



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

박만식교수지도

석사학위 청구논문

주기도를 이용한 시계열자료의
군집분석

2016

성신여자대학교 대학원

통계학과

권수현

주기도를 이용한 시계열자료의 군집분석

박만식교수지도

이 논문을 석사학위 논문으로 제출함

2015년 11월

성신여자대학교 대학원

통계학과

권수현

인준서

권수현의 석사학위 논문으로 인준함.

심사위원 _____인

심사위원 _____인

심사위원 _____인

성신여자대학교 대학원

논문개요

시계열 자료를 군집화할 때, 시계열 자료의 계열수가 많은 경우에는 자료의 특성이 유사한 시계열들을 군집화한 후 이를 사전정보로 활용하여 모형설정 및 예측을 수행한다면 훨씬 효율적일 것이다. 시계열 자료의 군집분석 방법에는 다양한 거리가 정의될 수 있으며, 크게 두 가지로 구분한다. 첫째, 시간 영역에서의 거리로서 시계열 자료의 특성인 자기상관함수, 부분자기상관함수로부터 계열 간의 유사성을 측정한다. 둘째, 주파수 영역에서의 거리로서 자기공분산함수를 통해 얻어지는 주기도를 이용하여 계열 간의 유사성을 측정한다.

본 연구에서는 기존 시계열 자료를 군집화하는 데 사용된 거리들을 간략히 소개하고, 주기도 간의 상관계수를 이용한 거리를 제안한 후 모의실험을 통해 그 성능을 측정하였다. 또한, 1990년 1월부터 2015년 4월까지 전국 월평균 제조업생산지수 자료를 대상으로 기존의 거리와 본 연구에서 제안한 거리를 이용하여 지수자료를 군집화하였다.

주요어 : 군집분석, 유사성, 스펙트럼, 평활창, 주파수 영역, 주기도, 상관계수, 정상시계열

목 차

I.	서론	1
II.	유사성 거리척도 정의	7
2.1	시간 영역에서의 거리	7
2.1.1	자기상관에 근거한 거리	7
2.2	주파수 영역에서의 거리	9
2.2.1	주기도에 근거한 거리	9
2.2.2	평활된 주기도에 근거한 거리	13
2.2.3	주기도의 상관성에 근거한 거리	16
III.	모의실험	19
3.1	모의실험 방법	19
3.2	모의실험 결과	23
IV.	실증연구	33
4.1	실증 자료	33
4.2	실증연구 방법	35
4.3	실증연구 결과	35
V.	결론	42
	참고 문헌	44
	Abstract	47

부록 48

그림 목 차

그림 1.	$m = 2$ 일 때 $\tilde{J}_n(v_k)$ 의 계산과정	13
그림 2.	평활창에 따른 가중치 $W_n(h)$	15
그림 3.	시나리오별 이론적인 스펙트럼 밀도함수	22
그림 4.	24개 산업의 월평균 제조업생산지수 자료의 시도표. 34	
그림 5.	실증자료 군집분석 덴드로그램 : 유클리드 거리 . .	36
그림 6.	실증자료 군집분석 덴드로그램 : 피어슨 상관계수 .	37
그림 7.	실증자료 군집분석 덴드로그램 : 급내 상관계수 . .	38
그림 8.	1% $N\tilde{J}_U$ 를 이용한 군집분석 결과.	39
그림 9.	1% $LN\tilde{J}_p$ 를 이용한 군집분석 결과.	40
그림 10.	1% \tilde{J}_C 를 이용한 군집분석 결과.	41
그림 11.	J_U 정분류율	49
그림 12.	NJ_U 정분류율	50
그림 13.	LNJ_U 정분류율	51
그림 14.	KD 정분류율	52
그림 15.	J_p 정분류율	53
그림 16.	LNJ_p 정분류율	54
그림 17.	J_C 정분류율	55
그림 18.	NJ_C 정분류율	56
그림 19.	LNJ_C 정분류율	57

표 목 차

표 1.	모의실험 시나리오	20
표 2.	시나리오1의 모의실험 결과	25
표 3.	시나리오2의 모의실험 결과	26
표 4.	시나리오3의 모의실험 결과	27
표 5.	시나리오4의 모의실험 결과	28
표 6.	시나리오5의 모의실험 결과	29
표 7.	시나리오6의 모의실험 결과	30

제 1 장

서론

군집분석(clustering analysis)은 주어진 개체 중에서 유사한 것들을 몇몇의 집단으로 군집화하여, 각 집단의 성격을 파악함으로써 자료 전체의 구조에 대한 이해를 돕는 통계 분석방법이다. 군집분석은 자료 그 자체에만 의존하여 자료 탐색과 자료를 요약하는, 사전에 정의된 어떠한 특수한 목적이 없으며, 전체 자료를 군집을 통해 잘 구분하는 것이 분석 목적이다. 따라서, 동일한 군집 내의 개체들은 유사한 성질을 갖도록, 또한 서로 다른 군집에 속한 개체들 사이에는 상대적으로 서로 다른 성격을 갖도록 군집이 형성되어야 한다. 군집분석에서는 군집의 개수 혹은 구조에 대한 가정이 없으며, 개체들 사이에 정의된 유사성(similarity)에 대한 거리(distance)에 의하여 군집을 형성하고, 형성된 군집의 특성을 파악하여 군집들 사이의 관계를 분석한다는 점에서 분류 및 판별분석과 차이가 있다. 따라서 군집분석에서는 크게 두 가지 사항, 즉 거리의 정의 방법 및 군집 형성의 알고리즘 방법이 주요 연구 대상이 된다. 특히, 시계열 자료(time-series data)를 분석하는 과정에서 확률모형설정 및 예측을 수행해야 할 계열수가 많은 경우, 주어진 시계열 자료를 각각 예측하는 대신 유사한 특성을 가진 시계열들을 군집화한 후, 각 그룹으로부터 대표현된 것에 기초하여 예측을 수행한다면 훨씬 효율적일 것이다.

시계열 자료의 군집분석은 경영, 경제, 지리학, 기상학, 의학, 환

경학, 생물학 등의 여러 분야에서 접할 수 있다. 경제 분야에서는 시계열의 경향이 유사한 거래 종목들을 군집화하여 자료의 변동성에 관한 여러 가지 특성을 설명하고 주식투자자 및 금융관계자들에게 유용한 정보를 제공할 수 있으며, 산업생산지수나 소비지수들에 따라 산업분야를 군집화할 수 있다. 박만식 등 (2008)은 실제 거래되고 있는 최근 3년 국내 11개 기업의 일일 주가자료를 이용하여 금융 자료의 주요한 특성인 수익률의 변동성과 구조의 유사성을 통해 군집화하였다. Fu *et al.* (2001)은 홍콩증권거래소에 상장된 종목 중 상위 우량종목을 대상으로 산출하는 시가총액 가중지수(항생지수, HSI)를 이용하여 시간에 따른 지수의 패턴 변화 및 구조를 탐색하여 종목을 군집화하였다.

생물학 및 의학에서는 마이크로어레이(microarray) 유전자 발현 자료(gene expression data)에서 비슷한 성질을 갖는 유전자들을 군집화함으로써 자료의 차원 수를 줄이거나 유전자를 선별하여 그 유전자만을 분석하는 방법이 있다. Goutte *et al.* (1998)은 기능적 자기공명영상(fMRI) 자료에서 복셀(voxel)들의 시간에 따른 신호의 변화를 바탕으로 활성화 정도가 유사한 복셀을 군집화하여 질환이 발생하는 위치 및 크기를 추출하고 해부학적 구조를 보다 정확하게 인지하여 환자의 시술이나 치료 및 질병을 예측하는데 정보를 제공할 수 있다. Goldstein *et al.* (2002)은 유전자들의 시간에 따라 발현 수준의 변화를 고려함으로써 발현패턴이 다른 유전자들을 군집화하여 암 전이에 영향을 주는 유전자들의 그룹을 찾는 방법을 연구하였다.

지리학분야에서는 천광일 (2014)은 계방산의 15년간 산림식생 자료를 이용하여 시계열적 변화를 통해 군집화하였고 이는 산림생태계 경영 기반 정보를 구축하고 생물다양성 증진 및 보전에 의미있는 정보를 제공할 수 있다. 기상학분야에서는 기후에 의해 지역을 구분하고자

하는 연구가 진행되었다. 고정웅 등 (2006)은 일 최고 기온, 일 최저 기온, 일교차 자료를 이용하여 한반도의 기온 분포의 특성을 분석하고 지역을 구분하고자 하는 연구가 있으며 이는 국지 규모에서의 기후 시나리오 작성을 위한 중요한 기초 자료로 활용될 수 있다. 경민수 등 (2007)은 표준강수지수인 SPI(standard precipitation index)를 이용하여 시간에 따른 변동성을 기준으로 우리나라 전 지역을 가뭄의 특성에 따라 분류하고, 각각의 지역별로 가뭄심도-가뭄면적-가뭄지속기간 곡선을 구축하고 이를 통해 지역적 특성을 고려한 가뭄관리대책을 마련할 수 있다. 정준모 등 (2015)은 잣나무 조사지역을 월평균기온과 월강수량의 특성을 바탕으로 유사성이 높은 몇 개의 군집으로 분류한 후 군집별 연륜 성장(annual ring growth) 예측모델을 개발하여 연륜 성장과 기후와의 관계를 규명하였고 이는 기후변화에 대한 산림 분야에서의 대응방안을 마련하는데 정보를 줄 수 있다. Liao (2005)는 시계열 군집 분석 방법을 연구한 기존의 논문들을 모든 분야에 대해 조사하여, 각 논문에서 사용한 거리, 군집화 알고리즘, 군집결과를 평가하는 방법을 정리하였다. 보다 다양한 적용분야에 관해서 Liao (2005)의 논문을 참고하기 바란다.

시계열 자료들을 군집화하는 연구에서 시간에 따라 관측된 자료의 속성을 충분히 반영하는 거리를 정의하는 것은 상당히 중요한 연구이다. 군집분석에서는 일반적인 개체들간의 거리로 유클리드(Euclidean distance)를 이용할 수 있다. 그러나, 유클리드 거리는 시계열 자료의 특성 중 주요한 자기상관(autocorrelation)를 설명하지 못하므로 시계열 자료의 개체들간의 거리로는 바람직하지 않다(자세한 내용은 Galeano and Peña (2000)의 논문을 참고하기 바란다). 자기상관은 시간 영역(time domain)에서 자기공분산함수(autocovariance function)로

측정되고 이를 주파수 영역(frequency domain)의 주기도(periodogram)로 변환할 수 있다. 이때 서로 다른 주기도들 간의 주기도로 서로 독립임을 증명할 수 있다. 시계열 자료분석에서 모수의 추정, 미래시점의 예측 등 거의 모든 추론은 자기 공분산 함수를 이용한 식으로부터 유도된다. 따라서, 이론적으로 자기공분산함수를 통해 정의되는 스펙트럼(spectrum) 밀도함수(density function)의 추정량인 주기도를 이용하여 두 계열들 사이의 유사성을 측정하는 것이 보다 바람직할 수 있다.

이미 언급한 바와 같이, 시계열 분석은 시간 영역과 주파수 영역에서의 분석으로 나누어진다. 시간 영역에서의 분석은 자기공분산함수를 이용하여 과거와 현재의 확률구조에 대한 정보를 기초로 시계열의 자기상관성을 구축하는 모수적 모형을 적합하고, 이 모형을 토대로 예측값을 생성한다. 그리고 주파수 영역에서의 분석은 정상성(stationarity)을 만족하는 시계열 자료의 자기공분산함수를 푸리에 변환(Fourier transform)을 통해 서로 다른 주기를 갖는 삼각 함수(trigonometric function)와 같은 강한 주기함수의 일차결합으로 나타내고, 주파수 별로 변동을 분리한 후 각 주기성분들의 기여도를 이용하여 그 특성을 설명하는 방법이다.

시계열 자료의 군집분석 방법에는 다양한 거리가 정의된다. 시간 영역에서의 분석 방법을 적용한 경우, Piccolo (1990)는 정상성을 만족하는 시계열 자료에 대해 자기회귀이동평균(auto-regressive moving-average; ARMA) 모형을 적합하고 이로부터 유사성을 측정하는 방법을 제안하였다. Galeano and Peña (2000)는 정상성을 만족하는 시계열 자료의 자기상관함수(autocorrelation function)를 이용하여 군집화하는 방법을 제안하였고, 자기상관함수와 유사하게 Chatfield (1979)는 정상성을 만족하는 시계열 자료의 역자기상관함수(inverse autocorrelation

function)로부터 유사성을 측정하는 방법을 정의하였다. Caiado *et al.* (2006)는 정상성을 만족하는 시계열 자료의 부분자기상관함수(partial autocorrelation function)를 이용하여 유사성을 측정하는 방법을 정의하였다. Corduas and Piccolo (2008)는 정상성을 만족하지 않는 시계열 자료의 시계열을 설명하는 확률 모형의 추론으로부터 유사성을 측정하는 방법과 자기상관함수, 부분자기상관함수와 같은 시계열이 갖고 있는 특성을 정의하여 유사성을 측정하는 방법에 대해 모의실험을 진행하였다. D'Urso and Maharaj (2009)는 정상성을 만족하는 시계열 자료의 자기상관함수를 이용하여 퍼지 군집분석(fuzzy clustering analysis)을 제안하였다.

주파수 영역의 분석 방법을 적용한 경우, Kakizawa *et al.* (1998)은 지진과, 지뢰폭발과동 자료를 이용하여 군집화하기 위해 스펙트럼 밀도함수를 추정하고 Kullback-Leibler(KL) 거리를 이용하여 유사성을 측정하는 연구를 진행하였다. Caiado *et al.* (2006)는 주파수 영역에서 시계열의 표준화한 주기도를 이용한 거리를 제안하고, 정상 시계열(stationary time-series) 모형과 비정상 시계열(non-stationary time-series) 모형으로 군집화하는 모의실험을 진행하여 기존의 거리 척도와 그 성능을 비교하였다. 더불어, 미국의 계절조정된 산업생산지수 자료를 이용하여 군집화하였다. 박만식과 김희영 (2008)은 우리나라 22개 지역에서 측정한 월별 강수량 자료를 대상으로 수정된 Daniell 평활창(smoothing window)을 적용한 평활된 주기도에 근거한 거리를 이용하여 지역을 군집화하였다. 박만식과 김희영 (2010)은 Bohman, Bartlett, 그리고 Parzen의 평활창을 이용하여 평활된 주기도에 의한 거리를 제안하고, 모의실험을 통하여 다양한 시나리오에서 그 성능을 비교하였다. 또한, Caiado *et al.* (2006)의 실증자료를 이용하여 산업을 군집화하였다.

본 논문에서는 주기도 간의 상관성을 근거로 한 거리를 제안한다. 이 때, 피어슨 상관계수(Pearson's correlation coefficient)와 급내 상관계수(Intra-class correlation coefficient)를 이용하여 거리를 정의하고자 한다. 또한, 기존 연구에서 고려한 평활창의 폭(bandwidth)보다 더 다양하게 고려하여 변화를 살펴보고 최적의 폭을 알아보하고자 한다. 더불어, 자료의 특성에 따라 낮은 주파수만의 주기도 값을 이용하면 좋은 군집분석 결과를 보일 것으로 예상하고 이를 알아보하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 유사성 거리척도로 시간 영역에서의 거리, 주파수 영역에서의 거리들을 소개하고 상관계수를 활용한 거리를 제안한다. 3장에서는 모의실험을 통해 거리척도 간의 성능을 비교하고, 4장에서 모의실험을 통해 우수한 거리척도를 선택하여 1991년 1월부터 2015년 4월까지의 전국 월평균 제조업생산지수 자료를 군집화한 결과를 살펴본다. 5장에서는 결론과 시계열 자료의 군집분석에 대한 추후 연구과제에 대해 언급한다.

제 2 장

유사성 거리척도 정의

2장에서는 관측된 두 개의 시계열 $\mathbf{x} = \{x_1, \dots, x_n\}$, $\mathbf{y} = \{y_1, \dots, y_n\}$ 사이의 거리를 정의하는 몇 가지 방법들을 소개한다.

거리(distance)는 개체들 간의 유사성(similarity) 정도를 의미하며, 표본크기 n 인 자료에서 i 번째와 j 번째 개체 사이의 거리 d_{ij} 는 일반적으로 다음과 같은 조건을 만족함을 가정한다. $i, j = 1, 2, \dots, n$ 에 대하여

$$\begin{cases} d_{ij} \geq 0, & d_{ii} = 0 \\ d_{ij} = d_{ji} \\ d_{ij} + d_{jk} \geq d_{ik} \end{cases}$$

2.1 시간 영역에서의 거리

시간 영역에서는 시간이 경과함에 따라 시계열이 변화하는 유형을 분석하는 것으로 시간 차원에서 접근하며 자기상관함수와 부분자기상관함수를 통해 거리를 정의할 수 있다.

2.1.1 자기상관에 근거한 거리

시계열 분석에서의 정상성이란 시간의 흐름에 따라 확률변수의 확률적 성질이 변하지 않는 것으로 미래는 확률적으로 과거와 동일하다는 것을 의미한다. 즉, 임의의 시점 t 에 대한 관측값을 Z_t 라고 할

때 시계열 자료 $\{Z_t; t = 1, \dots, n\}$ 에 대하여 $E[Z(t)] = \mu$ 로 일정한 상수이고 분산 $\text{Var}[Z(t)]$ 이 유한한 상수로 존재하며, 두 시점 사이의 공분산 $\text{Cov}(Z_t, Z_{t+s})$ 이 시차(time lag) s 에만 의존할 경우 정상성을 만족한다고 할 수 있다. 정상성을 만족하는 시계열 자료를 정상 시계열이라고 한다.

실수값만을 취하는 정상시계열 $\{X_t; t = 1, \dots, n\}$ 의 자기공분산은 다음과 같이 정의된다. $s \geq 0$ 에 대하여

$$\gamma_s^x = \text{Cov}(X_t, X_{t+s}) = E(X_{t+s}X_t) - E(X_{t+s})E(X_t) = \gamma_{-s}^x.$$

이때, $\{\gamma_s^x, s \geq 0\}$ 를 자기공분산함수(autocovariance function)라고 한다.

s 차 자기상관계수는 X_t 와 X_{t+s} 관측값 사이의 상관계수로 다음과 같이 정의된다. $s \geq 0$ 에 대하여

$$\rho_s^x = \text{Corr}(X_t, X_{t+s}) = \frac{\gamma_s^x}{\gamma_0^x}.$$

이때, $\{\rho_s^x, s \geq 0\}$ 를 자기상관함수(autocorrelation function; ACF)라고 한다.

시차 s 의 부분자기상관계수는 X_t 와 X_{t+s} 관측값 사이의 상관계수이며 두 시점 사이의 관측값의 효과를 보정한 상관계수로 다음과 같이 정의된다.

$$\phi_s^x = \text{Corr}(X_t, X_{t+s} | X_{t+1}, X_{t+2}, \dots, X_{t+s-1}).$$

이때, $\{\phi_s^x, s \geq 0\}$ 를 부분자기상관함수(partial autocorrelation function; PACF)라고 한다. 자기상관함수와 부분자기상관함수는 시차에서 관측

된 시계열 사이의 상관계수를 측정하여 확률모형에 대한 정보를 얻을 수 있으므로 유사성의 척도로 이용될 수 있다.

\mathbf{x} 와 \mathbf{y} 의 추정된 자기상관함수 $\hat{\boldsymbol{\rho}}^{\mathbf{x}} = (\hat{\rho}_1^{\mathbf{x}}, \dots, \hat{\rho}_L^{\mathbf{x}})'$, $\hat{\boldsymbol{\rho}}^{\mathbf{y}} = (\hat{\rho}_1^{\mathbf{y}}, \dots, \hat{\rho}_L^{\mathbf{y}})'$ 와 부분자기상관함수 $\hat{\boldsymbol{\phi}}^{\mathbf{x}} = (\hat{\phi}_1^{\mathbf{x}}, \dots, \hat{\phi}_L^{\mathbf{x}})'$, $\hat{\boldsymbol{\phi}}^{\mathbf{y}} = (\hat{\phi}_1^{\mathbf{y}}, \dots, \hat{\phi}_L^{\mathbf{y}})'$ 를 이용하여 다음과 같은 유클리드 거리를 정의할 수 있다(Galeano and Peña, 2000; Caiado *et al*, 2006).

$$d_{ACF}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^L (\hat{\rho}_i^{\mathbf{x}} - \hat{\rho}_i^{\mathbf{y}})^2}, \quad d_{PACF}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^L (\hat{\phi}_i^{\mathbf{x}} - \hat{\phi}_i^{\mathbf{y}})^2} \quad (2.1)$$

2.2 주파수 영역에서의 거리

주파수 영역에서는 시간의 차원이 아닌 변동주기의 차원에서 접근하며 이를 스펙트럼 분석(spectral analysis)이라고 하는데, 이는 시계열 자료를 주기별로 시계열의 변동성분을 분리하여 분석하는 과정을 말한다.

2.2.1 주기도에 근거한 거리

정상시계열 $\{X_t : t = 1, \dots, n\}$ 은 다음과 같은 스펙트럼 밀도함수가 정의된다.

$$f(\theta) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \gamma_s^{\mathbf{x}} \exp\{-i\theta s\}, \quad \theta \in [-\pi, \pi]. \quad (2.2)$$

스펙트럼 밀도함수의 특징을 살펴보면 다음과 같다.

- $f(\theta) \geq 0$ 이며,
- $f(\theta) = f(-\theta)$ 이다.

- 자기공분산함수 γ_s 와 스펙트럼 밀도함수 $f(\theta)$ 는 아래의 관계를 가진다.

$$\gamma_s^x = 2 \int_0^\pi f(v) \cos(vs) dv, \quad f(\theta) = \frac{1}{2\pi} \left[\gamma_0^x + 2 \sum_{s=1}^{\infty} \gamma_s^x \cos(\theta s) \right].$$

따라서, 주파수 영역에서의 스펙트럼 밀도함수는 시간 영역에서의 자기공분산함수를 고유하게 결정짓는 함수이다.

스펙트럼 밀도함수를 추정하는 방법은 관측된 시계열 자료 \mathbf{x} 로부터 다음과 같이 정의되는 주기도(periodogram)를 이용한다.

$$J_n^x(\mathbf{v}_k) = \frac{1}{n} \left| \sum_{t=1}^n X_t \exp(-it\mathbf{v}_k) \right|^2. \quad (2.3)$$

여기서, $\left\{ k \in \mathbf{Z} \mid -\pi < \mathbf{v}_k \equiv \frac{2\pi k}{n} \leq \pi \right\} = \{ -\lfloor (n-1)/2 \rfloor, \dots, \lfloor n/2 \rfloor \}$ 이고, $\lfloor x \rfloor$ 는 x 를 넘지않는 최대 정수이다. 주기도의 특징은 다음과 같다.

- $J_n^x(\mathbf{v}_k) = J_n^x(-\mathbf{v}_k)$ 가 된다.
- 표본 자기공분산함수를 이용하여 다음과 같이 표현된다.

$$J_n^x(\mathbf{v}_k) = \begin{cases} n \left| \frac{1}{n} \sum_{t=1}^n X_t \right|^2, & \mathbf{v}_k = 0 \text{이면,} \\ \sum_{|h| < n} \hat{\gamma}_h^x \exp\{-ih\mathbf{v}_k\}, & \text{그 이외.} \end{cases}$$

또는, 다음과 같이 실수영역에서 계산가능한 형태로 표현될 수 있다.

$$J_n^{\mathbf{x}}(\mathbf{v}_k) = 2 \left[\hat{\gamma}_0^{\mathbf{x}} + 2 \sum_{h=1}^{n-1} \hat{\gamma}_h^{\mathbf{x}} \cos(\mathbf{v}_k h) \right]. \quad (2.4)$$

Caiado *et al.* (2006)은 관측된 두 시계열 \mathbf{x}, \mathbf{y} 의 주기도

$$J_n^{\mathbf{x}}(\mathbf{v}_k) = \frac{1}{n} \left| \sum_{t=1}^n x_t \exp\{-it\mathbf{v}_k\} \right|^2, \quad J_n^{\mathbf{y}}(\mathbf{v}_k) = \frac{1}{n} \left| \sum_{t=1}^n y_t \exp\{-it\mathbf{v}_k\} \right|^2.$$

를 이용하여 \mathbf{x} 와 \mathbf{y} 사이의 거리를 다음과 같이 정의하였다.

$$d_{JU}(\mathbf{x}, \mathbf{y}) = \left[\sum_{k=1}^{\lfloor n/2 \rfloor} (J_n^{\mathbf{x}}(\mathbf{v}_k) - J_n^{\mathbf{y}}(\mathbf{v}_k))^2 \right]^{1/2}, \quad (2.5)$$

$$d_{NJU}(\mathbf{x}, \mathbf{y}) = \left[\sum_{k=1}^{\lfloor n/2 \rfloor} (NJ_n^{\mathbf{x}}(\mathbf{v}_k) - NJ_n^{\mathbf{y}}(\mathbf{v}_k))^2 \right]^{1/2}, \quad (2.6)$$

$$d_{LNJU}(\mathbf{x}, \mathbf{y}) = \left[\sum_{k=1}^{\lfloor n/2 \rfloor} (\ln NJ_n^{\mathbf{x}}(\mathbf{v}_k) - \ln NJ_n^{\mathbf{y}}(\mathbf{v}_k))^2 \right]^{1/2}. \quad (2.7)$$

여기서, $NJ_n(\mathbf{v}_k)$ 는 주기도 $J_n(\mathbf{v}_k)$ 를 표본분산(sample variance)으로 나눈 것으로 $NJ_n(\mathbf{v}_k) = J_n(\mathbf{v}_k) / \hat{\gamma}_0$ 이다.

또한, Caiado *et al.* (2006)은 Kullback-Leibler(K-L) information 거리를 제안하였으며 이는 다음과 같이 정의하였다.

$$d_{KL}(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{\lfloor n/2 \rfloor} \left[\frac{NJ_n^{\mathbf{x}}(\mathbf{v}_k)}{NJ_n^{\mathbf{y}}(\mathbf{v}_k)} - \ln \frac{NJ_n^{\mathbf{x}}(\mathbf{v}_k)}{NJ_n^{\mathbf{y}}(\mathbf{v}_k)} - 1 \right].$$

위에서 정의한 거리는 $d_{KL}(\mathbf{x}, \mathbf{y}) \neq d_{KL}(\mathbf{y}, \mathbf{x})$ 이므로, 박만식과 김희영 (2010)은 대칭적인 K-L divergence 거리를 제안하였다.

$$d_{KD} = d_{KL}(\mathbf{x}, \mathbf{y}) + d_{KL}(\mathbf{y}, \mathbf{x}). \quad (2.8)$$

2.2.2 평활된 주기도에 근거한 거리

평활(smoothing)은 최근 값에 큰 가중치를 주고 과거 값에는 작은 가중치를 주는 것이다. 시계열 자료가 갖고 있는 불규칙 변동(또는 잡음)을 제거함으로써 추세를 더 명확하게 보기 위함이다. 그러나, 평활의 폭을 크게 할 경우 시계열이 갖고 있던 추세가 사라질 수 있으므로 원 시계열에서의 추세를 유지하는 평활의 폭을 정하는 것이 중요하다.

주기도를 통해 시계열의 자기상관구조를 부각시키기 위해 평활을 적용한 스펙트럼 밀도함수의 추정량은 다음과 같이 표현된다.

$$\hat{f}(\nu_k) = \sum_{|h| \leq m} W_n(h) J(\nu_{k+h}) = \tilde{J}_n(\nu_k). \quad (2.9)$$

$W_n(\cdot)$ 는 평활창(smoothing window)이고 $2m+1$ 은 평활창의 폭(bandwidth)이라고 한다. 평활창의 폭이 커지면 더 많은 주기도의 값이 평활에 사용되므로 스펙트럼 밀도함수의 추정량은 점점 부드러운 형태를 가지게 된다.

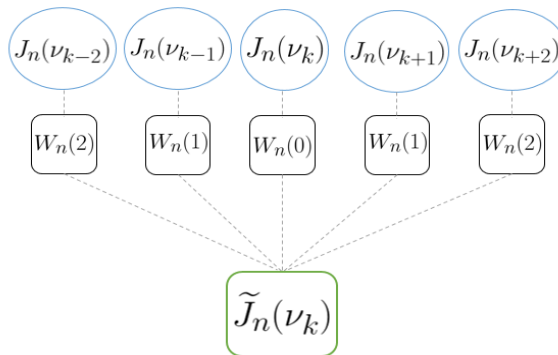


그림 1: $m = 2$ 일 때 $\tilde{J}_n(\nu_k)$ 의 계산과정

김희영과 박만식 (2008)은 위에서 제시한 평활된 주기도를 이용하

여 다음과 같은 거리를 제안하였다.

$$d_{J_U}^*(\mathbf{x}, \mathbf{y}) = \left[\sum_{k=1}^{\lfloor n/2 \rfloor} \left(\tilde{J}_n^{\mathbf{x}}(\mathbf{v}_k) - \tilde{J}_n^{\mathbf{y}}(\mathbf{v}_k) \right)^2 \right]^{1/2}, \quad (2.10)$$

$$d_{NJ_U}^*(\mathbf{x}, \mathbf{y}) = \left[\sum_{k=1}^{\lfloor n/2 \rfloor} \left(NJ_n^{\mathbf{x}}(\mathbf{v}_k) - NJ_n^{\mathbf{y}}(\mathbf{v}_k) \right)^2 \right]^{1/2}, \quad (2.11)$$

$$d_{LNJ_U}^*(\mathbf{x}, \mathbf{y}) = \left[\sum_{k=1}^{\lfloor n/2 \rfloor} \left(\ln NJ_n^{\mathbf{x}}(\mathbf{v}_k) - \ln NJ_n^{\mathbf{y}}(\mathbf{v}_k) \right)^2 \right]^{1/2}. \quad (2.12)$$

여기서, $N\tilde{J}_n(\mathbf{v}_k) = \tilde{J}_n(\mathbf{v}_k)/\hat{\gamma}_0$ 이다. 식 (2.8)에서 평활된 주기도 \tilde{J}_n 를 이용한 거리를 정의하면 다음과 같다.

$$d_{KD}^* = d_{KL}^*(\mathbf{x}, \mathbf{y}) + d_{KL}^*(\mathbf{y}, \mathbf{x}). \quad (2.13)$$

여기서,

$$d_{KL}^*(\mathbf{x}, \mathbf{y}) = \sum_{k=1}^{\lfloor n/2 \rfloor} \left[\frac{NJ_n^{\mathbf{x}}(\mathbf{v}_k)}{NJ_n^{\mathbf{y}}(\mathbf{v}_k)} - \ln \frac{NJ_n^{\mathbf{x}}(\mathbf{v}_k)}{NJ_n^{\mathbf{y}}(\mathbf{v}_k)} - 1 \right].$$

본 논문에서는 Bartlett, Bohman, Parzen 평활창을 고려하였다. 평활창의 자세한 수식에 관하여는 박만식과 김희영 (2010)의 논문을 참고하기 바란다. 세 가지 평활창의 특징을 그래프를 통해 알아보자. 그림 2은 식 (2.12)에서 $m = 50$ 으로 가정하여 본 논문에서 고려하고자 하는 세가지 평활창의 시점변화에 따른 가중치의 변화를 그래프로 나타낸것이다. Bartlett 평활창은 현 시점보다 멀어질수록 가중치가 선형적으로 감소하는 형태를 갖는다. Bohman과 Parzen 평활창은 현 시점보다 멀어질수록 가중치가 지수적으로 감소하는 형태를 보이나 Parzen 평활

창이 Bohman 평활창에 비해 가파르게 감소한다. 즉, Parzen 평활창이 Bohman 평활창 보다 더 가까운 시점에 가중치를 높게 준다는 것을 알 수 있다. 그러나 Bohman 평활창과 Parzen 평활창은 큰 차이가 없어 보인다.

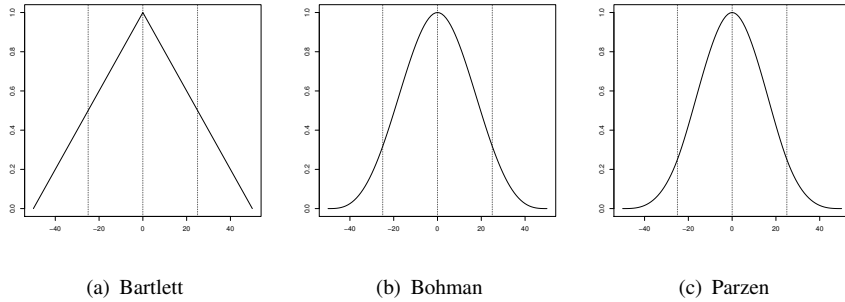


그림 2: 평활창에 따른 가중치 $W_n(h)$

2.2.3 주기도의 상관성에 근거한 거리

상관분석(correlation analysis)은 통계학에서 두 변수간에 어떤 선형적 관계를 갖고 있는 지를 분석하는 방법이다. 두변수는 서로 독립적인 관계로부터 서로 상관된 관계일 수 있으며 이때 두 변수간의 관계의 강도를 상관관계라 한다. 상관관계의 정도를 파악하는 상관계수(correlation coefficient)는 두 변수간의 연관된 정도를 나타낼 뿐 인과관계를 설명하는 것은 아니다.

두 시계열 $\mathbf{x} = \{x_1, \dots, x_n\}$, $\mathbf{y} = \{y_1, \dots, y_n\}$ 의 주기도를 이용하여 피어슨 상관계수(Pearson's correlation coefficient)를 다음과 같이 정의할 수 있다.

$$P_J = \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{\left(J_n^{\mathbf{x}}(\mathbf{v}_k) - \overline{J_n^{\mathbf{x}}(\cdot)} \right) \left(J_n^{\mathbf{y}}(\mathbf{v}_k) - \overline{J_n^{\mathbf{y}}(\cdot)} \right)}{\sqrt{\widehat{\text{Var}}(J_n^{\mathbf{x}}) \widehat{\text{Var}}(J_n^{\mathbf{y}})}}. \quad (2.14)$$

여기서,

$$\overline{J_n^{\mathbf{x}}(\cdot)} = \frac{1}{\lfloor n/2 \rfloor} \sum_{k=1}^{\lfloor n/2 \rfloor} J_n^{\mathbf{x}}(\mathbf{v}_k), \quad \overline{J_n^{\mathbf{y}}(\cdot)} = \frac{1}{\lfloor n/2 \rfloor} \sum_{k=1}^{\lfloor n/2 \rfloor} J_n^{\mathbf{y}}(\mathbf{v}_k).$$

식 (2.14)에서 J_n 대신에 NJ_n 또는 $\ln NJ_n$ 으로 계산하면 각각 P_{NJ} , P_{LNJ} 라고 정의하겠다. 평활된 주기도를 이용하여 피어슨 상관계수를 표현하면 다음과 같다.

$$P_{\tilde{J}} = \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{\left(\tilde{J}_n^{\mathbf{x}}(\mathbf{v}_k) - \overline{\tilde{J}_n^{\mathbf{x}}(\cdot)} \right) \left(\tilde{J}_n^{\mathbf{y}}(\mathbf{v}_k) - \overline{\tilde{J}_n^{\mathbf{y}}(\cdot)} \right)}{\sqrt{\widehat{\text{Var}}(\tilde{J}_n^{\mathbf{x}}) \widehat{\text{Var}}(\tilde{J}_n^{\mathbf{y}})}}. \quad (2.15)$$

여기서,

$$\overline{\tilde{J}_n^{\mathbf{x}}(\cdot)} = \frac{1}{\lfloor n/2 \rfloor} \sum_{k=1}^{\lfloor n/2 \rfloor} \tilde{J}_n^{\mathbf{x}}(\mathbf{v}_k), \quad \overline{\tilde{J}_n^{\mathbf{y}}(\cdot)} = \frac{1}{\lfloor n/2 \rfloor} \sum_{k=1}^{\lfloor n/2 \rfloor} \tilde{J}_n^{\mathbf{y}}(\mathbf{v}_k).$$

식 (2.15)에서 \tilde{J}_n 대신에 $N\tilde{J}_n$ 또는 $\ln N\tilde{J}_n$ 으로 계산하면 각각 $P_{N\tilde{J}}$, $P_{LN\tilde{J}}$ 라고 정의하겠다.

급내 상관계수(Intra-class correlation coefficient; ICC)는 신뢰도 연구에서 측정값이 정량적일 때 쓰이는 지표이다. 신뢰도 연구는 한 명의 평가자가 연구 대상을 반복 측정하거나 여러 명의 평가자가 한 대상을 평가할 때 평가자 내 혹은 평가자 간 일치도를 알아보는 연구이다. 즉, 특정 집단의 평가결과가 얼마나 재현 가능한지를 알아보는 지표이다.

두 시계열 $\mathbf{x} = \{x_1, \dots, x_n\}$, $\mathbf{y} = \{y_1, \dots, y_n\}$ 의 주기도를 이용하여 급내 상관계수는 다음과 같이 표현한다.

$$C_J = \frac{1}{\lfloor n/2 \rfloor s_J^2} \sum_{k=1}^{\lfloor n/2 \rfloor} (J_n^{\mathbf{x}}(\mathbf{v}_k) - \bar{J}_n) (J_n^{\mathbf{y}}(\mathbf{v}_k) - \bar{J}_n). \quad (2.16)$$

여기서,

$$\bar{J}_n = \frac{1}{n} \sum_{k=1}^{\lfloor n/2 \rfloor} (J_n^{\mathbf{x}}(\mathbf{v}_k) + J_n^{\mathbf{y}}(\mathbf{v}_k)).$$

$$s_J^2 = \frac{1}{2\lfloor n/2 \rfloor} \sum_{k=1}^{\lfloor n/2 \rfloor} (J_n^{\mathbf{x}}(\mathbf{v}_k) - \bar{J}_n)^2 + (J_n^{\mathbf{y}}(\mathbf{v}_k) - \bar{J}_n)^2.$$

식 (2.16)에서 J_n 대신에 NJ_n 또는 $\ln NJ_n$ 으로 계산하면 각각 C_{NJ} , C_{LNJ} 라고 정의하겠다. 평화된 주기도를 이용하여 급내 상관계수를 표현하면

다음과 같다.

$$C_{\tilde{J}} = \frac{1}{[n/2]s_{\tilde{J}}^2} \sum_{k=1}^{[n/2]} \left(\tilde{J}_n^{\mathbf{x}}(\mathbf{v}_k) - \tilde{J}_n \right) \left(\tilde{J}_n^{\mathbf{y}}(\mathbf{v}_k) - \tilde{J}_n \right). \quad (2.17)$$

여기서,

$$\begin{aligned} \tilde{J}_n &= \frac{1}{n} \sum_{k=1}^{[n/2]} \left(\tilde{J}_n^{\mathbf{x}}(\mathbf{v}_k) + \tilde{J}_n^{\mathbf{y}}(\mathbf{v}_k) \right). \\ s_{\tilde{J}}^2 &= \frac{1}{2[n/2]} \sum_{k=1}^{[n/2]} \left(\tilde{J}_n^{\mathbf{x}}(\mathbf{v}_k) - \tilde{J}_n \right)^2 + \left(\tilde{J}_n^{\mathbf{y}}(\mathbf{v}_k) - \tilde{J}_n \right)^2. \end{aligned}$$

식 (2.17)에서 \tilde{J}_n 대신에 $N\tilde{J}_n$ 또는 $\ln N\tilde{J}_n$ 으로 계산하면 각각 $C_{N\tilde{J}}$, $C_{LN\tilde{J}}$ 라고 정의하겠다.

위에서 정의한 피어슨 상관계수, 급내 상관계수를 이용하여 다음과 같은 거리를 제안한다.

$$d_{J_P}(\mathbf{x}, \mathbf{y}) = 1 - P_J, \quad d_{LNJ_P}(\mathbf{x}, \mathbf{y}) = 1 - P_{LNJ} \quad (2.18)$$

$$d_{J_P}^*(\mathbf{x}, \mathbf{y}) = 1 - P_{\tilde{J}}, \quad d_{LNJ_P}^*(\mathbf{x}, \mathbf{y}) = 1 - P_{LN\tilde{J}} \quad (2.19)$$

$$d_{J_C}(\mathbf{x}, \mathbf{y}) = 1 - C_J, \quad d_{NJ_C}(\mathbf{x}, \mathbf{y}) = 1 - C_{NJ}, \quad d_{LNJ_C}(\mathbf{x}, \mathbf{y}) = 1 - C_{LNJ} \quad (2.20)$$

$$d_{J_C}^*(\mathbf{x}, \mathbf{y}) = 1 - C_{\tilde{J}}, \quad d_{NJ_C}^*(\mathbf{x}, \mathbf{y}) = 1 - C_{N\tilde{J}}, \quad d_{LNJ_C}^*(\mathbf{x}, \mathbf{y}) = 1 - C_{LN\tilde{J}} \quad (2.21)$$

여기서, d 는 $0 \leq d \leq 2$ 의 값을 갖는다.

제 3 장

모의실험

3.1 모의실험 방법

본 절에서는 시계열 자료의 확률모형들 중 자기회귀이동평균 모형($ARMA(p, q)$; $X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$)을 갖는 정상 시계열 모의 자료를 이용하여 2.1절과 2.2절에서 제시한 여러가지 거리들을 적용한 군집분석에 대한 설명을 하고자 한다. 이를 위해 다른 시계열 확률모형을 가진 자료 12개를 생성하여 2개의 참 군집(true cluster)을 만들고 거리 척도를 적용하여 군집분석 결과를 정 분류율로 계산하였다. 모의자료는 각 시나리오별 시계열 확률모형의 참값의 모수(ϕ 또는 θ)로 이루어진 자료를 이용하였다.

표 1에서 시나리오 1, 2는 첫 6개의 계열은 모수 ϕ 인 자기회귀모형(AR model)을 따르고, 나머지 6개의 계열은 모수 θ 인 이동평균모형(MA model)을 따르는 자료로 서로 다른 시계열 확률모형 간의 분류 성능을 확인할 수 있다. 시나리오 2는 시나리오 1에 비해 모수의 범위를 좁혀 좀 더 극단적인 상황을 가정하였다. 또한, 시나리오 3, 4는 12개의 계열 모두 자기회귀모형을 따르는 자료로서 같은 시계열 확률모형 간의 분류 성능을 확인할 수 있다. 시나리오 4는 시나리오 3에 비해 극단적인 상황을 가정하였다. 마지막으로, 시나리오 5, 6은 12개의 계열 모두 이동평균모형을 따르는 자료이며, 시나리오 6은 시나리오 5에 비해 더 극단적인 상황을 가정하였다. 이는 그림 3를 통해 확인할 수

표 1: 모의실험 시나리오

Scenario	Model (p, q)	Parameters	
		Autoregression(ϕ)	Moving Average(θ)
1	(1,0)	0.10, 0.20, 0.40, 0.60, 0.80, 0.90	
	(0,1)	0.10, 0.20, 0.40, 0.60, 0.80, 0.90	
2	(1,0)	0.40, 0.50, 0.60, 0.70, 0.80, 0.90	
	(0,1)	0.40, 0.50, 0.60, 0.70, 0.80, 0.90	
3	(1,0)	0.80, 0.82, 0.84, 0.86, 0.88, 0.90	
	(1,0)	0.60, 0.62, 0.64, 0.66, 0.68, 0.70	
4	(1,0)	0.65, 0.70, 0.75, 0.80, 0.85, 0.90	
	(1,0)	0.10, 0.15, 0.20, 0.25, 0.30, 0.35	
5	(0,1)	0.80, 0.82, 0.84, 0.86, 0.88, 0.90	
	(0,1)	0.60, 0.62, 0.64, 0.66, 0.68, 0.70	
6	(0,1)	0.65, 0.70, 0.75, 0.80, 0.85, 0.90	
	(0,1)	0.10, 0.15, 0.20, 0.25, 0.30, 0.35	

시나리오 1~4는 Caiado *et al.* (2006)와 박만식과 김희영 (2010)을 참조하였다.

있다.

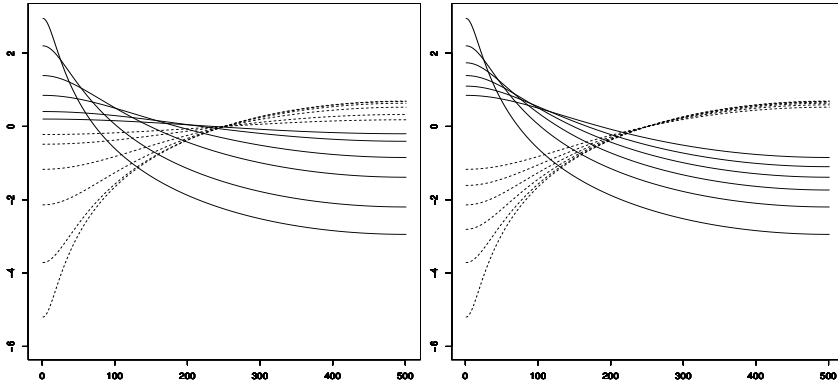
모의실험을 위해 사용된 프로그램은 R 3.1.2 (Program R)이다. 모의 자료를 생성하기 위해 ‘arima.sim’ 함수를 이용하였고, 2장에서 정의한 거리 식을 이용하여 12×12 거리행렬을 계산하는 알고리즘을 구축하였으며 ‘hclust’ 함수를 이용하였다.

모의실험의 절차는 다음과 같다.

- (1) ARMA(p, q)모형 하에서 참 군집(true cluster) 1, $C_1 = \{\mathbf{x}^1, \dots, \mathbf{x}^6\}$ 이고, 참 군집 2, $C_2 = \{\mathbf{x}^7, \dots, \mathbf{x}^{12}\}$ 을 가지도록 \mathbf{x}^1 에서 \mathbf{x}^{12} 자료를 임의로 생성한다.
- (2) 자료의 크기 n 은 100, 200, 500을 고려하였으며, ‘arima.sim’ 함수를 통한 안정적인 자료의 생성을 위해 10,001번째부터 자료 n 개를 추출한다.

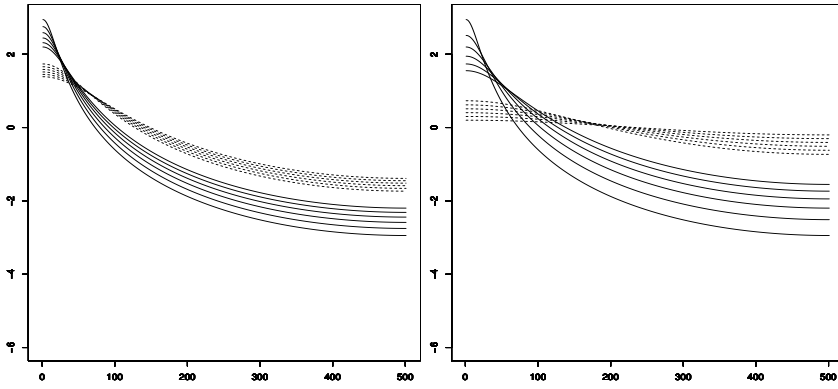
- (3) 단계 (2)에서 생성한 자료를 이용하여 ACF와 PACF를 계산한다.
시차 L 은 n 의 25%까지 고려한다.
- (4) 자기공분산 함수를 이용하여 식 (2.4)을 통해 주기도를 계산한다.
주과수 v_k 는 n 의 10%, 20%, 50%(최대값)를 고려한다.
- (5) 단계 (4)에서 계산한 주기도를 이용하여 식 (2.9)를 통해 평활된 주기도를 계산한다. 평활의 폭 m 은 n 의 1%에서 15%까지 고려한다. 예를 들어, $n = 500$ 인 경우 평활의 폭을 1%로 가정한다면 $m = 5$ 이며 평활시 사용하는 자료의 수는 $2m + 1 = 11$ 개이다.
- (6) 앞에서 정의한 거리 식 (2.1), 식 (2.5)~식 (2.8), 식 (2.10)~식 (2.13), 식 (2.18)~식 (2.21)을 이용하여 유사성 정도를 측정한다.
- (7) 단계 (6)에서 계산한 거리를 이용하여 2개의 군집으로 군집화 한 후, 참 군집과 같은 결과인 경우 1, 아니면 0의 값을 계산한다.
군집분석 알고리즘은 모든 개체 쌍의 거리 중 가장 먼 거리를 이용한 완전 연결법(complete linkage method) 이용한다.
- (8) 위의 과정을 1,000번 반복하여 정분류율을 계산한다.

각 모의실험 시나리오의 군집분석 결과를 거리척도에 따라 정분류율을 비교 해 보고자 한다.



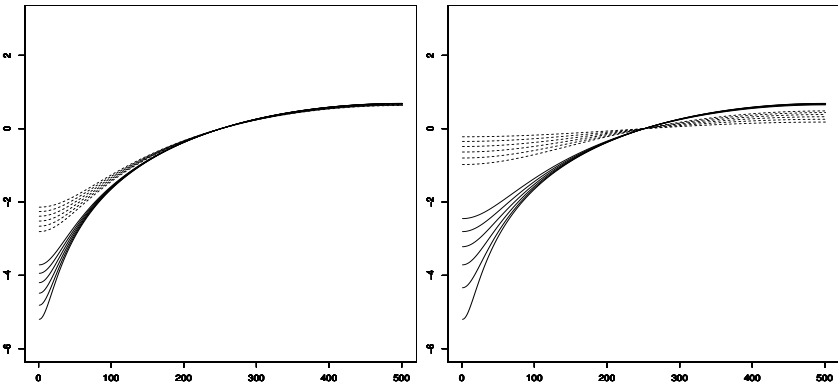
(a) 시나리오 1

(b) 시나리오 2



(c) 시나리오 3

(d) 시나리오 4



(e) 시나리오 5

(f) 시나리오 6

그림 3: 시나리오별 이론적인 스펙트럼 밀도함수

실선은 표 1에서 첫 6개 계열의 모형이고, 점선은 나머지 6개 계열의 모형이다.

3.2 모의실험 결과

모의실험을 위해 생성된 모의 자료에 대한 시나리오 1, 2의 군집 결과를 정리하면 표 2, 표 3과 같다. AR 모형과 MA 모형의 시계열 자료를 잘 분류하는지 알아보는 시나리오 1, 2를 살펴보자. 피어슨 상관 계수, 급내 상관계수를 이용한 거리는 유클리드 거리에 비해 정분류율이 높다. 유클리드 거리 중에서 LNJ_U 의 정분류율이 가장 높으며 그 다음은 KD_U 이며, J_U 와 NJ_U 은 평활을 하지 않은 주기도를 이용한 결과보다 평활된 주기도를 이용한 결과의 정분류율이 더 높다. 평활된 주기도의 경우, Bohman 평활창과 Parzen 평활창을 이용한 경우 비슷한 결과를 보이며, Bartlett 평활창은 다른 경향을 보인다. 원 주기도를 이용한 경우, 모든 주파수의 주기도 값을 이용하여 거리를 계산한 경우 정분류율이 높았으며, n 이 커질수록 정분류율이 높아지는 경향이 보이지만, 평활된 주기도는 n 이 커질수록 정분류율이 높아지는 경향이 보이지 않는다. 평활된 주기도를 이용한 NJ_U 는 m 이 작을수록 정분류율이 높으며, 이는 n 이 커질수록 더 두드러지게 드러난다. 단, $n = 100$, Bartlett 평활창에 근거한 평활된 주기도를 이용한 LNJ_U 또는 KD_U 는 n 이 커질수록 정분류율이 높아진다. 평활된 주기도를 이용한 J_p 또는 LNJ_p 는 다른 거리 척도와 비교하였을 때, 정분류율이 월등히 높으며, m 의 변화에 크게 영향을 받지 않는다. 또한, 원 주기도를 이용한 J_p 또는 LNJ_p 는 거리 계산시 주기도 크기를 n 의 50%로 설정한 경우가 그렇지 않은 경우에 비해 급격하게 성능이 좋아짐을 보여준다. 다른 거리 척도에서는 이 경향이 크게 드러나지 않는다. 그림 3 (a)와 (b)를 보면 AR 모형과 MA 모형의 주기도는 서로 반대 방향으로 움직이므로 상관계수를 사용할 경우 분류 성능이 훨씬 좋아짐을 알 수 있다. $n = 100$ 일 때,

Bohman 평활창에 근거한 평활된 주기도를 이용한 NJ_U 는 $m = 2\%$ 의 경우 정분류율이 26.8%로 두번째로 높았다. 자기상관을 이용한 거리 척도 중에서 PACF를 이용한 거리는 ACF를 이용한 거리보다 더 올바르게 두 군집으로 분류한다. 특히 $n = 200$ 이상일 때 그 결과를 주목할 만 하다. 평활된 주기도를 이용한 NJ_U 는 $n = 200$ 이며 $M = 1\%$ 인 경우 가장 좋은 결과를 보인다. 평활된 주기도를 이용한 LNJ_U 또는 KD의 경우, $n = 200$ 일 때, $M = 1\%$ 로 설정하여 Bohman 또는 Parzen 평활창을 이용하는 것이 좋은 성능을 보인다.

시나리오 1은 전반적으로 거리 척도의 성능이 좋지 않다. 그림 3(a)를 보면, $\phi = 0.1$ 인 AR 모형과 $\theta = 0.1$ 인 MA 모형을 분류하는 것은 사실상 쉽지 않음을 알 수 있다. 그림 3(b)를 보면, 두 군집이 시나리오 1에 비해서 더 잘 분류가 가능함을 보여준다. 평활된 주기도를 이용한 거리 척도가 원 주기도를 이용한 거리 척도에 비해 정분류율이 월등히 높아 성능이 좋다. $n = 500$ 일 때, 원 주기도의 LNJ_U 와 평활된 주기도의 LNJ_U 는 비슷한 성능을 보여준다. 평활된 주기도를 이용한 NJ_U 는 n 이 커짐에 따라서 정분류율이 m 과 반비례하는 경향이 매우 두드러진다. 즉, $m = 1\%$ 또는 2% 로 설정하면 약 100%에 가까운 정분류율을 보여준다. $m = 2\%$ 인 NJ_U , LNJ_U , KD는 $n = 100$ 인 경우에도 90%이상의 정분류율을 나타낸다. J_p 또는 LNJ_p 는 n 에 관계없이 100%의 성능을 보이며, 원 주기도의 경우, 거리 계산시 사용된 주기도의 크기에 큰 영향을 받으며 자료의 전체를 쓰는 것이 가장 좋은 성능을 보인다. 상관계수가 자료의 크기에 영향을 받는다는 점을 나타낸다. PACF는 $n = 100$ 일 때, 평활된 주기도의 NJ_U 만큼 성능이 좋으며, $n = 200$ 이상일 때, 평활된 주기도의 LNJ_U 또는 KD 만큼 정분류율이 높다.

표 2: 시나리오1의 모의실험 결과

n	metric	Raw periodogram										Smoothed periodogram										Correlation		
		Bartlett window					Bohman window					Parzen window					ACF	PACF						
		10%	20%	50%	1%	2%	5%	10%	1%	2%	5%	10%	1%	2%	5%	1%			2%	5%	10%	5%	10%	
100	J_U	0.0	0.0	0.0	-	0.0	0.1	0.0	-	0.0	0.1	0.0	0.0	-	0.0	0.1	0.0	-	0.0	0.1	0.0	0.0	1.4	17.8
	NJ_U	0.0	0.0	0.6	-	26.8	16.8	8.8	-	26.8	21.4	10.9	-	26.8	22.8	8.8	-	26.8	22.8	8.8	-	26.8	22.8	12.0
	LNJ_U	5.9	5.8	8.0	-	9.9	16.6	17.1	-	22.4	10.7	10.7	-	23.9	10.8	10.3	-	23.9	10.4	10.9	-	23.9	10.4	10.9
	KD	3.9	4.2	4.2	-	8.3	17.2	17.8	-	22.3	10.4	11.2	-	23.9	10.4	10.9	-	23.9	10.4	10.9	-	23.9	10.4	10.9
	J_p	0.6	5.6	36.4	-	63.2	61.9	60.8	-	63.2	62.4	61.7	-	63.2	62.9	62.0	-	63.2	62.9	62.0	-	63.2	62.9	62.0
	LNJ_p	0.3	3.3	25.2	-	63.2	62.4	61.8	-	63.2	63.5	62.4	-	63.2	63.6	62.0	-	63.2	63.6	62.0	-	63.2	63.6	62.0
	J_c	0.2	2.2	35.9	-	63.2	63.4	60.4	-	63.2	63.5	62.6	-	63.2	63.2	62.9	-	63.2	63.2	62.9	-	63.2	63.9	63.5
	NJ_c	0.2	1.3	36.0	-	63.2	63.3	61.8	-	63.2	63.7	63.4	-	63.2	63.9	63.5	-	63.2	63.9	63.5	-	63.2	63.9	63.5
	LNJ_c	0.3	3.0	25.7	-	63.2	63.3	62.4	-	63.2	63.4	62.5	-	63.2	63.6	62.1	-	63.2	63.6	62.1	-	63.2	63.6	62.1
	200	J_U	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	25.0
NJ_U		0.0	0.0	0.2	33.5	22.3	6.5	2.3	33.5	31.0	9.8	2.5	33.5	32.1	12.5	2.8	33.5	32.1	12.5	2.8	33.5	32.1	12.5	2.8
LNJ_U		6.2	4.8	8.8	10.8	19.5	19.5	18.6	28.0	9.3	13.4	14.9	8.3	31.4	13.2	15.0	8.3	31.4	13.2	15.0	8.3	31.4	13.2	15.0
KD		4.7	4.5	3.3	7.3	19.8	20.7	20.3	27.3	7.2	16.8	11.7	31.2	6.3	16.4	12.9	31.2	6.3	16.4	12.9	31.2	6.3	16.4	12.9
J_p		1.6	7.7	49.0	84.0	84.0	81.3	78.3	84.0	85.0	82.4	79.3	84.0	85.1	83.1	79.4	84.0	85.1	83.1	79.4	84.0	85.1	83.1	79.4
LNJ_p		0.7	5.4	37.6	84.0	84.7	82.9	80.3	84.0	84.7	83.3	81.0	84.0	84.8	83.5	81.4	84.0	84.8	83.5	81.4	84.0	84.8	83.5	81.4
J_c		0.1	2.7	47.9	84.0	84.7	82.6	79.0	84.0	84.9	82.9	79.5	84.0	84.5	83.7	80.7	84.0	84.5	83.7	80.7	84.0	84.5	83.7	80.7
NJ_c		0.1	2.5	50.5	84.0	84.5	82.7	78.8	84.0	84.9	83.0	78.8	84.0	84.5	83.3	79.5	84.0	84.5	83.3	79.5	84.0	84.5	83.3	79.5
LNJ_c		0.3	5.6	36.0	84.0	84.5	82.6	79.6	84.0	84.9	82.4	79.5	84.0	84.6	82.6	80.6	84.0	84.6	82.6	80.6	84.0	84.6	82.6	80.6
500		J_U	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	30.6
	NJ_U	0.0	0.0	0.2	24.2	2.3	0.1	0.0	29.9	6.2	0.0	0.0	32.6	9.8	0.0	0.0	32.6	9.8	0.0	0.0	32.6	9.8	0.0	0.0
	LNJ_U	4.2	4.8	11.5	21.6	23.3	19.3	19.7	6.4	11.7	17.4	15.6	5.3	10.3	17.0	15.8	5.3	10.3	17.0	15.8	5.3	10.3	17.0	15.8
	KD	5.7	3.4	5.1	23.8	27.2	22.0	17.2	5.8	18.3	10.3	6.6	4.1	17.9	10.1	6.6	4.1	17.9	10.1	6.6	4.1	17.9	10.1	6.6
	J_p	1.2	12.3	69.2	97.7	97.6	96.6	96.3	97.6	97.8	96.9	96.3	97.6	97.9	96.8	96.4	97.6	97.9	96.8	96.4	97.6	97.9	96.8	96.4
	LNJ_p	1.1	7.6	53.8	97.7	97.6	97.6	97.7	97.6	97.8	97.4	96.6	97.6	97.9	97.3	96.8	97.6	97.9	97.3	96.8	97.6	97.9	97.3	96.8
	J_c	0.0	6.7	68.9	97.6	97.6	97.2	96.0	97.6	97.7	97.1	96.2	97.6	97.7	97.1	96.3	97.6	97.7	97.1	96.3	97.6	97.7	97.1	96.3
	NJ_c	0.1	5.3	70.3	97.6	97.5	97.3	96.7	97.6	97.6	96.9	96.8	97.6	97.6	96.9	96.8	97.6	97.6	96.9	96.8	97.6	97.6	96.9	96.8
	LNJ_c	0.9	8.6	53.0	97.8	97.3	97.1	96.1	97.6	97.7	97.2	95.9	97.6	97.9	97.2	96.3	97.6	97.9	97.2	96.3	97.6	97.9	97.2	96.3

표 3: 시나리오2의 모의실험 결과

n	metric	Smoothed periodogram															Correlation					
		Raw periodogram					Bartlett window					Bohman window					Parzen window			ACF	PACF	
		10%	20%	50%	1%	2%	5%	10%	1%	2%	5%	10%	1%	2%	5%	1%	2%	5%	10%			
2	J_U	0.0	0.0	0.5	-	1.3	4.2	3.5	-	0.1	6.3	4.3	-	0.0	6.3	4.5	-	0.0	6.3	4.5	27.9	99.0
	NJ_U	0.1	0.0	14.8	-	98.8	96.7	84.1	-	98.8	98.2	89.3	-	98.8	98.0	91.8	-	98.8	98.0	91.8		
	LNJ_U	79.8	78.2	84.0	-	91.8	98.5	97.5	-	98.4	91.8	93.5	-	98.6	91.0	93.3	-	98.6	91.0	93.3		
	KD	63.1	55.9	42.0	-	87.6	98.9	98.4	-	98.4	91.3	96.4	-	98.6	90.0	96.0	-	98.6	90.0	96.0		
	J_p	2.2	43.2	99.3	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0		
	LNJ_p	0.8	29.2	95.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0		
	J_c	0.1	9.8	99.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0		
	NJ_c	0.1	7.5	99.4	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0		
	LNJ_c	0.3	27.9	94.4	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0	-	100.0	100.0	100.0		
	J_U	0.0	0.0	0.2	0.1	0.6	0.5	0.5	0.0	1.2	0.7	0.5	0.0	1.0	0.7	0.4	0.0	1.0	0.7	0.4	18.6	100.0
NJ_U	0.0	0.0	7.5	99.9	99.4	87.0	59.4	99.9	99.8	94.0	61.4	99.9	99.9	99.9	96.4	66.3	99.9	95.6	97.8	98.8		
LNJ_U	92.7	93.4	95.7	95.9	99.4	99.6	99.7	99.9	99.9	97.7	98.7	99.9	99.9	95.6	97.8	98.8	99.9	95.6	97.8	98.8		
KD	71.6	63.4	52.9	93.8	99.6	100.0	100.0	100.0	99.9	94.6	99.7	99.4	99.9	92.7	99.8	99.5	99.9	92.7	99.8	99.5		
J_p	6.2	70.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0		
LNJ_p	4.0	50.2	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0		
J_c	0.0	26.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0		
NJ_c	0.0	22.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0		
LNJ_c	0.9	48.9	99.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0		
500	J_U	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.6	100.0
	NJ_U	0.0	0.0	3.1	99.7	85.4	45.5	25.3	99.9	96.8	45.3	23.0	99.9	98.4	51.6	23.5	99.9	98.4	51.6	23.5		
	LNJ_U	98.4	97.8	99.7	99.9	99.9	100.0	99.9	98.6	99.8	100.0	99.9	98.6	99.8	100.0	99.9	98.6	99.8	100.0	99.9		
	KD	77.9	74.4	65.0	99.9	99.9	100.0	99.9	98.4	99.9	100.0	99.9	97.7	99.9	100.0	99.9	97.7	99.9	100.0	99.9		
	J_p	13.5	91.1	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
	LNJ_p	7.9	74.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
	J_c	0.0	66.7	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
	NJ_c	0.0	62.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	
	LNJ_c	2.4	72.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	

표 4: 시나리오3의 모의실험 결과

n	metric	Raw periodogram										Smoothed periodogram										Correlation		
		Bartlett window					Bohman window					Parzen window					ACF	PACF						
		10%	20%	50%	1%	2%	5%	10%	1%	2%	5%	10%	1%	2%	5%	10%			5%	10%				
3	J_U	0.1	0.1	0.1	-	4.0	4.5	4.2	-	4.0	4.7	4.4	-	3.8	5.0	4.1	-	3.8	5.0	4.1	1.7	0.9		
	NJ_U	0.4	0.5	0.3	-	15.3	12.4	8.4	-	15.3	14.0	10.6	-	15.3	14.3	12.0	-	15.3	14.3	12.0				
	LNJ_U	0.2	0.5	3.9	-	14.8	15.0	14.7	-	15.8	14.3	14.2	-	15.7	14.3	14.2	-	15.7	14.3	14.2				
	KD	0.0	0.5	0.9	-	14.8	15.0	14.6	-	15.8	14.3	14.1	-	15.7	14.5	14.1	-	15.7	14.5	14.1				
	J_p	0.1	0.0	0.1	-	0.0	3.1	1.5	-	0.0	3.4	2.9	-	0.0	3.5	2.6	-	0.0	3.5	2.6				
	LNJ_p	0.2	0.1	0.1	-	14.6	1.8	0.4	-	15.8	0.3	0.2	-	15.6	0.1	0.3	-	15.6	0.1	0.3				
	J_C	0.1	0.4	3.1	-	10.2	10.8	11.8	-	9.8	10.7	11.6	-	9.8	10.8	11.5	-	9.8	10.8	11.5				
	NJ_C	0.0	0.2	0.4	-	11.7	9.0	7.8	-	11.7	11.5	8.3	-	11.7	11.5	8.0	-	11.7	11.5	8.0				
	LNJ_C	0.0	0.1	0.7	-	14.6	14.4	13.4	-	13.5	14.8	14.3	-	12.8	14.6	13.8	-	12.8	14.6	13.8				
	J_U	0.0	0.0	0.0	8.0	9.0	7.8	6.8	8.0	8.8	8.2	7.0	8.1	8.7	9.3	7.3	3.3							
	NJ_U	0.7	0.9	1.0	39.7	37.6	24.5	17.1	39.7	39.4	29.9	20.9	39.7	39.6	30.7	21.7								
	LNJ_U	0.4	1.9	11.9	37.0	36.5	37.5	35.1	40.3	35.3	33.2	32.3	40.6	34.8	33.8	32.5								
KD	0.0	0.0	1.9	36.9	36.5	37.8	34.9	40.3	35.0	33.1	32.1	40.6	34.6	33.7	32.8									
J_p	0.0	0.0	0.1	0.0	14.2	9.2	5.6	0.0	15.5	12.6	7.1	0.0	15.8	12.4	7.8									
LNJ_p	0.3	0.0	0.1	36.5	12.2	3.9	1.8	40.4	2.6	1.4	2.2	40.5	1.9	1.7	2.2									
J_C	0.0	3.3	14.0	25.4	28.3	30.3	28.4	24.5	27.4	30.7	29.3	24.4	27.0	30.2	29.9									
NJ_C	0.0	0.4	1.3	35.9	34.7	24.0	18.7	35.9	36.6	28.2	22.3	35.9	36.8	29.5	22.8									
LNJ_C	0.2	0.2	0.9	38.7	37.7	37.7	37.3	36.9	37.2	37.7	36.5	36.5	37.5	38.0	36.6									
500	J_U	0.0	0.0	0.0	15.4	14.8	10.6	9.0	15.5	14.7	11.4	9.2	15.4	11.4	9.2	4.5								
	NJ_U	0.1	0.1	0.1	68.8	58.8	40.5	30.1	71.1	64.5	46.9	31.7	72.2	66.5	48.5	33.8								
	LNJ_U	1.0	8.2	38.3	68.9	70.5	67.2	67.0	64.5	63.0	63.1	63.5	64.8	62.8	63.6	63.6								
	KD	0.3	0.7	3.7	69.0	70.7	67.1	67.5	64.4	63.0	62.7	63.7	64.6	63.2	63.4	63.8								
	J_p	0.2	0.2	0.2	46.4	38.0	22.2	15.1	47.9	42.9	27.3	16.8	48.3	48.3	28.9	18.2								
	LNJ_p	0.4	0.1	0.1	32.0	17.4	10.1	4.3	5.1	8.8	11.3	6.1	3.1	8.4	10.9	7.0								
	J_C	0.9	19.1	39.9	63.2	64.4	61.3	58.7	63.0	63.9	61.6	59.2	62.6	63.8	61.8	59.9								
	NJ_C	0.4	1.2	4.9	68.6	60.5	48.7	41.3	70.0	64.8	51.1	44.0	71.1	65.9	51.5	44.1								
	LNJ_C	0.4	0.2	2.1	69.6	70.1	71.2	70.5	70.0	69.5	70.3	69.	70.1	69.7	70.3	70.7								

표 6: 시나리오5의 모의실험 결과

n	metric	Smoothed periodogram															Correlation								
		Raw periodogram					Bartlett window					Bohman window					Parzen window			ACF	PACF				
		10%	20%	50%	1%	2%	5%	10%	1%	2%	5%	10%	1%	2%	5%	10%	1%	2%	5%						
100	J_U	0.0	0.0	0.0	-	0.9	0.5	0.3	-	0.8	0.8	0.5	0.5	-	0.8	0.8	0.5	0.5	0.8	0.8	0.5	0.5	0.0	0.4	
	NJ_U	0.0	0.0	0.0	-	0.3	0.1	0.0	-	0.3	0.0	0.1	0.1	-	0.3	0.0	0.1	0.1	0.3	0.0	0.1	0.1	-	-	
	LNJ_U	1.3	1.0	0.0	-	0.3	0.2	0.9	-	0.2	0.2	1.2	1.2	-	0.3	0.4	0.9	0.9	0.3	0.4	0.9	0.9	-	-	
	KD	0.9	0.5	0.0	-	0.3	0.2	0.9	-	0.2	0.2	1.2	1.2	-	0.3	0.4	0.9	0.9	0.3	0.4	0.9	0.9	-	-	
	J_p	0.1	0.0	0.0	-	0.0	0.0	0.0	-	0.0	0.0	0.0	0.0	-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	LNJ_p	0.0	0.0	0.0	-	0.3	0.0	0.0	-	0.3	0.0	0.0	0.0	-	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	
	J_C	0.2	0.1	0.1	-	1.1	0.4	0.4	-	0.9	0.5	0.6	0.6	-	0.9	0.6	0.6	0.6	0.9	0.6	0.6	0.6	0.6	0.6	
	NJ_C	0.2	0.0	0.1	-	0.3	0.0	0.0	-	0.3	0.0	0.0	0.0	-	0.3	0.1	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.0	
	LNJ_C	0.0	0.1	0.0	-	0.3	0.4	0.5	-	0.3	0.1	1.2	1.2	-	0.3	0.1	1.3	1.3	0.3	0.1	1.3	1.3	0.0	0.0	
	200	J_U	0.0	0.0	0.0	2.5	2.1	1.7	1.3	2.8	2.1	1.7	0.9	0.9	2.6	2.0	1.7	1.1	0.9	2.6	2.0	1.7	1.1	0.0	1.3
		NJ_U	0.0	0.0	0.0	0.5	0.3	0.1	0.1	0.5	0.4	0.2	0.1	0.1	0.5	0.4	0.3	0.1	0.1	0.5	0.4	0.3	0.1	-	-
		LNJ_U	4.7	1.4	0.6	0.4	1.0	2.2	6.2	0.5	1.0	6.6	16.1	16.1	0.5	0.4	5.4	15.4	0.5	0.4	5.4	15.4	0.0	0.0	0.0
KD		2.1	0.8	0.0	0.4	1.0	2.2	6.0	0.5	1.0	6.7	16.8	16.8	0.5	0.4	5.7	16.2	0.5	0.4	5.7	16.2	0.0	0.0	0.0	
J_p		0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
LNJ_p		0.0	0.1	0.0	0.4	0.0	0.3	0.1	0.5	0.0	0.2	0.5	0.2	0.5	0.1	0.1	0.3	0.1	0.1	0.1	0.1	0.3	0.0	0.0	
J_C		0.0	0.0	0.1	2.4	2.1	2.2	1.6	2.7	2.1	1.8	1.2	1.2	2.7	1.9	1.7	1.3	1.2	2.7	1.9	1.7	1.3	0.0	0.0	
NJ_C		0.0	0.0	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.2	0.0	0.0	0.3	0.4	0.2	0.0	0.0	0.3	0.4	0.2	0.0	0.0	0.0	
LNJ_C		0.2	0.0	0.1	0.4	0.9	2.3	3.3	0.4	1.0	6.3	13.4	13.4	0.4	0.4	5.6	12.4	0.4	0.4	5.6	12.4	0.0	0.0	0.0	
500		J_U	0.0	0.0	0.0	10.8	9.0	6.4	3.5	10.4	9.6	6.4	3.3	3.3	10.5	9.5	6.4	4.1	3.3	10.5	9.5	6.4	4.1	0.0	11.0
		NJ_U	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0
		LNJ_U	23.8	12.3	5.0	7.9	22.5	44.2	53.2	5.4	33.6	63.6	63.5	63.5	3.8	30.2	61.8	62.9	3.8	30.2	61.8	62.9	0.0	0.0	0.0
	KD	5.9	2.0	0.1	7.9	22.6	44.1	53.8	5.4	34.2	64.2	64.8	64.8	3.8	30.5	62.7	64.1	3.8	30.5	62.7	64.1	0.0	0.0	0.0	
	J_p	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
	LNJ_p	0.2	0.0	0.1	0.1	0.0	0.4	0.4	0.3	1.4	5.4	4.2	4.2	0.3	1.3	5.2	4.5	0.3	1.3	5.2	4.5	4.5	0.0	0.0	
	J_C	0.1	0.3	0.3	10.5	9.7	6.1	3.8	10.7	9.8	6.6	4.5	4.5	10.4	10.2	6.7	4.3	10.4	10.2	6.7	4.3	4.3	0.0	0.0	
	NJ_C	0.1	0.1	0.0	0.2	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.0	
	LNJ_C	0.3	0.1	0.1	7.8	19.2	37.0	41.4	5.4	31.9	60.5	60.8	60.8	4.3	28.1	58.8	60.6	4.3	28.1	58.8	60.6	0.0	0.0	0.0	

표 7: 시나리오6의 모의실험 결과

n	metric	Raw periodogram										Smoothed periodogram										Correlation		
		Bartlett window					Bohman window					Parzen window					ACF	PACF						
		10%	20%	50%	1%	2%	5%	10%	1%	2%	5%	10%	1%	2%	5%	10%			1%	2%	5%			
100	J_U	0.1	0.1	0.1	-	18.1	20.7	18.6	-	13.5	21.9	20.8	-	12.6	21.9	20.8	-	12.6	21.9	20.8	0.4	36.9		
	NJ_U	0.0	0.0	0.2	-	21.9	16.3	9.3	-	21.9	18.1	12.5	-	21.9	19.8	13.5	-	21.9	19.8	13.5				
	LNJ_U	41.4	36.6	17.5	-	23.6	50.3	60.4	-	22.3	45.3	64.4	-	22.1	41.9	62.8	-	22.1	41.9	62.8				
	KD	25.0	15.6	4.3	-	23.5	50.7	60.9	-	22.3	45.7	65.7	-	22.1	42.3	63.7	-	22.1	42.3	63.7				
	J_p	0.4	0.7	0.0	-	0.0	0.0	0.0	-	0.0	0.0	0.0	-	0.1	0.0	0.0	-	0.1	0.0	0.0				
	LNJ_p	0.2	1.2	0.7	-	19.0	0.0	0.0	-	17.6	0.0	0.0	-	17.7	0.3	0.0	-	17.7	0.3	0.0				
	J_c	0.0	0.8	0.3	-	10.8	8.8	4.6	-	9.0	11.2	5.9	-	8.9	11.6	6.4	-	8.9	11.6	6.4				
	NJ_c	0.0	0.1	0.1	-	5.4	1.2	0.2	-	5.4	3.7	0.9	-	5.4	4.9	1.1	-	5.4	4.9	1.1				
	LNJ_c	0.3	1.2	0.8	-	6.6	7.7	3.2	-	5.7	11.5	9.0	-	5.8	11.9	9.6	-	5.8	11.9	9.6				
	200	J_U	0.0	0.1	0.0	43.8	46.5	48.5	45.5	36.9	47.5	49.2	47.7	34.4	47.1	49.7	47.4	2.5	47.1	49.7	47.4	69.8		
NJ_U		0.0	0.0	0.2	42.7	41.6	32.3	23.8	42.7	41.3	36.3	26.0	42.7	42.0	37.2	27.3		42.0	37.2	27.3				
LNJ_U		69.7	66.4	45.0	46.6	73.1	84.2	87.8	44.4	66.4	85.6	88.1	43.5	62.5	85.7	87.4		62.5	85.7	87.4				
KD		39.4	28.6	9.5	46.6	73.5	84.9	88.4	44.4	66.3	86.3	88.0	43.5	62.5	86.4	87.8		62.5	86.4	87.8				
J_p		0.6	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		0.0	0.0	0.0				
LNJ_p		0.2	1.8	0.2	41.9	0.2	0.0	0.0	40.4	1.1	0.1	0.0	0.0	39.9	1.3	0.1	0.0	39.9	1.3	0.1	0.0			
J_c		0.1	1.1	0.4	31.1	31.9	23.5	12.0	28.5	33.2	26.9	14.1	26.9	33.6	29.2	15.2		33.6	29.2	15.2				
NJ_c		0.0	0.5	0.0	9.5	9.7	2.5	0.5	9.5	9.6	4.4	0.7	9.5	9.4	4.6	1.1		9.4	4.6	1.1				
LNJ_c		0.3	1.1	0.2	11.5	18.3	17.1	7.4	10.7	19.7	29.1	18.7	10.4	17.3	28.5	20.0		17.3	28.5	20.0				
500		J_U	0.0	0.6	0.0	81.7	80.4	79.4	78.4	81.6	81.9	79.7	80.6	81.3	81.0	80.0	80.6	4.7	81.0	80.0	80.6	91.2		
	NJ_U	0.0	0.0	0.9	64.8	62.3	58.5	52.1	65.6	62.4	58.5	50.8	64.8	64.6	58.9	51.8		64.6	58.9	51.8				
	LNJ_U	89.3	89.2	81.5	93.1	96.9	97.2	97.5	93.9	97.5	97.4	97.5	92.4	97.7	97.2	97.8		97.7	97.2	97.8				
	KD	83.1	43.7	17.2	93.2	97.1	97.3	97.6	94.4	97.6	97.8	97.1	92.5	97.6	97.6	97.4		97.6	97.6	97.4				
	J_p	0.8	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		0.0	0.0	0.0				
	LNJ_p	1.8	2.8	0.0	0.6	0.0	0.0	0.0	2.8	1.2	0.0	0.0	3.8	1.2	0.0	0.0		1.2	0.0	0.0				
	J_c	0.2	0.8	0.6	64.2	60.2	46.6	25.5	63.2	60.7	48.4	28.7	63.9	61.0	50.5	32.3		61.0	50.5	32.3				
	NJ_c	0.1	0.7	0.0	11.4	10.6	3.0	0.5	10.9	11.6	4.2	1.2	10.8	11.3	4.9	1.3		11.3	4.9	1.3				
	LNJ_c	0.4	1.4	0.0	28.0	40.1	41.7	20.6	31.8	53.8	58.1	35.9	29.7	51.9	59.3	39.4		51.9	59.3	39.4				

다음으로, 시나리오 3과 4에서 두 군집은 큰 값의 모수를 갖는 AR 모형과 상대적으로 작은 값을 모수로 갖는 AR 모형으로 구성되어 있다. 평활창의 종류, 평활의 폭, 표본 크기에 관계없이 평활된 주기도를 이용한 거리 척도가 원 주기도를 이용한 거리 척도에 비해서 월등히 좋은 정분류율을 갖는다. 시나리오 3에서, 평활된 주기도를 이용한 거리 척도는 자기상관에 근거한 거리 척도보다 훨씬 더 성능이 뛰어나다. 시나리오 4에서 n 이 커질수록 평활된 주기도를 이용한 거리 척도의 정분류율만큼 높은 결과를 보여준다. 시나리오 3, 4에서 평활된 주기도의 NJ_U 가 가장 높은 정분류율을 보인다.

시나리오 4에서, PACF는 $n = 500$ 인 경우에만 $m = 1\%$ 인 Bohman 또는 Parzen 평활창을 이용하여 평활된 주기도의 LNJ_U 또는 KD보다 약 5% 더 높은 결과를 보이며, 평활된 주기도의 NJ_U 만큼 좋은 성능을 가진다. J_p 는 n 이 커질수록 정분류율이 높아지지만, $n = 200$ 이상이면 다른 거리 척도에 비해 현저히 낮은 결과를 보인다. $n = 200$ 이하이면 $m = 1\%$ 로 설정한 LNJ_p 가 J_p 보다 정분류율이 높으며 다른 거리 척도와 비슷한 성능을 보인다. 그림 3(c)를 보면, 양의 상관관계를 가지는 것을 알 수 있으며 이는 같은 AR 모형 간에는 분류하는 것이 쉽지 않다는 것을 알 수 있다. 평활된 주기도의 J_C 는 평활창의 종류에 관계없이 m 이 작을수록 높은 정분류율을 보인다. $m = 1\%$ 일 때 가장 좋은 성능을 가지며, 이는 Bohman 또는 Parzen 평활창에 근거한 평활된 주기도의 LNJ_U 또는 KD와 비슷한 수준이다.

마지막으로 시나리오 5, 6에서 설정한 두 군집은 큰 값의 모수를 갖는 MA 과정과 상대적으로 작은 값을 모수로 갖는 MA 과정으로 구성되어 있다. 시나리오 5, 6에서 주목해야할 점은 원 주기도의 LNJ_U 또는 KD는 n 의 10%만을 이용하여 거리를 계산하여 실험한 것이 50%

를 이용한 것에 비해 월등히 높은 정분류율을 보인다는 것이다. 이는 그림 3(e)와 그림 3(f)을 통해 알 수 있다. 그림 3의 나머지 4가지 그림과 달리 그림 3(e)와 (f)는 두 군집의 다른 그래프가 낮은 주파수에서 주기도 값의 차이가 크게 나타나며 주파수가 높아질수록 주기도 값이 비슷해 짐을 확인할 수 있다. 이러한 경향은 원 주기도의 KD 가 LNJ_U 에 비해 확연하다.

제 4 장

실증연구

4.1 실증 자료

본 장에서는 시계열 자료의 군집분석시 유사성 거리 척도를 실제 자료에 적용한 결과를 소개하고자 한다. 실증연구에서 군집분석을 위해 사용한 시계열 자료는 제조업생산지수 자료이다. 본 연구에서 사용한 제조업생산지수 자료는 통계청(<http://kosis.kr>)에서 얻은 1990년 1월부터 2015년 4월까지의 304개월 동안의 24개 산업에 대한 전국 월평균 자료를 이용한다.

그림 4에 나타낸 실증자료의 시도표를 살펴보면 변동폭이 거의 없이 꾸준히 증가하는 산업 또는 감소하는 산업이 있으며, 변동폭이 큰 산업이 있다. 또한, 특정 시기에 변화가 큰 형태를 보이는 산업도 있다. 예를 들면, C201, C203, C241, C242는 변동폭이 거의 유지되며 꾸준히 증가하는 산업이고, C151, C274, C333은 꾸준히 감소하는 산업임을 알 수 있다. C105, C112, C202는 다른 산업에 비해 변동폭이 매우 큰 산업이며, C243, C283, C292는 특정 시기에 큰 변화를 보이는 산업이다. 이처럼 다양한 형태를 갖는 산업 24개를 고려하여, 시계열이 갖는 성질에 맞게 군집을 구성하고자 한다.

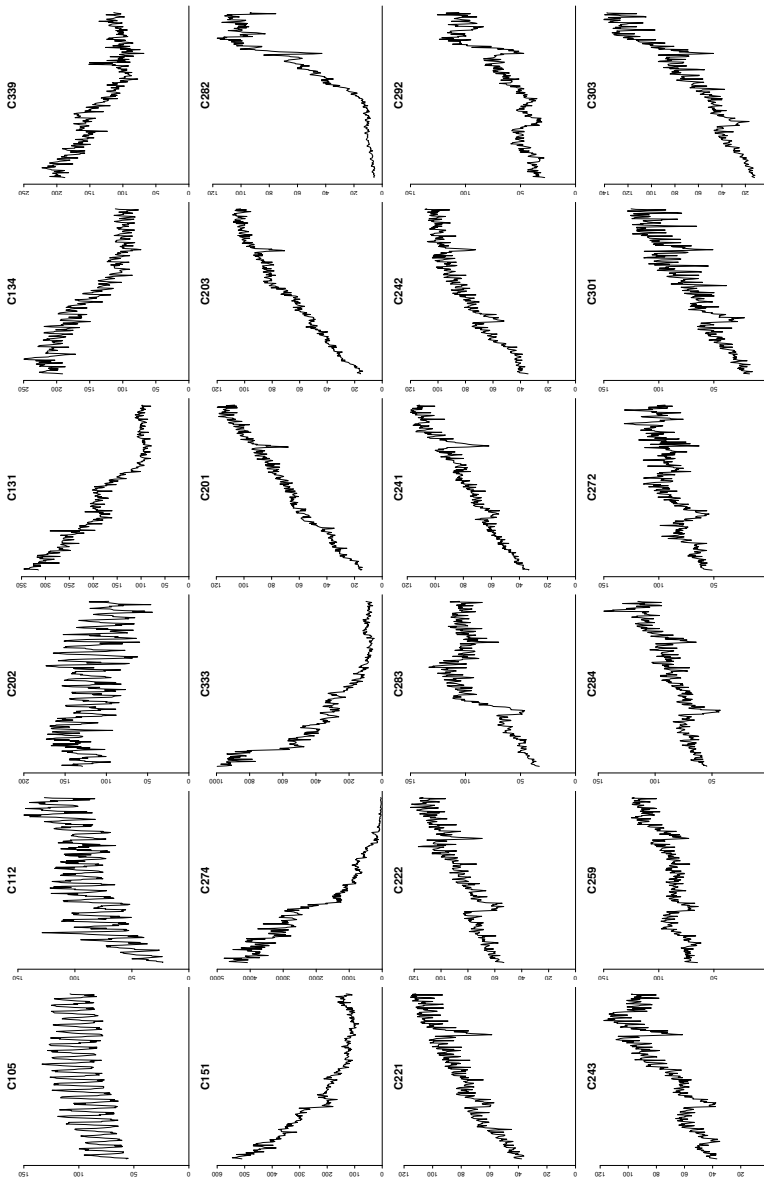


그림 4: 24개 산업의 월평균 제조업생산지수 자료의 시도표.

4.2 실증연구 방법

4.1절에서 소개한 실증 자료를 이용하여 유클리드 거리인 d_{NJ_U} , d_{LNJ_U} 와 d_{KD} , 평활된 주기도를 이용한 $d_{NJ_U}^*$, $d_{LNJ_U}^*$ 와 d_{KD}^* , 피어슨 상관 계수를 이용한 d_{J_p} 와 d_{LNJ_p} , 평활된 주기도를 이용한 $d_{J_p}^*$ 와 $d_{LNJ_p}^*$, 마지막으로 급내 상관계수를 근거로 한 d_{J_c} , d_{NJ_c} 와 d_{LNJ_c} , 평활된 주기도를 이용한 $d_{J_c}^*$, $d_{NJ_c}^*$ 와 $d_{LNJ_c}^*$ 총 8개의 척도를 이용하여 군집분석을 실시한다. 평활된 주기도를 계산할 때 Bohman 평활창을 이용하여 평활의 폭을 n 의 1%, 5%로 지정하였다. 분석에 고려된 24개 산업의 304개월 동안의 시계열자료의 시도표를 살펴보면 산업마다 다양한 계절 변동을 보이고 있다. 따라서, 계절 변동을 제거하기 위해 12차 차분후 분석에 사용하였으므로 $n = 292$ 이다.

4.3 실증연구 결과

4.1절에서 소개한 제조업생산지수 자료를 이용하여 실증연구를 실시하였다. 그림 5, 그림 6, 그리고 그림 7은 총 8가지 거리척도를 이용하여 군집분석을 실시하여 8개의 군집으로 나눈 덴드로그램을 보여준다. 실증연구에 사용된 총 8가지의 거리척도는 전반적으로 자료가 가지고 있는 추세에 따라 군집을 잘 분류하였으나, 변동폭에 대해서는 잘 구분하지 못하였다. 그 중 가장 자료의 구조가 유사한 결과는 평활의 폭을 n 의 1%로 지정한 $d_{NJ_U}^*$, $d_{LNJ_p}^*$, $d_{J_c}^*$ 척도를 이용한 결과이다. 이를 통해 그림 8, 그림 9, 그림 10에는 그림 3의 배열과 같으며 덴드로그램을 바탕으로 24개의 산업이 8개의 군집으로 분류된 결과를 시도표로 보여준다.

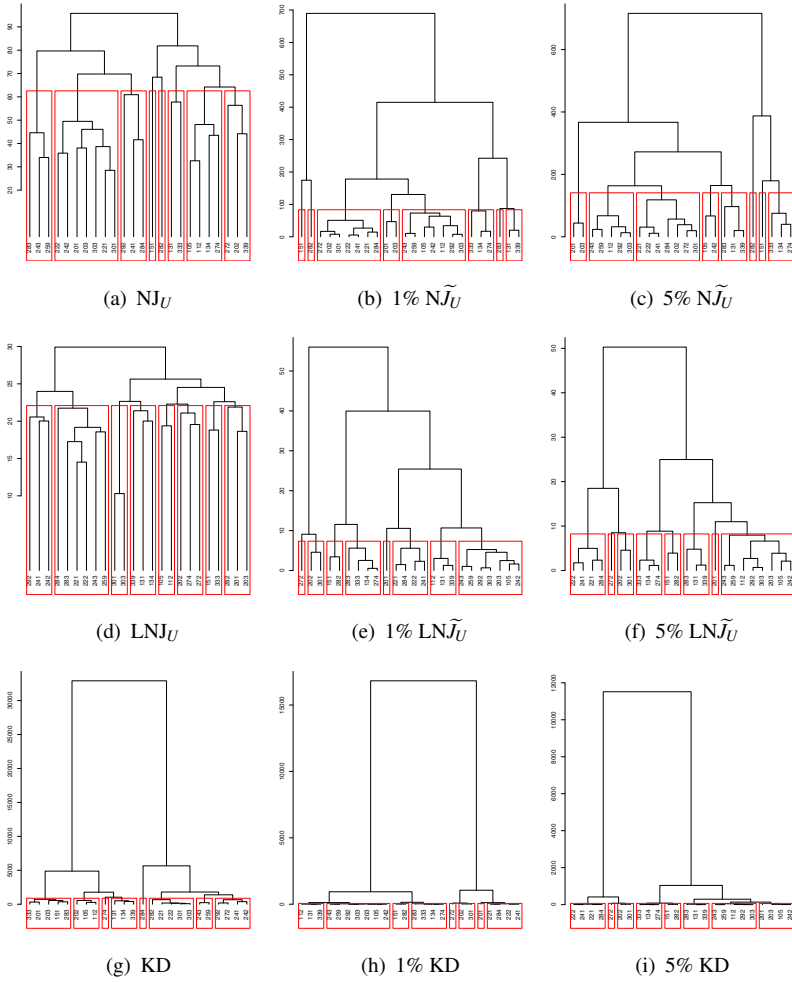


그림 5: 실증자료 군집분석 덴드로그램 : 유클리드 거리

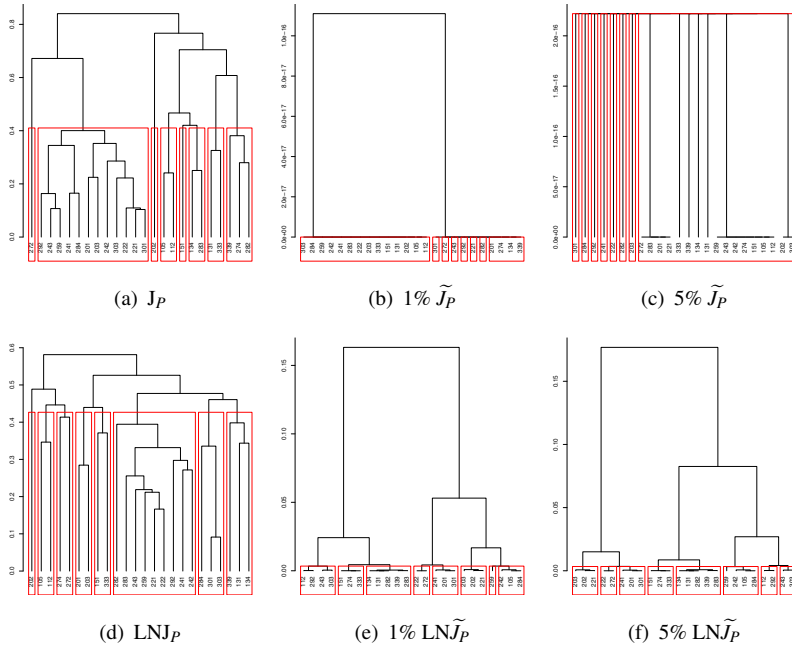


그림 6: 실증자료 군집분석 덴드로그램 : 피어슨 상관계수

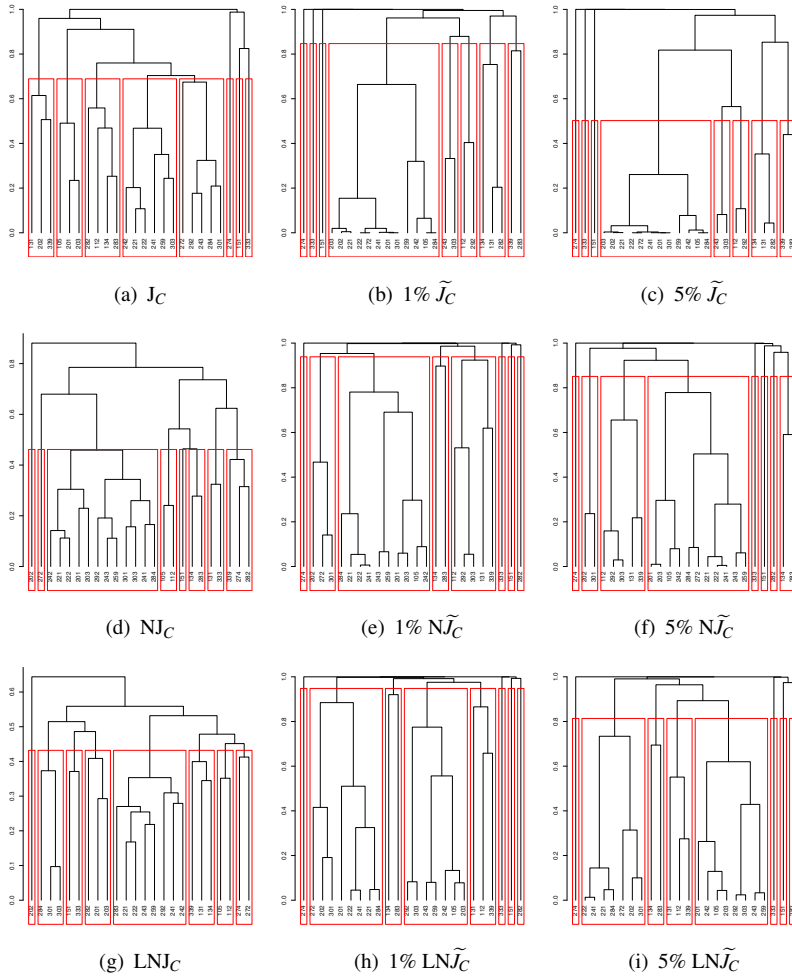


그림 7: 실증자료 군집분석 덴드로그램 : 급내 상관계수

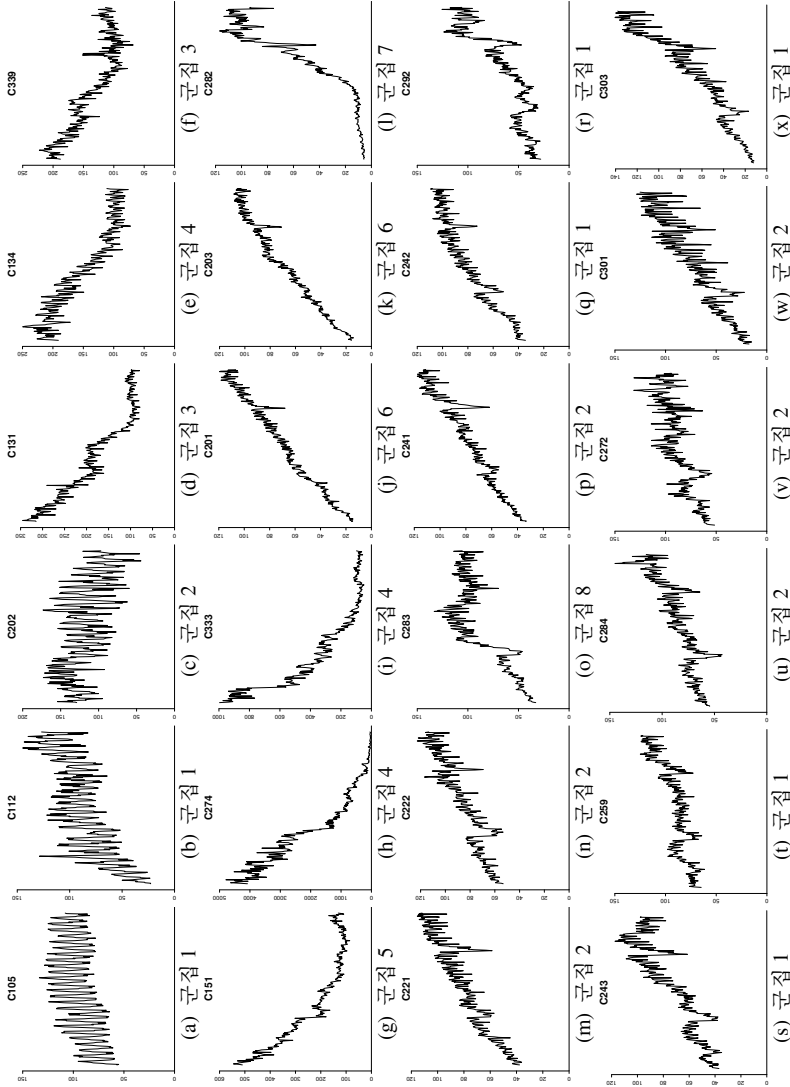


그림 8: 1% NJU 를 이용한 군집분석 결과.

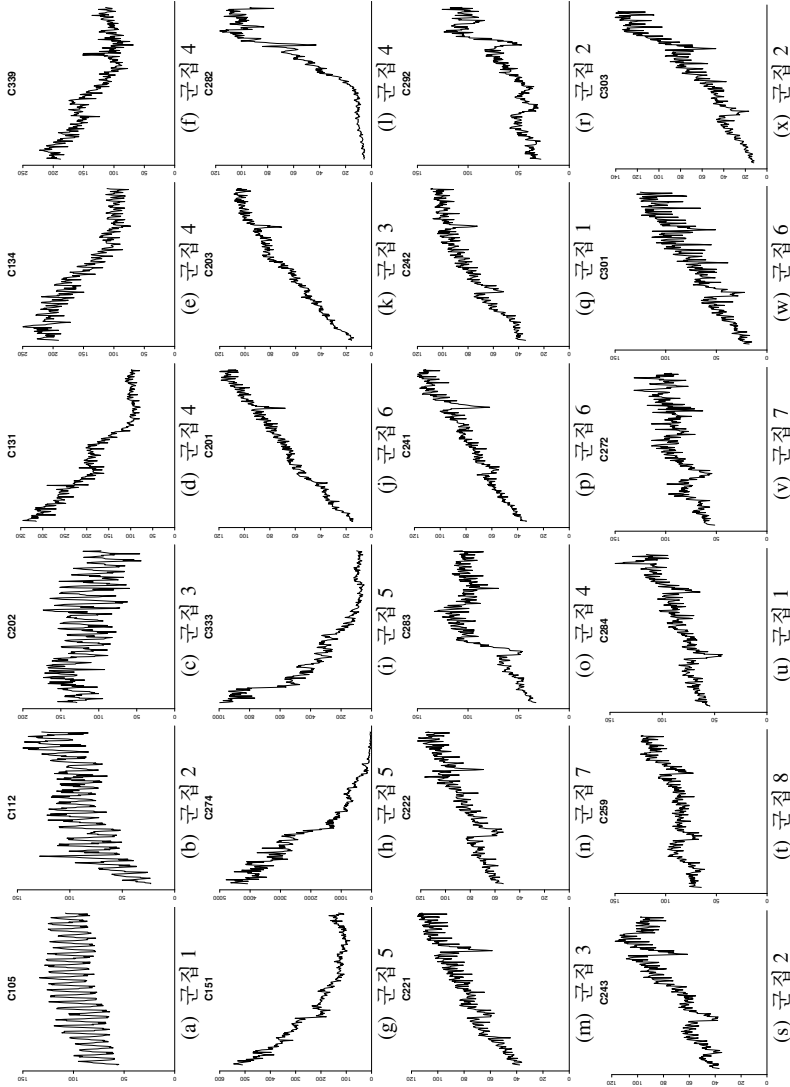


그림 9: 1% LNJ_p 를 이용한 군집분석 결과.

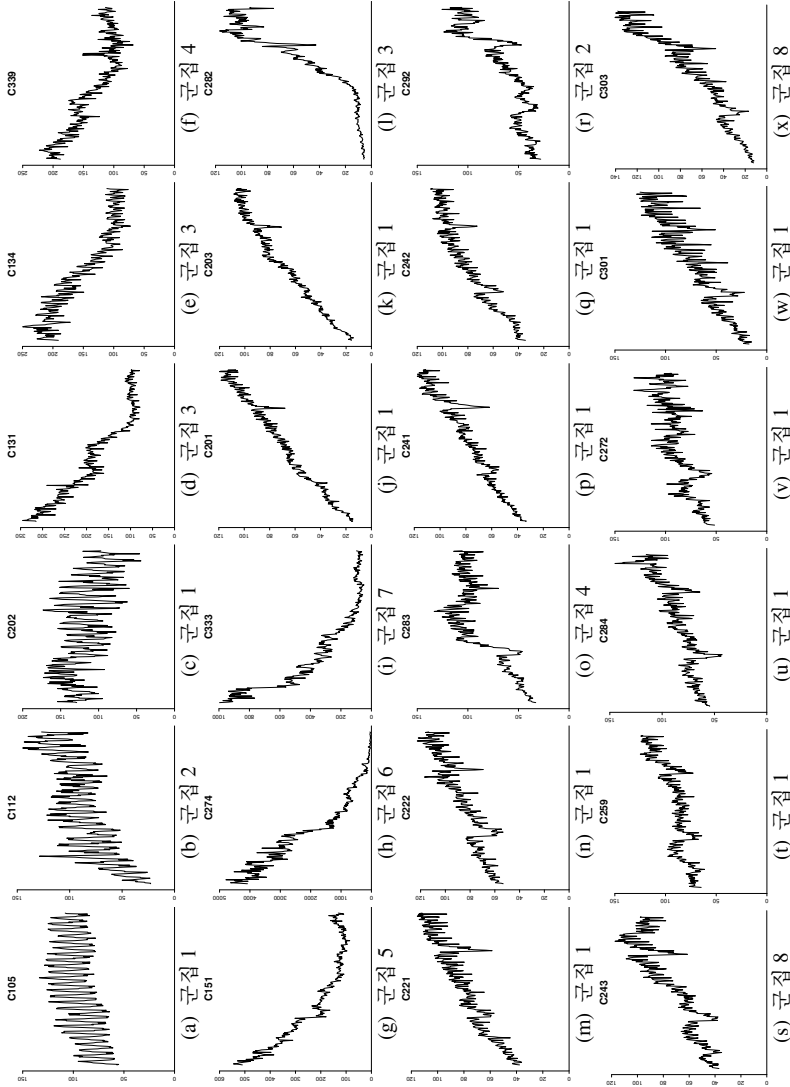


그림 10: 1% \tilde{J}_C 를 이용한 군집분석 결과.

제 5 장

결론

시계열 자료를 분석하는 과정에서 확률모형설정 및 예측을 수행해야 할 계열수가 많은 경우, 주어진 시계열을 각각 예측하는 대신 특징이 유사한 시계열들을 군집화한 후, 각 그룹으로부터 재표현된 것에 기초하여 예측을 수행한다면 상당히 효율적일 것이다. 이에 본 논문에서는 시계열 자료를 군집화하기 위하여 사용되는 거리를 시간 영역에서와 주파수 영역에서로 나누어 소개하였으며, 주파수 영역에 좀 더 초점을 맞추었다. 유클리드 거리뿐만 아니라 주기도 간의 상관성을 고려하여 피어슨 상관계수와 급내 상관계수를 이용한 거리를 제안하여 모의실험을 통해 그 성능을 비교하였다.

모의실험 결과, 주파수를 모두 사용한 것이 정분류율이 높았으며, 1% 평활된 주기도를 이용하면 평활하지 않은 주기도를 이용한 것보다 모든 거리 척도에서 정분류율이 월등히 높아진다. 그 중에서 LNJ_U , J_P , J_C 의 정분류율이 높았다. J_C 가 J_P 보다 높은 정분류율을 보였으며, n 이 커질수록 정분류율이 높아졌다.

실증연구에서 통계청에서 제공하는 1990년 1월부터 2015년 4월까지 전국 월평균 제조업생산지수 자료를 이용하여, 모의실험에서 우수한 성능을 보인 거리척도를 적용하여 8개의 군집으로 군집화하였다. 그 결과 평활의 폭을 1%로 적용한 평활된 주기도에 의한 거리가 시계열의 형태를 잘 구분하는 것을 알 수 있었다. J_P 를 이용한 경우, J_C 를

이용한 결과보다 비슷한 형태의 시계열들을 고르게 군집화하였다.

향후에는 제안된 주기도의 상관성에 근거한 거리를 이용하여 정상시계열과 비정상시계열을 분류하는 모의실험을 수행해볼 것이다. 또한, 주기도의 교차상관성에 근거한 거리에 대해 연구해보고자 한다.

참고 문헌

- [1] 경민수, 김상단, 김보경, 김형수 (2007). 군집분석을 통한 수문학적 가뭄의 가뭄심도-가뭄면적-가뭄지속기간 곡선의 작성, 대한토목학회논문집 B, 제 27권, 제 3B호, 267-276.
- [2] 고정웅, 백희정, 권원태, 박종열 (2006). 한반도의 기온 분포 특성과 지역 구분, 기후연구, 제 1권, 제 1호, 3-14.
- [3] 김해경, 이명숙 (2005). 경제 및 금융자료를 위한 시계열 분석, 경문사.
- [4] 박만식, 김나영, 김희영 (2008). 이분산 시계열모형을 이용한 국내주식자료의 군집분석, *Communications of the Korean Statistical Society*, Vol. 15, No. 6, 925-937.
- [5] 박만식, 김희영 (2008). 평활된 주기도를 이용한 강수량자료의 군집화, 응용통계연구, 제 21권, 제 3호, 547-560.
- [6] 박유성, 김기환 (2002). SAS/ETS를 이용한 시계열자료분석 I, 자유아카데미.
- [7] 손영숙, 백장선 (2005). 시간경로 유전자 발현자료의 군집분석에서 이질적인 시계열의 탐지를 위한 패턴일치지수, 응용통계연구, 제 18권, 제 2호, 371-379.
- [8] 신태수 (2007). 스펙트럼 분석을 이용한 시계열 회귀분석 모델의 계절성 변수 선정, 서강대학교 대학원, 석사학위논문.
- [9] 이종협 (2007). 시계열분석과 응용, 자유아카데미.
- [10] 정준모, 김현섭, 이상태, 이경재, 김미숙, 전영우 (2015). 잣나무 연륜생장과 기후요소와의 상관관계 분석 및 성장예측, 한국농림기상학회지, 제 17권, 제 2호, 85-92.
- [11] 조신섭, 손영숙 (2009). SAS/ETS를 이용한 시계열분석, 율곡출판사.

- [12] 천광일 (2014). 계방산 산림식생의 공간적 분포 특성 및 시계열적 변화에 관한 연구, 경북대학교 대학원, 박사학위논문.
- [13] 통계청, 광업·제조업동향조사.
- [14] 한광중 (2015). 미래예측과 시계열 분석, 백산출판사.
- [15] 홍한음, 박민정, 조신섭 (2009). 시계열자료 군집방법의 비교연구, 응용통계연구, 제 22권, 제 6호, 1203-1214.
- [16] Caiado, J., Crato, N., and Peña, D. (2006). A periodogram-based metric for time series classification, *Computational Statistics & Data Analysis*, Vol. 50, 2668-2684.
- [17] Chatfield, C. (1979). Inverse autocorrelations, *Journal of Royal Statistical Society Series A*, Vol. 142, 363-377.
- [18] Corduas, M., and Piccolo, D. (2008). Time series clustering and classification by the autoregressive metric, *Computational Statistics and Data Analysis*, Vol. 52, 1860-1872.
- [19] D'Urso, P., and Maharaj, E. A. (2009). Autocorrelation-based fuzzy clustering of time series, *Fuzzy Sets and Systems*, Vol. 160, 3565-3589.
- [20] Fu, T. C., Chung, F. L, Ng, V., and Luk, R. (2001). Pattern discovery from stock time series using self-organizing maps, *KDD 2001 Workshop on Temporal Data Mining*, August 26-29, San Francisco, 27-37.
- [21] Galeano, P., and Pena, D. (2000). Multivariate analysis in vector time series, *Resenhas*, Vol. 4, No. 4, 383-403.
- [22] Goldstein, D. R., Ghosh, D., and Conlon, E. M. (2002). Statistical issue in the clustering of gene expression data, *Statistica Sinica*, Vol. 12, 219-240.
- [23] Goutte, C., Toft, P., Rostrup, E., Nielsen, F. Å., and Hansen, L. K. (1999). On Clustering fMRI Time Series, *NeuroImage*, Vol. 9, 298-310.

- [24] James H. S. (1980). Tests for Comparing Elements of a Correlation Matrix, *Psychological Bulletin*, Vol. 87, No. 2, 245-251.
- [25] Kakizawa, Y., Shumway, R. H., and Taniguchi, M. (1998). Discrimination and Clustering for Multivariate time series, *Journal of American Statistical Association*, Vol. 93, No. 441, 328-340.
- [26] Kalpakis, K., Gada, D., and Puttagunta, V. (2001). Distance measures for effective clustering of ARIMA time series, *Proceedings of the 2001 IEEE International Conference on Data Mining*, San Jose, CA, 29 Nov 2001-02 Dec 2001, 273-280.
- [27] Kardiyen, F., and Güney, H. (2015). A study on time series clustering, *Gazi University Journal of Science*, Vol. 28, No. 2, 331-347
- [28] Liao, T. W. (2005). Clustering of time series data – a survey, *Pattern Recognition*, Vol.38, 1857-1874.
- [29] Maharaj, E. A. (2000). Clustering of time series, *Journal of Classification*, Vol. 17, 297-314.
- [30] Park, M. S., and Kim, H-Y(2010). Classification of Time-Series Data Based on Several Lag Windows, *Communications of the Korean Statistical Society*, Vol. 17, 377-390.
- [31] Piccolo, D. (1990). A distance measure for classifying ARIMA models, *Journal of Time Series Analysis*, Vol.11, 153-164.
- [32] Wayne, A. F. (1979). Introduction to Statistical Time Series, John Wiley & Sons.

Abstract

Clustering Time-Series Data Based on Periodogram

Suhyun Kwon

Department of Statistics

The Graduate School

Sungshin Women's University

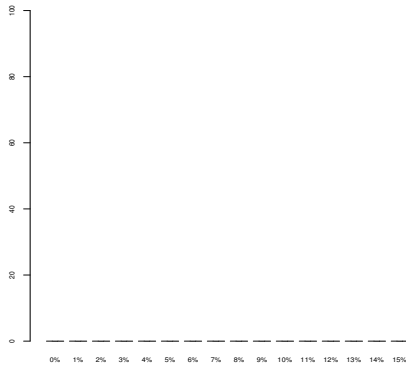
Classification and clustering time series is becoming an important area of research in various fields. Clustering time-series data is to clarify how similarity between time-series can be measured. There are two different approaches for identifying the similarity. The first approach focuses on the distances based on autocorrelation and partial autocorrelation inherent in given time-series measurements. The second approach considers the time series clustering based on the periodogram and its transformations realized in the frequency domain. In this paper, we propose the metrics based on relationship between (smoothed) periodograms for the time-series data classification. A real-data analysis with economic time series data is also presented.

Keywords : clustering, similarity, spectrum, smoothing window, frequency domain, periodogram, correlation, stationary time series

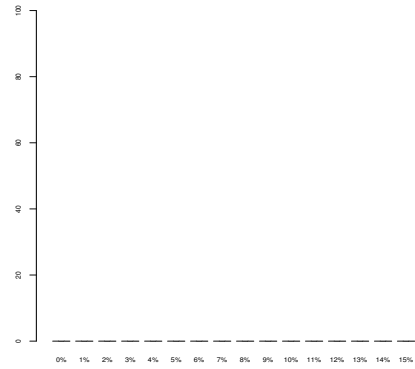
부록

부록에는 $n = 500$ 인 경우 총 9가지의 거리척도에서 평활의 폭을 1%에서 15%로 변화했을 때 정분류율이 어떻게 변화하는지 살펴보기 위해 막대 그래프를 추가로 실었다. 표에서 0%는 평활하지 않은 주기도를 이용한 결과이며, 흰색 막대는 Bartlett, 진회색 막대는 Bohman, 연한 회색 막대는 Parzen 평활창을 이용한 결과이다.

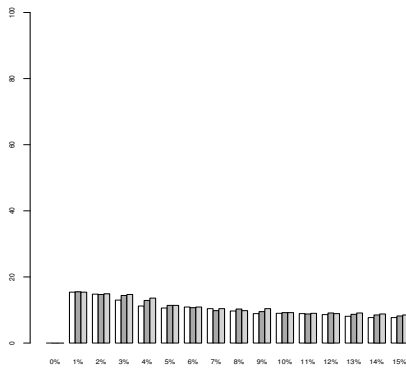
막대 그래프를 실음으로써 모의실험의 결과를 표로 나타냈을 때보다 그 경향을 더 쉽게 찾을 수 있다. 평활을 하지 않았을 때보다 1% 평활을 했을 때 정분류율이 급격하게 높아졌다. 평활의 폭을 1%, 2%로 적게하는 것이 높으며 폭을 크게 할수록 정분류율이 점차 작아지는 것을 대부분의 경우 볼 수 있다. Bohman 평활창과 Parzen 평활창은 거의 비슷하며, Bartlett 평활창보다 정분류율이 높은 경향을 보인다.



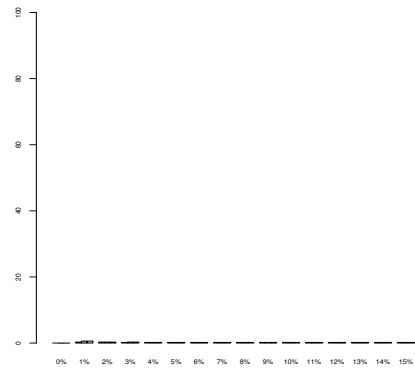
(a) 시나리오 1



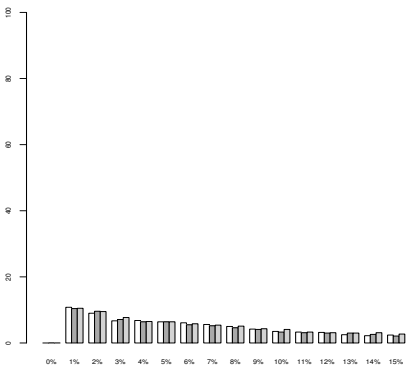
(b) 시나리오 2



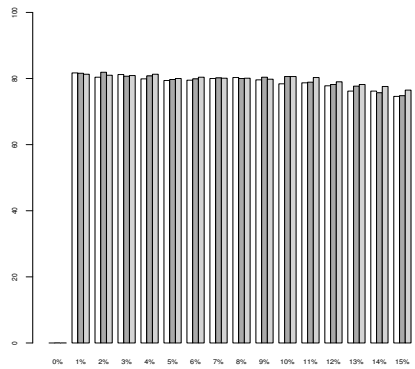
(c) 시나리오 3



(d) 시나리오 4



(e) 시나리오 5



(f) 시나리오 6

그림 11: J_U 정분류율

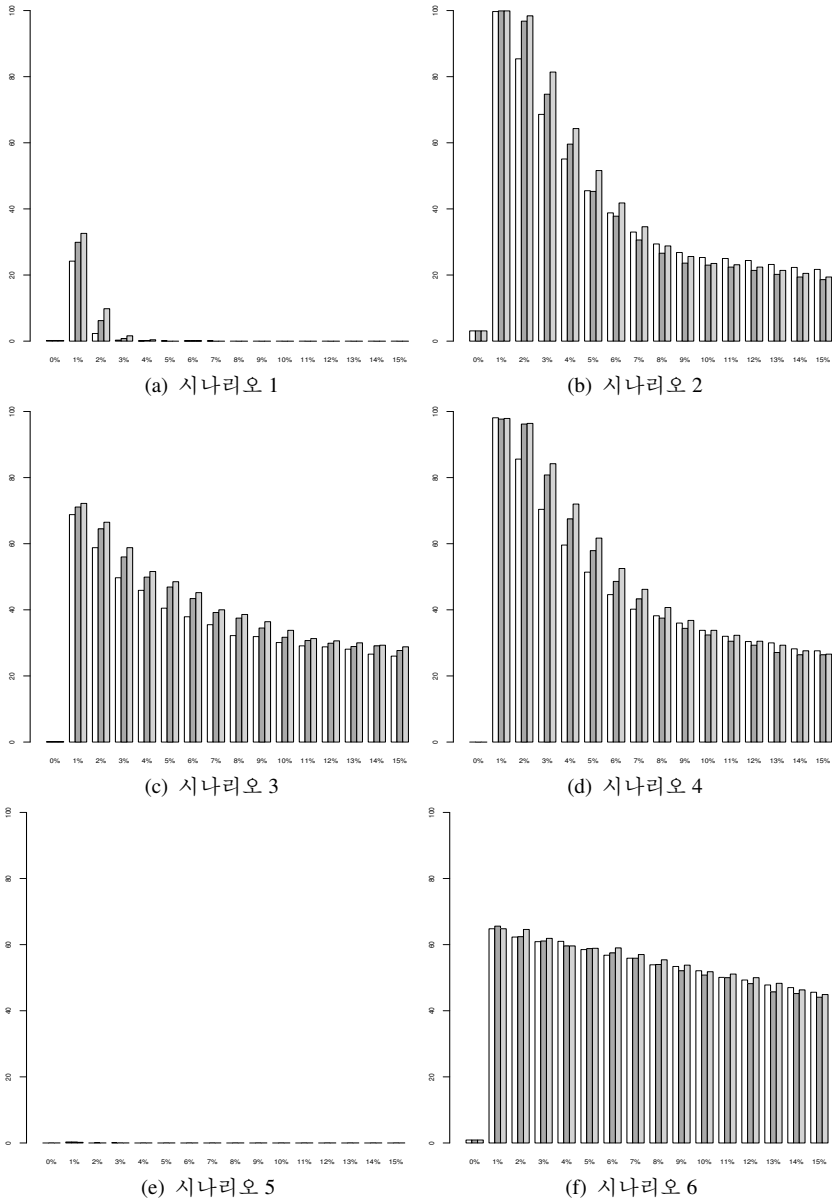


그림 12: NJ_U 정분류율

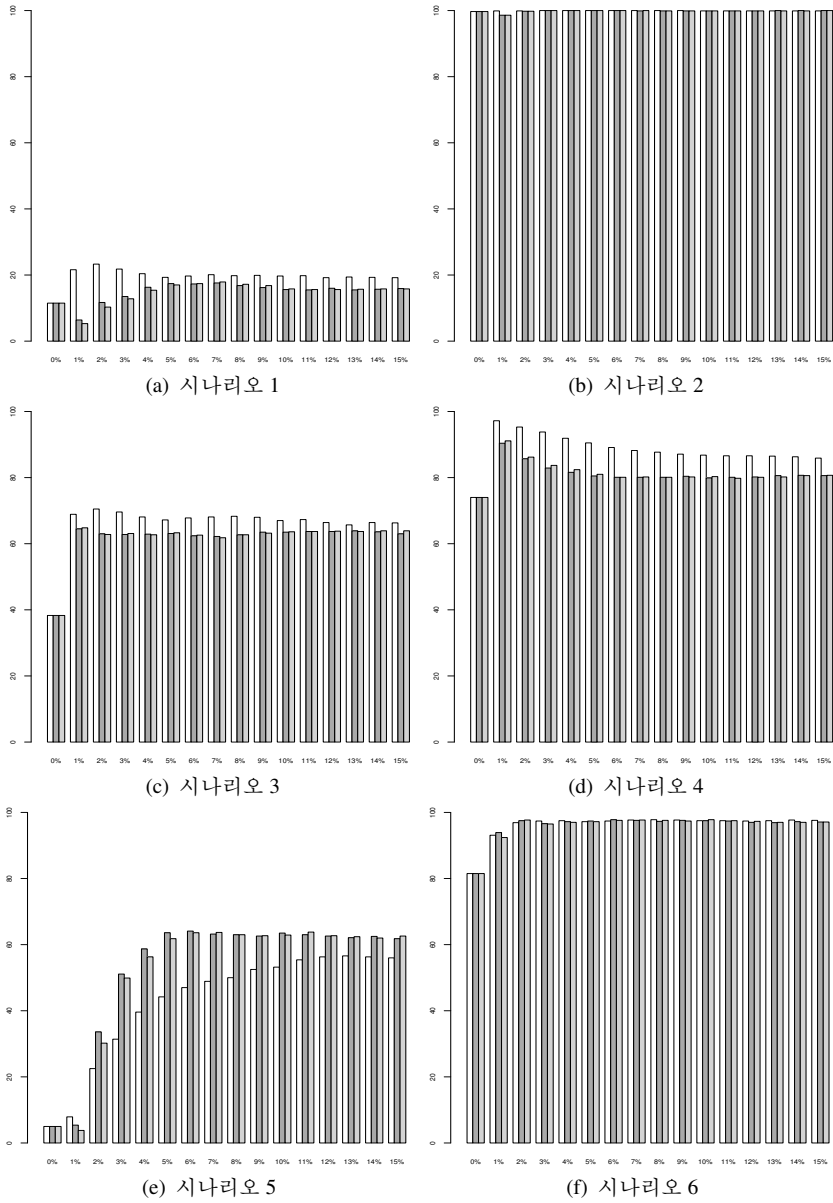


그림 13: LNJ_U 정분류율

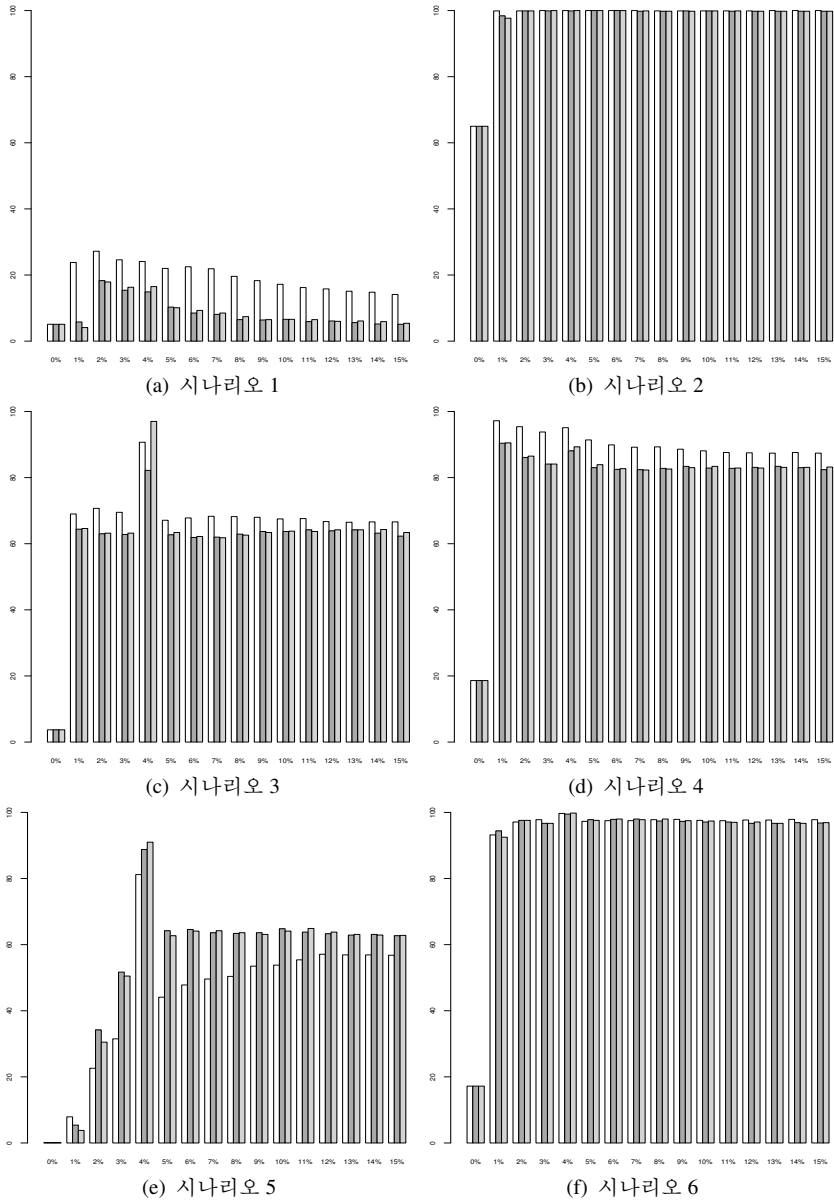


그림 14: KD 정분류율

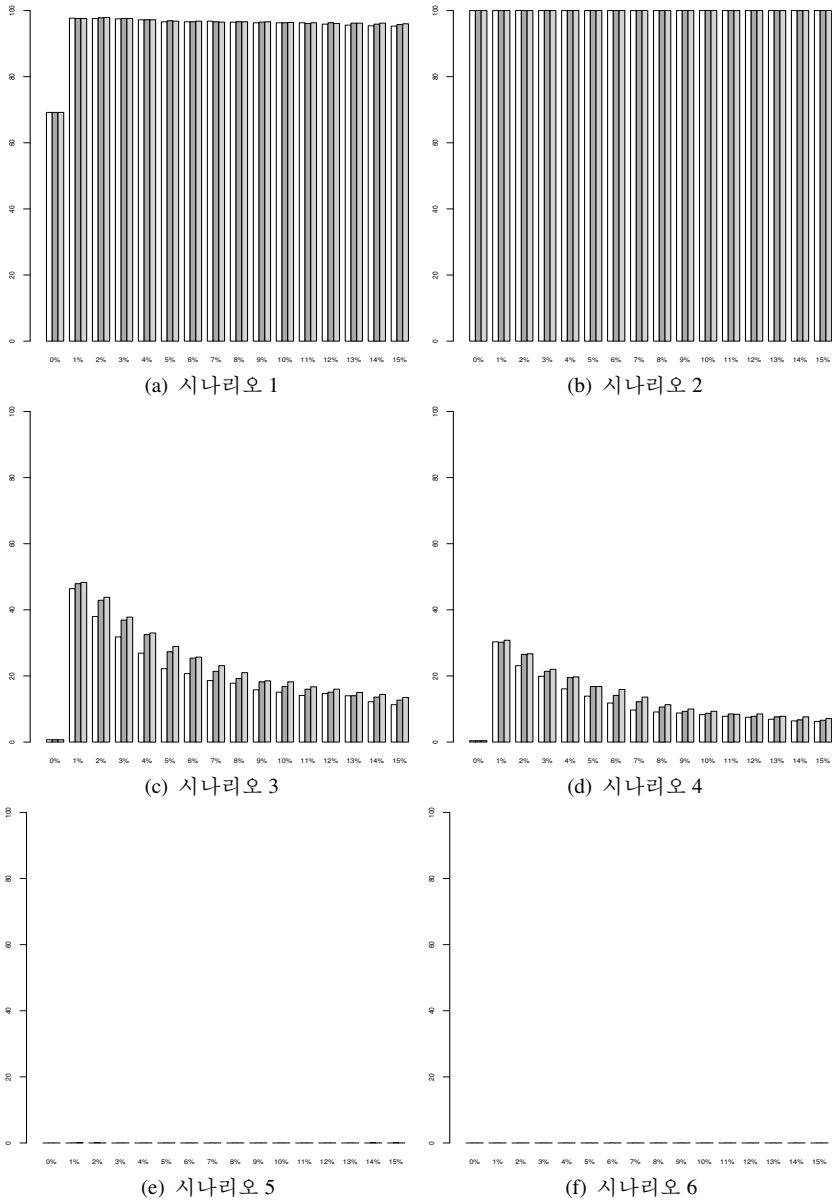
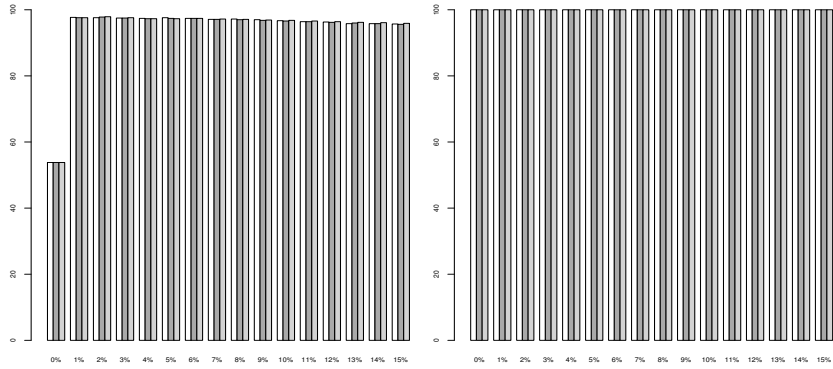
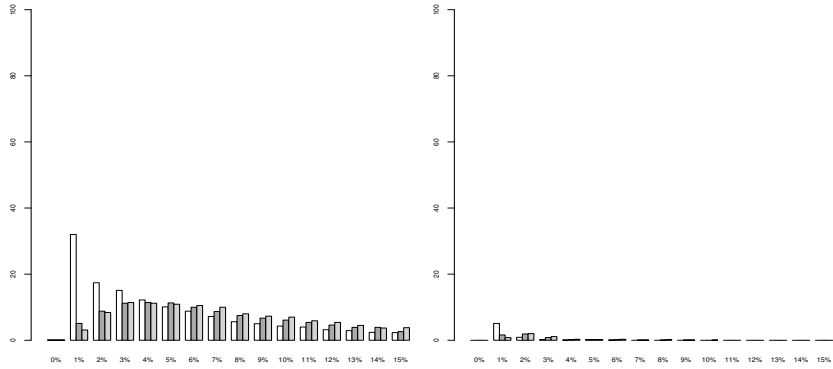


그림 15: J_p 정분류율



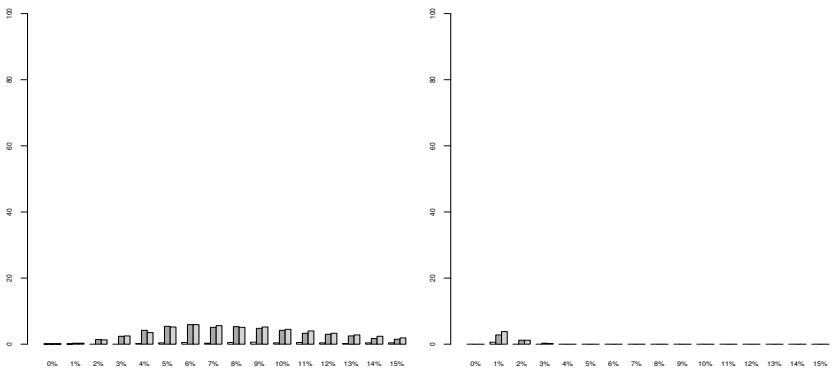
(a) 시나리오 1

(b) 시나리오 2



(c) 시나리오 3

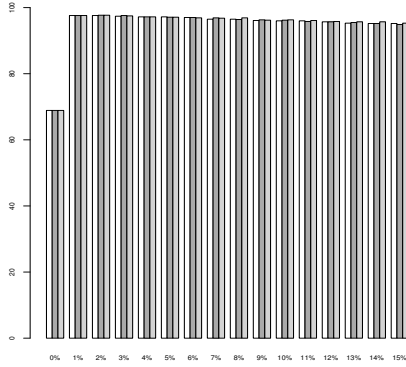
(d) 시나리오 4



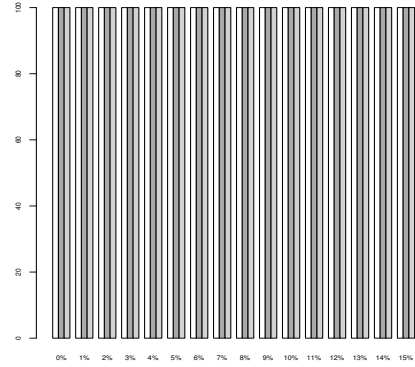
(e) 시나리오 5

(f) 시나리오 6

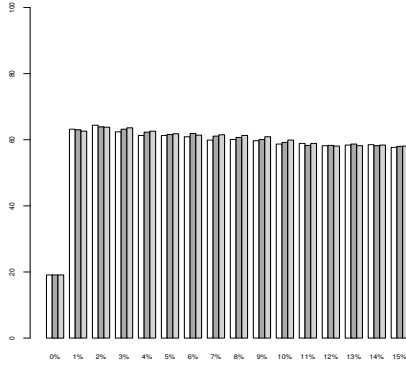
그림 16: LNJ_p 정분류율



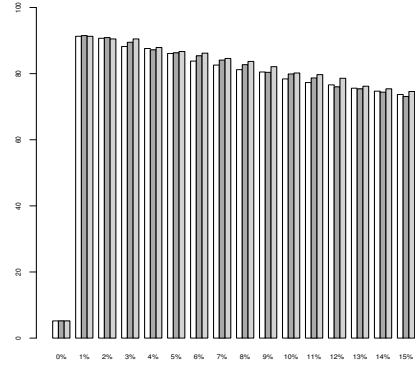
(a) 시나리오 1



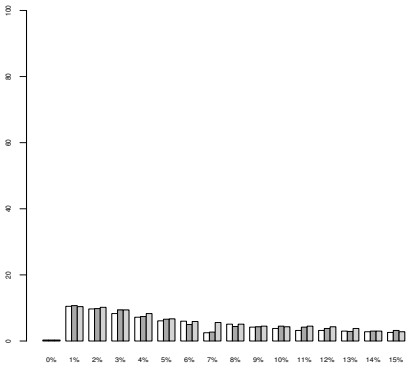
(b) 시나리오 2



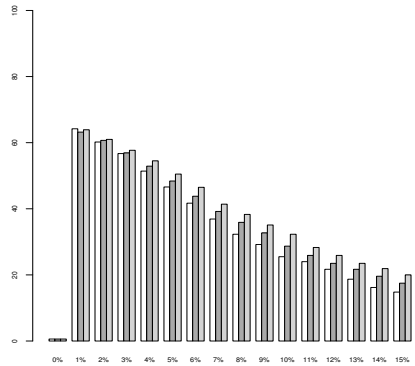
(c) 시나리오 3



(d) 시나리오 4

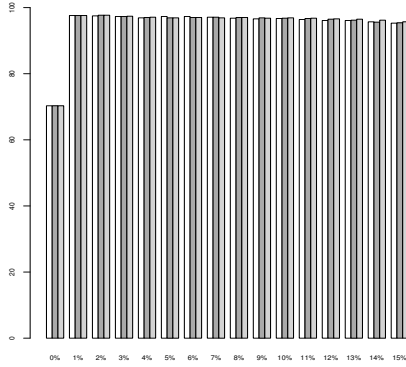


(e) 시나리오 5

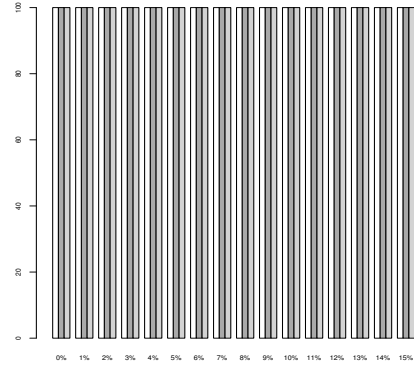


(f) 시나리오 6

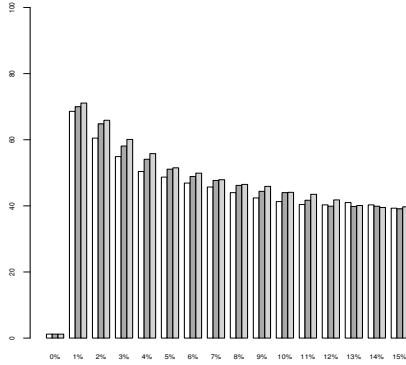
그림 17: J_C 정분류율



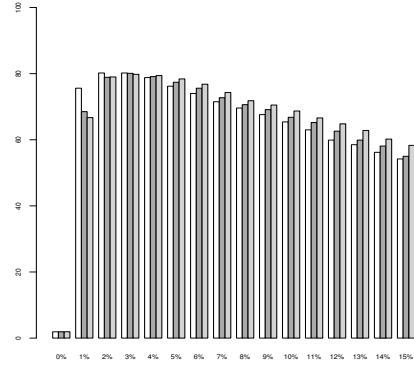
(a) 시나리오 1



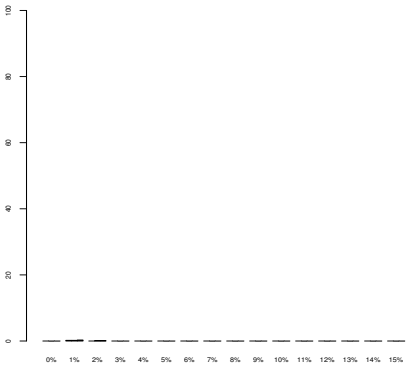
(b) 시나리오 2



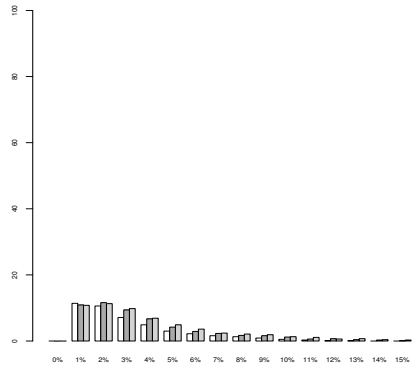
(c) 시나리오 3



(d) 시나리오 4



(e) 시나리오 5



(f) 시나리오 6

그림 18: NJ_C 정분류율

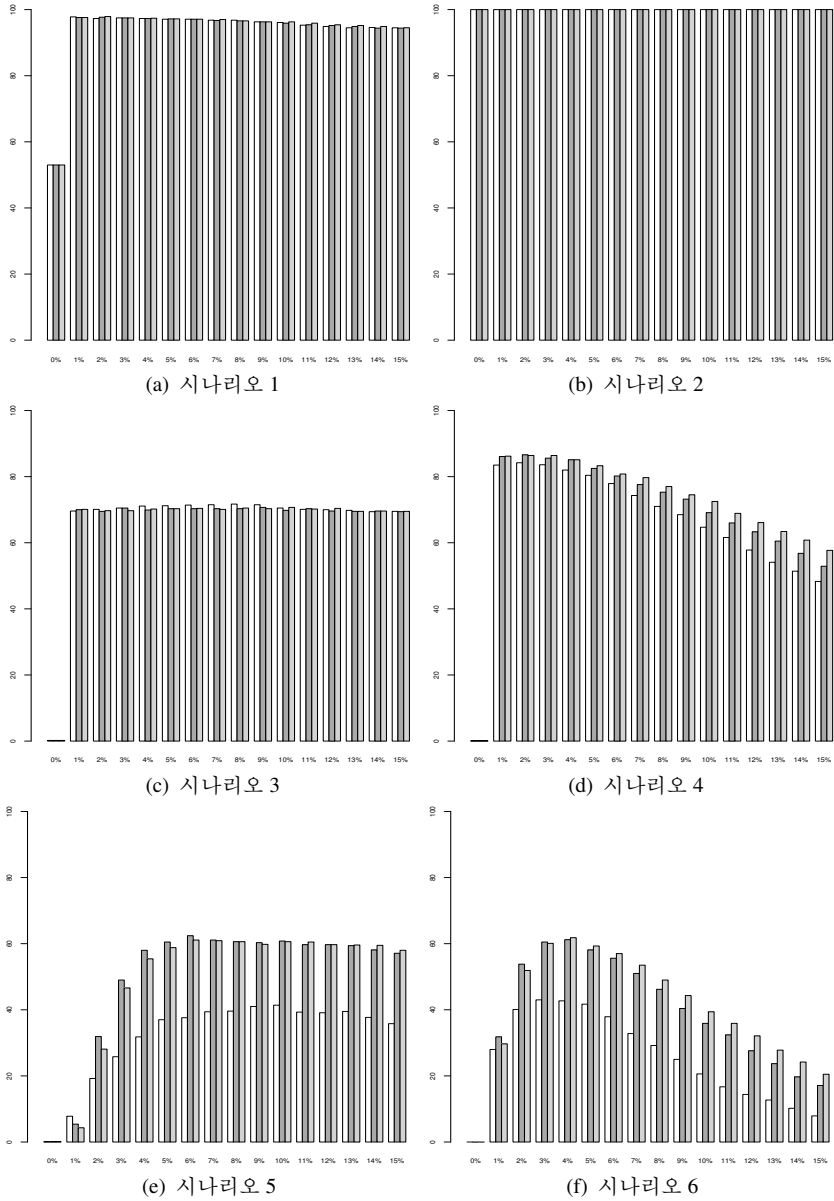


그림 19: LNJ_C 정분류율