

혼합원형분포를 이용한 지방국도의 시간교통량 추정모형

나종화¹ · 장영미²

¹충북대학교 정보통계학과, ²한국보건복지정보개발원

(2011년 3월 접수, 2011년 4월 채택)

요약

본 논문에서는 우리나라 지방국도의 특정지점에서 수집된 교통량 자료를 이용하여 일일 시간교통량 추정모형을 개발하였다. 본 연구의 특징은 일일 24시의 시간변수를 원형변수로 취급하고, 지방부 교통량 자료의 특성상 출·퇴근 시간에 교통량이 집중되는 이봉형의 현상을 감안하여 원형분포의 혼합모형을 고려하였다. 또한 시간대별 교통량의 분포가 요일에 따라 유사한 패턴을 가지는 데 착안하여 요일별 모형을 제시하였다. 혼합원형분포의 모수추정에는 EM 알고리즘이 사용되었으며, 모형의 성능비교를 위해 가변수 회귀모형과의 비교를 실시하였다. 제시된 요일별 지방국도의 시간교통량 적합모형은 계측기의 손상 등으로 인한 교통량 결측자료의 추정에 효과적으로 사용될 수 있다.

주요어: 시간교통량, 원형분포, von Mises 분포, 겹친왜정규 분포, EM 알고리즘.

1. 서론

본 논문에서는 우리나라 지방국도에서 조사된 시간교통량 자료를 원형자료로의 전환을 통해 이를 모형화하는 새로운 분석을 수행하고자 한다. 분석에 사용될 자료는 한국건설기술연구원에서 2004년과 2005년 전국 일반국도의 특정지점에 설치된 상시조사장비를 통해서 수집된 교통량 자료이다. 여기서 교통량이란 도로의 한 지점 또는 단면을 단위시간 동안 통과한 차량의 수를 의미한다. 도로를 통과하는 단위 시간당의 교통량은 도로 시설물의 효율성 척도로 사용되며, 다른 지점과의 상대적 비교를 통하여 각 도로 구간의 역할을 추정 또는 평가할 수 있는 지표로도 사용된다. 또한, 교통량 자료는 도로 계획 및 설계와 도로 운영 등에 폭 넓게 이용되며, 교통 계획, 관리 계획 수립과 관련된 여러 분야에서 활용 빈도가 높은 중요한 자료중의 하나이다. 본 논문의 분석에 사용된 교통량 자료의 측정 지점에 대한 정보는 표 1.1과 같다.

표 1.2는 위의 조사지점에서 2004년 1년 동안 수집된 자료를 토대로 시간대별 교통량 자료의 일부를 보여주는 것으로, 매일 한 시간 단위로 측정된 시간대별 교통량 자료의 총 수는 8784개(24시간×366일)이다. 이 자료에서 시간대 변수는 $0^\circ \sim 360^\circ$ 사이의 각(또는 원형) 변수로 변환되어 분석에 이용될 것이다.

본 연구에서는 위의 자료를 이용하여 특정도로의 일일교통량에 대한 통계적 모형을 적합하고자 한다. 2절에서는 원형자료에 대한 기초분석과 원형모형을 소개하고, 3절에서는 다양한 대칭 및 비대칭의 원형모형과 그의 혼합분포를 통한 적합결과를 제시하였다. 4절에서는 구축모형의 성능에 대한 평가를 수행하였으며, 5절은 결론으로 구성되었다.

이 논문은 2010년도 충북대학교 학술연구지원사업의 연구비 지원에 의하여 연구되었음.

¹교신저자: (361-763) 충북 청주시 흥덕구 개신동 12, 충북대학교 정보통계학과, 교수.

E-mail: cherin@cbu.ac.kr

표 1.1. 교통량측정 도로 정보

도로구분	소재지	관측구간	도로유형	차로수
101 도로	경기도 평택시	현덕면~인덕면	지방부도로	4차선

표 1.2. 101 국도의 연간 시간대별 교통량 자료(2004년)

일련번호	월	일	시간대	각(degree)	교통량
1	1	1	0~1	0°~15°	138
2	1	1	1~2	15°~30°	111
3	1	1	2~3	30°~45°	101
⋮	⋮	⋮	⋮	⋮	⋮
8783	12	31	22~23	330°~345°	426
8784	12	31	23~24	345°~360°	291

2. 원형자료에 대한 기초분석과 혼합원형모형 소개

2.1. 기초분석

먼저 각(angle)으로 주어지는 n 개의 원형자료를 $\theta_1, \dots, \theta_n$ 이라 하자. 이 때, 원형자료의 평균방향을 나타내는 원형평균(circular mean) $\bar{\theta}$ 는

$$\bar{\theta} = \arctan^*(S/C) = \begin{cases} \arctan(S/C), & \text{if } C > 0, S \geq 0, \\ \pi/2, & \text{if } C = 0, S > 0, \\ \arctan(S/C) + \pi, & \text{if } C < 0, \\ \arctan(S/C) + 2\pi, & \text{if } C \geq 0, S < 0, \\ \text{undefined}, & \text{if } C = 0, S = 0, \end{cases}$$

으로 정의된다. 위 식에서 C 와 S 는 합성벡터 R 의 성분으로 다음과 같이 정의된다.

$$R \equiv (C, S) = \left(\sum_{i=1}^n \cos\theta_i, \sum_{i=1}^n \sin\theta_i \right), \quad i = 1, \dots, n.$$

또한 원형자료의 평균으로부터 흩어진 정도를 나타내는 원형분산(circular variance) V_0 는

$$V_0 = 1 - r, \quad 0 \leq V_0 \leq 1$$

으로 정의된다. 위 식에서 $r = \|R\|/n$ 으로 평균합성벡터의 크기를 나타낸다.

원형자료의 분석과 관련된 대표적 문헌으로는 Mardia (1972), Batschelet (1981), Fisher (1993), Mardia와 Jupp (1999) 그리고 Jammalamadaka와 SenGupta (2001)를 들 수 있다. 특히 Jammalamadaka와 SenGupta (2001)에서는 R 을 이용한 다양한 최근 분석방법을 소개하고 있다.

2.2. 원형분포와 혼합원형분포 소개

선형에서의 정규분포와 마찬가지로 원형자료의 적합에는 von Mises 분포가 매우 중요한 역할을 한다. 원형정규(circular normal) 분포라고도 불리는 von Mises 분포는 모수추정이 용이할 뿐 아니라 검정과 관련된 다양한 표본분포의 유도가 쉬운 장점 때문에 폭 넓게 응용되고 있다. von Mises 외의 대칭형 원

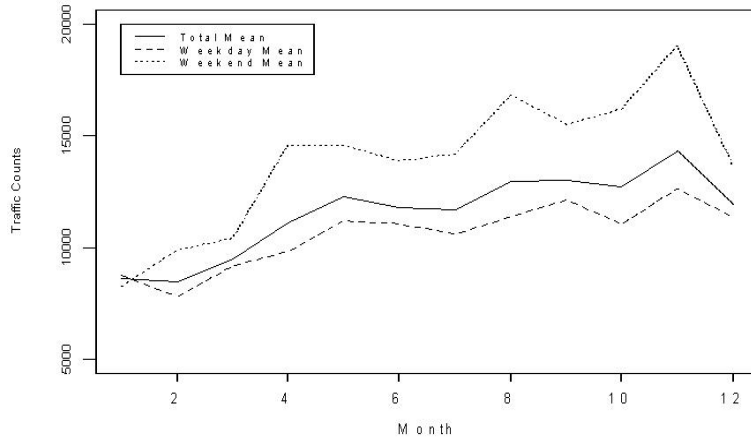


그림 3.1. 월별 평균일일교통량

형분포로는 겹친정규(wrapped normal), 겹친코쉬(wrapped Cauchy)를 비롯하여 Cardioid 및 원형균일(circular uniform) 분포 등이 있다. 최근에는 비대칭의 원형분포에 대한 연구가 활발하게 진행되고 있으며, 대표적인 분포로는 겹친라플라스(wrapped Laplace)와 겹친왜정규(wrapped skew-normal) 분포를 들 수 있다. 그 외의 고전적인 비대칭 분포로는 Batschelet (1981)과 Papakonstantinou (1979) 및 겹친(wrapped) α -stable 분포 등을 들 수 있다. 이 가운데 Pewsey (2000)가 소개한 겹친왜정규분포는 Azzalini (1985)에 의한 선형의 왜정규분포로부터 겹침(wrapping)의 원리를 통해 유도되는 분포로 그 활용성이 뛰어나다.

본 논문에서 다루어지는 지방부 국도의 특성(출, 퇴근 시간대에 교통량이 집중되는 현상)을 반영하기 위한 다분형 원형분포 모형으로 다음과 같이 정의되는 혼합분포 모형을 고려하기로 한다. 즉, 분포 $f_l (l = 1, 2, \dots, k)$ 을 모수가 λ_l 인 원형분포라 할 때, 혼합원형분포는 다음과 같이 정의된다.

$$f(\theta | \Lambda) = \sum_{l=1}^k \alpha_l f_l(\theta | \lambda_l). \tag{2.1}$$

여기서 Λ 는 모수집합으로 $\Lambda = \{\alpha_1, \alpha_2, \dots, \alpha_k, \lambda_1, \lambda_2, \dots, \lambda_k\}$ 이고, α_l 은 각 단일분포의 혼합비율로 $\sum_{l=1}^k \alpha_l = 1$ 을 만족한다.

일반적으로 혼합모형에서의 최대가능도추정은 가능도함수의 최대화 과정에 계산상의 어려움을 내포하고 있다. 이를 극복하기위해 본 논문에서는 EM 알고리즘을 이용하여 다양한 원형혼합분포에 대한 최대가능도추정을 실시하였다. 선형에서의 혼합분포와 이의 모수 추정에 관한 대표적 문헌으로는 Titterton 등 (1985)과 McLachlan과 Peel (2000)을 들 수 있으며, 특히 후자의 경우 EM 알고리즘의 활용에 대해 자세히 다루고 있다. 최근 Jang 등 (2007)과 Na와 Jang (2010a, 2010b)에서는 EM 알고리즘을 이용하여 von Mises, 겹친왜정규 및 겹친 라플라스 혼합분포에 대한 모수추정을 제시한 바 있다.

3. 시간교통량 예측모형

3.1. 기초분석

먼저 시간교통량 자료의 면밀한 분석에 앞서 일일교통량의 기초분석을 통해 도로의 특성을 파악하기로 한다. 그림 3.1은 표 1.2에 소개된 101 도로의 교통량에 대한 변화를 월별로 나타낸 그림이다. 자료가

표 3.1. 월별/요일별 평균일일교통량

구분	월요일	화요일	수요일	목요일	금요일	토요일	일요일	월평균	주중 평균	주말 평균
1월	7,421	8,697	10,389	9,065	8,259	9,472	6,781	8,617	8,757	8,276
2월	7,728	7,862	7,683	7,582	8,053	9,976	9,834	8,438	7,782	9,897
3월	11,143	8,924	8,381	8,635	8,573	11,585	9,239	9,496	9,177	10,412
4월	12,065	9,127	9,570	9,302	9,319	12,787	16,394	11,096	9,825	14,590
5월	11,325	11,408	11,395	10,988	10,817	13,933	15,193	12,280	11,193	14,563
6월	10,806	10,925	10,868	11,122	11,492	13,934	13,832	11,790	11,029	13,884
7월	11,050	10,456	10,608	10,449	10,633	14,214	14,130	11,660	10,630	14,177
8월	11,672	10,742	10,754	11,422	12,459	15,005	18,254	12,964	11,391	16,810
9월	10,126	13,543	15,327	10,169	11,197	13,344	17,636	13,029	12,134	15,490
10월	11,706	10,768	10,850	10,940	11,008	14,584	17,817	12,713	11,052	16,200
11월	13,247	12,492	12,369	12,156	12,879	17,870	20,216	14,355	12,650	19,043
12월	11,322	10,978	11,014	11,134	12,180	14,017	13,270	11,935	11,341	13,644
요일 평균	10,881	10,515	10,816	10,226	10,545	13,367	14,452	11,537	10,595	13,909

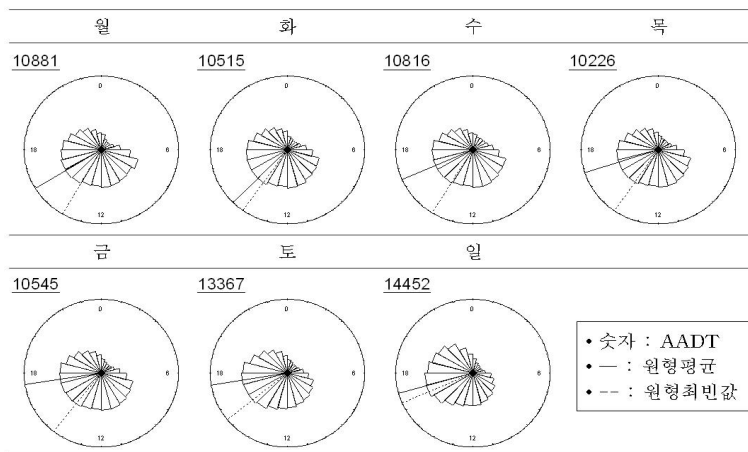


그림 3.2. 요일별 시간교통량에 대한 로즈다이어그램

수집된 101 도로상의 조사지점은 전형적인 우리나라의 지방부 국도의 특성을 가지는 곳으로 분류되며, 아래의 그림에서 알 수 있듯이 주중보다 주말의 교통량이 많은 특성을 지니고 있다. 평균일일교통량에 대한 보다 자세한 요약은 표 3.1에 정리되어 있다. 표 3.1에서는 월평균 일일교통량을 보다 자세하게 요일별로 구분하여 제시하고 있다. 이후의 모든 분석과정에서 도로의 특성에서 크게 벗어나는 국정공휴일이나 명절기간 등은 제외하고 분석하였다.

이상에서 제시된 일일교통량에 대한 분석결과는 시간의 경과에 따른 월별 또는 요일별 교통량 변화의 추이를 파악하는데 도움이 될 수 있으나 하루 24시간 내에서의 교통량 변화에 대한 분포도를 알아보기 위해서는 시간대별 교통량 자료에 대한 분석이 필수적이다. 아래의 그림 3.2는 101 도로의 요일 및 시간대별 교통량의 분포를 로즈다이어그램(rose diagram)으로 시각화한 것이다. 이 그림에서 숫자는 연평균일일교통량(AADT; Annual Average Daily Traffic)을 의미하며, 실선과 점선은 각각 원형평균과 원

표 3.2. 요일별 시간교통량에 대한 기술통계량

통계량	요일						
	월	화	수	목	금	토	일
전체교통량	565828	546770	562424	541976	558874	695088	751484
평균교통량	10881	10515	10816	10226	10545	13367	14452
원형평균(시간)	14.01	14.35	14.14	14.30	14.53	15.41	16.41
원형최빈값(시간)	15.91	14.98	16.43	16.80	17.41	17.37	16.96
원형분산	0.60	0.58	0.61	0.59	0.59	0.53	0.50

표 3.3. 모형적합에 사용된 혼합 원형분포 모형

대칭 혼합모형	비대칭 혼합모형
von Mises 혼합모형(vMM)	Papakonstantinou 혼합모형(PM)
겹친코쉬 혼합모형(WCM)	겹친라플라스 혼합모형(WLM)
	겹친왜정규 혼합모형(WSNM)

형최빈값을 나타내며, 자세한 수치 정보는 표 3.2에 제시되어 있다. 원형최빈값(circular mode)에 대한 정의는 Mardia (1972)를 참고하기 바란다.

그림 3.2와 표 3.2를 통해 주중(월~금)의 교통량 분포 간에는 큰 차이가 없으나, 주중과 주말(토, 일) 간에는 상당히 다른 패턴을 보이고 있으며, 특히 주말의 경우에는 토요일과 일요일 간에도 뚜렷한 차이가 있음을 알 수 있다.

3.2. 원형분포모형의 적합

이 절에서는 101 도로의 시간교통량 자료에 대해 요일별로 원형분포 모형을 적합하였다. 아래의 히스토그램(그림 3.3~그림 3.5)에서도 나타나듯이 지방부 도로로 분류되는 101 도로의 경우 주중자료의 특징을 살펴보면 출·퇴근 시간을 전후로 교통량이 집중되는 현상이 약하게 나타나고 있다. 이러한 이봉형의 자료 패턴은 도시부 도로의 경우에는 더욱 뚜렷하게 나타나는 현상으로 이해된다. 따라서 본 연구에서는 다봉형 자료에 대한 설명력의 제고와 도시부 도로에의 확장성 등을 고려해 식 (2.1)의 혼합모형에서 ($k = 2$)인 이봉형을 고려하여 모형적합을 실시하였다. 주말 자료의 경우 역시 약간의 굴곡현상이 포함하는 자료로 단봉형 보다는 혼합모형을 통해 더욱 설명력을 높일 수 있다. 모형적합에 사용된 분포를 요약하면 표 3.3과 같다.

위의 표 3.3에서 제시된 여러 가지 원형분포의 밀도함수는 다음과 같이 정의된다.

I. von Mises 분포:

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, \quad 0 \leq \theta < 2\pi.$$

여기서, μ 와 κ 는 각각 평균방향과 집중도를 나타내는 모수로 $0 \leq \mu < 2\pi$, $\kappa \geq 0$ 을 만족한다. 또한 $I_0(\kappa)$ 는 일종의 정규화 상수로써, 차수 영의 제1종 수정된 Bessel 함수로 다음과 같이 정의된다.

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos \theta) d\theta = \sum_{r=0}^{\infty} \left(\frac{\kappa}{2}\right)^{2r} \left(\frac{1}{r!}\right)^2.$$

위 분포를 기호로는 $vM(\mu, \kappa)$ 로 나타낸다.

II. 겹친코쉬 분포:

$$g(\theta) = \frac{1}{2\pi} \left\{ 1 + 2 \sum_{k=1}^{\infty} \rho^k \cos k(\theta - \mu) \right\} \\ = \frac{1 - \rho^2}{2\pi \{1 + \rho^2 - 2\rho \cos(\theta - \mu)\}}, \quad 0 \leq \theta < 2\pi.$$

위 식에서 μ 는 위치모수로 $0 \leq \mu < 2\pi$ 이며, ρ 는 척도모수로 $\rho = e^{-\sigma}$ 이고 $0 \leq \rho < 1$ 을 만족한다. 위 분포를 기호로는 $WC(\mu, \rho)$ 로 나타낸다.

III. Papakonstantinou (1979) 분포:

$$f(\theta; \kappa, \nu) = \frac{1}{2\pi} + \frac{\kappa}{2\pi} \sin(\theta + \nu \sin \theta), \quad 0 \leq \theta < 2\pi, \quad |\kappa| \leq 1, \quad |\nu| < 1.$$

여기서, 모수 ν 는 왜도모수로 분포의 치우침을 결정한다. 위 분포를 기호로는 $\text{Papa}(\kappa, \nu)$ 로 나타낸다.

IV. 겹친라플라스 분포:

$$f_w(\theta) = p \frac{\lambda_1 e^{-\lambda_1 \theta}}{1 - e^{-2\pi \lambda_1}} + (1-p) \frac{\lambda_2 e^{\lambda_2 \theta}}{e^{2\pi \lambda_2} - 1}, \quad \theta \in [0, 2\pi)$$

또는

$$f_w(\theta) = \frac{\lambda \kappa}{1 + \kappa^2} \left(\frac{e^{-\lambda \kappa \theta}}{1 - e^{-2\pi \lambda \kappa}} + \frac{e^{(\lambda/\kappa)\theta}}{e^{-2\pi \lambda/\kappa} - 1} \right), \quad \theta \in [0, 2\pi).$$

위 식에서, $\lambda_1, \lambda_2 > 0$ 이고 $p \in [0, 1]$ 이다. 또한

$$p = \frac{1}{\kappa^2 + 1}, \quad \lambda_1 = \lambda \kappa, \quad \lambda_2 = \lambda/\kappa$$

의 관계가 성립한다 (Jammalamadaka와 Kozubowski, 2003).

V. 겹친왜정규 분포:

$$f(\theta; \xi, \eta, \lambda) = \frac{2}{\eta} \sum_{r=-\infty}^{\infty} \phi \left(\frac{\theta + 2\pi r - \xi}{\eta} \right) \Phi \left\{ \lambda \left(\frac{\theta + 2\pi r - \xi}{\eta} \right) \right\}, \quad 0 \leq \theta < 2\pi.$$

위 식에서 $-\infty < \xi < \infty$, $\eta > 0$ 이고 $-\infty < \lambda < \infty$ 이다. 위 분포를 기호로는 $\text{WSN}(\xi, \eta, \lambda)$ 로 나타낸다. 위 분포는 선형의 왜정규분포 (Azzalini, 1985)

$$f(y; \xi, \eta, \lambda) = \frac{2}{\eta} \phi \left(\frac{y - \xi}{\eta} \right) \Phi \left\{ \lambda \left(\frac{y - \xi}{\eta} \right) \right\}, \quad -\infty < y < \infty$$

에서 $\Theta = Y \pmod{2\pi}$ 변환을 통해 유도된다.

표 3.4와 그림 3.3은 주중(편의상 수요일) 시간교통량 자료에 대해 다양한 혼합모형을 적합한 결과를 제시한 것이다. 표 3.4에서 적합모형 PM과 WLM의 경우, 모수 ξ_1 과 ξ_2 는 혼합된 두 분포의 최빈값을 나타내며, 자료의 상위 두 최빈값(봉우리값)을 통해 추정된다. 본 연구에서 혼합모형의 모수추정에는 EM 알고리즘이 사용되었으며, 자세한 과정은 Jang 등 (2007), Na와 Jang (2010a, 2010b)을 참고하기 바란다.

표 3.4. EM 알고리즘을 통한 모수추정 결과

적합모형	모수						로그우도	
vMM	α	μ_1	μ_2	κ_1	κ_2		-17925.45	
	0.34	2.53	4.38	2.32	1.35			
WCM	α	μ_1	μ_2	ρ_1	ρ_2		-18113.87	
	0.39	2.62	4.36	0.57	0.51			
PM	α	κ_1	κ_2	ν_1	ν_2	ξ_1	ξ_2	-17980.85
	0.25	0.69	-0.98	0.38	0.05	1.05	-0.80	
WLM	α	λ_{11}	λ_{12}	λ_{21}	λ_{22}	ξ_1	ξ_2	-17975.68
	0.29	1.38	1.17	1.67	0.81	2.49	4.39	
WSNM	α	ξ_1	ξ_2	η_1	η_2	λ_1	λ_2	-18012.03
	0.22	1.67	5.23	1.38	1.71	4.86	-1.26	

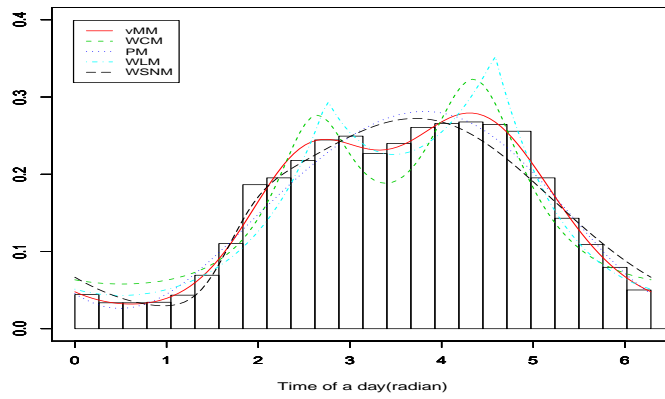


그림 3.3. 혼합원형분포를 통한 시간교통량 모형적합(수요일)

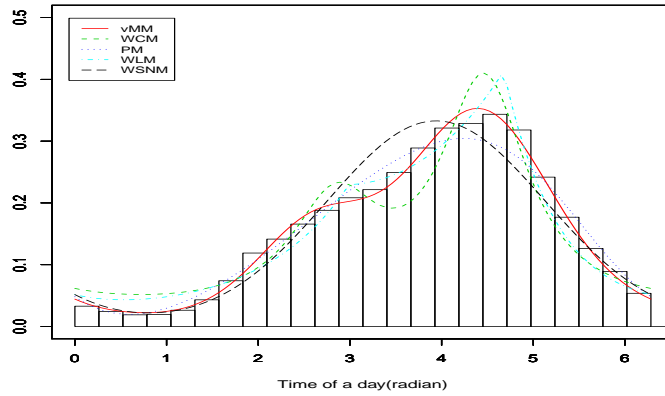


그림 3.4. 혼합원형분포를 통한 시간교통량 모형적합(토요일)

그림 3.4와 그림 3.5는 주말(토요일, 일요일) 자료에 대한 적합 결과이며, 추정된 모수 값은 지면 관계상 생략하였다. 다만, 적합이 뛰어난 대표적 모형에 대한 추정 결과는 표 3.5에 제시되어 있다.

그림 3.3~그림 3.5로부터 시간교통량 자료에 대한 적합모형으로 von Mises 혼합모형과 겹친왜정규 혼

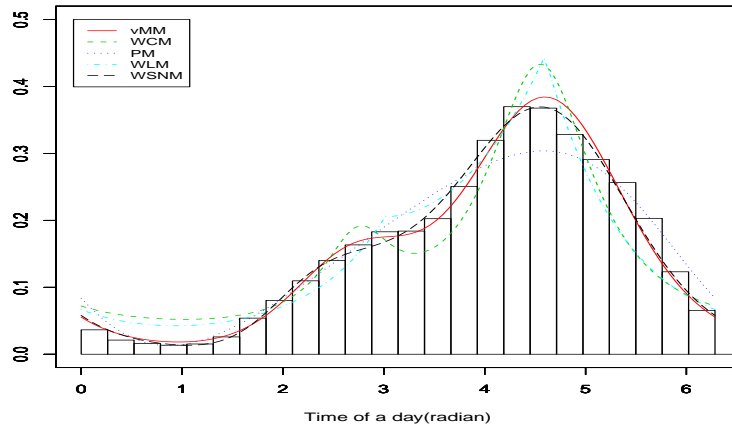


그림 3.5. 혼합원형분포를 통한 시간교통량 모형적합(일요일)

표 3.5. 시간교통량에 대한 요일별 적합 모형

요일	적합모형	적합도검정	
		χ^2	p -값
월	$0.29 \cdot vM(2.39, 3.40) + 0.71 \cdot vM(4.29, 1.38)$	8241.63	0
	$0.28 \cdot WSN(1.58, 1.20, 5.09) + 0.72 \cdot WSN(5.11, 1.40, -1.08)$	5300.09	0
화	$0.30 \cdot vM(2.53, 2.81) + 0.70 \cdot vM(4.34, 1.39)$	3634.15	0
	$0.23 \cdot WSN(1.79, 1.33, 5.04) + 0.77 \cdot WSN(5.16, 1.64, -1.18)$	6854.07	0
수	$0.34 \cdot vM(2.53, 2.32) + 0.36 \cdot vM(4.38, 1.35)$	1793.01	0
	$0.22 \cdot WSN(1.67, 1.38, 4.86) + 0.78 \cdot WSN(5.23, 1.71, -1.26)$	10027.79	0
목	$0.30 \cdot vM(2.50, 2.97) + 0.70 \cdot vM(4.35, 1.41)$	2406.38	0
	$0.24 \cdot WSN(1.89, 1.25, 5.05) + 0.76 \cdot WSN(5.11, 1.59, -1.08)$	11584.07	0
금	$0.32 \cdot vM(2.53, 2.62) + 0.68 \cdot vM(4.42, 1.51)$	3375.46	0
	$0.23 \cdot WSN(1.89, 1.27, 5.09) + 0.77 \cdot WSN(5.36, 1.75, -1.23)$	18801.73	0
토	$0.23 \cdot vM(2.60, 2.59) + 0.77 \cdot vM(4.42, 1.62)$	1717.69	0
	$0.71 \cdot WSN(-2.01, 1.29, -0.54) + 0.29 \cdot WSN(-1.32, 1.29, -0.76)$	16062.39	0
일	$0.20 \cdot vM(2.70, 2.96) + 0.80 \cdot vM(4.61, 1.75)$	5661.07	0
	$0.45 \cdot WSN(1.97, 2.31, 5.43) + 0.55 \cdot WSN(4.25, 0.86, 0.72)$	4715.25	0

합모형이 전반적으로 자료를 잘 적합함을 알 수 있다. 특히 von Mises 혼합모형은 모형의 단순성의 측면에서 장점을 가지며, 겹친왜정규 혼합모형은 비대칭성이 뚜렷한 경우에 더욱 효과적이다. 반면 겹친 라플라스 혼합분포는 붐우리가 매우 급격한 자료의 적합에 적당한 분포라 말할 수 있다 (Na와 Jang, 2010b).

표 3.5는 101 도로의 시간교통량 자료에 대해 다른 모형에 비해 뛰어난 결과를 보이는 두 개 모형(von Mises 혼합모형과 겹친왜정규 혼합모형)의 요일별 적합결과이다.

4. 모형의 성능비교

교통량은 교통 계획과 관리 계획 수립에 관련된 여러 분야에서 활용빈도가 높은 중요한 자료 중의 하나이다. 그러나 측정장비의 오작동, 악천후 또는 주변환경 등에 의해 결측이 발생할 수 있으며, 이 경우 교

통량 예측모형을 통해 결측 문제를 해결할 수 있다. 이 절에서는 기존의 분석 방법과 3절에서 구축된 시간교통량 예측 모형을 비교 평가한 결과를 제시하였다. 모형의 평가의 순서는 먼저, 2004년 자료를 이용하여 각각의 모형을 구축하고, 구축된 모형을 통해 얻어지는 시간교통량의 예측값과 실제의 관측값과의 차이를 비교하였다. 또한 동일한 모형을 2005년 자료에 대해서도 비교 하였다. 비교대상으로 사용될 기존 모형으로는 가변수 회귀모형을 선정하였으며, 평가 지표로는 MAPE와 RMSE를 사용하였다.

4.1. 가변수 회귀모형

시간교통량은 월별, 요일별, 시간대별 요인에 의해 영향을 받는 것으로 가정하고 시간교통량을 종속변수로 월, 요일, 시간을 설명변수로 하여 모형을 구축하였다. 단 월, 요일, 시간변수는 자료의 특성상 가변수로 변환하여 모형식에 포함하였다. 월, 요일, 시간에 따른 교통량에 대해 분산분석 결과, 3차 교호효과(월×요일×시간)는 통계적으로 유의하지 않은 것으로 나타났으며 2차 교호효과(월×요일/월×시간/요일×시간)는 유의한 것으로 나타났다. 따라서 본 연구에서는 주효과와 2차 교호효과를 고려한 가변수 회귀모형과 모형의 복잡성을 줄이기 위해 주효과만 고려한 가변수 회귀모형을 구축하였다. 편의상, 전자를 가변수 회귀모형 I 이라 하고 후자를 가변수 회귀모형 II라 부르기로 한다. 각 모형의 모수의 개수는 전자가 457개, 후자가 40개로 이루어져 있으며 단계적 변수선택법을 이용하여 최적모형을 선정하였다. 가변수 회귀모형 I 은 변수선택 결과, 225개의 모수가 선택되었고 결정계수 R^2 은 0.8272로 나타났다. 가변수 회귀모형 II는 40개 모수가 모두 포함된 모형을 구축하였다. 모형에 대한 설명력은 R^2 가 0.7238으로 가변수 회귀모형 I 에 비해 다소 낮게 나타났다. 적합된 두 회귀식의 구체적인 표현은 복잡성에 비해 큰 의미가 없으므로 생략하기로 한다.

4.2. 성능비교

모형의 추정 결과에 기초한 해당 모형의 설명력 또는 예측력은 추정치와 관측치의 차이, 즉 잔차에 대한 여러 가지 통계량을 산출함으로써 파악할 수 있다. 일반적으로 MAPE(Mean Absolute Percentage Error) 및 RMSE(Root Mean Squared Error) 등이 사용되며, 이러한 통계량의 산출을 통해 모형의 추정 및 예측 과정에서 발생하는 오차의 크기와 방향을 살펴볼 수 있다. 각 평가 지표는 다음과 같이 정의된다.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\text{관측값}_i - \text{예측값}_i}{\text{관측값}_i} \right| \times 100(\%),$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\text{관측값}_i - \text{예측값}_i)^2}{N - 1}}.$$

여기서 N 은 전체 자료의 개수를 의미한다. 이를 기초로 한 가변수 회귀모형과 원형모형에 대한 MAPE와 RMSE를 계산한 결과는 표 4.1과 같다.

표 4.1의 결과를 요약하면 다음과 같다. 먼저 MAPE의 관점에서 von Mises 혼합모형의 성능이 가장 뛰어나며, 겹친왜정규 혼합분포도 가변수 회귀모형 I 과 대등한 수준을 보이고 있다. 또한 가변수 회귀모형 I 의 경우는 모수가 224개나 되는 매우 복잡한 모형인 반면, 제시된 원형모형의 경우는 모수의 수가 5개 또는 7개인 것을 감안할 때 시간교통량의 모형이 크게 단순화되었다고 말할 수 있다. RMSE의 관점에서도 제시된 원형모형이 가변수 회귀모형 I 보다 다소 떨어지나, 여전히 모수가 40개인 가변수 회귀모형 II 보다 뛰어난 성능을 보이고 있다.

표 4.1. 시간교통량 예측모형의 MAPE와 RMSE

모형		도로별평가지표			
		2004년		2005년	
		MAPE(%)	RMSE	MAPE(%)	RMSE
가변수	가변수 회귀 I	27.42	146.99	27.67	219.53
회귀모형	가변수 회귀 II	41.13	189.50	37.06	239.44
원형분포모형	vMM	25.90	159.93	22.94	221.23
	WSNM	30.73	167.12	25.90	226.97

5. 맺음말

본 연구에서는 우리나라 지방국도 가운데 지방부 도로에 속하는 101 도로의 특정 지점에서 일년간 측정된 시간대별 교통량 자료를 이용하여 일일교통량에 대한 모형화 분석을 시도하였다. 특히 본 연구에서는 시간교통량 자료를 원형자료로 전환하여 분석을 수행하였으며, 그 결과 기존의 회귀모형을 크게 개선할 수 있음을 보였다. 또한 제시된 원형모형은 기존의 가변수 회귀모형에 비해 매우 단순하며, 교통량의 변화와 도로의 형태(도시부, 지방부, 관광부)에 따라 유연하게 적합될 수 있다. 따라서 제시된 방법을 이용하여 각 도로의 지점별 모형을 구축한다면, 교통량 예측 등의 도로의 효율적 운영에 기여할 수 있을 것으로 기대된다.

참고문헌

- Azzalini, A. (1985). A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, **12**, 171-178.
- Batschelet, E. (1981). *Circular Statistics in Biology*, Academic Press, London.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*, Cambridge University Press.
- Jammalamadaka, S. R. and Kozubowski, T. J. (2003). A new family of circular models : The wrapped Laplace distributions, *Advances and Application in Statistics*, **3**, 77-103.
- Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*, World Scientific.
- Jang, Y. M., Yang, D. Y., Lee, J. Y. and Na, J. H. (2007). Modelling on multi-modal circular data using von Mises mixture distribution, *The Korean Communications in Statistics*, **14**, 517-530.
- Mardia, K. V. (1972). *Statistics of Directional Data*, Academic Press, New York.
- Mardia, K. V. and Jupp, P. E. (1999). *Directional Statistics*, Wiley.
- McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*, Wiley.
- Na, J. H. and Jang, Y. M. (2010a). Modeling on asymmetric circular data using wrapped skew-normal mixture, *Journal of the Korean Data & Information Science Society*, **21**, 241-250.
- Na, J. H. and Jang, Y. M. (2010b). Modeling sharply peaked asymmetric multi-modal circular data using wrapped Laplace mixture, *Journal of the Korean Data & Information Science Society*, **21**, 863-871.
- Papakonstantinou, V. (1979). *Bieträge zur Zirkulären Statistik*, PhD Dissertation, University of Zurich, Switzerland.
- Pewsey, A. (2000). The wrapped skew-normal distribution on the circle, *Communications in Statistics: Theory and Methods*, **29**, 2459-2472.
- Titterton, D. M., Smith, A. F. M., and Makov, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, Chichester.

Modeling on Daily Traffic Volume of Local State Road Using Circular Mixture Distributions

Jong-Hwa Na¹ · Young-Mi Jang²

¹Department of Information & Statistics, Chungbuk National University

²Korea Health and Welfare Information Service

(Received March 2011; accepted April 2011)

Abstract

In this paper we developed a statistical model for traffic volume data which collected from a spot of specific local state road. One peculiar property of daily traffic data is that it has bimodal shape which have two peaks on times of both going to office and coming back to home. So, various mixture models of circular distribution are suggested for bimodal traffic data and EM algorithms are applied to estimate the parameters of the suggested models. To compare the accuracy of the suggested models, classical regressions with dummy variables are also considered. The suggested models for traffic volume data can be effectively used to estimate missing values due to measuring instrument disorder.

Keywords: Traffic volume, circular distribution, von Mises, wrapped skew-normal, EM algorithm.

This work was supported by the research grant of the Chungbuk National University in 2010.

¹Corresponding author: Professor, Department of Information & Statistics, Chungbuk National University, Cheong-ju, Chungbuk 361-763, Korea. E-mail: cherin@cbu.ac.kr