응용통계연구 (2009)
22(6), 1153–1165

난수의 임의성을 평가하기 위한 탐색적 그림도구로서의 재현그림

장대흥[1]

[1]부경대학교 수리과학부 통계학전공

(2009년 7월 접수, 2009년 10월 채택)

요 약

의사난수발생기로부터 얻어지는 난수의 임의성을 검정하기 위한 통계방법들이 기존에 많이 제시되었다. 수열의 임의성을 평가하기 위한 탐색적 그림도구로서 우리는 재현그림을 이용할 수 있다.

주요용어: 의사난수발생기, 임의성검정, 재현그림.

1. 서론

난수는 복잡한 현상에 대한 시뮬레이션이나 모델링, 암호학, 대규모 모집단에서의 표본추출 등에 유용한 수단이다. 우리는 통상 이러한 난수를 구하기 위하여 의사난수발생기(pseudo-random number generator)를 이용한다. 의사난수발생기에서 얻어지는 난수가 임의성(randomness)을 만족하는 지를 검정하기 위한 임의성 검정방법들이 많이 개발되었다(도수검정(frequency test), 계열검정(serial test), 포커검정(Poker test), 연검정(run test), 충돌검정(collision test), 자기상관검정(autocorrelation test), 간격검정(gap test) 등이 자주 쓰인다.). 최근까지도 많은 학자들이 새로운 임의성검정 방법에 대하여 연구하고 있다 (Chatterjee 등, 2000; Marsaglia와 Tsang, 2002; Ryabko 등, 2004; Castro 등, 2005; Katos, 2005; Ryabko와 Monarev, 2005; Hamano와 Kaneko, 2007; Hamano와 Yamamoto, 2008; Kim 등, 2008; Rukhin과 Volkovich, 2008; Wang 등, 2008; Tan과 Guan, 2009).

난수의 임의성을 평가하기 위한 탐색적 그림도구로서 우리는 자기상관함수그림(autocorrelation plot)를 이용할 수 있고 장대흥 (2002)은 임의성와 카오스를 구별하기 위한 탐색적 그래픽 도구로서 시차(lag)를 이용한 산점도행렬(시차도)를 사용할 것을 제안하였다. 이 산점도행렬(시차도)를 이용하면 우리는 난수의 임의성을 탐색적으로 평가할 수 있다.

난수의 임의성을 평가하기 위한 또 다른 탐색적 그림도구로서 우리는 재현그림(recurrence plot)을 사용할 수 있다. Eckmann 등 (1987)은 동적 시스템의 위상공간(phase space) 상태의 재현을 시각화하는 재현그림을 제안하였다. Marwan 등 (2007)은 재현그림에 대한 리뷰 페이퍼를 잘 정리하여 발표하였다. 이러한 재현그림을 이용하면 우리는 의사난수발생기로부터 얻어지는 난수의 임의성을 탐색적으로 검토하여 볼 수 있다. 2절에서 재현그림에 대하여 언급을 하고 3절에서는 이러한 재현그림을 통하여 어떻게 의사난수발생기로부터 얻어지는 난수의 임의성을 검토할 수 있는 지에 대하여 언급하고 4절에서 결론을 내렸다.

[1](608-737) 부산광역시 남구 대연3동 599-1 부경대학교 수리과학부 통계학전공, 교수.
E-mail: dhjang@pknu.ac.kr

2. 재현그림

시계열자료가 주어졌을 때 위상공간은 테이켄스정리에 의하여 다음과 같이 재구성될 수 있다.

$$\mathbf{x}_i = \left(u_i, u_{i+\tau}, \ldots, u_{i+\tau(m-1)}\right),$$

여기서 $\mathbf{x}_i, i = 1, 2, \ldots, N$은 위상공간 궤도(trajectory)이고, N은 위상공간 상태의 개수, u_i는 시계열이고 m은 매립차원(embedding dimension)이며 τ는 시간지연(time delay)이다. 재현그림을 통하여 우리는 m-차원 위상공간 궤도를 2차원으로 표시하여 탐구할 수 있다. 재연그림은 다음과 같은 행렬로 표현할 수 있다.

$$\mathbf{R}_{i,j}(\epsilon) = \Theta(\epsilon - \|\mathbf{x}_i - \mathbf{x}_j\|),$$

여기서 $\mathbf{x}_i \in R^m$, $i, j = 1, 2, \ldots, N$이고, ϵ은 분계점(threshold) 거리, $\|\cdot\|$은 노름, Θ는 헤비사이드(Heaviside) 함수이다. 두 개의 위상공간 $\mathbf{x}_i$와 $\mathbf{x}_j$의 거리가 ϵ보다 작거나 같으면 위상공간 궤도가 같은 위치에 있다고 보아 $\mathbf{R}_{i,j}(\epsilon) = 1$이 되어 정사각형 모양의 재현그림에서 x축에서의 i-번째 시점과 y축에서의 j-번째 시점에 해당하는 위치에 점으로 나타나게 된다. 모든 $i, j = 1, 2, \ldots, N$에 대하여 $\mathbf{R}_{i,j}(\epsilon) = 1$일 때만 (i, j) 위치에 점을 찍으면 재현그림이 완성된다.

재현그림에서 중요한 모수가 분계점 ϵ의 값이다. ϵ이 너무 작으면 정사각형의 재현그림에서 점들을 거의 볼 수 없게 되어 빈 공간에 가까운 재현그림이 되고, ϵ이 너무 크면 재현그림에서 거의 모든 점들이 다른 점들의 이웃이 되어 정사각형의 공간을 모든 점들이 채워 거의 까맣게 되는 재현그림이 된다. Mindlin과 Gilmore (1992), Zbilut와 Webber (1992), Matassini 등 (2002), Thiel 등 (2002), Zbilut 등 (2002)이 분계점 ϵ의 값을 선택하는 방법들에 대하여 연구하였다.

우리는 이러한 재현그림에 나타나는 점들의 패턴을 살펴봄으로써 의사난수발생기로부터 얻어지는 난수의 임의성을 탐색적으로 검토하여 볼 수 있다. 난수가 임의성을 만족한다면 정사각형의 재현그림에서 점들이 어떤 패턴이나 집락을 이루지 못하고 정사각형 공간 전체에 무작위로 흩어져 나타날 것이다. 반대로 난수가 임의성을 만족하지 못하면 이러한 시계열 자료가 갖고 있는 고유한 구조 때문에 정사각형의 재현그림에서 점들이 어떤 패턴이나 집락을 이룰 것이다. 우리는 이러한 패턴이나 집락을 보며 시계열 자료의 어느 부분에서 임의성을 만족하지 못하는 지를 확인할 수 있다. 그러므로 우리가 자료분석의 확증적 단계에서 임의성검정을 시행하기 전에 자료분석의 탐색적 단계에서 재현그림을 그려봄으로써 난수가 임의성의 성질을 갖고 있는 지를 탐색적으로 검토하여 볼 수 있다.

3. 임의성을 평가하기 위한 탐색적 그림도구로서의 재현그림

우리가 어떻게 재현그림을 임의성을 평가하기 위한 탐색적 그림도구로서 쓸 수 있는 지를 몇 가지 사례들을 이용하여 살펴볼 수 있다. 본 연구의 목적이 난수의 임의성을 평가하는 것이므로 매립차원 m과 시간지연 τ을 고려할 필요없이 시계열 자체를 이용하면 된다. 즉 $m = 1$로 놓으면 된다. 또한 노름은 유클리드 거리를 이용하기로 하고 모든 예제에서 특별한 언급이 없는 한 ϵ으로 시계열 자료의 표준편차 값의 0.1배를 이용하기로 한다.

예제 3.1: 다음 그림 3.1은 R 패키지에서의 의사난수발생기에서 구한 난수 300개를 이용하여 구한 꺽은선그래프, 히스토그램, 자기상관함수그림을 나타낸다. 완전난수는 아니나 의사난수의 임의성을 볼 수 있다.

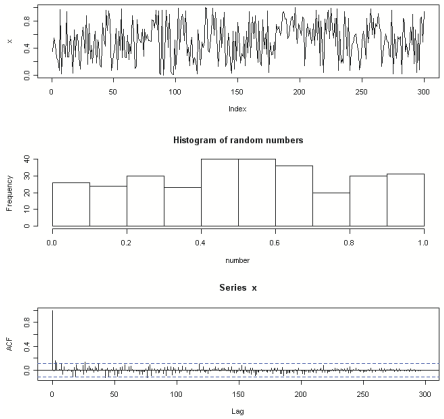

그림 3.1. 난수 300개를 이용하여 구한 꺽은선그래프, 히스토그램, 자기상관함수그림

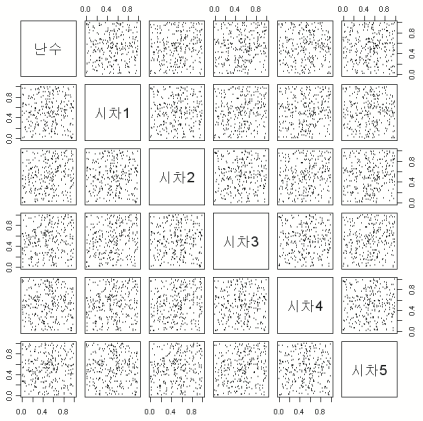

그림 3.2. 난수 300개를 이용하여 구한 산점도행렬

앞에서 구한 난수 300개를 이용하여 산점도행렬(시차도)을 그리면 다음 그림 3.2와 같다. 모든 산점도에서 점들이 무작위로 흩어져 나타남으로 이 난수가 임의성을 만족함을 알 수 있다.

히스토그램, 자기상관함수그림, 산점도행렬 외에 난수의 임의성을 평가하기 위한 또 다른 탐색적 그림도구로서 우리는 재현그림을 사용할 수 있다. 앞에서 구한 난수 300개를 이용하여 재현그림을 그리면 다음 그림 3.3과 같다. 정사각형의 재현그림에서 점들이 어떤 패턴이나 집락을 이루지 못하고 정사각형 공간 전체에 무작위로 흩어져 나타나므로 난수가 임의성을 만족한다고 볼 수 있다.

앞에서의 난수와 비교하기 위하여 다음과 같은 전형적인 카오스 현상을 일으키는 로지스틱맵을 이용하여 시계열 자료 10,000개(초기값: 0.1)를 구한다.

$$x_{n+1} = 4x_n(1 - x_n)$$

이 시계열 자료를 이용하여 꺽은선그래프(10,000개 중 끝에 있는 300개(9,701~10,000)만 그림), 히스토그램, 자기상관함수그림을 그리면 다음 그림 3.4와 같다. 자기상관함수그림에서는 모든 시차에 걸쳐

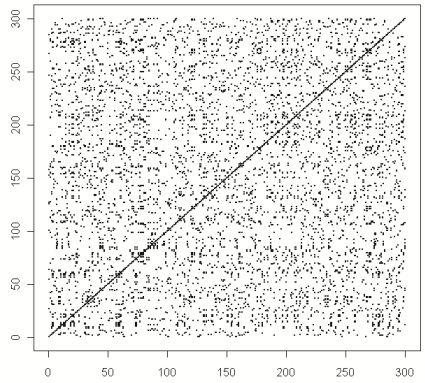

그림 3.3. 난수 300개를 이용하여 구한 재현그림

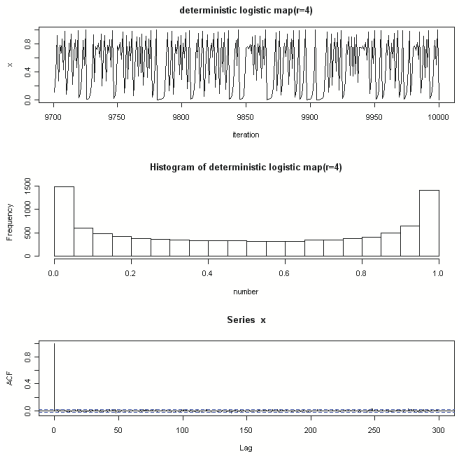

그림 3.4. 로지스틱맵에서 구한 시계열 자료 300개를 이용하여 구한 꺽은선그래프, 히스토그램, 자기상관함수그림

자기상관이 거의 없음을 알 수 있으나 히스토그램에서 시계열 자료가 임의성을 만족하지 못하고 U자형 패턴을 이루어 0이나 1에 가까운 숫자가 많은 반면 0.5에 가까운 숫자는 적은 것을 확인할 수 있다.

산점도행렬(시차도)을 그리면 다음 그림 3.5와 같다. 모든 산점도에서 점들이 무작위로 흩어져 나타나지 못하고 특이한 패턴(이차다항식이 시차가 커짐에 따라 거듭제곱되는 합성함수 형태)을 나타냄으로 이 시계열 자료가 임의성을 만족하지 못함을 알 수 있다.

앞에서 구한 시계열 자료 10,000개 중 끝에 있는 300개(9,701∼10,000)를 이용하여 재현그림을 그리면 다음 그림 3.6과 같다. 정사각형의 재현그림에서 점들이 여러 개의 띠모양의 패턴과 빈공간을 형성하여 집락을 이룸을 알 수 있어 이 시계열 자료가 임의성을 만족하지 못함을 알 수 있고 이러한 패턴이 나타나는 부분에서 임의성이 파괴되고 있음을 알 수 있다.

ϵ으로 300개 시계열 자료의 표준편차 값의 0.5배를 이용하면 다음 그림 3.7을 얻을 수 있다. 그림 3.6보다 더 또렷하게 이 정사각형의 재현그림에서 점들이 여러 개의 띠모양의 패턴과 빈공간을 형성하여 집락을 이룸을 알 수 있어 이 시계열 자료가 임의성을 만족하지 못함을 알 수 있고 이러한 패턴이 나타나는 부분에서 임의성이 파괴되고 있음을 알 수 있다.

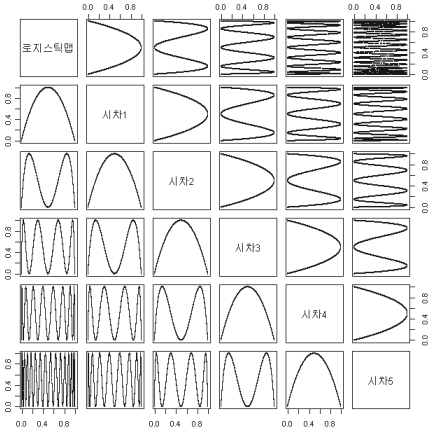

그림 3.5. 로지스틱맵에서 구한 시계열 자료 300개를 이용하여 구한 산점도행렬

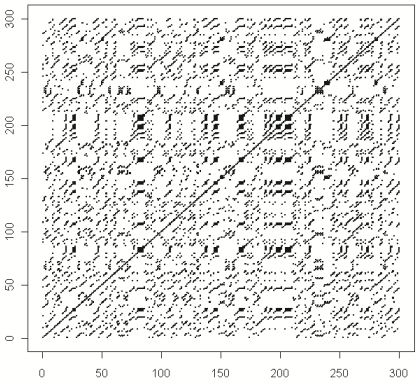

그림 3.6. 로지스틱맵에서 구한 시계열 자료 300개를 이용하여 구한 재현그림(I)

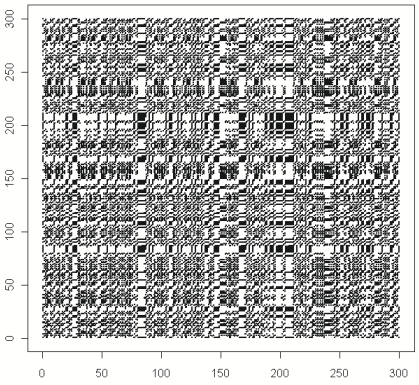

그림 3.7. 로지스틱맵에서 구한 시계열 자료 300개를 이용하여 구한 재현그림(II)

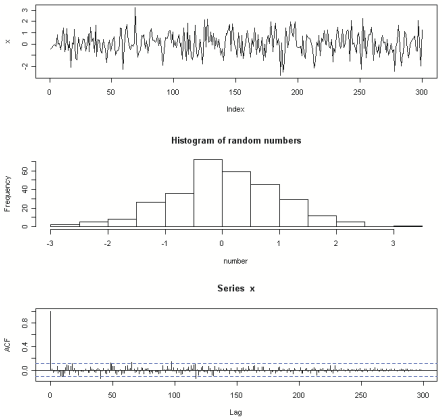

그림 3.8. 표준정규난수 300개를 이용하여 구한 꺽은선그래프, 히스토그램, 자기상관함수그림

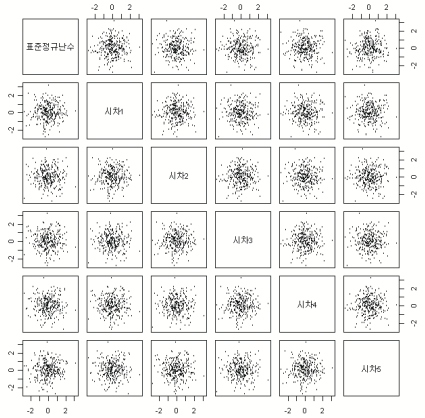

그림 3.9. 표준정규난수 300개를 이용하여 구한 산점도행렬

예제 3.2: 다음 그림 3.8은 R 패키지에서의 의사난수발생기에서 구한 표준정규난수 300개를 이용하여 구한 꺽은선그래프, 히스토그램, 자기상관함수그림을 나타낸다. 완전표준정규난수는 아니나 의사표준정규난수의 임의성을 볼 수 있다.

앞에서 구한 표준정규난수 300개를 이용하여 산점도행렬(시차도)을 그리면 다음 그림 3.9와 같다. 모든 산점도에서 점들이 둥그런 모양을 이루며 무작위로 흩어져 나타남으로 이 표준정규난수가 임의성을 만족함을 알 수 있다.

앞에서 구한 표준정규난수 300개를 이용하여 재현그림을 그리면 다음 그림 3.10과 같다. 정사각형의 재현그림에서 점들이 어떤 패턴이나 집락을 이루지 못하고 정사각형 공간 전체에 무작위로 흩어져 나타나므로 표준정규난수가 임의성을 만족한다고 볼 수 있다.

예제 3.3: 동전을 300개 던졌을 때 그림면을 1로, 숫자면을 0으로 놓았을 때의 수열을 컴퓨터로 시뮬레이션해 보자. 다음 그림 3.11은 R 패키지에서의 의사난수발생기에서 구한 성공률이 0.5인 베르누이

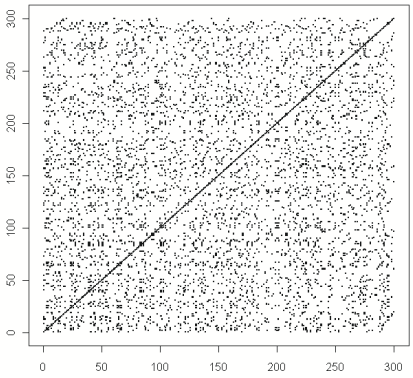

그림 3.10. 표준정규난수 300개를 이용하여 구한 재현그림

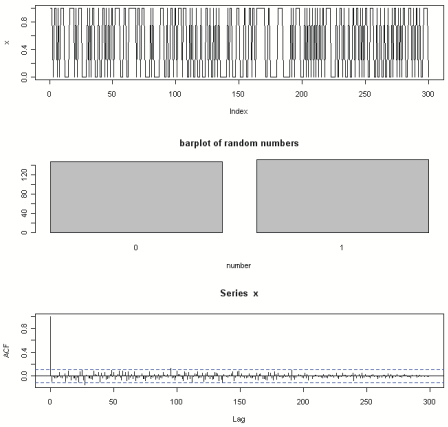

그림 3.11. 베르누이난수 300개를 이용하여 구한 꺾은선그래프, 막대그래프, 자기상관함수그림

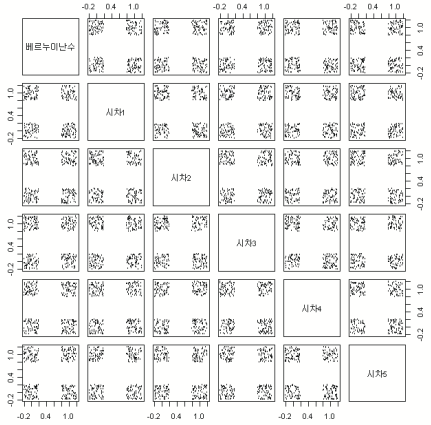

그림 3.12. 베르누이난수 300개를 이용하여 구한 산점도행렬

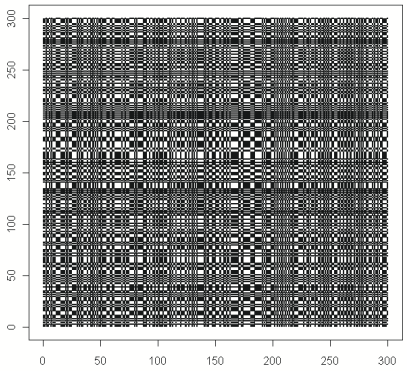

그림 3.13. 베르누이난수 300개를 이용하여 구한 재현그림

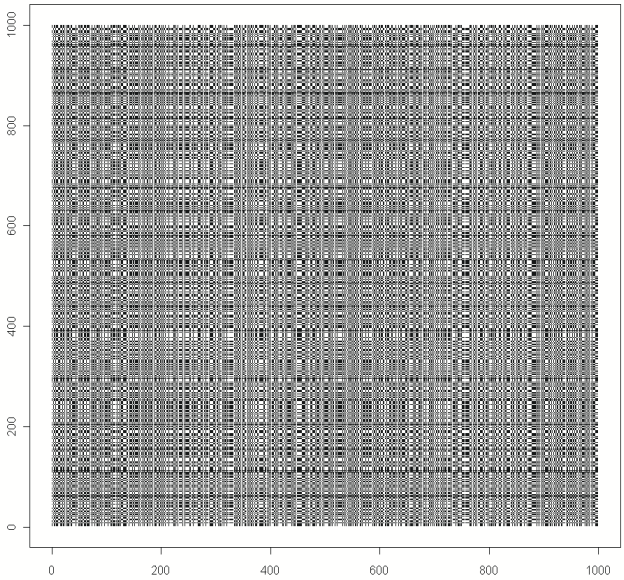

그림 3.14. 베르누이난수 1,000개를 이용하여 구한 재현그림

난수 300개를 이용하여 구한 꺽은선그래프, 막대그래프, 자기상관함수그림을 나타낸다. 완전베르누이난수는 아니나 의사베르누이난수의 임의성을 볼 수 있다.

앞에서 구한 베르누이난수 300개를 이용하여 산점도행렬(시차도)을 그리면 다음 그림 3.12와 같다. 숫자가 0과 1로 만 이루어져 있어 300개의 점들이 겹쳐 4개의 점들로만 나타남으로 좌우, 위아래로 조금 흔들어(흐트림(jittering)) 300개의 점들을 구별하였다. 모든 산점도에서 점들이 4개의 점 (0,0), (0,1), (1,0), (1,1)을 중심으로 둥그런 모양을 이루며 4개의 집락을 이루며 그 크기도 비슷하므로 이 베르누이난수가 임의성을 만족함을 알 수 있다.

앞에서 구한 베르누이난수 300개를 이용하여 재현그림을 그리면 다음 그림 3.13과 같다. 정사각형의 재현그림에서 점들이 중간 중간 뭉쳐 있기는 하나 아주 뚜렷한 패턴이나 집락을 이루지 못하고 정사각형

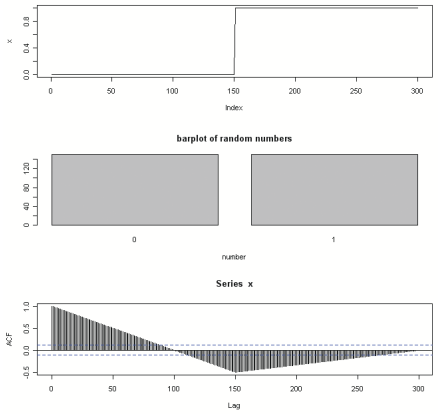

그림 **3.15.** 케이스 A의 경우에 대한 꺽은선그래프, 막대그래프, 자기상관함수그림

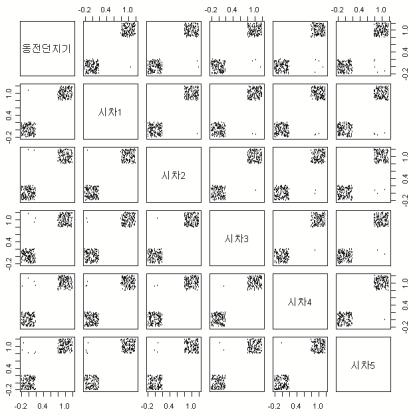

그림 **3.16.** 케이스 A의 경우에 대한 산점도행렬

공간 전체에 무작위로 흩어져 나타나므로 베르누이난수가 임의성을 만족한다고 볼 수 있다. 베르누이난수에 대한 재현그림이 예제 1이나 예제 2에서 언급한 재현그림(균등난수나 표준정규난수인 경우)과 다르게 느껴지는 이유는 베르누이난수는 단지 두 개의 값(0이나 1) 만을 가지기 때문이다.

좀 더 자세히 보기 위하여 또다른 베르누이난수 1,000개를 이용하여 재현그림을 그리면 다음 그림 3.14와 같다. 그림 3.13보다 더 또렷하게 재현그림에서 점들이 중간 중간 뭉쳐 있기는 하나 아주 뚜렷한 패턴이나 집락을 이루지 못하고 정사각형 공간 전체에 무작위로 흩어져 나타나므로 베르누이난수가 임의성을 만족한다고 볼 수 있다.

참고로 동전을 300개 던졌을 때 다음과 같이 극단적으로 임의적이지 않은 두 가지 경우를 생각해 보자.

1. 0이 150개, 1이 150개 연속으로 나오는 경우(케이스 A)
2. 0, 1이 150번 반복되는 경우(케이스 B)

케이스 A의 경우 꺽은선그래프, 막대그래프, 자기상관함수그림을 나타내어 보면 다음 그림 3.15와 같다. 자기상관함수그림이 특이하게 나타남을 알 수 있다.

케이스 A의 경우 산점도행렬(시차도)을 그리면 다음 그림 3.16과 같다. 숫자가 0과 1로 만 이루어져 있

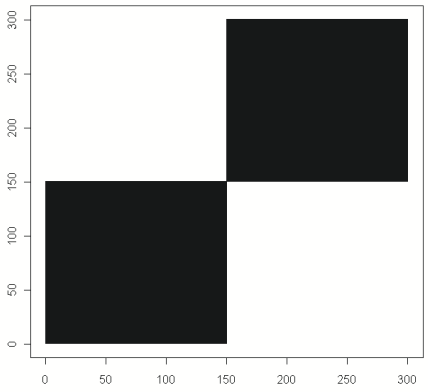

그림 **3.17.** 케이스 A의 경우에 대한 재현그림

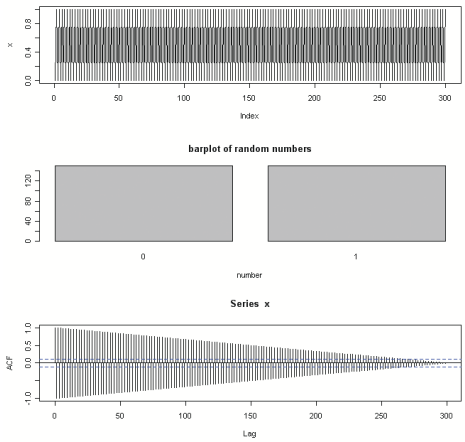

그림 **3.18.** 케이스 B의 경우에 대한 꺽은선그래프, 막대그래프, 자기상관함수그림

어 흐트림을 사용하였다. 모든 산점도에서 점들이 두 개의 점 (0,0), (1,1)을 중심으로 둥그런 모양을 이루며 2개의 집락을 이루므로 이 경우 임의성을 만족하지 못함을 알 수 있다.

케이스 A의 경우 재현그림을 그리면 다음 그림 3.17과 같다. 아주 특이한 모습을 갖으므로 임의성을 만족하지 못함을 알 수 있다.

케이스 B의 경우 꺽은선그래프, 막대그래프, 자기상관함수그림을 나타내어 보면 다음 그림 3.18과 같다. 자기상관함수그림이 특이하게 나타남을 알 수 있다.

케이스 B의 경우 산점도행렬(시차도)을 그리면 다음 그림 3.19와 같다. 숫자가 0과 1로 만 이루어져 있어 흐트림을 사용하였다. 각 산점도에서 점들이 두 개의 점 (0,0), (1,1)을 중심으로 둥그런 모양을 이루며 2개의 집락을 이루는 경우와 두 개의 점 (0,1), (1,0)을 중심으로 둥그런 모양을 이루며 2개의 집락을 이루는 경우 두 가지 패턴이 번갈아 나타난다. 그러므로 임의성을 만족하지 못함을 알 수 있다.

케이스 B의 경우 재현그림을 그리면 다음 그림 3.20과 같다. 아주 특이한 모습(검은 점과 흰 공간이 번갈아 나타나는)을 갖으므로 임의성을 만족하지 못함을 알 수 있다.

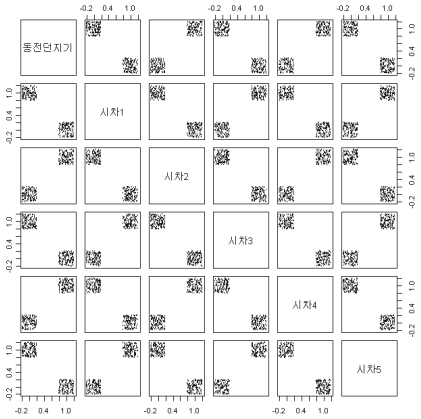

그림 **3.19.** 케이스 B의 경우에 대한 산점도행렬

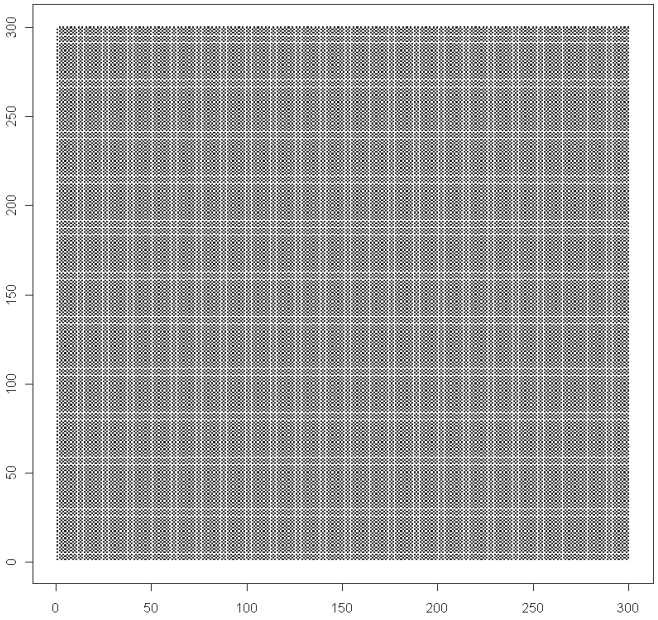

그림 **3.20.** 케이스 B의 경우에 대한 재현그림

4. 결론

우리는 자료분석의 확증적 단계에서 주어진 난수의 임의성검정을 시행하기 전에 자료분석의 탐색적 단계에서 난수의 임의성을 평가하기 위하여 그래픽도구들을 이용하여 그림들을 그려봄으로써 난수가 임의성의 성질을 갖고 있는 지를 탐색적으로 검토하여 볼 수 있다. 난수의 임의성을 평가하기 위한 탐색적 그림도구로서 우리는 자기상관함수그림과 산점도행렬(시차도)를 사용할 수 있다. 난수의 임의성을 평가하기 위한 또 다른 탐색적 그림도구로서 재현그림을 본 논문에서 제안하였고 몇 가지 예들을 보였다.

참고문헌

장대흥 (2002). 탐색적 자료분석시 그래프의 활용에 대한 연구, <응용통계연구>, **15**, 433–448.

Castro, J. C. H., Sierra, J. M., Seznec, A., Izquierdo, A. and Ribagorda, A. (2005). The strict avalanche criterion randomness test, *Mathematics and Computers in Simulation*, **68**, 1–7.

Chatterjee, S., Yilmaz, M., Habibullah, M. and Laudato, M. (2000). An approximate entropy test for randomness, *Communications in Statistics-Theory and Methods*, **29**, 655–675.

Eckmann, J. P., Kamphorst, S. O. and Ruelle, D. (1987). Recurrence plots of dynamical systems, *Europhysics Letters*, **5**, 973–977.

Hamano, K. and Kaneko, T. (2007). Correction of overlapping template matching test included in NIST randomness test suite, *IEICE Transaction on Fundamentals of Electronics, Communications and Computer Sciences*, **E90-A**, 1788–1792.

Hamano, K. and Yamamoto, H. (2008). A randomness test based on T-codes, *Proceedings in International Symposium on Information Theory and its Applications*, 2008.

Katos, V. (2005). A randomness test for block ciphers, *Applied Mathematics and Computation*, **162**, 29–35.

Kim, C., Choe, G. H. and Kim, D. H. (2008). Tests of randomness by the gambler's ruin algorithm, *Applied Mathematics and Computation*, **199**, 195–210.

Marsaglia, G. and Tsang, W. W. (2002). Some difficult-to-pass tests of randomness, *Journal of Statistical Software*, **7**, 3.

Marwan, N., Romano, M. C., Thiel, M. and Kurths, J. (2007). Recurrence plots for the analysis of complex systems, *Physics Reports*, **438**, 237–329.

Matassini, L., Kantz, H., Holyst, J. and Hegger, R. (2002). Optimizing of recurrence plots for noise reduction, *Physical Review E*, **65**, 021102.

Mindlin, G. M. and Gilmore, R. (1992). Topological analysis and synthesis of chaotic time series, *Physica D*, **58**, 229–242.

Rukhin, A. L. and Volkovich, Z. (2008). Testing randomness via aperiodic words, *Journal of Statistical Computation and Simulation*, **78**, 1133–1144.

Ryabko, B. Y. and Monarev, V. A. (2005). Using information theory approach to randomness testing, *Journal of Statistical Planning and Inference*, **133**, 95–110.

Ryabko, B. Y., Stognienko, V. S. and Shokin, Y. I.(2004). A new test for randomness and its application to some cryptographic problem, *Journal of Statistical Planning and Inference*, **123**, 365–376.

Tan, S. K. and Guan, S. (2009). Randomness quality of permuted pseudorandom binary sequence, *Mathematics and Computers in Simulation*, **79**, 1618–1626.

Thiel, M., Romano, M. C., Kurths, J.,. Meucci, R., Allaria, E. and Arecchi, F. T. (2002). Influence of observational noise on the recurrence quantification analysis, *Physica D*, **171**, 138–152.

Wang, K., Pei, W., Xia, h. and Cheung. Y. (2008). Pseudo-Random number generator based on asymptotic deterministic randomness, *Physics Letters A*, **372**, 4388–4394.

Zbilut, J. P. and Webber Jr., C. L. (1992). Embeddings and delays as derived from quantification of recurrence plots, *Physics Letters A*, **171**, 199–203.

Zbilut, J. P., Zaldivar-Commenges, J. M. and Strozzi, F. (2002). Recurrence quantification based Liapunov exponents for monitoring divergence in experimental data, *Physics Letters A*, **297**, 173–181.

Recurrence Plots as an Exploratory Graphical Tool for Evaluating Randomness

Dae-Heung Jang[1]

[1]Division of Mathematical Sciences, Pukyong National University

(Received July 2009; accepted October 2009)

Abstract

There are many traditional statistical tests for randomness. We can consider recurrence plots as an exploratory graphical tool for evaluating randomness.

Keywords: Pseudo-random number generator, randomness test, recurrence plots.

[1]Professor, Division of Mathematical Sciences, Pukyong National University, 599-1 Daeyeon-dong, Nam-gu, Busan 608-737, Korea. E-mail: dhjang@pknu.ac.kr