

Modeling Circular Data with Uniformly Dispersed Noise

Hye-Kyung Yu¹ · Kyoung-Ho Jun² · Jong-Hwa Na³

¹Department of Information & Statistics, Chungbuk National University

²Office of Admission Management, Chungbuk National University

³Department of Information & Statistics, Chungbuk National University

(Received May 22, 2012; Revised June 22, 2012; Accepted July 23, 2012)

Abstract

In this paper we developed a statistical model for circular data with noises. In this case, model fitting by single circular model has a lack-of-fit problem. To overcome this problem, we consider some mixture models that include circular uniform distribution and apply an EM algorithm to estimate the parameters. Both von Mises and Wrapped skew normal distributions are considered in this paper. Simulation studies are executed to assess the suggested EM algorithms. Finally, we applied the suggested method to fit 2008 EHFRS(Epidemic Hemorrhagic Fever with Renal Syndrome) data provided by the KCDC(Korea Centers for Disease Control and Prevention).

Keywords: Circular data, von Mises, wrapped skew-normal, mixture distribution, EM algorithm.

1. 서론

원형자료(circular data)는 크기가 무시된 연속형 자료로, 2차원의 방향자료를 의미한다. 지금까지 많은 응용분야에서 원형자료의 적합과 관련된 연구가 진행되어 왔다. 단봉형 자료에 대한 원형분포로는 대칭형의 경우 von Mises 분포를 중심으로 겹친정규, 겹친코쉬, 원형균일분포 등이 있으며, 비대칭의 경우에는 Batschelet, α -stable 분포와 겹친왜정규분포 등을 들 수 있다. 이 가운데 겹친왜정규분포는 Pewsey (2000)에 의해 처음 소개된 분포로, 분포가 가지는 유연성으로 인해, 최근 많은 응용분야에서 연구가 이루어지고 있다 (Pewsey, 2004). 또한 다봉형 원형자료의 적합에 대한 연구로는 Jennifer 등 (2003), Jang 등 (2007)과 Na와 Jang (2010, 2011)의 연구를 들 수 있으며, 이들 연구는 주로 von Mises와 겹친왜정규 등 동일 계열의 분포간의 혼합분포의 적합 과정을 다루었다. 원형자료의 분석과 관련된 주요 문헌으로는 Mardia (1972), Batschelet (1981), Fisher (1993), Jammalamadaka와 Sen-Gupta (2001) 등이 있다.

한편 원형자료가 범위의 전 영역(방향)에서 일정한 빈도로 산재하며, 동시에 특정방향으로의 집중현상을 보이는 경우를 생각해보자. 이러한 경우는 일정 기간 동안에 발생하는 특정 질병의 발생 빈도의 자

This work was supported by the research grant of the Chungbuk National University in 2011.

³Corresponding author: Professor, Department of Information & Statistics, Chungbuk National University, Cheong-ju, Chungbuk 361-763, Korea. E-mail: cherin@cbu.ac.kr

료 등에서 흔히 나타난다. 예를 들어, 본 논문에서 다루어지는 신증후군출혈열 자료의 경우 한해의 후반부(10월 12월)에 집중되는 현상과 함께, 일정비율로 연간 꾸준히 발생하는 패턴을 동시에 가진다. 이 경우 단일 원형분포만으로는 적합은 적합결여의 문제를 야기하며, 원형균일분포와의 혼합분포를 고려하는 것이 바람직할 것이다.

따라서 본 논문에서는 원형균일(Circular Uniform; CU) 분포와 대표적인 단일 원형분포와의 혼합분포(mixture distribution)를 제안하고, 모수추정 과정을 비롯하여 실제 자료의 적합문제를 다루기로 한다. 원형균일과의 혼합분포에 고려되는 분포로는 대칭형의 von Mises 분포와 최근 많은 연구가 이루어지고 있는 비대칭의 겹친왜정규분포를 고려하였다. 2절에서는 이들 분포와 원형균일과의 혼합분포에 대해 소개하고, 3절에서는 EM 알고리즘을 통한 모수추정 문제를 다루었다. 4절에서는 모의실험을 통한 모수추정 결과와 실증분석을 다루고, 5절은 결론으로 구성되었다.

2. 원형균일(CU)과의 혼합분포

2.1. vM와 CU 혼합분포

대칭인 von Mises(vM) 분포와 원형균일(CU)과의 혼합분포는 그 밀도함수가 다음과 같이 주어진다.

$$f(\theta; p, \mu, \kappa) = pf_{VM}(\theta; \mu, \kappa) + (1-p)f_{CU}(\theta),$$

여기서 $p \in [0, 1]$ 는 혼합확률이며, $f_{VM}(\theta; \mu, \kappa)$ 와 $f_{CU}(\theta)$ 는 각각 von Mises와 원형균일분포의 밀도함수이다. 위의 식을 보다 구체적으로 표현하면 다음과 같다.

$$f(\theta; p, \mu, \kappa) = p \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\} + (1-p) \frac{1}{2\pi}, \quad 0 \leq p \leq 1, \kappa > 0. \quad (2.1)$$

위 식에서 $I_0(\kappa)$ 는 차수 0의 제1종 수정된 Bessel 함수로 다음과 같이 정의된다.

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp\{\kappa \cos(\theta)\} d\theta.$$

2.2. WSN과 CU 혼합분포

겹친왜정규(WSN)와 원형균일(CU) 분포의 혼합분포는 그 밀도함수가 다음과 같이 주어진다.

$$f(\theta; p, \xi, \eta, \lambda) = pf_{WSN}(\theta; \xi, \eta, \lambda) + (1-p)f_{CU}(\theta),$$

여기서 $p \in [0, 1]$ 는 혼합확률이며, $f_{WSN}(\theta; \xi, \eta, \lambda)$ 와 $f_{CU}(\theta)$ 는 각각 겹친왜정규와 원형균일분포의 밀도함수이다. 위의 식을 자세히 표현하면 다음과 같다.

$$f(\theta; p, \xi, \eta, \lambda) = p \frac{2}{\eta} \sum_{r=-\infty}^{\infty} \phi\left(\frac{\theta + 2\pi r - \xi}{\eta}\right) \Phi\left\{\lambda\left(\frac{\theta + 2\pi r - \xi}{\eta}\right)\right\} + (1-p) \frac{1}{2\pi}, \quad (2.2)$$

$$0 \leq p \leq 1, \quad -\infty < \xi < \infty, \quad \eta > 0, \quad -\infty < \lambda < \infty.$$

3. EM 알고리즘을 통한 혼합분포의 모수추정

이 절에서는 앞서 소개한 두 종류의 혼합분포에 대한 최다가능도추정치를 구하는 방법으로 EM 알고리즘을 제시한다.

3.1. EM 알고리즘 소개

Dempster 등 (1977)에 의해 소개된 EM 알고리즘은 특정분포의 모수에 대한 최대가능도 추정에 유용한 알고리즘이다. 특히 관측 과정의 제약에 의해 결측자료가 발생하는 경우, 또는 추정해야 할 모수의 수가 많아지거나 우도함수의 형태가 복잡하여 직접 최대화가 어려운 경우에 대한 효과적인 해결책이 된다. EM 알고리즘은 먼저 결측자료를 관측자료에 보완하여 최대화가 보다 쉬운 완비우도함수를 정의하고, E-단계(Expectation Step)와 M-단계(Maximization Step)로 불리는 두 단계를 반복적으로 수행하여 모수를 추정하는 수치 해석적 방법이라 할 수 있다.

EM 알고리즘의 절차는 다음과 같다. 먼저 관측자료(또는 불완비 자료)를 X , 결측자료(또는 잠재자료)를 Y 라 하면 완비자료는 (X, Y) 로 표현할 수 있으며, 추정해야 할 모수의 집합을 θ 라 하면 관측자료의 우도함수는 $L(X|\theta)$ 으로, 완비자료의 우도함수는 $L(X, Y|\theta)$ 으로 나타낼 수 있다. EM 알고리즘의 E-단계는 관측된 우도함수를 보완하기 위해 결측자료를 보완(augment)하는 단계로 다음과 같이 정의되는 조건부 기댓값인 Q함수를 계산한다.

$$Q(\theta, \theta^r) \equiv E_{Y|X, \theta^r} [\ln L(X, Y|\theta) | X = x, \theta = \theta^r], \quad (3.1)$$

여기서 θ^r (θ^0 은 초기치임)은 현 단계에 주어진 모수 추정치의 집합(초기값의 집합)을 나타내며, Q함수는 조건부 분포 $Y|X = x, \theta = \theta^r$ 에 대한 (조건부)기댓값을 의미한다. M-단계에서는 $Q(\theta, \theta^r)$ 를 최대로 하는 θ 를 계산하여 새로운 모수 추정치 θ^{r+1} 로 갱신한다. 이제 새로 갱신된 모수값을 이용하여 E-단계를 다시 수행하고, 계속해서 M-단계를 수행하게 된다. 이 과정을 θ 가 특정한 값으로 수렴할 때까지 반복 수행한다. 즉, 충분히 작은 ϵ 에 대해, 다음의 정지규칙

$$Q(\theta^{r+1}, \theta^r) - Q(\theta^r, \theta^{r-1}) \leq \epsilon \quad (3.2)$$

를 만족할 때까지 반복수행하여 최우추정치를 구하게 된다.

3.2. 혼합분포의 모수추정

일반적으로 k 개의 단일분포를 가지는 혼합분포의 형태는 다음과 같다.

$$f(\theta|\lambda) = \sum_{l=1}^k p_l f_l(\theta|\lambda_l),$$

여기서 λ 는 모수집합으로 $\lambda = \{p_1, p_2, \dots, p_k, \lambda_1, \lambda_2, \dots, \lambda_k\}$ 이고, p_l 은 각 단일분포의 혼합비율로 $\sum_{l=1}^k p_l = 1$ 을 만족한다. 위 혼합분포로부터의 표본을 $\theta = \{\theta_1, \theta_2, \dots, \theta_k\}$ 이라 할 때, 로그우도함수는 다음과 같다.

$$\ln L(\lambda|\theta) = \ln \prod_{i=1}^n f(\theta_i|\lambda) = \sum_{i=1}^n \ln \left(\sum_{l=1}^k p_l f_l(\theta_i|\lambda_l) \right).$$

먼저 관측자료 θ 를 불완비자료로, 관측되지 않은 $y = \{y_1, y_2, \dots, y_n\}$ 를 잠재자료라고 하자. 각 i 에 대해 $y_i \in 1, 2, \dots, k$ 이고, i 번째 관측치 θ_i 가 l 번째 분포 f_l 로부터 생성되었다면 $y_i = l$ 로 주어진다. 잠재자료 y 의 값을 아는 경우 완비자료 (θ, y) 의 로그우도함수는

$$\ln L(\lambda|\theta) = \ln(f(\theta, y|\lambda)) = \sum_{i=1}^n \ln(p_{y_i} f_{y_i}(\theta_i|\lambda_{y_i}))$$

으로 표현되며, 위 식에 대한 최대화는 쉽게 수행될 수 있게 된다. 그러나 일반적으로 각 자료 θ_i 에 대응하는 잠재 자료 y_i 의 값을 모르기 때문에 $Y|(\theta, \lambda)$ 의 분포를 추정한 후 이로부터 식 (3.1)의 조건부 기댓값인 Q함수를 구하고(E-단계), Q함수를 모수 λ 에 대해 최대화하는 단계(M-단계)를 거쳐 모수 λ 의 값을 갱신하게 된다.

위 과정에서 $Y|(\theta, \lambda)$ 의 분포는 $\lambda^g = (p_1^g, p_2^g, \dots, p_k^g, \lambda_1^g, \lambda_2^g, \dots, \lambda_k^g)$ 가 주어질 때, 각 i 와 l 에 대해 다음과 같이 주어진다.

$$p(l|\theta_i, \lambda^g) = \frac{p_l^g f_l(\theta_i|\lambda_l^g)}{f(\theta_i|\lambda^g)} = \frac{p_l^g f_l(\theta_i|\lambda_l^g)}{\sum_{l=1}^k p_l^g f_l(\theta_i|\lambda_l^g)}.$$

식 (3.1)을 이용하여 조건부 기댓값 Q를 계산하면

$$\begin{aligned} Q(\lambda, \lambda^g) &= \sum_{i=1}^n E_{Y_i|\theta_i, \lambda^g}[\ln p_{y_i} f_{y_i}(\theta_i|\lambda_{y_i})] \\ &= \sum_{l=1}^k \sum_{i=1}^n \ln(p_l f_l(\theta_i|\lambda_l)) p(l|\theta_i, \lambda^g) \\ &= \sum_{l=1}^k \sum_{i=1}^n \ln(p_l) p(l|\theta_i, \lambda^g) + \sum_{l=1}^k \sum_{i=1}^n \ln(f_l(\theta_i|\lambda_l)) p(l|\theta_i, \lambda^g) \end{aligned} \quad (3.3)$$

이 되고, Q함수의 최대화를 통해 모수 λ 를 갱신하게 된다. 이 과정을 식 (3.2)의 정지규칙이 만족될 때까지 반복하여 모수 λ 에 대한 최다가능도추정치를 얻게 된다.

3.3. 원형혼합분포의 모수추정

여기서는 2절에서 소개된 두 종류의 원형혼합분포에 대한 모수추정 과정을 다루기로 한다. 두 경우 모두 앞 절의 전개에서 $k=2$ 인 경우에 해당한다.

3.3.1. 혼합비율의 추정 EM 알고리즘의 각 단계에서 혼합비율 p_1, p_2 에 대한 추정(또는 갱신)은 다음과 같다. 혼합비율의 추정은 식 (3.3)의 첫 번째 항에만 관련되며, 제약조건 $p_1 + p_2 = 1$ 하에서, 라그랑지 방법을 통해 다음과 같이 추정된다.

$$\hat{p}_1 = \frac{1}{n} \sum_{i=1}^n p(y_i = 1|\theta_i, \lambda^g), \quad \hat{p}_2 = 1 - \hat{p}_1. \quad (3.4)$$

3.3.2. vM와 CU 혼합분포의 모수추정 식 (2.1)의 혼합분포에서 EM 알고리즘의 각 단계에서 혼합비율을 제외한 나머지 모수들에 대한 추정 과정은 다음과 같다. 모수 μ 와 κ 의 추정은 식 (3.3)의 두 번째 항의 최대화를 통해 구해진다. 식 (3.3)의 두 번째 항은

$$T = \sum_{i=1}^n [\log(f_{vM}(\theta_i; \mu, \kappa))p(y_i = 1|\theta_i, \lambda^0) + \log(f_{CU}(\theta_i))p(y_i = 2|\theta_i, \lambda^0)]$$

으로 표현되며, 이를 최대로 하여주는 모수를 추정하기 위해 μ 와 κ 로 미분하여 0으로 놓으면

$$\begin{aligned} \frac{\partial T}{\partial \mu} &= \sum_{i=1}^n \frac{\frac{\partial}{\partial \mu} f_{vM}(\theta_i; \mu, \kappa)}{f_{vM}(\theta_i; \mu, \kappa)} p(y_i = 1|\theta_i, \lambda^0) \equiv 0, \\ \frac{\partial T}{\partial \kappa} &= \sum_{i=1}^n \frac{\frac{\partial}{\partial \kappa} f_{vM}(\theta_i; \mu, \kappa)}{f_{vM}(\theta_i; \mu, \kappa)} p(y_i = 1|\theta_i, \lambda^0) \equiv 0 \end{aligned}$$

Table 4.1. Design for ML estimation of mixture model via EM algorithm

mixture of vM and CU	mixture of WSN and CU
$0.6 \cdot \text{vM}(3, 5) + 0.4 \cdot \text{CU}$	$0.7 \cdot \text{WSN}(3, 2, 3) + 0.3 \cdot \text{CU}$

Table 4.2. ML estimation via EM algorithm(mixtur of vM and CU)

parameter	p	μ	κ
true value	0.6	3	5
$n = 50$	0.623	3.162	5.467
$n = 100$	0.600	3.134	5.350
$n = 500$	0.608	2.959	4.914
$n = 1000$	0.620	3.016	5.008

이 된다. 위 식의 근을 구해 모수를 갱신하게 되고, 이 과정의 반복을 통해 최대가능도추정치를 구할 수 있다.

3.3.3. WSN과 CU 혼합분포의 모수추정 식 (2.2)의 혼합분포에서 EM 알고리즘의 각 단계에서 혼합비율을 제외한 나머지 모수들에 대한 추정 과정은 다음과 같다. 모수 ξ, η 및 λ 의 추정은 식 (3.3)의 두 번째 항의 최대화를 통해 구해진다. 식 (3.3)의 두 번째 항은

$$T = \sum_{i=1}^n [\log(f_{WSN}(\theta_i; \xi, \eta, \lambda))p(y_i = 1|\theta_i, \lambda^0) + \log(f_{CU}(\theta_i))p(y_i = 2|\theta_i, \lambda^0)]$$

으로 표현되며, 이를 최대로 하는 모수를 추정하기위해 ξ 와 η, λ 로 미분하여 0으로 놓으면

$$\begin{aligned} \frac{\partial T}{\partial \xi} &= \sum_{i=1}^n \frac{\frac{\partial}{\partial \xi} f_{WSN}(\theta_i; \xi, \eta, \lambda)}{f_{WSN}(\theta_i; \xi, \eta, \lambda)} p(y_i = 1|\theta_i, \lambda^0) \equiv 0, \\ \frac{\partial T}{\partial \eta} &= \sum_{i=1}^n \frac{\frac{\partial}{\partial \eta} f_{WSN}(\theta_i; \xi, \eta, \lambda)}{f_{WSN}(\theta_i; \xi, \eta, \lambda)} p(y_i = 1|\theta_i, \lambda^0) \equiv 0, \\ \frac{\partial T}{\partial \lambda} &= \sum_{i=1}^n \frac{\frac{\partial}{\partial \lambda} f_{WSN}(\theta_i; \xi, \eta, \lambda)}{f_{WSN}(\theta_i; \xi, \eta, \lambda)} p(y_i = 1|\theta_i, \lambda^0) \equiv 0 \end{aligned}$$

이 된다. 위 식의 근을 구해 모수를 갱신하게 되고, 이 과정의 반복을 통해 최대가능도추정치를 구할 수 있다.

4. 모의실험과 실제자료분석

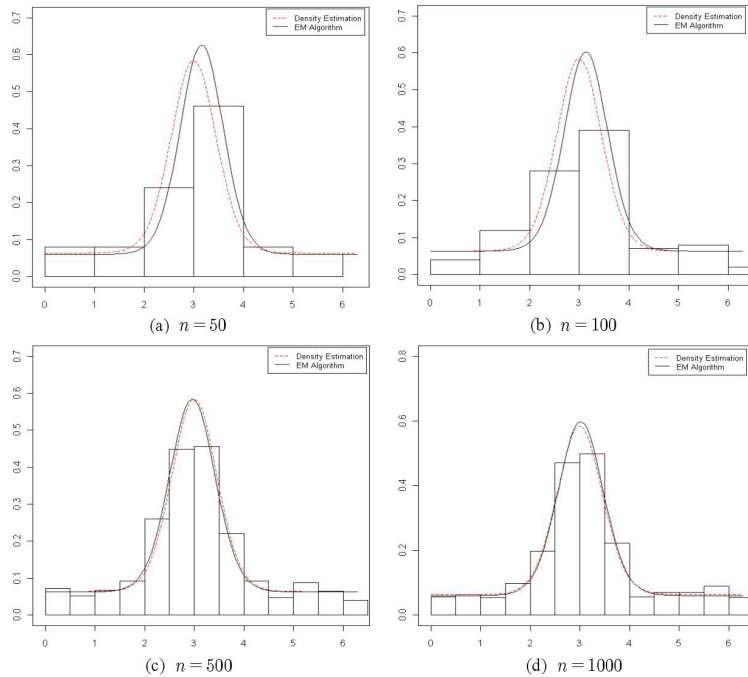
4.1. 모의실험

앞서 소개한 두 혼합모형의 EM 알고리즘을 이용한 최대가능도 추정에 대한 모의실험을 수행하였다. 모의실험에서 사용된 혼합 모집단의 분포는 Table 4.1과 같다.

Table 4.1의 혼합분포로부터 표본의 수가 $n = 50, 100, 500, 1000$ 인 난수를 발생하고, 각각의 경우에 대해 EM 알고리즘을 이용한 최대가능도 추정을 실시하였다. 아래의 Table 4.2~Table 4.3과 Figure 4.1~Figure 4.2는 각 혼합모형에 대해 EM 알고리즘을 통한 모수추정 결과를 나타낸다.

Table 4.3. ML estimation via EM algorithm(mixture of WSN and CU)

parameter	p	ξ	η	λ
true value	0.7	3	2	3
$n = 50$	0.665	2.639	1.896	3.294
$n = 100$	0.736	2.842	1.945	3.157
$n = 500$	0.715	2.980	1.859	2.943
$n = 1000$	0.705	2.975	1.940	2.893

**Figure 4.1.** Fitted model for mixture of vM and CU

4.2. 실제자료분석

본 연구에서 사용된 자료는 한국질병관리본부에서 2008년도에 전국에 걸쳐 조사한 신증후군출혈열(Epidemic Hemorrhagic Fever with Renal Syndrome; EHFRS) 자료이다. 신증후군출혈열은 고열을 동반한 단백뇨나 전신성의 출혈 경향 등을 보이는 바이러스성 전염병으로써 유행성출혈열이라는 이름으로 널리 알려져 있다. 한국에서 매년 발생하고 있는 무서운 제 3군 법정전염병으로 사망률은 7% 나 된다. Table 4.4는 2008년도 1월 1일부터 12월 31일까지 366일 동안 전국에 걸쳐 조사된 신증후군출혈열의 일별 발생빈도를 나타낸다. 본 연구에서는 1년에 걸친 시간 자료를 원형자료로 취급하여 분석을 수행하기로 한다. 원형자료로의 전환은 1년 366일을 $0^\circ \sim 360^\circ$ 에 대응시켜 구하게 되며, t 번째 일은

$$360^\circ \times \left(\frac{t}{366} \right)$$

을 통해 원형자료로 변환된다. Table 4.4에서 맨 우측 열은 원형자료로의 변환결과를 나타낸다.

위 자료를 월별 빈도로 요약한 결과는 Table 4.5와 같다. 또한 위 자료를 원형자료로 취급하여 로즈

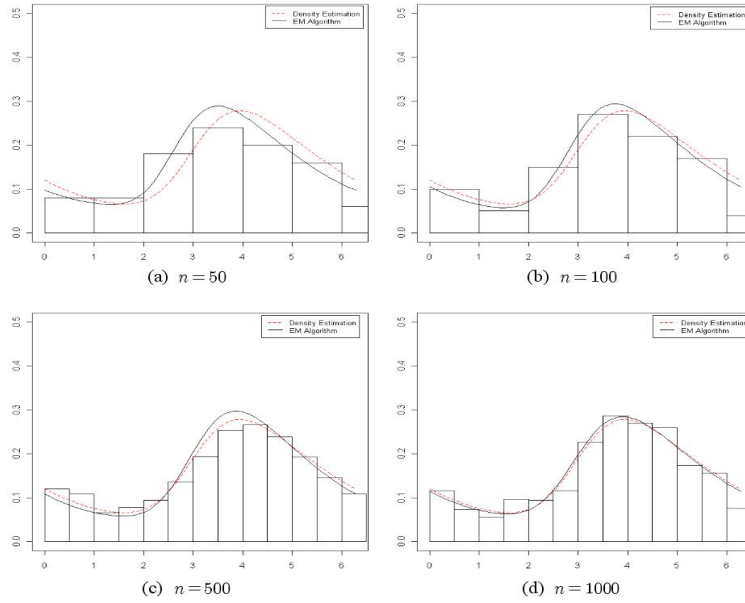


Figure 4.2. Fitted model for mixture of WSN and CU

Table 4.4. Epidemic hemorrhagic fever with renal syndrome data

No	month	day	frequency	angle
1	1	1	0	0.98°
2	1	2	1	1.96°
3	1	3	2	2.95°
4	1	4	3	3.93°
5	1	5	1	4.91°
⋮	⋮	⋮	⋮	⋮
364	12	29	7	358.03°
365	12	30	1	359.01°
366	12	31	1	360.00°

Table 4.5. Monthly frequency of EHFRS data

month	Jan	Feb	Mar	Apl	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
freq.	27	8	6	5	10	13	10	11	17	80	130	58	375

다이어그램(rose diagram)을 포함한 순환 점도표(circular dot plot)와 히스토그램으로 요약한 결과는 Figure 4.3과 같다. 그 결과 신증후군출혈열 자료는 특정 시점(10월 중순~12월 하순)에 집중되는 현상과 함께, 전 범위에 걸쳐 꾸준히 발생하는(잡음이 개입된) 패턴을 보이고 있다. 따라서 이 자료의 적합에는 본 논문에서 제시한 균일분포와의 혼합모형과 함께 비대칭의 경향도 어느 정도 고려하는 것이 좋을 것으로 판단된다.

Table 4.6은 위 자료에 대해 기초통계를 구한 것으로, 평균방향(mean direction) 및 원형분산(circular variance)을 나타낸다. 이들 통계량에 대한 자세한 정의는 Jammalamadaka와 SenGupta (2001)를 참

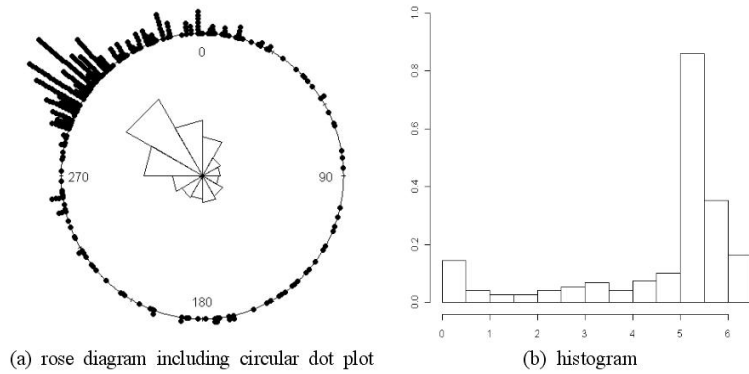


Figure 4.3. Summary of EHFRS data

Table 4.6. Descriptive statistics for EHFRS data

N	mean direction	circular variance
375	5.452 (312.41°)	0.351

Table 4.7. Monthly frequency of EHFRS data

Fitted model	estimates				log likelihood	χ^2 p -value
	\hat{p}	$\hat{\mu}$	$\hat{\kappa}$	$\hat{\lambda}$		
vM	—	5.452(312.41°)	1.726	—	-509.00	0.000
(vM+CU) mixture	0.696	5.464(312.86°)	7.186	—	-464.29	0.061
WSN	—	6.347(364.39°)	1.374	-1.290	-539.07	0.000
(WSN+UC) mixture	0.645	5.067(290.34°)	0.533	5.575	-444.76	0.073

고하기 바란다.

Table 4.7은 위 자료를 다양한 원형분포로 적합한 결과를 나타낸다. 먼저 χ^2 검정 결과 두 종류의 단일 모형은 모두 적합결여를 나타내며, 두 종류의 혼합모형은 모두 자료를 잘 적합하는 것으로 나타났다. 또한 로그우도의 값을 통해 비대칭의 특징을 반영한 겹친왜정규와의 혼합분포가 (대칭인) von Mises와의 혼합분포의 경우 보다 더욱 개선된 모형임을 알 수 있다.

Figure 4.4는 위의 적합결과를 그림으로 나타낸 것이다.

5. 결론

본 논문에서는 잡음이 개입된 원형자료의 모형화 문제를 다루었다. 이 경우 단일 원형모형의 적합시 발생하는 적합결여의 문제를 해결하기 위해 원형균일분포와 대칭형의 von Mises 및 비대칭의 겹친왜정규 분포와의 혼합분포를 제안하고, 이들 혼합분포의 모수추정을 위해 EM 알고리즘을 제안하였다. 모의실험을 통해 제시된 EM 알고리즘의 정도를 확인하였으며, 실제자료(신증후군출혈열 자료)의 분석을 통해 그 유용성과 함께 다양한 분야의 적용 가능성을 확인하였다.

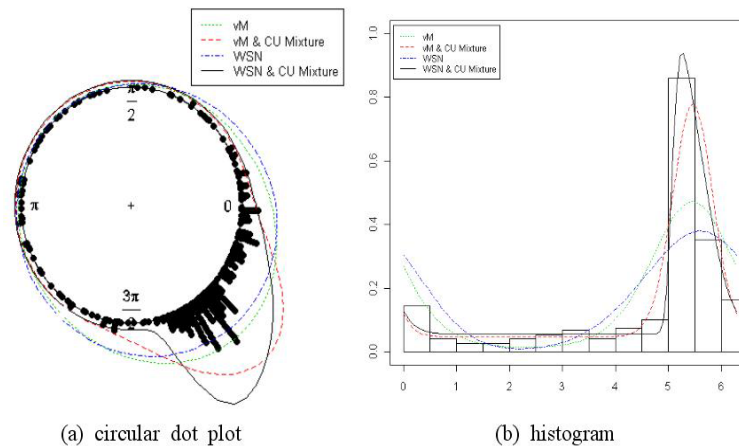


Figure 4.4. Fitted model for EHFRS data

References

- Batschelet, E. (1981). *Circular Statistics in Biology*, Academic Press, London.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.
- Fisher, N. I. (1993). *Statistical Analysis of Circular Data*, Cambridge University Press.
- 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.
- Jennifer, A. M., Peter, J. H. and Ian, T. J. (2003). Fitting mixtures of von Mises distribution: A case study involving sudden infant death syndrome, *Computational Statistics & Data Analysis*, **41**, 505–513.
- Mardia, K. V. (1972). *Statistics of Directional Data*, Academic Press, New York.
- Na, J. H. and Jang, Y. M. (2010). 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. (2011). Modeling on daily traffic volume of local state road using circular mixture distributions, *The Korean Journal of Applied Statistics*, **24**, 547–557.
- Pewsey, A. (2000). The wrapped skew-normal distribution on the circle, *Communications in Statistics: Theory and Methods*, **29**, 2459–2472.
- Pewsey, A. (2004). Modelling asymmetrically distributed circular data using the wrapped skew-normal distribution, *Environmental and Ecological Statistics*, **13**, 257–269.