_i) + V(\lambda_i)} \right\} y_i \\ &= \left\{ \frac{1}{1 + E(\lambda_i)\phi} \right\} E(\lambda_i) + \left\{ \frac{E(\lambda_i)\phi}{1 + E(\lambda_i)\phi} \right\} y_i \\ &= wE(\lambda_i) + (1-w)y_i \end{aligned} \quad (6)$$

전술한 바와 같이 기대교통사고건수의 추정량인 감마사후평균은 관측교통사고건수, 감마사전분포의 평균, 감마사전분포의 분산을 이용해서 얻을 수 있다. 포아송-감마 혼합모형에서 관심대상이 되는 포아송모수의 감마사전분포의 초모수는 주변확률분포인 음이항 분포를 이용하여 역추정될 수 있으며, 이러한 역추정기법을 경험적 베이저안 방법(empirical Bayes method)이라 한다(Maritz, 1970; Hauer, 1997).

전술한 음이항분포의 모멘트는 식 (7)와 같으며, 이를 이용하여 감마사전분포의 초모수는 역추정될 수 있다.

$$E(y_i) = E(\lambda_i), \quad V(y_i) = E(\lambda_i) + \phi E(\lambda_i)^2 \quad (7)$$

식 (7)에서 제시된 음이항분포의 모멘트를 구성하는 모수 $E(\lambda)$ 와 ϕ 는 주어진 자료의 로그우도함수(log-likelihood function)를 최대화시켜 추정할 수 있으며, 이는 음이항모형(negative binomial model)에서 모수를 추정하는 통상적인 절차를 따를 수 있다(Cameron and Trivedi, 1998; Cameron and Trivedi, 2005; Greene, 2003; Washington et al., 2003). 구체적으로 음이항모형에서 분석지점 i 의 로그우도함수는 아래와 같다.

$$\ln L = \sum_{i=1}^n [\ln \Gamma(\phi^{-1} + y_i) - \ln \Gamma(\phi^{-1}) - \ln \Gamma(y_i + 1) + \phi^{-1} \ln u_i + y_i \ln(1 - u_i)] \quad (8)$$

식 (8)에서 u_i 는 $\phi^{-1}\{\phi^{-1} + E(\lambda)\}^{-1}$ 과 같으며 사전평균 $E(\lambda_i)$ 는 식 (9)와 같이 분석지점 i 의 특성을 반영하는 설명변수 벡터(\mathbf{x}_i)와 회귀계수벡터(β)를 가지는 지수평균함수(exponential mean function)로 제시할 수 있다.

$$E(\lambda_i) = \mu_i = \exp(\mathbf{x}_i' \beta) \quad (9)$$

따라서 식 (8)에 제시된 음이항모형의 로그우도함수를 최

대화시켜 회귀계수(β)와 과분산계수(ϕ)를 추정할 수 있으며, 이 추정치들을 이용하여 기대교통사고건수의 사후분포의 평균인 경험적 베이스 추정치는 식 (10)과 같이 쉽게 얻을 수 있다(Hauer, 1997).

$$\hat{E}(\lambda_i|y_i) = \left\{ \frac{1}{1 + \hat{E}(\lambda_i)\hat{\phi}} \right\} \hat{E}(\lambda_i) + \left\{ \frac{\hat{E}(\lambda_i)\hat{\phi}}{1 + \hat{E}(\lambda_i)\hat{\phi}} \right\} y_i \quad (10)$$

3. 모의실험

본 장에서는 기대교통사고건수를 추정하기 위해 사용될 수 있는 경험적 베이스 추정치와 관측교통사고건수의 추정오차를 비교·분석하기 위해 수행된 모의실험의 과정을 설명하였다.

3.1 모의실험 시나리오

본 연구에서는 전장에서 설명한 포아송-감마 혼합모형의 특성을 반영하기 위해 기대교통사고건수의 평균과 과분산계수를 아래와 같이 각각 10개의 수준으로 설정하여 총 100개(10×10)의 모의실험 시나리오를 설정하였다.

• 기대교통사고건수의 평균(μ)

1, 2, 3, 4, 5, 6, 7, 8, 9, 10

• 기대교통사고건수의 과분산계수(ϕ)

0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0

분석 시나리오에서 기대교통사고건수의 평균의 최대값은 “10”으로 설정하였는데 이는 어떤 지점의 기대교통사고건수가 연간 10건 이하라는 가정을 충족시키기 위함이며, 과분산계수의 최대값은 기대교통사고건수의 표준편차가 10건 이하¹⁾라는 가정하에 “1.0”으로 설정하였다.

3.2 모의실험 과정

본 모의실험에서는 각 분석 시나리오 별로 1,000개의 분석지점별 기대교통사고건수와 이에 기반한 관측교통사고건수를 모의 발생시킨 후 음이항 모형을 이용해 추정된 경험적 베이스 추정치와 모의 발생된 관측교통사고건수를 이용한 추정방법의 평균제곱오차(MSE: mean squared error)를 비교 분석하였다. 전술한 모의실험을 통한 교통사고건수 추정방법 절차는 아래와 같다.

Step 0: 1,000개의 교통사고 발생지점 발생($i=1, 2, \dots, 1000$)

Step 1: 시나리오 별 감마사전분포의 평균과 분산을 이용하여 각 분석지점의 기대교통사고건수(λ_i) 발생: $\lambda_i \sim \text{Gamma}(\mu_i, \mu_i^2 \phi_i)$

Step 2: 각 지점의 기대교통사고건수(λ_i)를 이용하여 관측교통사고건수(y_i) 발생: $y_i \sim \text{Poisson}(\lambda_i)$

Step 3: 관측교통사고건수를 이용하여 음이항모형 추정: $\hat{E}(y_i) = \hat{\mu}_i, \hat{V}(y_i) = \hat{\mu}_i + \hat{\mu}_i^2 \hat{\phi}_i$

Step 4: 감마사후평균(경험적 베이스 추정치: $\hat{E}[\lambda_i|y_i]$) 추정(식 (10) 참조)

Step 5: 각 추정방법의 평균제곱오차 계산(식 (11) 참조)

Step 6: Step 0부터 Step 5까지 반복(반복횟수=30회)

1) 기대교통사고건수의 분산 ($V[\lambda_i] = \mu_i^2 \phi$)의 최대값은 $10^2 \times 1 = 100$.

$$MSE_{EB} = \frac{\sum_{i=1}^{30000} \{\lambda_i - \hat{E}[\lambda_i|y_i]\}^2}{30000}, \quad MSE_y = \frac{\sum_{i=1}^{30000} (\lambda_i - y_i)^2}{30000} \quad (11)$$

식 (11)에서 MSE_{EB} 와 MSE_y 는 각각 경험적 베イズ 추정치와 관측교통사고건수의 평균제곱오차를 의미한다. 모의실험은 100개의 시나리오 별로 동일한 과정을 거쳤으며, 모의 실험의 모든 과정은 통계패키지인 STATA version 10을 이용하여 수행되었다.

4. 모의실험 결과

모의실험 과정을 통해 발생시킨 총 3,000,000개(1000×30×100)의 교통사고 자료를 이용하여 관측교통사고건수와 경험적 베이스 추정치의 평균제곱오차를 구했으며, 그 결과는 표 1에 제시하였다. 분석결과 모든 시나리오에서 경험적 베이스 추정치의 평균제곱오차가 관측교통사고건수의 평균제곱오차보다 작은 것으로 나타났다. 따라서 현재 국내의 교통안전연구에서 사용되고 있는 기대교통사고건수 추정방법론(i.e. 관측교통사고건수 기반 추정방법론)은 경험적 베이지안 방법을 이용해 개선되어야 할 것으로 판단된다. 또한 이러한 결과는 기존의 베이지안 결정이론(Bayesian decision theory)을 뒷받침하는 것으로 기대손실(expected loss)를 최소화하는 추정치는 사후평균이라는 사실을 나타낸다(Lee, 2004; Cameron and Trivedi, 2005). 한편 각 추정방법의 평균제곱오차는 과분산계수가 커지거나 기대교통사고건수의 평균이 커질 때 증

가하는 현상을 보이는데 이는 모의 발생된 각 지점의 기대교통사고건수의 값이 커짐에 따라 평균제곱오차의 값이 커지기 때문이다.

표 2는 두 추정방법의 평균제곱오차를 비교하기 위해 계산된 시나리오 별 상대평균제곱오차(MSE_y/MSE_{EB})의 값으로 그 값이 모든 시나리오에서 1보다 크게 나타나 전술한 바와 같이 관측교통사고건수를 이용하여 기대교통사고건수를 추정했을 경우 발생하는 오차가 경험적 베이스 추정치를 이용하여 기대교통사고건수를 추정했을 때 발생하는 오차보다 큰 것을 보여준다.

한편 경험적 베이스 추정치의 추정오차는 관측교통사고건수의 추정오차에 비해 항상 작지만 상대추정오차는 그림 1과 같이 기대교통사고건수의 평균이 커지거나 과분산계수의 값이 커질수록 감소하는 것으로 나타났다. 이러한 현상은 기대교통사고건수의 평균이 커지거나 과분산계수의 값이 커질수록 베이지안 가중치의 값이 작아져 경험적 베이스 추정치와 관측교통사고건수의 값이 유사해지기 때문인 것으로 판단된다(식 (6) 참조).

표 3은 시나리오 별 베이지안 가중치의 기대값을 나타내는 것으로 전술한 바와 같이 사전평균에 대한 베이지안 가중치는 기대교통사고건수의 평균이 커지거나 과분산계수가 커질 때 감소하는 것으로 나타났다. 이는 식 (6)에 제시된 바와 같이 사후평균의 값은 사전평균과 관측교통사고건수로 구성되어 있으나 사전분포에 대한 불확실성이 높아질 경우(i.e. 베이지안 가중치가 커질 경우) 사후평균은 관측교통사고

표 1. 추정방법 별 평균제곱오차

MSE _{EB}		E(λ_i)									
		1	2	3	4	5	6	7	8	9	10
ϕ_i	0.1	0.09	0.34	0.71	1.17	1.70	2.26	2.86	3.55	4.26	4.96
	0.2	0.17	0.58	1.13	1.80	2.56	3.26	4.13	5.01	5.82	6.79
	0.3	0.23	0.74	1.46	2.18	3.00	3.87	4.72	5.70	6.72	7.50
	0.4	0.29	0.90	1.64	2.46	3.31	4.28	5.07	5.96	7.15	8.06
	0.5	0.34	1.01	1.81	2.68	3.54	4.49	5.51	6.49	7.32	8.34
	0.6	0.38	1.09	1.91	2.83	3.78	4.74	5.64	6.67	7.63	8.50
	0.7	0.40	1.15	2.03	2.97	3.86	4.85	5.80	6.84	7.81	8.78
	0.8	0.45	1.22	2.17	3.07	3.96	4.90	6.01	6.94	8.06	8.79
	0.9	0.48	1.30	2.18	3.16	4.09	5.08	6.09	7.10	8.13	9.13
	1	0.50	1.34	2.25	3.19	4.33	5.20	6.03	7.30	8.14	9.14
MSE _y		E(λ_i)									
		1	2	3	4	5	6	7	8	9	10
ϕ_i	0.1	1.00	1.98	3.00	4.06	5.01	5.97	7.02	7.97	8.95	9.97
	0.2	0.98	1.99	3.02	4.01	5.05	5.96	7.04	8.11	9.04	10.16
	0.3	1.00	2.00	3.02	3.95	4.95	5.91	6.95	8.06	9.16	10.03
	0.4	1.00	2.03	3.03	3.96	4.90	6.03	6.98	7.92	9.19	10.09
	0.5	1.01	2.00	3.03	3.98	4.92	5.99	7.07	8.11	8.88	10.06
	0.6	1.01	2.00	2.96	3.99	5.07	5.95	6.97	8.02	9.00	10.01
	0.7	1.00	1.94	3.01	3.98	5.01	6.00	6.90	8.04	9.06	10.01
	0.8	1.02	1.99	3.06	4.03	5.02	5.91	7.06	8.10	9.20	9.93
	0.9	1.02	2.01	2.96	4.01	5.03	6.05	6.97	8.08	9.09	10.15
	1	0.97	2.01	3.00	3.98	5.13	6.04	6.92	8.18	9.05	9.97

표 2. 상대평균제곱오차(MSE_y/MSE_EB)

상대평균제곱오차	$E(\lambda_i)$										
	1	2	3	4	5	6	7	8	9	10	
ϕ	0.1	10.59	5.82	4.24	3.47	2.94	2.64	2.46	2.24	2.10	2.01
	0.2	5.74	3.45	2.67	2.23	1.97	1.83	1.71	1.62	1.55	1.50
	0.3	4.30	2.70	2.07	1.81	1.65	1.53	1.47	1.41	1.36	1.34
	0.4	3.44	2.27	1.84	1.61	1.48	1.41	1.38	1.33	1.28	1.25
	0.5	2.97	1.99	1.67	1.49	1.39	1.33	1.28	1.25	1.21	1.21
	0.6	2.63	1.83	1.55	1.41	1.34	1.25	1.24	1.20	1.18	1.18
	0.7	2.51	1.69	1.48	1.34	1.30	1.24	1.19	1.18	1.16	1.14
	0.8	2.26	1.63	1.41	1.31	1.27	1.21	1.17	1.17	1.14	1.13
	0.9	2.15	1.55	1.36	1.27	1.23*	1.19	1.14	1.14	1.12	1.11
	1	1.96	1.50	1.33	1.25	1.18	1.16	1.15	1.12	1.11	1.09

*상대평균제곱오차가 1.25보다 작은 분석 시나리오를 나타냄

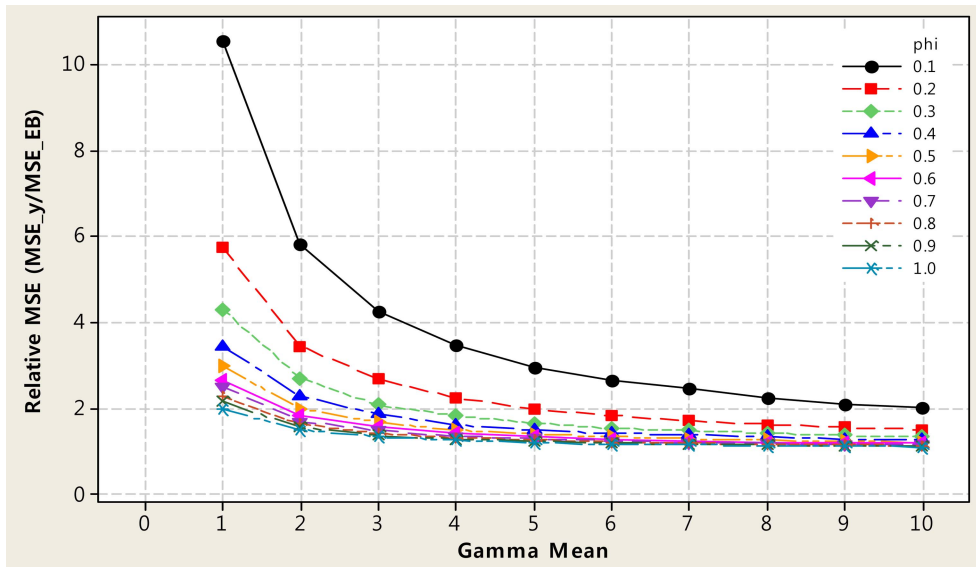


그림 1. 기대교통사고건수의 평균과 과분산계수에 따른 상대평균제곱오차 변화

표 3. 시나리오별 베이지안 가중치

베이지안 가중치	$E(\lambda_i)$										
	1	2	3	4	5	6	7	8	9	10	
ϕ	0.1	0.91	0.83	0.77	0.71	0.67	0.63	0.59	0.56	0.53	0.50
	0.2	0.83	0.71	0.63	0.56	0.50	0.45	0.42	0.38	0.36	0.33
	0.3	0.77	0.63	0.53	0.45	0.40	0.36	0.32	0.29	0.27	0.25
	0.4	0.71	0.56	0.45	0.38	0.33	0.29	0.26	0.24	0.22	0.20
	0.5	0.67	0.50	0.40	0.33	0.29	0.25	0.22	0.20	0.18	0.17
	0.6	0.63	0.45	0.36	0.29	0.25	0.22	0.19	0.17	0.16	0.14
	0.7	0.59	0.42	0.32	0.26	0.22	0.19	0.17	0.15	0.14	0.13
	0.8	0.56	0.38	0.29	0.24	0.20	0.17	0.15	0.14	0.12	0.11
	0.9	0.53	0.36	0.27	0.22	0.18*	0.16	0.14	0.12	0.11	0.10
	1	0.50	0.33	0.25	0.20	0.17	0.14	0.13	0.11	0.10	0.09

*베이지안 가중치가 0.2보다 작은 분석 시나리오를 나타냄

건수와 유사해지는 특성을 반영한다. 예를 들어 기대교통사고건수의 평균이 10이고 과분산계수가 1일 때 사전분포의 표준편차는 $10(\sqrt{100 \times 1})$ 으로 나타낼 수 있는데, 이 경우 사전평균에 대한 베이지안 가중치는 0.09로 추정된 사전분포의

높은 불확실성을 반영한다. 반면 기대교통사고건수의 평균이 10이고 과분산계수가 0.1일 때 사전분포에 대한 베이지안 가중치의 기대값은 0.91로 나타났는데 이는 사전분포의 표준편차가 $0.32(\sqrt{1 \times 0.1})$ 로 작아 사전분포의 불확실성이 상대적으

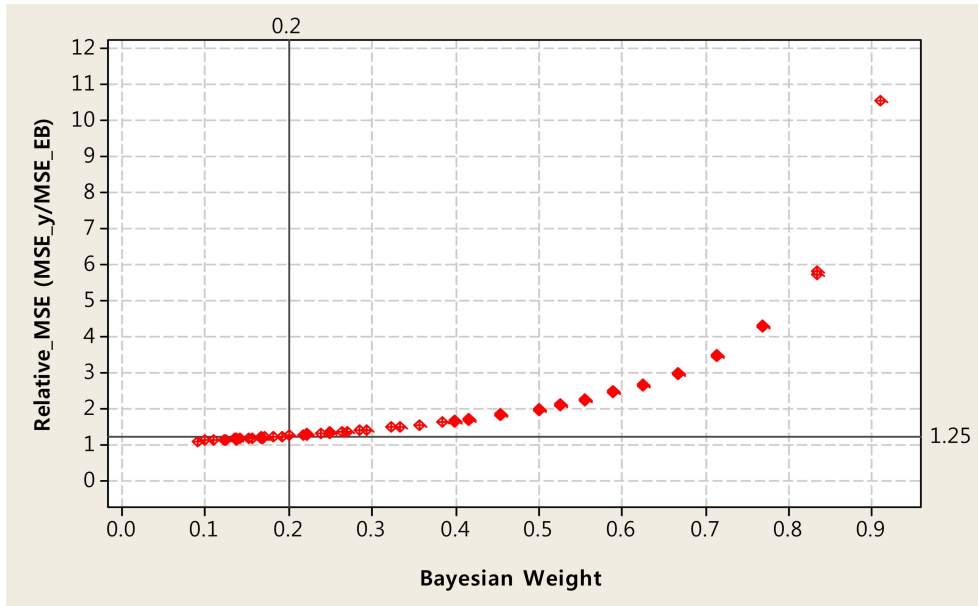


그림 2. 베이지안 가중치에 따른 상대평균제곱오차의 변화

로 낮기 때문이다.

따라서 기대교통사고건수를 추정함에 있어 사전분포의 불확실성이 매우 클 경우(i.e. 사전평균에 대한 베이지안 가중치가 작을 경우) 경험적 베이스 추정치와 관측교통사고건수의 추정오차는 유사해지는 특성을 갖는다. 그림 2는 분석 시나리오 별 베이지안 가중치의 기대값에 따른 두 추정방법의 상대평균제곱오차의 변화를 나타내는데, 모든 시나리오에서 상대평균제곱오차의 값은 1보다 크지만(i.e. 베이지안 추정치의 추정오차가 관측교통사고건수의 추정오차에 비해 작음) 베이지안 가중치가 작아질수록 상대평균제곱오차의 값은 1에 가까워지는 것을 알 수 있다. 즉 사전분포의 불확실성이 클 때 베이지안 가중치의 값은 작아지며, 이 경우 관측교통사고건수를 이용하여 기대교통사고건수를 추정한다 할지라도 그 추정오차는 크지 않다는 것을 나타낸다. 본 연구에서 모의 발생된 교통사고자료를 기반으로 이러한 특성을 살펴봤을 때 베이지안 가중치가 0.2보다 작을 경우에는 관측교통사고건수를 이용해 기대교통사고건수를 추정한다 할지라도 그 추정오차는 경험적 베이스 추정치의 추정오차의 1.25배를 넘지 않는 경향이 있음을 알 수 있다(그림 2 참조). 그러나 베이지안 가중치가 0.2보다 클 경우에는 관측교통사고건수만을 이용해 기대교통사고건수를 추정했을 때 발생하는 추정오차는 경험적 베이스 추정치의 추정오차에 비해 최대 10.6배 가량 커질 수 있음을 보였다.

5. 결론 및 향후 과제

본 연구는 관측교통사고건수를 이용한 기대교통사고건수 추정방법과 경험적 베이지안 추정치를 이용한 기대교통사고건수 추정방법의 추정오차를 비교·분석하여 현재 국내에서 주로 사용되고 있는 관측교통사고건수 기반 교통안전연구 방법론의 개선방향을 제시하기 위해 수행되었다. 이를 위해 포아송-감마혼합모형을 따르는 교통사고자료를 기대교통사고건수의 평균과 과분산계수에 따른 100가지의 분석 시나리오

별로 모의 발생시켰으며 모의 발생된 총 3,000,000개 지점의 교통사고 자료를 분석하여 다음과 같은 결과를 얻었다.

1. 한 분석지점의 기대교통사고건수를 경험적 베이스 추정방법을 이용해 추정했을 경우 그 추정오차는 관측교통사고건수만을 이용했을 때 발생하는 추정오차에 비해 항상 작다. 이러한 모의실험 결과는 기존 베이지안 결정이론을 확장하는 것으로 기대교통사고건수를 추정하기 위해 현재 국내에서 주로 사용되고 있는 방법인 관측교통사고건수 기반 추정방법은 경험적 베이지안 추정방법을 통해 개선되어야 한다는 것을 나타낸다.
2. 음이항 모형을 이용해 추정된 사전분포의 불확실성이 매우 높을 경우에는 경험적 베이지안 추정방법의 추정오차와 관측교통사고건수를 이용한 추정방법의 추정오차는 유사해진다. 이러한 결과는 사전분포의 불확실성이 커질 때 사전평균에 대한 베이지안 가중치가 작아져 경험적 베이스 추정치와 관측교통사고건수가 점점 유사해지기 때문이다 (모의실험 결과 베이지안 가중치가 0.2보다 작을 경우 관측교통사고건수의 추정오차는 경험적 베이스 추정치의 추정오차의 1.25배를 넘지 않음). 따라서 음이항 모형을 이용해 추정된 사전분포의 불확실성이 높을 경우에는 추정된 음이항 모형에 대한 종합적인 검증을 수행하여 맹목적인 경험적 베이지안 방법의 적용을 지양해야 할 것이다.
3. 전술한 바와 같이 사전분포의 불확실성이 높아질수록 두 가지 추정방법의 추정오차의 차이는 감소하나, 기대교통사고건수를 추정함에 있어 경험적 베이지안 방법의 추정오차는 관측교통사고건수만을 이용했을 때보다 항상 작다는 측면에서 국내의 교통안전연구 가이드라인에 경험적 베이지안 추정방법의 도입이 필요하다고 할 수 있다.
4. 따라서 한 지점 (또는 구간)의 기대교통사고건수 추정신뢰성 있는 음이항 모형을 통해 추정된 초모수를 이용한 경험적 베이지안 방법을 적용하는 것이 바람직하다고 할 수 있으며, 이러한 경험적 베이스 추정치를 이용해 교통사고 잦은 곳의 선정, 교통안전 개선사업의 수행계획 수

립, 교통안전 개선사업의 평가와 같은 교통안전연구가 수행되어야 할 것으로 판단된다.

본 연구는 기대교통사고건수를 추정하기 위해 사용되고 있는 2가지 방법론의 추정오차를 모의실험을 통해 비교·분석하여 교통안전연구에 경험적 베이지안 방법의 적용 당위성을 밝혔다는 측면에서 큰 의의를 갖는다고 할 수 있다. 그러나 본 연구는 횡단면적 교통사고자료의 특성만을 고려하여 2가지 방법론의 추정오차를 비교·분석한 바 향후 교통사고의 패널특성을 반영하여 2가지 방법론의 추정오차를 비교할 필요가 있다. 또한 본 연구는 각 추정방법 별 기대교통사고건수의 구간 추정치(interval estimate)의 오차를 분석하지 못해 이에 대한 보완 연구가 추진되어야 할 것으로 판단된다. 끝으로 향후 교통안전연구의 실증분석 수행 시 베이지안 추정치와 관측교통사고건수만을 이용해 추정된 기대교통사고건수의 차이를 본 연구결과인 과분산계수와 기대교통사고건수의 추정치의 크기에 따라 비교 분석하여 제시할 필요가 있다고 판단된다.

감사의 글

본 연구는 2009학년도 경성대학교 공학기술연구소 연구지원사업으로 이루어졌음.

참고문헌

도로교통공단(2009) **교통사고 잦은 구간 선정 및 효과분석 기법 연구**.

박규영, 김태희, 김성욱, 이수범(2006) 도로안전개선사업의 공간적 범위에 따른 도로안전시설물의 효과도 비교. **대한교통학회지**, 대한교통학회, 제24권, 제4호, pp. 31-42.

박민호, 박규영, 장일준, 이수범(2006). 중앙분리대 설치에 따른 사고전환효과 분석. **대한교통학회지**, 대한교통학회, 제24권, 제2호, pp. 113-124.

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.

Cameron, A. C. and Trivedi, P. K. (1998) *Regression Analysis of Count Data*, Cambridge University Press.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics*, Cambridge University Press, New York.

Elvik, R. (1997) Effects on Accidents of Automatic Speed Enforcement in Norway. *Transportation Research Record*, No. 1595, pp. 14-19.

Greene, W. H. (2003) *Econometric Analysis*, Prentice Hall, Upper Saddle River, New Jersey.

Hauer, E. (1980a) Bias-by-selection: Overestimation of the effectiveness of safety countermeasures caused by the process of selection for treatment. *Accident Analysis and Prevention*, Vol. 12, No. 2, pp. 113-117.

Hauer, E. (1980b). Selection for treatment as a source of bias in before-and-after studies. *Traffic Engineering and Control*, Vol. 21, pp. 419-421.

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., Oxford, U.K.

Hauer, E., Harwood, D. W., Council, F. M., and Griffith, M. S. (2002) Estimating safety by the empirical bayes method: A tutorial. *Transportation Research Record No. 1784*, pp. 126-131.

Hauer, E. and Persaud, B. (1983) Common bias in before-and-after accident comparisons and its elimination. *Transportation Research Record No. 905*, pp. 164-174.

Hauer, E. and Persaud, B. N. (1987) How to estimate the safety of rail-highway grade crossings and the safety effects of warning devices. *Transportation Research Record No. 1114*, pp. 131-140.

McMillen, R. D. (1999) *Statistical Evaluation in Traffic Safety Studies*, Institute of Transportation Engineers.

Lee, P. M. (2004) *Bayesian Statistics: An Introduction*. Oxford University Press Inc., New York.

Maritz, J. S. (1970) *Empirical Bayes Methods*. Methuen, London.

Shin, K. and Washington, S. P. (2007) The impact of red light cameras on safety in arizona. *Accident Analysis and Prevention*, Vol. 39, pp. 1212-1221.

Shin, K., Washington, S., and van Schalkwyk, I. (2009) Evaluation of the city of scottsdale loop 101 automated speed enforcement demonstration program. *Accident Analysis and Prevention*, Vol. 41, No. 3, pp. 393-403.

Washington, S. P., Karlaftis, M. G., and Mannering, F. L. (2003) *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall/CRC, New York.

(접수일: 2010.3.23/심사일: 2010.4.30/심사완료일: 2010.6.10)