

로지트 모델 시뮬레이션에 의한 도시교통안전계획에 관한 연구

김 용 수 (국립서울산업대학 산업안전 공학과)
이 근 철 (" ")

1. 머리말

최근 도시교통계획분야에 있어서 교통수요의 예측모델 또는 교통정책의 평가 모델로서 비집계(非集計) 모델에 관한 연구가 진행되고 있다. 이 비집계모델은 개개의 의사결정단위에 있어서 선택행동을 영역마다 집계하는 것이 아니고 의사결정레벨의 데이터를 그대로 모델로 구성할 경우 이 데이터를 이용함으로써 의사결정단위의 선택행동을 모델화 할 수 있다는 특징을 갖고 있다.

한편 비집계모델의 사용목적은 여러가지로 생각되나 이것을 교통수요의 목적이나 교통정책의 평가를 위한 모델로 이용할 때는 파라미터의 추정값이나 선택비율의 추정값에 대한 안정성이 매우 중요한 문제가 되고 있다.

본 연구에서는 비집계로지트모델(선택 확률비를 얻기 위한 식)에 의한 추정치 안정성문제에 대하여 효용값의 분포특성, 데이터사이즈 및 집계화방법에 대하여 검토를 행하고 구체적으로 효용값의 분포특성을 미리 설정 하였다.

또한 이 분포특성에 대응한 가상관측치의 데이터를 시뮬레이션으로 작성하고 작성된 데이터를 기초로 하여 검토를 행한 것이다.

즉 효용값을 갖는 분포특성에 있어서 추정정도가 양호한 선택비율의 추정값을 얻기 위하여 어떠한 추정방법으로서 어느정도의 데이터사이즈가 필요하며 그리고 데이터사이즈는 효용값의 분포특성이나 선택비율의 추정방법에 따라서 어떻게 변화하는가를 효용값의 분포특성으로부터 미리 설정하고 아울러 시뮬레이션으로서 가상관측치의 데이터를 작성한 것이다.

그리고 작성된 가상관측치의 데이터를 기초로 하여 그림 1과 같이 검토를 진행하고 앞으로 비집계로지트모델을 형성하는데 목적이 있는 것이다.

2. 추정정도에 미치는 선택비율의 영향요인

1) 효용치의 분포특성

2개의 선택 가능한 대체안 ℓ, m 이 있는 경우 개인 i 가 대체안 ℓ, m 을 선택 하였을 때의 효용 $u_{i\ell}, u_{im}$ 은 각각 (1), (2)식으로 나타내진다.

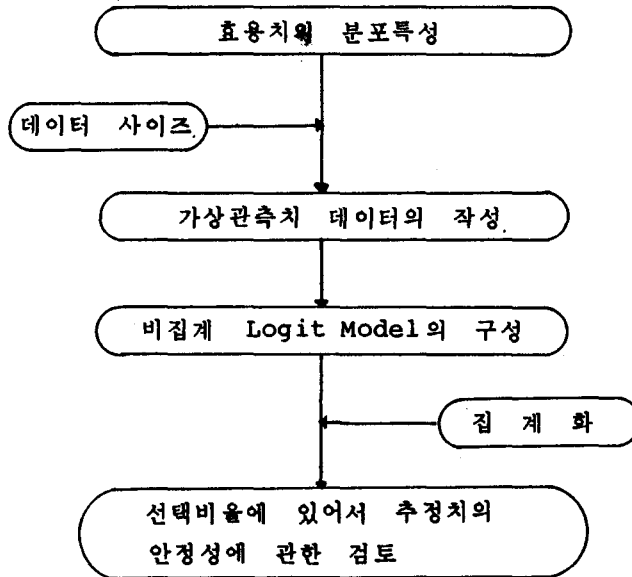


그림 1. 연구의 흐름도

$$\begin{aligned}
 u_{i\ell} &= v_{i\ell} + \varepsilon_{i\ell} \\
 &= \beta_{0\ell} + \sum_{h=1}^L \beta_h X_{i\ell h} + \varepsilon_{i\ell} \dots\dots\dots (1)
 \end{aligned}$$

$$\begin{aligned}
 u_{im} &= v_{im} + \varepsilon_{im} \\
 &= \beta_{0m} + \sum_{h=1}^L \beta_h X_{imh} + \varepsilon_{im} \dots\dots\dots (2)
 \end{aligned}$$

(1), (2)식에 있어서 $u_{i\ell}, u_{im}$ 은 대체안 ℓ, m 의 특성이나 개인 i 의 사회 경제적 속성등으로 구성되는 부분으로서 각각 변수 $X_{i\ell h}, X_{imh}$ 의 일차결합으로 나타내진다. 또한 $\varepsilon_{i\ell}, \varepsilon_{im}$ 은 오차항으로서 확률적 변동부분이며 각각 상호 독립적으로 (3)식으로 표현되는 확률밀도함수 $f(\varepsilon)$ 을 갖는 감벨분포를 따른다고 한다.

$$f(\varepsilon) = \text{EXP}(-\varepsilon) \cdot \text{EXP}\{-\text{EXP}(-\varepsilon)\} \dots\dots\dots (3)$$

식(1)(2)에서 $u_i = v_{il} - v_{im}$, $v_i = v_{il} - v_{im}$, $\beta_0 = \beta_{0l} - \beta_{0m}$, $\chi_{ik} = \chi_{ilk} - \chi_{imk}$ 로 놓으면 (4)식을 얻는다.

$$\begin{aligned} u_i &= v_i + \varepsilon_{il} - \varepsilon_{im} \\ &= \beta_0 + \sum_{k=1}^p \beta_k \chi_{ik} + \varepsilon_{il} - \varepsilon_{im} \dots\dots\dots (4) \end{aligned}$$

이 경우에 개인 i 에 의해서 대체안 θ 이 선택되는 확율을 P_i 라 하면 P_i 는 (5)식으로 표현된다.

$$\begin{aligned} P_i &= P_{r_0} b [v_{il} > v_{im}] \\ &= P_{r_0} b [v_i > 0] \\ &= \frac{1}{1+e^{-v_i}} \dots\dots\dots (5) \end{aligned}$$

여기서 (4) 식의 χ_{ik} 가 연속적으로 변화 한다고 하면 v_i, u_i 도 똑같이 연속적으로 변화 함으로 연속적으로 변화하는 χ_{ik} 를 χ_k 로 나타내고, v_i, u_i 을 각각 (6)(7)식과 같이 v, u 로 나타낸다. 각 설명변수 χ_k 의 기대치와 분산을 각각 $E[\chi_k]$, $V[\chi_k]$ 로 하고, 그리고 각 설명변수간의 상관계수를 $\gamma_{kk'}$ 로 하면 (6)식의 효용치 v 의 기대치 $E[v]$ 와 분산 $V[v]$ 는 각각 (8), (9)식과 같이 된다.

$$v = \beta_0 + \sum_{k=1}^p \beta_k \chi_k \dots\dots\dots (6)$$

$$u = v + \varepsilon_l - \varepsilon_m \dots\dots\dots (7)$$

$$E[v] = \beta_0 + \sum_{k=1}^p \beta_k E[\chi_k] \dots\dots\dots (8)$$

$$\begin{aligned} V[v] &= [\beta_0 + \sum_{k=1}^p \beta_k \chi_k] \\ &= \sum_{k=1}^p \sum_{k'=1}^p \beta_k \beta_{k'} \text{COV}[\chi_k, \chi_{k'}] \\ &= \sum_{k=1}^p \sum_{k'=1}^p \beta_k \beta_{k'} \gamma_{kk'} \sqrt{V[\chi_k] \cdot V[\chi_{k'}]} \dots\dots\dots (9) \end{aligned}$$

v_i 는 (6)식과 같이 v 로 표시 되므로 이에 따라서 (5)식의 P_i 를 (10)식과 같이 $P(v)$ 로 나타내고 v 의 확율 밀도함수를 $g(v)$ 로 하면 대상으로 하는 집단 전체에 있어서 대체안 θ 의 선택비율 $E[P(v)]$ 는 (11)식과 같이 된다. 여기서 $E[P(v)]$ 로 바꾸고 $P[E(v)]$ 을 대상으로 하는 집단 전체에 있어서 대체안 θ 의 선택비율로 한다면 $P[E(v)]$ 는 (12)식과 같이 된다.

$$P(v) = \frac{1}{1 + e^{-v}} \dots\dots\dots (10)$$

$$E[P(v)] = \int_{-\infty}^{\infty} P(v)g(v)dv$$

$$= \int_{-\infty}^{\infty} \frac{g(v)}{1 + e^{-v}} dv \dots\dots\dots (11)$$

$$P[E(v)] = \frac{1}{1 + e^{-E(v)}}$$

$$= \frac{1}{1 + e^{-\beta \sigma^{-2} \sum_{k=1}^n \beta \beta E(x_k)}} \dots\dots\dots (12)$$

$v, g(v), E(v), P[E(v)]$ 의 관계를 그림으로 표시하면 그림 2와 같으며 $E[x_k], V[x_k], \gamma_{kk}'$ 와(6)식으로 규정되는 분포특성이 효율치 V 를 갖는 경우에 (6)식의 v 와 (7)식의 u 가 동일부호인 확율을 S 라 하면 S 는 (13)식으로 표현 된다.

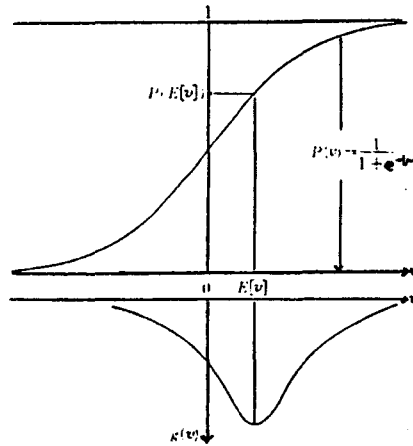


그림 2. $v, g(v), E(v)$ 및 $P[E(v)]$

(13)식에서 $\varphi(\epsilon^*)$ 는 (7)식에 있어서 2개의 오차항 ϵ_l, ϵ_m 의 차이 $\epsilon^* = \epsilon_l - \epsilon_m$ 의 확율 밀도 함수이다. (13)식의 S 는 정의로부터 (6)식과 $g(v)$ 에 의해서 규정되는 분포특성을 효율치 v 가 갖는 경우 로지트모델의 적합도를 나타낸다고 말할 수 있다.

$$S = \int_{-\infty}^{\infty} [\int_{-\infty}^{\infty} \varphi(\epsilon^*) d\epsilon^*] g(v) dv$$

$$+ \int_{-\infty}^0 [\int_{-\infty}^{-v} \varphi(\epsilon^*) d\epsilon^*] g(v) dv$$

$$= \int_{-\infty}^0 \frac{g(v) + g(-v)}{1 + e^v} dv \dots\dots\dots (13)$$

2. 집계화 방법

대상이 되는 집단전체에 있어서 선택비율을 구하기 위하여는 대상집단의 개인마다의 선택확률 $P(v)$ 을 구하고 이들의 기대값을 (11)식과 같이 집단전체로 취하는 방법을 생각 할 수 있다. 이 방법을 실제 샘플법이라고 부르고 있으며 전체의 개인에 대하여 설명변수의 값을 준비할 필요가 있다.

실제로 예측작업을 고려한다면 이 방법으로 예측을 행하는 것은 어려우므로 대상이 되는 집단 전체의 선택비율을 구하는 방법중 평균치법과 모멘트법을 들어서 검토를 하면 다음과 같다.

평균치법이란 (12)식과 같이 효용값 (v) 의 기대값 $E[v]$ 에 있어서 선택확률 $P(E(v))$ 을 갖고서 $E[P(v)]$ 로 대체한 것을 말한다. 다시 말하면 평균적 개인에 있어서 선택확률을 갖고 대상이 되는 단체집단에 있어서의 선택비율을 정하는 방법이다. 이에 대하여 모멘트법은 (10)식의 $P(v)$ 를 (14)식과 같이 $E(v)$ 로 대체하고 테일러로 전개하여 2항까지 근사화한 것으로서 근사화된 기대치 P^* 을 갖고서 (11)식의 $E[P(v)]$ 로 바꾼것이다. 따라서 모멘트 법이란 (15)식의 값을 갖고서 대상이 되는 집단전체에 있어서 선택비율을 정하는 방법을 말한다.

$$P(v) \approx P(E[v]) + (v - E[v]) \frac{dP(E[v])}{dv} + \frac{1}{2} (v - E[v])^2 \frac{d^2P(E[v])}{dv^2} \dots \dots \dots (14)$$

$$P^* = E\left[P(E[v]) + (v - E[v]) \frac{dP(E[v])}{dv} + \frac{1}{2} (v - E[v])^2 \frac{d^2P(E[v])}{dv^2} \right]$$

$$= P(E[v]) - \frac{1}{2} V[v] P(E[v]) \{1 - P(E[v])\} \{2P(E[v]) - 1\} \dots (15)$$

이 모멘트법은 선택확률의 집계화의 한가지 방법으로서 현재까지 몇가지의 연구사례가 적용되고 있으나 모멘트법의 적용가능범위에 대해서는 명시되어 있지 않고 이에 대한 유용성도 명확하지 않다고 한다.

이상과 같이 대상이 되는 집단 전체에 있어서 선택비율을 나타내는 3가지방법, 즉 $E[u]$, $P(E[u])$, P^* 에 대하여 각각 설명하였으나 실제로 비집계로지트 모델의 파라메타를 추정한후에 각각 다음과 같은 식에 의해서 선택비율을 추정하고 있다.

$E[u]$ 의 추정 ;

$$\begin{aligned} \overline{P[u]} &= \frac{1}{n} \sum_{i=1}^n P(\hat{u}_i) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{1 + e^{-u_i}} \dots\dots\dots (16) \end{aligned}$$

$P(E[u])$ 의 추정 ;

$$\begin{aligned} \overline{P(\hat{u})} &= P\left(\frac{1}{n} \sum_{i=1}^n u_i\right) \\ &= \frac{1}{1 + e^{-\frac{1}{n} \sum_{i=1}^n \hat{u}_i}} \dots\dots\dots (17) \end{aligned}$$

P^* 의 추정 ;

$$\hat{P}^* = P(\hat{u}) - \frac{1}{2} \hat{V}[u] P(\hat{u}) \{1 - P(\hat{u})\} \{2P(\hat{u}) - 1\} \dots\dots\dots (18)$$

(16)~(18) 식에 있어서 \hat{u}_i 는 효용치 u_i 의 추정치로서 (19) 식으로 나타낸다.

$$\hat{u}_i = \hat{\beta}_0 + \sum_{k=1}^K \hat{\beta}_k \chi_{ik} \dots\dots\dots (19)$$

2.3 추정정도에 미치는 영향요인

대상이 되는 집단 전체에 있어서 선택비율의 추정정도에 영향을 주는 요인으로서 효용값의 분포특성과 집계화의 방법 즉 추정방법과 데이터사이즈를 생각할 수 있다.

효용값의 분포특성은 (13) 식의 S등으로 나타내는 모델의 고유적합도를 규정하고 있으며 모델의 적합도는 데이터사이즈와 더불어 파라메타의 추정치 $\hat{\beta}_0, \hat{\beta}_k$ 의 추정정도에 영향을 준다. 또한 파라메타의 추정치 $\hat{\beta}_0, \hat{\beta}_k$ 의 추정정도는 (19) 식을 통해서 효용값 u_i 의 추정치 \hat{u}_i 에 영향을 주고 있다.

앞에서 말한바와같이 $\overline{P(\hat{u})}$, $P(\hat{u})$, \hat{P}^* 중 어느방법으로도 선택비율을 추정해도 효용값 u_i 의 추정치 \hat{u}_i 의 추정정도가 데이터사이즈와 함께 선택비율의 추정정도에 영향을 주는것은 (16),(17), (18)식으로 부터 명백하다.

특히 평균치법으로 선택비율을 추정하는 경우에 데이터사이즈가 적으면 (17)식의 \overline{V} 의 추정정도가 나쁘게되고 이것이 선택비율 $P(\hat{u})$ 의 추정정도에도 영향을 준다고 생각된다.

한편 효용치의 분포특성과 데이터사이즈 및 집계화방법이 선택비율의 추정정도에 어느정도 영향을 미치는가를 파악하기 위하여는 시뮬레이션으로서 가상관측치의 데이터를 작성하고 이것을 기초로하여 검토를 행하면 된다.

3. 가상관측치 데이터의 작성

가상관측치데이터의 작성은 그림 3과 같은 순서에 따라서 행하며 이 경우에 필요한 것은 (1) 설명변수군의 작성방법과 (2) 감벨난수의 발생방법등으로써 다음에 간단히 설명한다.

(1) 설명변수군의 작성방법

$E(x_k)$, $V(x_k)$, $r_{kk'}$ 에 의해서 규정된 구조를 갖는 설명변수군의 작성은 그림 4와 같은 순서

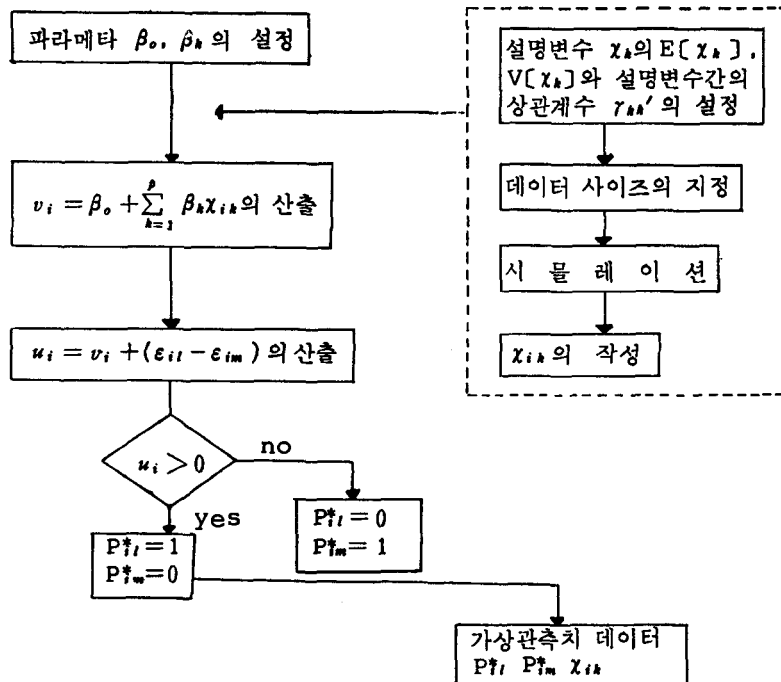


그림 3. 가상관측치 데이터의 작성순서

에 따라서 행한다. 단 여기서 설명변수군은 정규분포에 따르는 것으로 하며 처음에 설정된 $E[x_k], V[x_k], r_{kk}'$ 를 기초로해서 설명변수군의 분산공분산행렬 Σ 의 계산을 행하고 이 분산공분산행렬 Σ 를 이용해서 (20)식을 만족시키는 비특이한 하측삼각행렬 A 를 구한다.

$$AA^T = \Sigma \dots\dots\dots (20)$$

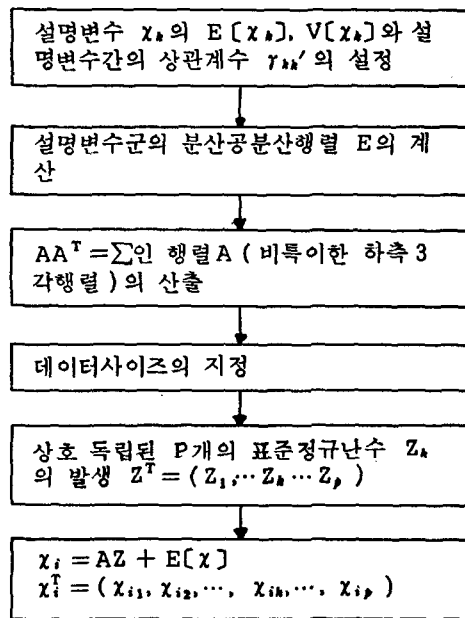


그림 4. 설명변수군의 작성순서

다음에 상호독립된 P개의 표준정규난수 $Z_k (K = 1, 2, \dots, P)$ 를 발생시킴으로서 (21)식에 의해서 설명변수군의 작성을 행한다.

$$x_i = AZ + E[x] \dots\dots\dots (21)$$

여기서

로지트 모델 시뮬레이션에 의한 도시교통안전계획에 관한, 연구

$$x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix}, \quad z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{pmatrix}, \quad E[x] = \begin{pmatrix} E[x_1] \\ E[x_2] \\ \vdots \\ E[x_p] \end{pmatrix}$$

(2) 감벨난수의 발생방법

확률적으로 변동하는 오차항을 작성하기위한 감벨난수는 감벨분포의 확률밀도 함수 $f(\epsilon)$ 가 (3) 식으로 표시됨으로 이 분포함수 $F(\epsilon)$ 를 (22) 식에 의해서 구하고, 이식을 (23) 식과 같이 변형하고 이 변형된 (23) 식의 $F(\epsilon)$ 에 $[0, 1]$ 의 일정한 난수를 대입시킴으로서 발생된다.

$$F(\epsilon) = \int_{-\infty}^{\epsilon} f(\epsilon) d\epsilon = e^{-\epsilon^2} \dots \dots \dots (22)$$

$$\epsilon = -I_n \{ -I_n [F(\epsilon)] \} \dots \dots \dots (23)$$

3.2 가상관측치데이터의 작성

여기에서는 모델에 사용된 설명변수가 x_1, x_2, x_3, x_4, x_5 의 5변수로서 파라메타는 표 1과 같은 값을 파라메타의 진값으로 설정한 것이다. 모델에 사용된 각 설명변수는 x_k ($k=1,2 \dots 5$)

표 1. 파라메타의 설정치

파라메타	설정치
β_0	0.0
β_1	0.1
β_2	0.2
β_3	0.3
β_4	0.4
β_5	0.5

표 2. $E[x_k]$, $V[x_k]$, $r_{kk'}$ 의 설정치

	CASE I	CASE II	CASE III	CASE IV	CASE V	CASE VI
$E[x_1]$	0	0	0	0	1	1
$V[x_1]$	0.5	1	5	5	5	5
$E[x_2]$	0	0	0	0	1	1
$V[x_2]$	0.5	1	5	5	5	5
$E[x_3]$	0	0	0	0	1	1
$V[x_3]$	0.5	1	10	10	10	10
$E[x_4]$	0	0	0	0	1	1
$V[x_4]$	0.5	1	15	15	15	15
$E[x_5]$	0	0	0	0	1	2
$V[x_5]$	0.5	1	20	20	20	20
$r_{kk'}$	0.0	0.0	0.0	0.8	0.0	0.0

는 각각 정규분포 $N(E[x_k], V[x_k])$ 에 따르는 것으로 한다. 각 설명변수 x_k 의 설정인 기대치 $E[x_k]$, 분산 $V[x_k]$ 및 설명변수간의 상관관계수 $r_{kk'}$ 의 값을 표 2에 나타내었다. 표 2에 있어서 설정치를 기초로하여 (8), (9), (11), (12), (15), (13)에 따라서 산출한 $E[v]$, $V[v]$, $E[P(v)]$, $P(E(v))$, P^* , S 의 값을 표 3에 나타내었다.

표 2에 있어서 CASE I, CASE II, CASE III,는 각 설명변수 x_k 가 상호 독립적이며 $r_{kk'}=0.0$ 으로 두고 있다. CASE IV에서는 설명변수간에 강한 상관관계가 있는 경우에 선택비율의 추정치의 안정성이 어떻게 변화하는가를 파악하기 위한 것이며 CASE III를 토대로해서 $r_{kk'}=0.8$ 로 취하고 있다.

표 3. 효용치의 분포특성

	CASE I	CASE II	CASE III	CASE IV	CASE V	CASE VI
$E[v]$	0.0	0.0	0.0	0.0	1.5	2.0
$V[v]$	0.275	0.550	8.550	25.079	8.550	8.550
$E[P(v)]$	0.500	0.500	0.500	0.500	0.670	0.721
$P(E(v))$	0.500	0.500	0.500	0.500	0.818	0.881
P^*	0.500	0.500	0.500	0.500	0.413	0.539
S	0.600	0.637	0.833	0.895	0.849	0.861

또한 CASE V, CASE VI는 $E[\lambda_k]$ 의 변화가 선택비율 추정치의 안정성에 어떠한 영향을 주는가를 파악하기 위한 것이며 CASE III와 함께 $E[\lambda_k]$ 를 변화시키고 있다.

한편 $E[v]$ 가 0인 경우에는 $V[v]$ 가 크면 클수록 S의 값은 커지므로 CASE I, CASE II, CASE III, CASE IV에서는 (13)식에 의한 S의 값이 표 3에 나타낸바와같이 (24)식과 같이 되어 있다.

$$S_I < S_{II} < S_{III} < S_{IV} \dots\dots\dots (24)$$

또한 $V[v]$ 가 같다면 $E[v]$ 가 0에서 멀어지는 쪽이 S의 값이 크게 됨으로 CASE III, CASE V, CASE VI에서는 S의 값이 표 3에 나타낸바와 같이 (25)식으로 되어 있다.

$$S_{III} < S_V < S_{VI} \dots\dots\dots (25)$$

가상관측치데이터는 설정된 CASE마다 데이터사이즈를 100에서 1,000까지 100씩 증가시키고 각각의 데이터사이즈에 대해서 5조씩 작성하고 있다. 따라서 하나의 CASE에 대해서는 50조의 가상관측치데이터를 작성하게 되는 것이다.

4. 선택비율에 있어서 추정치의 안정성

4.1 선택비율의 추정결과

앞에서 작성한 가상관측치데이터로부터 비집계로지트모델의 파라메타에 대하여 추정을 행한후에 구한 선택비율의 추정결과에 대해서 설명한다. 오히려 파라메타의 추정은最尤度法(Maximum Likelihood)으로 행한 것이다. 선택비율의 추정결과에 대한 일부분(CASE I, III, V)를 그림 5-13에 나타내었다. 그림 5~7은 효용함수의 파라메타를 추정한후에 효용치 v 의 추정치 \hat{v} 마다 선택확률 $P(\hat{v})$ 를 구하고 이 평균치 $\overline{P(\hat{v})}$ 를 갖고서 선택비율로 취한 것이다. 또한 그림 8~10은 효용치 v 의 추정치 \hat{v} 의 평균치 $\overline{\hat{v}}$ 에 있어서 선택확률 $P(\overline{\hat{v}})$ 를 갖고서 선택비율로 정한 것이다.

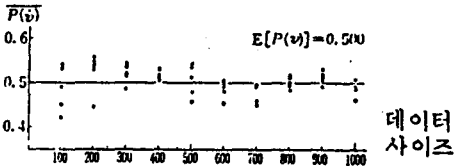


그림 5. 선택비율의 추정치: Case I의 $\overline{P(\hat{v})}$

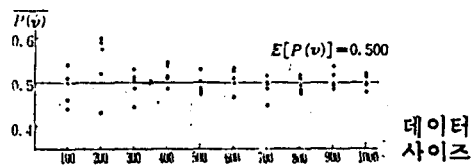


그림 6. 선택비율의 추정치: Case III의 $\overline{P(\hat{v})}$

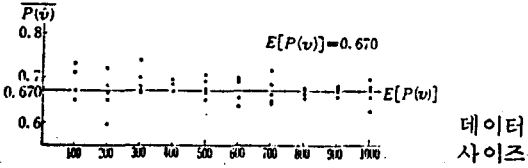


그림 7. 선택비율의 추정치 : Case V의 $\overline{P(\hat{v})}$

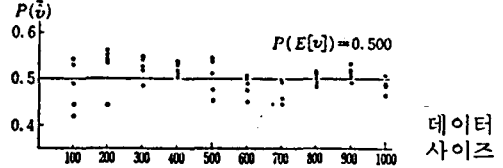


그림 8. 선택비율의 추정치 : Case I의 $\overline{P(\hat{v})}$

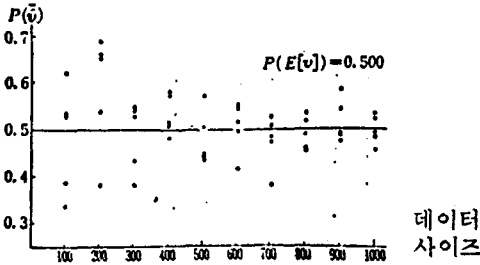


그림 9. 선택비율의 추정치 : Case III의 $\overline{P(\hat{v})}$

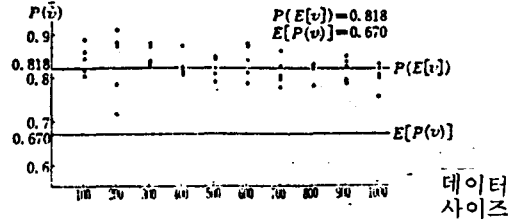


그림 10. 선택비율의 추정치 : Case V의 $\overline{P(\hat{v})}$

또한 그림 11~13은 모멘트법에 의해서 산출한 비율이다. 즉, 그림 5~7은 (16)식에 의해서, 그림 8~10은 (17)식에 의해서 그리고 그림 11~13은 (18)식에 의해서 각각 선택비율을 구한 것이다. CASE V에서는 표 3으로부터 명백한 바와같이 $E[P(v)]$, $P(E[v])$, P^* 의 값에 차이가 있으므로 그림 7에는 $E[P(v)]$, 그림 10에는 $E[P(v)]$ 와 $P(E[v])$, 그리고 그림 13에는 $E[P(v)]$ 와 P^* 의 값이 그 위치를 나타내고 있다.

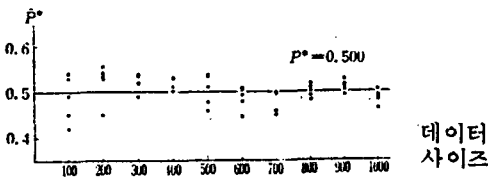


그림 11. 선택비율의 추정치 : Case I의 \hat{P}^*

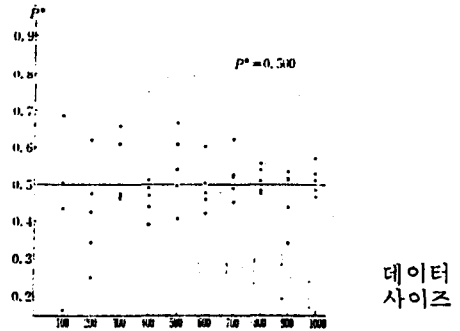


그림 12. 선택비율의 추정치 : Case III의 \hat{P}^*

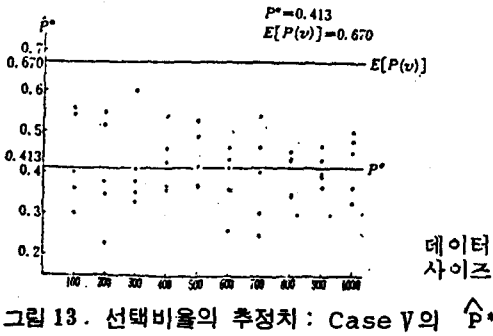


그림 13. 선택비율의 추정치: Case V의 \hat{P}^*

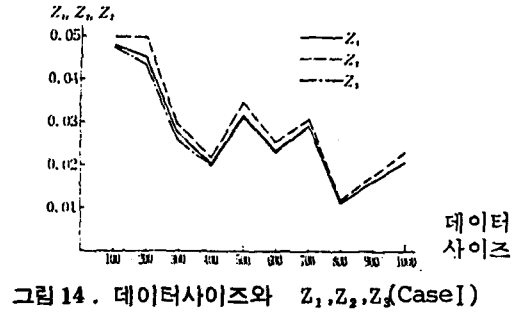


그림 14. 데이터사이즈와 Z_1, Z_2, Z_3 (Case I)

선택비율에 있어서 3가지 추정방법 $\overline{P(\hat{v})}$, $\overline{P(\hat{v})}$, \hat{P}^* 에 의한 추정결과와 더불어 CASE 마다 추정정도에 대해서 비교하면 CASE I (그림 5, 8, 11)에서는 어떠한 추정방법에 의해서도 동일한 안정성을 갖는 추정결과를 얻을 수 있으나 CASE III (그림 6, 9, 12)에서는 $\overline{P(\hat{v})}$, \hat{P}^* 에 의한 추정치는 데이터사이즈가 적은 경우에 매우 뚜렷하게 안정성이 결여되어 있음을 알수있다.

또한 CASE V (그림 7, 10, 13)에서는 $\overline{P(\hat{v})}$, \hat{P}^* 에 의한 추정치가 어느정도 큰 편증을 갖는다는 것을 알수있다. CASE II, IV, VI의 추정결과에 대해서는 그림에 나타내지 않았으나 CASE II, VI는 각각 CASE I, V와 같은 경향을 나타내고 있으며 CASE IV에서는 CASE III의 경향이 재차 뚜렷하게되어 \hat{P}^* 의 추정정도가 극단적으로 저하되고 있다.

4.2 데이터사이즈와 추정정도

추정치와 더불어 (26), (27), (28) 식에 의해서 정의된 추정오차의 값을 각 CASE마다 그리면 그림 14~19와 같이 된다.

$$\begin{aligned}
 Z_1 &= Z[\overline{P(\hat{v})}] \\
 &= R\left(\sum_{i=1}^5 \{ \overline{P_i(\hat{v})} - E[\overline{P(\hat{v})}] \}^2 / 5\right) \dots\dots\dots (26)
 \end{aligned}$$

$$\begin{aligned}
 Z_2 &= Z[\overline{P(\hat{v})}] \\
 &= R\left(\sum_{i=1}^5 \{ \overline{P_i(\hat{v})} - P(E[v]) \}^2 / 5 \right. \\
 &\quad \left. + \{ P(E[v]) - E[P(v)] \}^2 \right) \dots\dots\dots (27)
 \end{aligned}$$

$$\begin{aligned}
 Z_3 &= Z[\hat{P}^*] \\
 &= R\left(\sum_{i=1}^5 \{ \hat{P}_i^* - P^* \}^2 / 5 + \{ P^* - E[P(v)] \}^2 \right) \dots\dots\dots (28)
 \end{aligned}$$

여기서 $R(A) = \sqrt{A}$ 이다.

(26), (27), (28) 식에 있어서 ℓ 은 각 데이터사이즈에 있어서 선택비율의 추정번호 ($\ell = 1, 2 \dots 5$) 이며, n 은 데이터사이즈 ($n = 100, 200, \dots, 1,000$) 이다. 또한 $\overline{P(\hat{\theta})}$, $\overline{P(\hat{\theta})}$, \hat{P}^* 은 각각 데이터사이즈 n , 추정번호 ℓ 일때의 (16), (17)(18) 식에 의한 추정비율의 추정치이다.

CASE I (그림 14), CASE II (그림 15) 에서는 Z_1, Z_2, Z_3 와 더불어 동일한 값을 나타내나 CASE III (그림 16) 에서는 앞에서 말한바와 같이 데이터사이즈가 적은 경우는 $\overline{P(\hat{\theta})}$, \hat{P}^* 가 불안정하게 되므로 Z_2, Z_3 는 Z_1 에 비교하면 큰값을 가지게 된다.

CASE IV (그림 17) 에서는 이러한 경향이 보다 더우기 현저하게 되어있다. 따라서 $P(E(v))$ 와 $E[P(v)]$, P^* 와 $E[P(v)]$ 에 차이가 없는 CASE에도 $V(v)$ 가 큰 경우에는 데이터사이즈가 적지않으면 $\overline{P(\hat{\theta})}$, \hat{P}^* 가 큰 추정오차를 갖게 된다.

여기서 데이터사이즈가 적은 경우에는 S 의 값이 크고 모델의 적합도가 양호한 경우에도 추정방법에 의해서 선택비율의 추정정도가 현저하게 저하됨을 알수있다. CASE V (그림 18), CASE VI (그림 19) 에서는 $P(E(v))$ 와 $E[P(v)]$, P^* 와 $E[P(v)]$ 와의 사이에 차이가 있으며 $P(E(v))$ 와 P^* 가 각각 편중되어있기 때문에 Z_2, Z_3 는 데이터사이즈를 크게해도 적어지게 된다.

여기서 어떤 데이터사이즈이상의 경우에는 Z_1, Z_2, Z_3 가 어느수준 이하로 되는 데이터사이즈에 주목해서 정리하면 표 4~6 과 같이 된다. 표 4 에서 선택비율을 $\overline{P(\hat{\theta})}$ 로서 구하는 경우에는 어

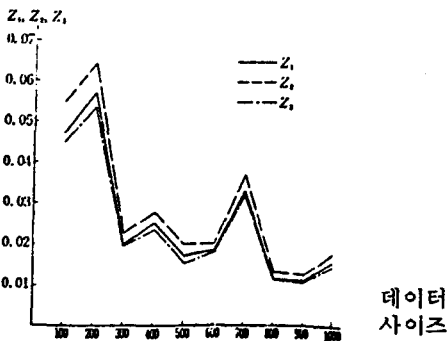


그림 15. 데이터사이즈와 Z_1, Z_2, Z_3 (Case II)

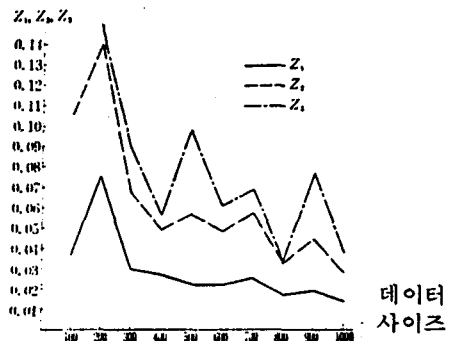


그림 16. 데이터사이즈와 Z_1, Z_2, Z_3 (Case III)

면 CASE에도 같은 정도의 데이터사이즈로서 만족함을 알수있다. CASE III를 예로든다면 Z_1 을 0.025 이하로 하면 800 이상의 데이터사이즈가 필요하며 0.050 이하의 경우에는 300 이상, 0.075

로지트 모델 시뮬레이션에 의한 도시교통안전계획에 관한 연구

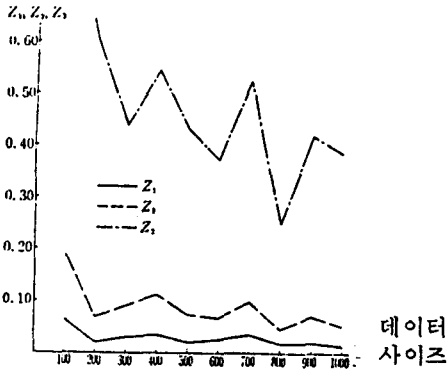


그림 17. 데이터사이즈와 Z_1, Z_2, Z_3 (Case IV)

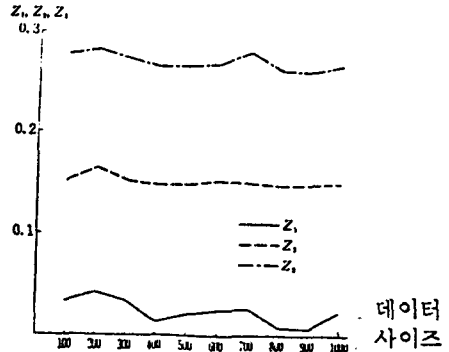


그림 18. 데이터사이즈와 Z_1, Z_2, Z_3 (Case V)

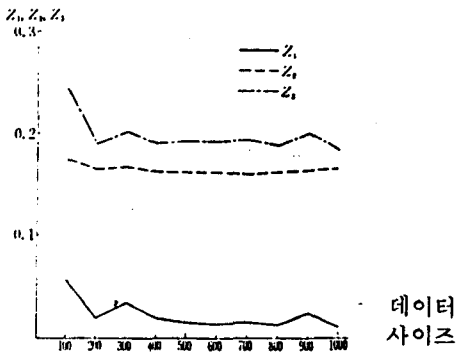


그림 19. 데이터사이즈와 Z_1, Z_2, Z_3 (Case VI)

이하의 경우에도 0.050 이하의 경우와 똑같이 300 이상, 그리고 0.100 이하가 되면 100 정도의 데이터사이즈가 필요하게 된다.

표 5에 의하면 선택비율을 $P(\hat{v})$ 로서 구하는 경우에 Z_2 를 0.025 이하로 하면 CASE I, II에서는 800 이상의 데이터사이즈가 필요하게 되고 0.050 이하가 되면 300 이상의 데이터사이즈로서 만족함을 알 수 있다. CASE V, VI에서는 데이터사이즈를 1000으로 해도 Z_2 를 0.100 이하로서 선택비율을 구한다는 것은 곤란하다.

그리고 표 6에 의해서 선택비율을 \hat{P}^* 로서 구하는 경우에는 CASE I, II에서 $P(\hat{v})$, $P(\hat{v})$ 의 경우와 동일한 정도의 데이터사이즈로서 만족되나 CASE III와 CASE IV에서는 $P(\hat{v})$ 경우보다도 큰 데이터사이즈가 필요함을 알 수 있다.

표 4. 허용오차와 데이터사이즈 ($P(\hat{v})$ 의 경우)

	0.025	0.050	0.075	0.100
CASE I	800	100	100	100
CASE II	800	300	100	100
CASE III	800	300	300	100
CASE IV	800	200	100	100
CASE V	800	100	100	100
CASE VI	400	200	100	100

표 5. 허용오차와 데이터사이즈 ($P(\hat{v})$ 의 경우)

	0.025	0.050	0.075	0.100
CASE I	800	300	100	100
CASE II	800	300	100	100
CASE III	-	800	300	300
CASE IV	-	-	800	500
CASE V	-	-	-	-
CASE VI	-	-	-	-