

이질적 과분산계수가 기대 교통사고건수 추정에 미치는 영향

신강원*

¹경성대학교 도시공학과

Impact of Heterogeneous Dispersion Parameter on the Expected Crash Frequency

Kangwon Shin^{1*}

¹Department of Urban Design and Development Engineering, Kyungsoong University

요약 본 연구는 기대 교통사고건수 추정을 위해 사용되는 SPF의 이질적 분산계수의 유의성이 이질적 사전분포에 직접적인 영향을 받는다는 가설을 검증하고, 이질적 사전분포에 대한 모형 오설정이 교통 안전개선 사업의 평가결과에 주는 영향의 특성을 분석하기 위해 수행되었다. 구체적으로 본 연구에서는 이질적 분산계수의 유의성과 이질적 사전분포의 연관성을 검증하기 위해 모의실험을 통해 이질적 사전분포를 발생시킨 후 이를 NB모형과 HNB모형을 이용하여 SPF를 추정하여 이질적 과분산계수가 SPF의 평균함수 및 분산함수에 주는 영향을 분석하였다. 또한 추정된 계수추정치들을 이용한 사전분포의 초모수 추정치의 오차특성과 이질적 과분산계수를 고려하지 않았을 경우 발생하는 교통사고감소계수(CRF)의 오차 부호와 크기를 상세 분석하여 제시하였다. 모의실험 자료 분석결과 이질적 분산계수의 오추정은 포아송 사전분포의 평균에는 큰 영향을 주지 않으나 분산의 크기를 변화시켜 궁극적으로는 기대교통사고건수의 추정량인 사후평균의 값에 오차를 발생시킬 수 있으며, 구체적으로 이질적 분산함수를 NB모형으로 오설정할 경우 CRF의 값은 참값에 비해 최대 120%의 오차를 발생시키는 것으로 나타났다.

Abstract This study tested the hypothesis that the significance of the heterogeneous dispersion parameter in safety performance function (SPF) used to estimate the expected crashes is affected by the endogenous heterogeneous prior distributions, and analyzed the impacts of the mis-specified dispersion parameter on the evaluation results for traffic safety countermeasures. In particular, this study simulated the Poisson means based on the heterogeneous dispersion parameters and estimated the SPF's using both the negative binomial (NB) model and the heterogeneous negative binomial (HNB) model for analyzing the impacts of the model mis-specification on the mean and dispersion functions in SPF. In addition, this study analyzed the characteristics of errors in the crash reduction factors (CRFs) obtained when the two models are used to estimate the posterior means and variances, which are essentially estimated through the estimated hyper-parameters in the heterogeneous prior distributions. The simulation study results showed that a mis-estimation on the heterogeneous dispersion parameters through the NB model does not affect the coefficient of the mean functions, but the variances of the prior distribution are seriously mis-estimated when the NB model is used to develop SPF's without considering the heterogeneity in dispersion. Consequently, when the NB model is used erroneously to estimate the prior distributions with heterogeneous dispersion parameters, the mis-estimated posterior mean can produce large errors in CRFs up to 120%.

Key Words : dispersion parameter, Poisson-gamma mixture, crash frequency, heterogeneous negative binomial model

1. 서론

경험적 베이지안(EB, empirical Bayesian)방법은 교통

안전연구에서 도로의 한 지점 또는 구간의 기대 교통사고건수(expected crashes) 추정을 위해 널리 사용되고 있다. EB방법은 감마 사전분포와 포아송 우도를 갖는 포아

이 논문은 2014학년도 경성대학교 학술연구비지원에 의하여 연구되었음.

*Corresponding Author : Kangwon Shin(Kyungsoong Univ.)

Tel: +82-51-663-4793 email: kangwon@ks.ac.kr

Received August 11, 2014

Revised September 10, 2014

Accepted September 11, 2014

송-감마혼합모형에서 도출되는 사후평균을 기대 교통사고건수로 추정하는데 사용하며, 사후평균은 포아송-감마혼합모형의 주변확률분포인 음이항 분포를 수용하여 음이항(NB, negative binomial)모형의 평균함수와 분산계수로 역추정된다. 따라서 교통안전연구에서 NB모형은 분석 대상지와 유사한 특성을 갖는 지점 및 구간의 교통사고건수를 종속변수로 갖고 도로기하구조 변수, 교통량 변수 등을 독립변수로 갖는 교통안전성능함수(SPF, safety performance function) 구축 시 핵심 모델링 방법으로 사용되고 있다.

이처럼 NB모형은 EB 적용의 구조 상 사후평균의 역추정에 쉽게 적용될 수 있을 뿐 아니라 교통사고자료의 과분산을 효율적으로 반영할 수 있어 선행연구에서 널리 적용되어 왔다. 그러나 NB모형은 SPF 개발 시 사용되는 교통사고자료가 하나의 분산계수를 갖는다는 등분산 가정에 기반하고 있어 분석자료에 이질적인 과분산계수가 내재되어 있을 경우에는 기대 교통사고건수가 오추정될 가능성이 있다. 따라서 이질적 과분산계수를 고려할 수 있는 SPF의 구조는 꾸준히 논의되어 왔으며, 이질적 과분산계수는 이질적 음이항(HNB, heterogeneous negative binomial) 모형을 통해 고려될 수 있음을 몇몇 선행연구는 밝혀왔다[1-6].

본 연구는 이질적 과분산계수의 유의성이 이질적 사전분포에 직접적인 영향을 받는다는 실증적 분석 결과에 주목하여, 교통사고 자료가 이질적 형태의 분산을 내재하고 있을 때 HNB모형이 실제 그 이질성을 이질적 과분산계수로서 통계적으로 유의하게 추정할 수 있는지를 검증하고 이러한 이질적 분산계수를 고려하지 않았을 때 발생하는 기대교통사고건수 추정 오차를 분석하여 교통안전연구에서 HNB모형의 적용 당위성을 제시하였다.

2. HNB 기반 교통사고건수 추정방법

교통안전연구에서 EB 방법을 이용한 기대 교통사고건수 추정은 1980년대 초반부터 널리 사용되어왔다. EB 방법은 분석지점 i 의 관측 교통사고건수 y_i 가 평균이 λ_i 인 포아송 분포를 따르며 λ_i 는 초모수 α, β 를 갖는 감마 사전분포를 따른다고 가정했을 때, 주변분포인 음이항분포와 관측치를 이용해 λ_i 의 기대값인 $E[\lambda_i|y_i]$ 를 구하여 이를 기대 교통사고건수(expected crashes)의 추정량으로

이용한다. 구체적으로 EB 방법을 통해 최종적으로 얻어지는 기대 교통사고건수 $E[\lambda_i|y_i]$ 의 구조는 Eq. 1과 같다 [7-9].

$$E[\lambda_i|y_i] = \frac{(y_i + \alpha)\beta}{(\beta + 1)} = \left\{ \frac{1}{1 + E[\lambda_i]\phi} \right\} E[\lambda_i] + \left\{ \frac{E[\lambda_i]\phi}{1 + E[\lambda_i]\phi} \right\} y_i, \quad \text{Eq. 1}$$

여기서 $E[\lambda_i]$ 는 감마 사전분포의 두 초모수 α, β 의 곱으로 표현되며, 초모수 α 는 과분산계수 ϕ 의 역수로도 표현된다($E[\lambda_i] = \alpha\beta = \phi^{-1}$). EB 추정치인 $E[\lambda_i|y_i]$ 의 특징 중의 하나는 베이저안 가중치 ($(1/(1 + E[\lambda_i]\phi))^{-1}$)이 값이 0과 1사이의 값을 가져 사후평균 $E[\lambda_i|y_i]$ 는 항상 관측치 y_i 와 사전분포의 평균인 $E[\lambda_i]$ 값 사이의 값을 취한다는 것이다. 따라서 사후평균 $E[\lambda_i|y_i]$ 는 감마 사전분포의 초모수인 α, β 의 값을 추정하면 얻을 수 있으며, 이 값들은 관측치 y_i 를 종속변수로 갖는 NB모형을 이용하여 아래와 같이 추정된다[9-11].

$$E[y_i] = \exp(\mathbf{x}_i\boldsymbol{\beta}) = E[\lambda_i], \quad \text{Eq. 2}$$

여기서 \mathbf{x}_i 는 NB모형을 구성하는 각 지점의 설명변수 벡터, $\boldsymbol{\beta}$ 는 로그우도함수를 최대화하여 추정된 회귀계수 벡터를 의미하며, Eq. 2는 NB모형의 평균함수로 이는 감마 사전분포의 평균과 같다. 또한 NB모형은 로그우도함수 최대화를 통해 과분산계수 ϕ 를 추정하며 전술한 두 추정치 $E[\lambda_i]$ 와 ϕ 는 $E[\lambda_i|y_i]$ 를 역추정하는데 사용된다.

전술한 NB모형의 구조에서 분석자료의 분산 $V[y_i]$ 는 $V[y_i] = E[y_i] + \phi E[y_i]^2$ 로 평균에 의존하는 형태이나 과분산계수 ϕ 의 값은 고정되어 있다. 그러나 분석자료의 과분산 또한 평균함수를 결정짓는 설명변수 \mathbf{x}_i 로 설명될 수 있으며, 이 때 과분산함수는 Eq. 3과 같이 정의된다.

$$\phi_i = \exp(\mathbf{x}_i\boldsymbol{\gamma}), \quad \text{Eq. 3}$$

위 식에서 $\boldsymbol{\gamma}$ 는 과분산함수의 회귀계수벡터로 Heydecker and Wu(2001)는 과분산함수의 회귀계수 또한 NB구조의 로그우도함수를 최대화하여 추정할 수 있음을 보였고[1], Hilbe(2007)는 과분산계수가 설명변수의 함수로 구성되는 NB모형을 HNB모형 (NB-H모형)으로 명명하여 상용소프트웨어인 STATA에서 사용 가능한

명령어 *gmbreg*를 제공하였다[12]. 교통안전연구에서 HNB모형이 적용 가능한 자료형태는 흔히 접할 수 있는데 교통량에 따른 과분산계수의 비선형 변화, 신호교차로와 비신호교차로가 혼재되어 있는 자료에서의 과분산계수의 이질성 등이 대표적이다[1-6].

이처럼 HNB모형은 자료 내에 내재되어 있는 이질적 과분산계수를 다양한 설명변수의 함수로 구성할 수 있으며, 적용 또한 어렵지 않아 기대 교통사고 건수 추정의 오차를 줄일 수 있는 방법이다. 구체적으로 HNB모형을 통해 추정되는 이질적 과분산계수 ϕ 는 NB모형을 통해 관측할 수 없었던 다양한 형태의 감마 사전분포의 모수 추정이 가능하게 한다.

3. 기대 교통사고건수 추정오차 분석

본 장에서는 기대 교통사고건수 추정 시 과분산계수의 이질성(heterogeneous dispersion)이 교통안전 개선사업의 평가결과에 미치는 영향을 비교·분석하기 위해 수행된 모의실험 시나리오 및 분석절차와 분석결과를 제시하였다. 구체적으로 모의실험 결과 분석은 이질적 과분산계수가 내재된 교통사고 자료 분석 시 두 가지 사후평균 추정방법(NB모형, HNB모형)에 따른 SPF의 추정 오차 및 사전분포의 형태 변화와 관측교통사고건수와 두 가지 사후평균 추정방법 적용 시 발생하는 기대 교통사고건수 추정오차 및 교통사고감소계수의 오추정 특징의 상세 비교를 통해 수행하였다.

3.1 분석 시나리오 및 분석절차

본 연구에서는 모의실험을 통해 관측교통사고건수와 경험적 베이저안 추정치의 추정오차를 비교 분석한 선행연구[10,11]의 모의실험 절차를 이질적 과분산계수가 내재된 교통사고 자료 분석 시 사용할 수 있는 여러 기대 교통사고건수 추정방법의 오차특성 분석에 적합한 형태로 수정하였다. 구체적으로 본 연구에서는 포아송 분포를 따르는 교통사고건수의 사전분포를 평균 $E[\lambda_i]$ 가 10개 수준(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)이고 과분산계수(ϕ)를 2개 수준(0.1, 10)으로 설정하여 각 사전분포를 따르는 10,000개 지점의 교통사고건수를 모의 발생시켰다.

따라서 동일한 사전평균 $E[\lambda_i]$ 를 갖는 총 20,000개의 관측치는 불확실성이 낮은 사전분포(low heterogeneity)

에서 발생된 10,000개의 관측치와 불확실성이 높은 사전분포(high heterogeneity)에서 발생된 10,000개의 관측치로 구성되며, 본 연구에서는 각 감마 사전분포를 따르는 포아송 교통사고건수의 분포발생 오차를 안정화시키기 위해 30회의 반복수행을 실시하였다(각 모의실험 별 관측치=600,000개). 한편 본 연구에서는 불확실성이 높은 사전분포에서 모의 발생된 관측치에 지시변수 값 '1'을 부여하여 HNB모형 기반 SPF를 추정한 후 결과분석을 수행하였으며, NB모형은 동일한 사전평균 $E[\lambda_i]$ 를 갖는 관측치의 과분산계수가 동일하다는 가정하에 SPF를 추정한 후 결과분석을 수행하였고 전술한 모의실험 절차는 아래와 같이 요약할 수 있다.

- Step 0 : 10,000개의 교통사고 발생지점 발생 ($i=1, 2, \dots, 10,000$)
- Step 1: 각 지점의 기대교통사고건수 (λ_i) 발생, $\lambda_i \sim \text{Gamma}(E[\lambda_i], E[\lambda_i]^2 \phi)$
- Step 2: 각 지점의 관측 교통사고건수(y_i)발생, $y_i \sim \text{Poisson}(\lambda_i)$
- Step 3: 이질적 관측교통사고건수 통합 및 불확실성 지시변수 생성 ($i=1, 2, \dots, 20,000$)
- Step 4: 관측교통사고건수를 이용하여 SPF 추정 (NB, HNB 모형)
- Step 5: 추정모형 별 사후평균 추정
- Step 6: 추정오차(평균제곱오차) 계산
- Step 7: Step 0부터 Step 6까지 반복(반복횟수=30회)

3.2 분석결과

3.2.1 교통안전성능함수

Table 1은 각 모의실험 시나리오 별 교통안전성능함수(SPF)의 추정결과로, 추정결과는 반복실험에 따른 평균함수와 분산함수의 계수에 대한 점추정치와 그 점추정치가 모수와 같다는 영가설을 검정한 결과를 포함하고 있다. 교통안전 성능함수 추정 결과 평균함수 계수값은 분산계수 추정 시 적용된 모형에 관계없이 모수와 동일한 것으로 나타났다.

이러한 분석결과는 선행연구[6]에서 제시한 바와 같이 분산계수의 이질성을 NB모형으로 오설정했다 할지라도 평균함수의 값은 영향을 크게 받지 않는다는 것을 의미한다. 그러나 분산계수의 값은 내재되어 있는 이질성

[Table 1] Safety Performance Function Estimates by Model

E(λ)	NB model estimates			HNB model		
	Mean : ln(λ)	Dispersion : ln(ϕ)		Mean : ln(λ)	Dispersion : ln(ϕ)	
	Constant	Constant	Group indicator	Constant	Constant	Group indicator***
1.0	-0.0012* (0.4910)**	-0.6701 (<0.001)	-	-0.0005 (0.7149)	-2.2923 (0.9797)	2.2929 (0.9768)
2.0	0.6935 (0.8294)	-0.6872 (<0.001)	-	0.6933 (0.9298)	-2.3101 (0.4535)	2.3210 (0.1614)
3.0	1.0962 (0.0884)	-0.6986 (<0.001)	-	1.0965 (0.1031)	-2.2965 (0.7407)	2.2986 (0.8936)
4.0	1.3858 (0.6195)	-0.6979 (<0.001)	-	1.3862 (0.9315)	-2.3085 (0.4493)	2.3145 (0.1720)
5.0	1.6093 (0.8630)	-0.6977 (<0.001)	-	1.6088 (0.4648)	-2.3009 (0.9197)	2.3023 (0.9433)
6.0	1.7913 (0.6338)	-0.6918 (<0.001)	-	1.7920 (0.7989)	-2.2987 (0.6845)	2.3029 (0.8870)
7.0	1.9461 (0.8028)	-0.6947 (<0.001)	-	1.9461 (0.8867)	-2.3111 (0.1259)	2.3066 (0.4856)
8.0	2.0801 (0.5853)	-0.6812 (<0.001)	-	2.0794 (0.9477)	-2.2988 (0.6530)	2.3024 (0.9348)
9.0	2.1966 (0.5447)	-0.6794 (<0.001)	-	2.1970 (0.8255)	-2.3107 (0.0996)	2.3150 (0.1710)
10.0	2.3026 (0.9981)	-0.6700 (<0.001)	-	2.3020 (0.2634)	-2.3016 (0.9242)	2.3082 (0.3422)

* Mean of coefficient parameters

** p-values are in parentheses (H0:parameter estimate of each model=true coefficient)

*** Prior uncertainty indicator (low heterogeneity=0, high heterogeneity=1)

(heterogeneity)을 HNB모형을 이용해 추정할 경우에만 그 값이 모수와 동일하게 나타나, 분산함수의 오설정은 분산계수 추정에는 큰 영향을 미치고 있음을 알 수 있다 (NB모형 적용 시 평균함수 추정치는 모수와 동일하다는 영가설을 기각하지 않으나, 분산함수 추정치는 모수와 동일하다는 영가설을 기각함; 유의수준=0.05).

3.2.2 사전분포

분산함수의 오설정은 분산계수의 추정치에 영향을 미쳐 기대 교통사고건수의 추정량인 사후평균을 구성하는 감마 사전분포(Gamma prior distribution)의 형태에도 큰 변화를 가져온다. Table 2는 분산계수 추정 모형(NB, HNB모형)에 따른 감마 사전분포의 평균과 분산의 추정치를 나타내는데, 분산함수의 이질성이 존재할 경우 NB 모형을 적용한 분산함수의 오설정은 기대교통사고건수의 감마사전분포의 분산을 크게 변화시킴을 알 수 있다. 이는 감마 사전분포의 평균은 NB 또는 HNB모형의 평균함수 계수와 같으나 ($E[\lambda_i]=E[y_i]$), 감마 사전분포의 분산은 NB 또는 HNB모형을 통해 추정된 분산계수 ϕ 의 함수로 구성되기 때문이다($V[\lambda_i]=E[y_i]^2\phi$). 예를 들어 기대 교통사고건수의 평균 $E[\lambda_i]=5.0$ 일 때, 추정되는 감마 사전분포의 평균 $E[\lambda_i]$ 은 NB 또는 HNB모형의 평균 (e.g., NB : $E[\lambda_i]=\exp(1.6093) \approx 5.0$, HNB : $E[\lambda_i] = \exp(1.6088)$

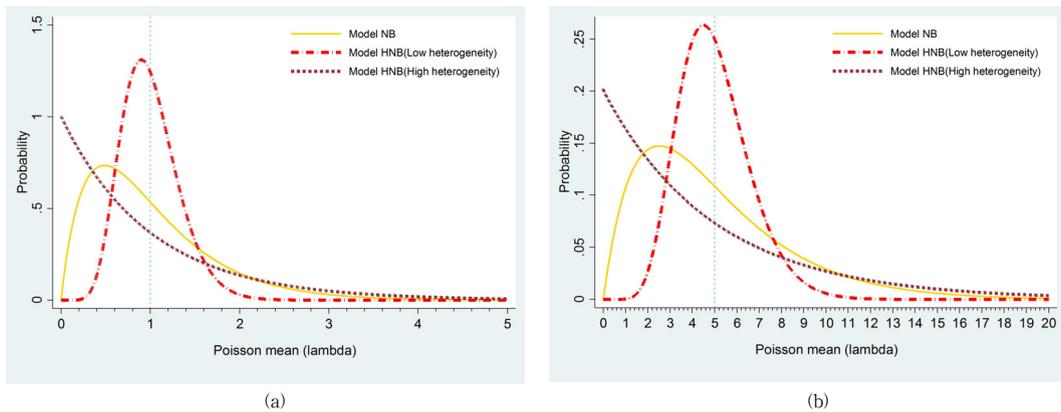
≈ 5.0)값을 보여 적용된 모형에 관계없이 유사한 값을 가지나, 사전분포의 분산은 적용모형의 따라 각기 다른 값이 추정된다(e.g., NB : $V[\lambda_i]$

$$= \{\exp(1.6093)\}^2 \times \exp(-0.6977) = 12.4398.$$

Fig. 1은 $E[\lambda_i]$ 가 각각 1과 5일 때, 이질적 분산계수를 무시하고 NB 모형을 이용해 추정된 사전분포와 이질적 분산계수를 HNB모형을 이용해 추정된 사전분포 형태의 차이를 보여준다. 각 모형을 적용하여 추정된 사전분포를 비교해보면, NB모형과 HNB모형을 적용하여 추정된 사전분포의 형태는 다르나, 각 사전분포의 평균은 모의실험 지정값인 ' $E[\lambda_i]=1$ '과 ' $E[\lambda_i]=5$ '로 같음을 알 수 있다. 그러나 사전분포의 분산을 살펴보면 HNB 모형을 적용하여 추정된 사전분포 중 $\phi=0.1$ (low heterogeneity)인 포아송 평균은 평균 $E[\lambda_i]$ 을 중심으로 밀집되어 있는 반면, $\phi=1.0$ (high heterogeneity)인 포아송 평균은 긴 오른쪽 꼬리분포를 보이고 있어 모의실험 시 사용된 가정인 ϕ 값에 따른 사전분포의 분산은 HNB 모형 적용 시에만 올바르게 추정되는 것으로 나타났다. 즉 이질적인 분산계수를 무시하고 NB 모형을 이용해 사전분포의 분산을 추정했을 경우에는 추정된 사전분포의 분산은 각기 다른 이질적 사전분포의 분산값의 사이에 위치하여 NB 모형을 이용해 추정된 사전분포의 모양은 $\phi=0.1$ 인 사전분포보다 오른쪽으로 치우친 형태를 띤다.

[Table 2] Gamma prior estimates by model

E(λ)	NB model estimates		HNB model estimates			
	$\hat{E}(\lambda)$	$\hat{V}(\lambda)$	Low heterogeneity		High heterogeneity	
			$\hat{E}(\lambda)$	$\hat{V}(\lambda)$	$\hat{E}(\lambda)$	$\hat{V}(\lambda)$
1.0	0.9988	0.5104	0.9995	0.1009	0.9995	0.9996
2.0	2.0007	2.0133	2.0003	0.3971	2.0003	4.0451
3.0	2.9928	4.4540	2.9937	0.9017	2.9937	8.9809
4.0	3.9980	7.9542	3.9996	1.5903	3.9996	16.0933
5.0	4.9993	12.4398	4.9968	2.5010	4.9968	25.0031
6.0	5.9972	18.0077	6.0014	3.6158	6.0014	36.1689
7.0	7.0013	24.4713	7.0013	4.8603	7.0013	48.7985
8.0	8.0053	32.4273	7.9997	6.4237	7.9997	64.2255
9.0	8.9944	41.0093	8.9980	8.0309	8.9980	81.3125
10.0	10.0001	51.1724	9.9942	9.9981	9.9942	100.5445



[Fig. 1] Example of Gamma prior distribution by Model
(a) $E[\lambda_i]=1$ (b) $E[\lambda_i]=5$

3.2.3 추정오차 및 CRF

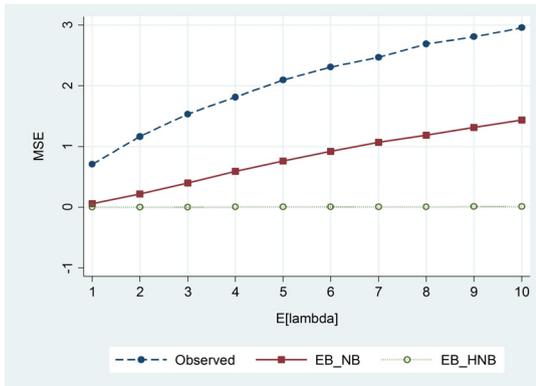
본 소절에서는 NB모형 및 HNB모형을 적용해 추정된 경험적 베이스 추정치 및 관측교통사고건수 y_i 의 추정오차를 비교분석하고 각 추정치를 사용하여 제시되는 교통사고감소계수(CRF, crash reduction factor)의 변화를 제시하였다.

Table 3은 모의실험결과를 집계하여 정리한 관측치 (y_i) 및 NB모형과 HNB모형을 이용해 추정된 경험적 베이스 추정치와 사후평균간 오차에 대한 평균제곱오차 (MSE, mean squared error)를 나타낸다. 각 추정치의 MSE를 살펴보면 모든 추정치는 $E(\lambda)$ 의 값이 커질수록 증가하는데, 이는 선행연구[10,11]에서 밝힌 바와 같이 사전평균 $E(\lambda)$ 의 값이 커질수록 사전분산 $V[\lambda]$ 가 증가하기 때문에 발생하는 현상이다. 그러나, 사전평균이 동일한 분포에서 추정된 기대 교통사고건수의 추정치를 비교해보면, 이질적 분산계수를 고려한 HNB모형 기반 경험

적 베이스 추정치의 MSE가 3가지 추정치 중 가장 낮을 뿐 아니라 그 값은 거의 '0'에 근접한 것으로 나타났다 (Fig. 2 참조).

[Table 3] MSEs of expected crashes by Model

E(λ)	Observed Value : y	EB estimate : $E[\lambda y]$	
		NB model	HNB model
1.0	0.7048	0.0597	0.0001
2.0	1.1665	0.2165	0.0003
3.0	1.5338	0.3992	0.0012
4.0	1.8145	0.5887	0.0018
5.0	2.0962	0.7617	0.0021
6.0	2.3102	0.9217	0.0032
7.0	2.4678	1.0667	0.0049
8.0	2.6905	1.1890	0.0067
9.0	2.8078	1.3116	0.0077
10.0	2.9573	1.4373	0.0119



[Fig. 2] MSEs by Model

한편 이질적 분산계수를 무시한 NB모형 기반 경험적 베이스 추정치의 MSE는 관측치의 MSE에 비해 낮으나 HNB 모형 기반 경험적 베이스 추정치의 MSE보다는 큰 것으로 나타났다. 이러한 분석결과는 분석자료가 다양한 지점 또는 구간 특성으로 인해 이질적인 분포형태를 띠며 과분산계수를 HNB모형을 이용하여 올바르게 추정하지 않고 통상적인 NB모형만을 이용해 추정할 경우에는 EB 추정오차는 증가할 수 있음을 보여준다.

전술한 각 추정치의 추정오차인 MSE는 그 값이 크지 않아 무시할만한 수준으로 보여지나 사후평균 ($E[\lambda|y]$)의 구조 상 기대교통사고건수의 추정오차는 관측치 y 의 값과 분산계수의 값에 따라서도 달라지는 바 본 연구에서는 교통안전개선사업의 효과를 비율로 추정한 값인 교통사고감소계수(CRF)의 값이 각 추정방법에 따라 어떤 형태로 변화되는지를 살펴보았다. CRF는 교통안전개선사업 전과 후의 기대 교통사고건수(expected crashes)인 B와 A를 이용해 아래식과 같이 추정되며, 해당 교통안전개선사업이 교통사고건수를 감소시켰을 경우 CRF의 값은 양의 값을 갖는다.

$$CRF(\%) = \left(1 - \frac{A}{B}\right) \times 100 \quad \text{Eq. 4}$$

본 연구에서는 각 추정치가 CRF의 크기 및 부호에 주는 영향을 직관적으로 분석하기 위해, 기대 교통사고건수에 아무런 영향을 주지 않는 교통안전개선사업(ineffective countermeasure)의 CRF를 전술한 3가지의 추정치(관측치, NB모형을 이용한 경험적 베이스 추정치, HNB모형을 이용한 경험적 베이스 추정치)를 이용하여 추정하고 이를 비교 분석하였다. CRF 추정 시 교통

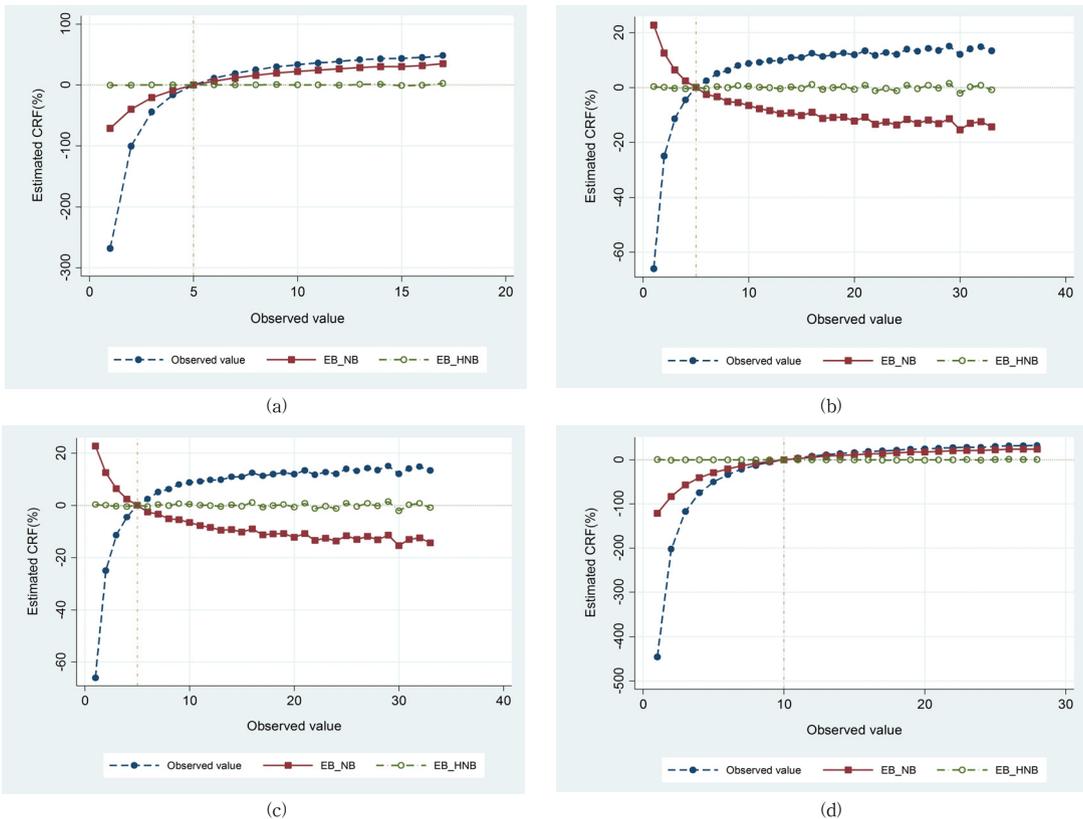
안전 개선사업 후의 기대 교통사고건수 A는 모의실험 과정에서 정의한 기대 교통사고건수의 참값을 사용했으며, 교통안전 개선사업 전의 기대교통사고 건수 B는 각 추정치를 적용하였다. Table 4는 각 추정치를 이용한 CRF 추정결과에 대한 요약 통계치를 나타내는데, 관측치와 NB모형 기반 경험적 베이스 추정치는 CRF의 참값이 '0'인 교통안전개선사업의 효과를 심각하게 오추정하고 있는 반면, 이질적 분산계수를 고려한 HNB모형 기반 경험적 베이스 추정치를 이용한 CRF는 '0'에 근접한 값을 보이고 있어 기설정된 CRF의 값을 적절하게 추정하고 있음을 알 수 있다 (구체적으로 CRF의 참값은 '0'임에도 관측치를 이용해 추정한 CRF는 최대 445.78%의 교통사고 증가 효과를, NB모형 기반 CRF는 최대 120.87%의 교통사고 증가 효과를 보이고 있다).

[Table 4] Summary Statistics for CRFs

Variable	Observed value : y	EB Estimate	
		NB Model	HNB Model
Mean	2.5925	-2.653	-0.0054
Std.dev.	50.7711	20.291	0.7489
Min	-445.7815	-120.8683	-3.0165
P25	4.0513	-10.0658	-0.3194
P50	8.6623	-5.7644	0.0058
P75	19.0174	8.2889	0.3032
Max	75.7682	45.9651	6.3028

Note: True CRF=0

관측치의 크기에 따라 달라지는 추정방법 별 CRF의 크기 및 부호의 특징을 살펴보기 위해 본 연구에서는 사전평균이 각각 5와 10인 모의실험 결과로 도출된 CRF의 특징을 Fig. 3에 제시하였다. 먼저 추정방법에 따른 CRF의 변화를 살펴보면, HNB모형 기반 CRF의 값은 관측치의 크기에 관계없이 '0'의 값에 수렴하고 있음을 알 수 있으나 그 외 추정방법을 통해 추정된 CRF의 값은 사후평균에 영향을 주는 관측치가 사전평균에서 멀어질수록 오추정이 심각하게 발생하고 있음을 알 수 있다. 구체적으로 관측교통사고건수만 사용했을 경우 CRF의 값은 관측치가 사전평균보다 작을 때 음(-)의 값을 갖고, 관측치가 사전평균보다 클 때 양(+)의 값을 보이는데 이는 선행연구[9-11]에서 밝힌 바와 같이 관측교통사고건수는 평균으로의 회귀오차를 갖기 때문이다. 한편 NB모형 기반 CRF는 $\phi=0.1$ 일 때 오차 발생 특성이 관측교통사고건수의 오차발생 특성과 유사하게 나타나나, $\phi=1.0$ 일 때 나타나는 오차의 부호는 관측교통사고건수의 오차부호와 정반대로 나타남을 알 수 있다 (예를 들어 $\phi=1.0$ 이고 y_i 의



[Fig. 3] CRF estimates by heterogeneity level and model ($E[I]=5, 10$)

(a) $E[\lambda]=5$ (low heterogeneity) (b) $E[\lambda]=5$ (high heterogeneity) (c) $E[\lambda]=10$ (low heterogeneity) (d) $E[\lambda]=10$ (high heterogeneity)

값이 사전평균보다 클 때, 관측교통사고건수 기반 CRF는 음(-)의 오차의 값을, NB기반 CRF는 양(+)의 오차의 값을 가짐). 이러한 현상은 관측교통사고건수의 추정오차에는 평균으로서의 회귀 오차가 강하게 작용하나, NB모형 기반 추정오차에는 오추정된 과분산계수의 값이 사후평균 추정 시 베이저안 가중치에 각기 다른 방향으로 작용하고 있기 때문이다.

구체적으로 $\phi=0.1$ 일 때 본 연구의 NB모형을 통해 추정된 과분산계수의 값은 참값인 $\phi=0.1$ 보다 큰 값이 추정되고, 이는 사전평균에 대한 베이저안 가중치의 값을 감소시켜 최종 사후평균 추정치를 관측치의 값에 좀 더 가깝게 이동시켜 CRF의 오차방향을 관측치 기반 CRF의 오차방향과 유사하게 만들어낸다. 그에 반해 $\phi=1.0$ 일 때, 본 연구의 NB모형을 통해 추정된 과분산계수의 값은 참값인 $\phi=1.0$ 보다 적은 값이 추정되고, 이는 사후평균 추정 시 사전평균에 대한 베이저안 가중치의 값을 증가시켜 최종 사후평균 추정치는 관측치보다는 사전평균의 값에

좀 더 가깝게 이동시켜 NB모형 기반의 CRF의 오차부호를 관측치의 오차부호와 반대로 나타내게 만든다.

한편 관측교통사고건수가 사전평균에 근접할 때 CRF의 값은 추정방법에 관계없이 그 오차가 가장 적는데, 이는 관측치와 사전평균의 함수로 구성되는 사후평균 추정 시 관측치와 사전평균이 동일하면 모형별로 추정되는 베이저안 가중치의 값에 관계없이 사후평균 추정치는 사전평균값에 수렴하기 때문이다. 이처럼 세 가지 추정방법 중 HNB모형 기반 CRF의 값은 전술한 추정오차 분석결과와 동일하게 분산계수, 관측치, 사전평균의 값에 관계없이 '0'에 근접한 값을 보임을 재확인할 수 있었다.

4. 결론 및 향후 과제

본 연구는 기대 교통사고건수 추정을 위해 사용되는 SPF의 이질적 분산계수의 유의성이 이질적 사전분포에

직접적인 영향을 받는다는 가설을 검증하고, 이질적 사전분포에 대한 모형 오설정이 교통 안전개선 사업의 평가결과에 주는 영향의 특성을 분석하기 위해 수행되었다. 구체적으로 본 연구에서는 이질적 분산계수의 유의성과 이질적 사전분포의 연관성을 검증하기 위해 모의실험을 통해 이질적 사전분포를 발생시킨 후 이를 NB모형과 HNB모형을 이용하여 SPF를 추정하여 이질적 사전분포가 SPF의 평균함수 및 분산함수에 주는 영향을 분석하였다. 또한 추정된 계수추정치를 이용한 사전분포의 초모수 추정치의 오차특성과 이질적 사전분포를 고려하지 않았을 경우 발생하는 교통사고감소계수(CRF)의 오차 부호와 크기를 상세 분석하여 제시하였으며, 전술한 연구과정을 통해 도출된 연구결론은 아래와 같다.

- 분석하고자 하는 교통사고 자료가 이질적 형태의 분산을 내재하고 있을 때(신호교차로 및 비신호교차로 교통사고, 연속교통류 및 단속교통류 교통사고 등으로 분석자료가 구성되어 있을 경우), 이질적 분산계수는 유의한 값을 보이며 이는 이질적 음이항(HNB)모형을 통해 모델링 할 수 있다.
- 교통안전연구에서 널리 사용되는 일반적인 음이항(NB)모형을 적용했을 경우 이질적 분산계수는 오추정되나, 분산함수 오추정은 평균함수 추정치에 편의(bias)를 주지는 않는다.
- 즉 이질적 분산계수의 오추정은 포아송 사전분포의 평균에는 큰 영향을 미치지 않으나 분산의 크기를 변화시켜 궁극적으로는 기대교통사고건수의 추정량인 사후평균의 값에 오차를 발생시킨다(구체적으로 이질적 분산함수를 NB모형으로 오설정할 경우 CRF의 값은 참값에 비해 최대 120%의 오차를 발생시킴).
- 따라서 분석하고자 하는 교통사고 자료가 매우 이질적인 특성을 갖는 개체들로 구성되어 있을 경우 HNB모형을 이용하여 이질적 분산함수의 유의성을 검토하여 교통안전 개선사업의 평가 등을 수행해야 한다.

본 연구는 교통안전 연구 수행 시 사용할 교통사고자료가 매우 다양한 특성을 갖는 개체들에서 수집되었을 경우, SPF의 평균함수 뿐 아니라 분산함수 또한 각 개체의 다양한 특성을 적절히 반영할 수 있도록 HNB모형을

통해 구축되어야 함을 나타낸다. 본 연구는 그간 국외에서 주로 제시되어 왔던 HNB모형의 적용 타당성과 오차의 특성을 모의실험을 밝혔다다는 점에서 선행연구와 차별성을 갖는다. 그러나 본 연구의 결론은 모의실험을 통해 도출됐다는 한계점을 갖고 있어 향후 오차특성 분석연구는 국내의 여러 실증자료를 통해 좀 더 다양한 조건 하에서 나타나는 이질적 분산함수의 오설정의 효과가 분석되어야 할 것이다. 또한 여러 실증자료 분석을 통해 도출할 수 있는 국내의 분산함수 형태도 향후 지속적으로 논의 되어야 할 것이다.

References

- [1] B. G. Heydecker, J. Wu, "Identification of Sites for Road Accident Remedial Work by Bayesian Statistical Methods: An Example of Uncertain Inference", *Advances in Engineering Software*, Vol. 32, No. 10-11, pp. 859-869, 2001.
DOI: [http://dx.doi.org/10.1016/S0965-9978\(01\)00037-0](http://dx.doi.org/10.1016/S0965-9978(01)00037-0)
- [2] K. El-Basyouny, T. Sayed, "Comparison of Two Negative Binomial Regression Techniques in developing Accident Prediction Models", *Transportation Research Record 1950*, pp. 9-16, 2006.
DOI: <http://dx.doi.org/10.3141/1950-02>
- [3] S. Mitra, S. P. Washington, "On the Nature of Over-Dispersion in Motor Vehicle Crash Prediction Models", *Accident Analysis and Prevention*, Vol. 39, No. 3, pp. 459-468, 2007.
DOI: <http://dx.doi.org/10.1016/j.aap.2006.08.002>
- [4] S. R. Geedipally, D. Lord, "Effects of the Varying Dispersion Parameter of Poisson- gamma models on the Estimation of Confidence Intervals of Crash Prediction Model", *Transportation Research Board 87th Annual Meeting*, Washington, D.C., 2008.
- [5] D. Lord, P. Y. Park, "Investigating the Effects of the Fixed and Varying Dispersion Parameters of Poisson-gamma Models on Empirical Bayes Estimates", *Accident Analysis and Prevention*, Vol. 40, No. 4, pp. 1441-1457, 2008.
DOI: <http://dx.doi.org/10.1016/j.aap.2008.03.014>
- [6] K. Shin, K. Choi, "Empirical Bayes Method in the Study of Traffic Safety via Heterogeneous Negative Binomial Model", *Conference of Korean Society of Civil Engineers*, Korean Society of Civil Engineers, pp. 54-57, 2009.
- [7] E. Hauer, "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, 1980.
DOI: [http://dx.doi.org/10.1016/0001-4575\(80\)90049-4](http://dx.doi.org/10.1016/0001-4575(80)90049-4)
- [8] E. Hauer, “Selection for Treatment as a Source of Bias in Before-and-After Studies,” *Traffic Engineering and Control*, Vol. 21, pp. 419-421, 1980.
- [9] Hauer E. *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Pergamon, 1997.
- [10] K. Shin, “Comparative Study on the Estimation Methods of Traffic Crashes : Empirical Bayes Estimate vs. Observed Crash”, *Journal of Korean Society of Civil Engineers*, Vol. 30, No. 5, pp. 453-459, 2010.
- [11] K. Shin, “Assessing Estimation Methods of the Expected Crashes using Panel Traffic Crash Data”, *Journal of Korean Society of Transportation*, Vol. 29, No. 1, pp. 103-111, 2011.
- [12] Hilbe, J. M. *Negative Binomial Regression*. Cambridge University Press, 2007.
DOI: <http://dx.doi.org/10.1017/CBO9780511811852>

신 강 원(Kangwon Shin)

[정회원]



- 2008년 12월 : 애리조나 주립대 토목환경공학과 (교통공학박사)
- 2009년 3월 ~ 현재 : 경성대학교 도시공학과 조교수

<관심분야>

TOD 교통계획, 교통계량분석, 교통환경정책