□ 論　　文 □

重力模型의 適合度 檢證

Testing Goodness of Fit of Gravity Models

金 亨 鎭
(국토개발연구원 교통연구실)

目　　次

ABSTRACT

This paper is concerned with assessing goodness of fit of gravity models. The Chi-square test, or one of its asymptotic equivalents, is usually recommended for the purpose. A difficulty that frequently arises, particularly when working with urban travel data, is that the expected number of trips for most origin-destination(O-D) pairs are small. In order to test goodness of fit of gravity model, a simple approach, which depends on the number of O-D pairs and certain trip totals being large, is proposed in this paper. In addition, derivation of variance of Chi-square ratio is proposed to test the confidence interval of Chi-square ratio and application of its results with simulated data set is made to verify the usefulness of the results.

Ⅰ. 서론

교통계획의 수요예측 단계에서 계획목표년도의 통행분포(trip distribution) 상태를 나타내는 기 종점 통행표(O-D trip matrix)는 장래 교통죤(traffic zone)간 토지이용의 유기적인 연관성을 나타낼 뿐만 아니라, 교통수요예측의 최종단계인 통행배분(network assignment) 분석을 위해서도 매우 중요한 데이터(data)이다. 이러한 장래 기 종점통행표는 일반적으로 현재의 통행특성이 장래에도 계속되리라는 가정 아래, 수집된 현재의 통행분포 데이터와 통행분포모형(trip distribution models)을 통하여 구해진다. 여기서 수집되는 데이터는 대상지역내의 모든 사람들에 대한 통행실태를 전수조사하는 것은 현실적으로 불가능하고 또한 많은 비용이 소요되므로 일반적으로 적정 비율의 표본조사를 통하여 표본기 종점통행표(sample O-D trip matrix)를 작성하여 얻어진다. 이렇게 구해진 표본기 종점통행표는 전수화(factoring)과정을 거치고 필요에 따라 보정을 통하여 목표년도의 기 종점통행표로 이용된다.

장래 통행분포 예측을 위한 모형으로는 중력모형(gravity model)이 많이 사용되고 있다. 중력모형의 파라메타(parameter)추정을 위하여 最尤推定方法(Maximum Likelihood Estimation Procedure)을 사용하는 커다란 이점은 거대관측량의 특성(large sample property)에 의한 파라메타 추정치가 갖는 일관성(consistency), 효율성(efficiency) 및 유용성(robustness)이다. 그러나 실제 적용에서 빈번히 발생되는 문제점은 표본수에 비하여 분석하고자 하는 기 종점(교통죤)의 수가 대부분 많으므로 표본 기 종점통행표 각 셀(cell)의 값이 너무 작거나 대부분의 셀 값이 0인 경우(very sparse O-D trip matrix)이다. 그러나 最尤推定方法에서는 충분통계량(sufficient statistics)을 이용하므로 이러한 문제점을 극복할 수 있으며, 몇몇 연구에서 모수추정치의 분산(variance)을 분석하거나 모형의 적합도를 검증함으로서 이를 입증하고 있다(Weber and Sen, 1985, Sen, 1986, 1991).

중력모형의 적합도 검증(goodness of fit test) 방법으로는 카이제곱검증(Chi-square test)방법이 널리 사용되고 있으며, 카이제곱비율(Chi-square ratio)을 적합도 검증의 기준으로 하고 있다. 그러나 특히 관측통행량이 작은 경우 모형의 적합도 검증을 위하여 카이제곱비율을 이용할 때 이의 유의수준에 대하여 고려하여야 한다. 그러나 지금까지 카이제곱비율의 유의수준을 분석하기 위한 카이제곱비율의 분산에 대한 연구는 거의 이루어지지 않고 있는 실정이다. 따라서 본 논문에서는 카이제곱비율의 유의수준을 분석하기 위하여 카이제곱비율의 분산식을 유도하였으며, 씨뮬레이션한 통행분포 데이터를 이용하여 통행분포를 추정하고, 추정치의 적합도 검증을 위하여 카이제곱비율을 산정 하였으며, 카이제곱비율의 유의수준을 분석하고 이의 유용성을 입증하였다.

Ⅱ. 중력모형

1. 중력모형

N_{ij}를 통행기점 $i \in I$ 와 통행종점 $j \in J$ 사이의 관측통행량(예; 사람통행, 차량통행, 등)으로 정의하고, $c_{ij} = (c_{ij}^{(1)}, \cdots c_{ij}^{(K)})$를 기점 i 와 종점 j 사이의 통행저항변수(통행시간, 통행비용, 통행거리, 등)를 나타내는 $c_{ij}^{(k)}$ 의 벡터(vector)라고 정의한다면, 일반화중력모형(Generalized Gravity Model)은 아래 식 (1)과 같이 정의된다.

$$T_{ij} = E[N_{ij}] = A(i)\ B(j)\ F(c_{ij})\ \forall\ I \in I,\ j \in J \quad (1)$$

식 (1)에서 A(i)는 기점 i 의 통행방출 정도를 나타내는 기점인자(origin factor)이며, B(j)는 종점의 통행흡인력 정도를 나타내는 종점인자(destination factor)이다. $F(c_{ij})$는 기 종점간의 통행저항인자(separation factor)이며, T_{ij}는 관측통행량 N_{ij}의 기대치(expectation)이다. 위의 중력모형에서 관측통행량 N_{ij}는 확률변수(random variable)로 어떤 분포(distribution)를 갖는다고 정의할 수 있으며, 만약 N_{ij}가 독립포아송분포(independent Poisson distribution)를 갖는다면 위의 모형은 포아송중력모형(Poisson Gravity Model)이라고 정의할 수 있다(Smith, 1987).

위의 중력모형을 실제로 적용하기 위하여 $F(c_{ij})$를 다시 정의 할 필요가 있으며, $F(c_{ij})$는 아래 식(2)와 같이 정의될 수 있다.

$$F(c_{ij}) = \exp[\theta' c_{ij}] = \exp[\sum_{k=1}^{K} \theta_k c_{ij}^{(k)}] \qquad (2)$$

여기서 $\theta = (\theta_1, \theta_2, \cdots, \theta_k, \cdots, \theta_K)'$이며, θ_k는 $k = 1, 2, 3, \cdots, K$ 에 대하여 관측통행 데이터로부터 구하여지는 파라메타이다. 위의 식 (2)에서 통행저항함수인 $F(c_{ij})$는 통행분포와 관련이 있는 통행비용변수들 c_{ij} 의 합의 모양으로 정의되고 있으므로 다양한 통행비용을 모형에서 적용할 수 있는 특성이 있다. 일반적으로 c_{ij}가 통행시간이나 통행비용(out-of-pocket cost)일 때 통행분포량(N_{ij})과는 반비례 관계이나, 통근통행분포의 경우 c_{ij}가 통근자의 居住地와 職場간의 職種適合度(occupational match)를 나타낼 경우 통근통행분포량과는 정비례 관계가 성립한다(Soot and Sen, 1991).

전통적인 중력모형에서 A(i) 와 B(j)는 기 종점의 인구, 직업수, 개발연상면적 등의 기 종점 통행유출 및 유입 변수와 직접적인 관계가 있다고 정의되고 있으나, 위의 중력모형에서 A(i) 와 B(j)는 전통적인 모형과는 달리 $F(c_{ij})$의 파라메타 θ와 함께 관측통행 데이터로부터 추정하여 구하여지는 파라메타이다.

2. 파라메타 추정(Parameter Estimation)

일반적으로 파라메타 추정(parameter estimation)은 기준년도의 관측통행 관련 데이터로부터 구하여진다. 위의 모형 (1)에서 파라메타를 추정하기 위하여 필요한 기준년도의 관측데이터는 N_{ij} 와 c_{ij} 이며, 이러한 데이터가 구하여진다면 아래 (3)과 같은 모형 (1)의 파라메타 벡터 ζ를 구하기 위한 몇 가지 방법이 가능하다.

$$\zeta = (A(1), A(2), \cdots, A(I), B(1), B(2), \cdots, B(J), \theta_1, \theta_2, \cdots, \theta_K)' \qquad (3)$$

그중 가장 널리 사용되고 있으며 그 통계학적 특성이 인정되고 있는 방법은 最尤推定方法이다. 중력모형의 파라메타 추정에서 最尤推定方法은 관측통행량이 작음에도 불구하고 偏倚(bias)가 매우 작은 파라메타 추정방법이다(Sen, 1991).

$A = (A(1), A(2), \cdots, A(I))'$, $B = (B(1), B(2), \cdots, B(J))'$ 그리고 θ를 포함한 일련의 벡터를 (A, B, θ)로 정의할 때 (A, B, θ)의 最尤推定値는 아래 식 (4)~(6)의 방정식의 해를 구함으로서 얻어질 수 있다(Sen, 1991).

$$\sum_i T_{ij} = \sum_i N_{ij},\ (\forall j \in J) \qquad (4)$$

$$\sum_j T_{ij} = \sum_j N_{ij},\ (\forall\ i \in I) \qquad (5)$$

$$\sum_{ij} c_{ij}^{(k)} T_{ij} = \sum_{ij} c_{ij}^{(k)} N_{ij},\ (\forall\ k \in K) \qquad (6)$$

最尤推定方法이 모형의 파라메타 추정에 널

리 사용되는 커다란 이점은 거대표본에 의한 특성(large sample property)에서 기인한다. Rao(1973)는 N_{ij}가 다항분포(multinomial distribution)일 때 最尤推定値의 일관성(consistency), 효율성(efficiency) 및 근사적 정규분포성(asymptotic normality)을 증명하였으며, 또한 이러한 특성은 N_{ij}가 포아송분포일 때도 변함없이 적용됨을 증명하였다. 그러나 이러한 결과들은 近似的인 결과치(asymptotic results)이며, 最尤推定方法을 적용하기 위한 거대표본을 만족하기에는 각각의 관측통행량(N_{ij})의 크기가 실재로 매우 작다. 따라서 最尤推定方法을 중력모형에 적용할 경우 偏倚한 추정치(biased estimates)를 얻게 된다. 그러나 그 偏倚는 매우 작을 것으로 추정되며, 이러한 추정의 이론적인 근거는 포아송분포나 다항분포를 갖는 N_{ij} 가정하의 지수형중력모형(exponential gravity model)의 最尤推定方法에 의한 파라메타 추정시 충분통계량(sufficient statistics)이 존재하기 때문이다. 여기서 정의하는 충분통계량은 기점총통행량(origin trip total) $\Sigma_j(N_{ij})$, 종점총통행량(destination trip total) $\Sigma_i(N_{ij})$ 및 총통행비용(total cost) $\Sigma_{ij}(c_{ij}^{(k)}N_{ij})$이며, 일반적으로 기 종점 총통행량 및 총통행비용은 각각의 통행량 N_{ij} 와 통행비용 $c_{ij}^{(k)}N_{ij}$ 보다 매우 큰 값을 가진다. 위의 식(4)~(6)에서 알 수 있듯이 이러한 충분통계량이 最尤推定方法 에 의한 중력모형의 파라메타 추정을 위하여 사용되고 있으므로 실제로 추정치의 偏倚는 무시할 수 있을 정도로 작을 것으로 예측된다.

Sen(1991)은 몬테카를로(Monte Carlo)방법을 이용하여 이론적으로 관측통행량의 크기가 매우 작은 경우의 O-D 통행메트릭스를 구축하였으며, 最尤推定方法을 사용하여 중력모형의 파라메타를 추정하였다. 그는 파라메타 추정치의 분산(variance)을 분석하여 관측통행량의 크기가 매우 작음에도 불구하고 파라메타 추정치의 偏倚가 매우 작음을 실증적으로 입증하고 있으며, 또한 모형의 적합도 검증결과 포아송중력모형이 통행분포 행태를 잘 반영하고 있음을 보여주고 있다.

Ⅲ. 모형의 적합도 검증

중력모형의 적합도 검증방법의 선택은 확률변수 N_{ij}의 분포에 따라서 결정되며, 본 논문에서는 N_{ij}가 독립포아송분포를 갖는다고 가정하였으며, 이에 따라 포아송중력모형의 적합도 검증방법에 관하여 논하고자 한다.

가장 널리 사용되고 있는 포아송중력모형의 적합도 검증방법은 아래 식 (7)과 같은 카이제곱 검증방법이다.

$$X^2 = \sum_{ij} \left(\frac{(N_{ij} - \hat{T}_{ij})^2}{\hat{T}_{ij}} \right) \tag{7}$$

식 (7)에서, $\hat{T}_{ij}$는 위의 모형 (1)의 파라메타를 추정한 후 구해진 T_{ij}의 추정치이며, T_{ij}가 클 때 X^2는 카이제곱분포(Chi-square distribution)를 갖는다(Madansky, 1988). 그러나 모형의 적합도 검증시 카이제곱 검증방법이나 이와 유사한 검증방법(Bishop, 1975)의 적용할 때의 문제점은 앞에서 언급한 바와 같이 대부분의 T_{ij} 가 매우 작은 크기의 값을 가질 수 있다는 데에서 기인한다. 일반적으로 총 관측통행량($\Sigma_{ij}N_{ij}$)은 충분히 큰 값을 가지나, 기 종점통행표에서 총 관측통행량이 각각의 O-D 셀에 분산되어지므로 각 셀의 N_{ij} 는 매우 작은 값을 가진다. 따라서 충분한 총 관측통행량과 많은 수의 O-D 셀에 의한 규칙성과 중심극한정리(central limit theorem)에 의거하여 실제와 유사한 통행분포 추정

결과를 얻으려는 노력이 시도되고 있다(Sen, 1986, 1991).

만약 충분한 총 관측통행자료가 활용 가능하다면 비록 몇 개의 T_{ij}가 작을지라도 $\tilde{T}_{ij} \approx T_{ij}$가 성립될 수 있다. 예를 들면, $F(c_{ij})$가 지수함수형태(exponential function)이면, T_{ij}를 추정하기 위하여 最尤推定方法을 사용하기 위한 유일한 방법은 다음 식 (8)과 같은 충분통계량을 활용하는 방법이다(Sen, 1986).

$$\sum_i N_{ij},\ \sum_j N_{ij},\ \text{그리고}\ \sum_{ij} (c_{ij}^{(k)}\ N_{ij}) \tag{8}$$

N_{ij}는 기 종점통행표 각 셀의 관측통행량이므로 작은 값을 가질 지라도, 식 (8)의 각각의 항은 기 종점통행표 각 셀의 가로 및 세로의 합(sum), 그리고 총합(total)을 나타내므로 대부분 N_{ij} 보다 매우 큰 값을 가진다(Weber and Sen, 1985). 이러한 조건하에,

$$\boldsymbol{Z}^2 = \sum_{ij} \left(\frac{(N_{ij}-T_{ij})^2}{T_{ij}} \right) \tag{9}$$

이면, $\tilde{T}_{ij} \approx T_{ij}$ 임으로 $\boldsymbol{X}^2 \approx \boldsymbol{Z}^2$ 이다.

식 (1)에서 정의된 바와 같이 $E[N_{ij}]=T_{ij}$ 이고, N_{ij}가 포아송분포를 가질 때 N_{ij}의 분산은 잘 알려진바 와 같이 평균값(mean) T_{ij} 임으로

$$\text{var}[N_{ij}] = E[N_{ij}-T_{ij}]^2 = T_{ij} \tag{10}$$

이며, 식 (10)에 의하여

$$E\left[\frac{(N_{ij}-T_{ij})^2}{T_{ij}}\right] = \frac{E[N_{ij}-T_{ij}]^2}{T_{ij}} = \frac{T_{ij}}{T_{ij}} = 1 \tag{11}$$

임으로, I와 J가 각각 I와 J의 기수 (cardinalities)일 때

$$E[\boldsymbol{Z}^2] = E\left[\sum_{ij}\left(\frac{(N_{ij}-T_{ij})^2}{T_{ij}}\right)\right] = I \times J \tag{12}$$

이다. 따라서 df를 자유도(degree of freedom)라고 정의 할 때 df $\approx I \times J$ 이며, $\boldsymbol{X}^2 \approx \boldsymbol{Z}^2 = I \times J$ 임으로 카이제곱비율(Chi-square ratio) $\boldsymbol{X}^2/\text{df}$ 의 近似期待値(asymptotic expectation)는

$$\boldsymbol{X}^2/\text{df} \approx (I \times J)/(I \times J) = 1 \tag{13}$$

위의 식 (13)에서 나타난 바와 같이, 관측통행분포가 포아송분포를 가질 때 중력모형의 최적 적합도는 카이제곱비율이 이론적인 近似値인 1 에 근접할 때임을 알 수 있다. 그러나 실재의 통행분포는 포아송분포의 이론적인 가정처럼 통행들이 완전히 독립적이 아니므로 카이제곱비율은 이론적인 近似値와 상당한 차이가 있음을 알 수 있으며 몇몇 연구의 중력모형 적합도평가에서 카이제곱비율이 2 보다 작으면 모형의 적합도가 매우 우수하다는 경험적인 사실을 보여주고 있다(Low, 1993).

Ⅳ. 카이제곱 분산식 유도

위의 식 (11)과 같이 $E[(N_{ij}-T_{ij})^2/T_{ij}]=1$이고, 2개의 독립확율변수(independent random variable)의 공분산(covariance)이 0 임으로 $\boldsymbol{Z}^2$의 분산 var[$\boldsymbol{Z}^2$]은 아래 식 (14)와 같이 나타낼 수 있다.

$$\begin{aligned} \text{var}[\boldsymbol{Z}^2] &= E\left[\sum_{ij}\frac{(N_{ij}-T_{ij})^2}{T_{ij}} - 1)\right]^2 \\ &= E\left[\sum_{ij}\frac{(N_{ij}-T_{ij})^2}{T_{ij}} - 1\ \right]^2 \end{aligned}$$

$$= E[\sum_{ij}(\frac{(N_{ij}-T_{ij})^4}{T_{ij}^2} - 2\frac{(N_{ij}-T_{ij})^2}{T_{ij}} + 1)]$$

$$= E[\sum_{ij}(\frac{(N_{ij}-T_{ij})^4}{T_{ij}^2} - 1)]$$

$$= E[\sum_{ij}(\frac{N_{ij}^4 - 4N_{ij}^3T_{ij} + 6N_{ij}^2T_{ij}^2 - 4N_{ij}T_{ij}^3 + T_{ij}^4}{T_{ij}^2} - 1)] \qquad (14)$$

위의 식 (14)를 정리하기 위하여 포아송분포의 1~4次積率(1~4 moment)을 정리하여 보면,

$$E[N_{ij}] = T_{ij},$$
$$E[N_{ij}^2] = T_{ij} + T_{ij}^2,$$
$$E[N_{ij}^3] = T_{ij} + 3T_{ij}^2 + T_{ij}^3,$$
$$E[N_{ij}^4] = T_{ij} + 7T_{ij}^2 + 6T_{ij}^3 + T_{ij}^4 \qquad (15)$$

따라서 식 (14)와 (15)로 부터 var[Z^2]은

$$var[Z^2] = \sum_{ij}[\frac{T_{ij} + 3T_{ij}^2}{T_{ij}^2} - 1]$$
$$= \sum_{ij}[\frac{1}{T_{ij}} + 2] \qquad (16)$$

식 (16)에 의하여 카이제곱비율의 분산 var[Z^2/(IJ)]은 아래 식 (17)과 같이 나타낼 수 있다.

$$var[Z^2/(IJ)] = \sum_{ij}[\frac{1}{T_{ij}(IJ)^2} + \frac{1}{2(IJ)^2}] \qquad (17)$$

위의 식 (17)에서 만약 T_{ij}가 유한값(bounded)을 가지고 I와 J가 비교적 클 때 우측 분모의 값이 매우 큰 값을 가지므로 var[Z^2/(IJ)] → 0 이다. 또한 앞에서 언급한 바와 같이 $\hat{T}_{ij} \approx T_{ij}$ 이고, T_{ij}가 유한값을 가질 때 var[X^2/(IJ)] → 0 이다. 따라서 만약 I와 J가 비교적으로 크고 T_{ij}가 작지 않을 때, 카이제곱비율의 분산은 무시할 정도로 작다. 그러나, T_{ij}가 매우 작고 I와 J가 비교적으로 클 경우, 카이제곱비율의 분산은 고려 대상이 되어야 한다. 여기서 T_{ij} 대신 추정치 $\hat{T}_{ij}$를 사용하여 카이제곱비율을 산정할 수 있으며, 다음장에서는 몇몇 실제 데이터를 사용하여 카이제곱비율의 분산을 고려하는 모형의 적합도 검증 평가방법을 제시하였다.

V. 적용사례

1. 데이터

이 장에서는 몇 가지의 관측데이터로 카이제곱비율의 분산을 분석하기 위하여 앞에서 유도된 식 (17)을 이용하였다. 중력모형의 적합도 검증을 위하여 사용된 데이터는 다음의 2가지 관측데이터를 기반으로 하여 씨뮬레이션(simulation)하여 구축하였다.

1) 스코키 데이터(Skokie data set) : 미국 일리노이주 시카고 북부 스코키 지역의 40개 기점과 18개의 종점 사이에 분포하는 25,247 통행으로 구성된 데이터.
2) 병원목적통행 데이터(Hospital Data set) : 1987년 미국 일리노이주 시카고 지역을 우편번호(Zip Code)에 의하여 250개의 기 종점으로 구분한 후, 250개의 환자출발존과 병원이 위치한 92개의 도착존 간에 분포하는 499,777 통행으로 구성된 데이터.

위의 통행분포자료를 토대로 하여 표본의 크기가 작은 데이터셋(data set)은 아래와 같은 씨뮬레이션방법에 의하여 구축하였다.

1) 중력모형을 이용하여 T_{ij}를 추정한 후, 추정

한 $\hat{T}_{ij}$를 실재의 T_{ij} (true value of T_{ij})로 간주하였다.

2) 위의 T_{ij}를 처음에는 10으로 나누고, 그 다음에는 100으로 나누었다. T_{ij}를 10 및 100으로 나눈 이유는 표본의 크기가 조금 작거나 또한 매우 작은 데이터를 얻기 위함이다.

3) T_{ij}, $T_{ij}/10$, $T_{ij}/100$ 각각의 값을 이용하여 새로운 N_{ij} 데이터셍을 얻기 위하여 포아송난수제조방법(Poisson random number generator)을 사용하였으며, 여기서 언어진 데이터셑을 각각 'N_{ij}', '$N_{ij}/10$', '$N_{ij}/100$'라고 정의하였다.

2. 모형의 적합도 검증 평가

위의 데이터에서 T_{ij}는 이미 알려져 있으므로 N_{ij}를 이용하여 카이제곱비율 $Z^2/[IJ]$을 산정 하였다. 그리고 새로 얻어진 N_{ij} ('N_{ij}', '$N_{ij}/10$', '$N_{ij}/100$')를 사용하여 最尤推定方法에 의하여 새로운 $\hat{T}_{ij}$('$\hat{T}_{ij}$','$\hat{T}_{ij}/10$','$\hat{T}_{ij}/100$')를 구한 후에 카이제곱비율 X^2/df 을 산정 하였으며, 그 결과는 〈표 1〉 및 〈표 2〉와 같다.

〈표1〉 Z^2 의 카이제곱비율(Z^2/df)

데이터 구분	병원 데이터	스코키 데이터
'N_{ij}'	0.98(0.10)	1.17(0.02)
'$N_{ij}/10$'	1.04(0.33)	1.10(0.06)
'$N_{ij}/100$'	0.80(1.07)	1.04(0.21)

주 : 괄호안은 표준편차

〈표2〉 X^2 의 카이제곱비율(X^2/df)

데이터 구분	병원 데이터	스코키 데이터
'$\hat{T}_{ij}$'	0.98(0.08)	1.04(0.01)
'$\hat{T}_{ij}/10$'	0.98(0.22)	0.91(0.07)
'$\hat{T}_{ij}/100$'	0.74(0.65)	0.87(0.17)

주: 괄호안은 표준편차

〈표 1〉과 〈표 2〉에서 나타난 카이제곱비율의 유의수준을 분석하기 위하여 분산을 이용하여 카이제곱비율의 표준편차(standard error)를 구하면 표의 괄호안과 같다. 앞에서 언급한 바와 같이 표본데이터가 완전한 포아송분포를 가지도록 재구성되었으므로 3가지 유형의 데이터를 이용한 중력모형 정산치의 카이제곱비율은 데이터수가 아주 작은 경우('$N_{ij}/100$')를 제외하고는 이론적인 근사치인 1에 접근하고 있다. 모형의 적합도 검증 평가해 보면, 〈표1〉 병원 데이터의 표본수가 아주 작은 경우를 제외하고는 모든 카이제곱비율의 값이 95% 신뢰구간(confidence interval)내에 존재한다. 따라서 실제 적용에서 관측통행량이 작은 경우에도 카이제곱비율은 이론적인 근사치인 1 에 접근하고 표준편차도 작으므로 중력모형의 적합도 검증 위한 매우 유용한 방법임을 알 수 있다.

Ⅵ. 결론

본 논문에서는 중력모형의 적합도 검증 위하여 널리 사용되고 있는 카이제곱비율에 의한 모형의 적합도 검증 평가하는 방법을 제시하였다. 지금까지의 연구는 특히 기 종점수에 비하여 관측통행량의 크기가 작은 경우에도 最尤推定方法은 모수추정치의 偏倚가 매우 작으며, 관측통행분포가 포아송분포일 때 모형의 적합도 검증결과 포아송중력모형이 통행분포 행태를 잘 반영하고 있음을 보여주고 있다. 그러나 위의 경우 카이제곱비율을 적합도 검증 위하여 이용할 때 이의 유의수준에 대한 연구가 거의 없는 실정이다. 따라서 본 논문에서는 카이제곱비율의 유의수준을 분석하기 위하여 카이제곱비율의 분산식을 유도하였으며, 씨뮬레이션한 데이터를 이용하여 카이제곱비율의 유의수준을

분석하였다.

관측통행량이 작은 경우를 위하여 씨뮬레이션한 데이터를 이용하여 중력모형의 파라메타를 추정하고, 또한 모형의 적합도 검증을 위하여 카이제곱비율을 산정한 결과 그 값은 이론적 근사치인 1에 접근하고 있으며, 그 값이 95% 신뢰구간내에 존재하므로 카이제곱비율은 기 종점수에 비하여 관측통행량의 크기가 작은 경우에도 중력모형의 적합도 검증 위한 매우 유용한 방법임을 알 수 있다. 그러나, 관측통행량이 아주 작을 경우에 카이제곱비율이 이론적인 근사치인 1 보다 크게 작으므로 모형의 적합도는 우수하다고 설명되어질 수 있으나, 이러한 값이 나타난 데 대한 이론적인 근거와 이의 실증적인 분석은 앞으로 연구되어야 할 과제로 생각된다.

참고문헌

1. Bishop, Yvonne M. M. et al. (1975), Discrete Multivariate Analysis: Theory and Practice, Cambridge, Massachusetts: The M.I.T. Press.
2. Low, John (1993), Gravity Model Estimation of Patient Flows in an Urban Hospital Market, Ph.D. Dissertation, School of Urban Planning and Policy, University of Illinois at Chicago.
3. Madansky, A. (1988), Prescription for Working Statisticians, New York: Springer-Verlag.
4. Rao, Calyampudi Radhakrishna. (1973), Linear Statistical Inference and Its Applications, New York: Wiley.
5. Sen, Ashish. (1986), "Maximum Likelihood Estimation of Gravity Model Parameters", Journal of Regional Science, 26, p.461-474.
6. Sen, Ashish and Matuszewski Zbigniew (1991), "Properties of Maximum Likelihood Estimates of Gravity Model Parameters" Journal of Regional Science, 31, p.469-486.
7. Smith, Toney E. (1987), "Poisson Gravity Models of Spatial Flows", Journal of Regional Science, 27, p.315-340.
8. Soot, Siim and Sen, Ashish (1991), A Spatial Employment and Economic Development Model, The journal of the RSAI, 70, 2, p149-166.
9. Weber, James and Sen, Ashish (1985), On the Sensitivity of Maximum Likelihood Estimation of Gravity Model Parameters, in B. G. Hutchinson, P. Nijkamp, and M. Batty, ed., Optimization and discrete Choice in Urban Systems, Springer-Verlag, p.148-161.