Journal of the Korean
Data & Information Science Society
1998, Vol. 9, No. 2, pp. 237 ~ 246

의사난수 생성기의 일양성과 독립성 검정

박 경 렬[1] · 권 기 창[2] · 권 영 담[3]

요 약

지금까지 알려진 의사난수 생성기에서는 혼합 합동 생성기, 승산 합동 생성기, 유니버셜 난수 생성기, 역함수 합동 생성기, 양의 역함수 난수 생성기 등 여러 가지가 있다. 이러한 의사난수 생성기에 대하여 각각 20, 40, 60, 80, 100개의 자료를 생성하여 유의수준(α) 0.1, 0.05, 0.01 기준으로 10,000번의 시행 과정을 통하여 난수의 특성인 일양성과 독립성을 만족하는지를 검정하였다.

주제어: 의사난수 생성기

1. 도 입

컴퓨터 상에서 생성되는 난수들은 수식을 이용하여 생성시키는데 이러한 수식으로 얻어진 난수 열들은 몇 가지 물리적인 처리의 모델링을 통하여 통계적으로 정화된 추상적인 개념인 진정한 난수(true random number) 개념과는 차이가 있다. 따라서 컴퓨터 상에서 수식을 사용하여 생성된 수열들을 의사난수(pseudo-random number)라고 한다.

의사난수를 활용하여 진정한 난수열로 대치할 수 있도록 더욱 타당하고 정밀한 수학적인 해결 방안을 모색한 많은 노력이 있었다. 이 중에서 임의성(randomness), 주기성(periodicity), 효율성 (efficiency), 반복성(repeatability), 일양성(uniformity), 독립성(independency), 호환성(portability) 등 난수의 특성을 최대로 만족하도록 개선되어져 왔다.

초기의 난수 생성기는 속도가 느리고 생성된 난수가 순환할 수 있다는 단점을 가지고 있었다. 이러한 단점을 보완하여 긴 주기성, 상대적으로 적은 법수를 사용하여 난수 생성을 성취할 수 있는 방법을 제안하였다. 본 논문에서는 지금까지 고안되어 알려진 의사난수 생성기로부터 생성된 20, 40, 60, 80, 100개의 자료를 가지고 일양성 검정과 독립성 검정을 실시한다. 이때 유의수준 α를 0.1, 0.05, 0.01 기준으로 10,000번의 시행 과정을 통하여 검정을 만족하는 누적빈도를 조사 하고, 이 자료에 대하여 일정 기준을 만족하는 의사난수 생성기를 선택한다.

[1]영남대학교 이과대학 통계학과 계산학전공 박사과정
[2]예천전문대학 사무자동화과 전임강사
[3]영남대학교 이과대학 통계학과 교수, (712-749) 경상북도 경산시 대동 214-1

2. 의사난수열 생성 방법

컴퓨터의 빠른 발전과 보급의 영향으로 수식을 활용한 의사난수 생성기가 여러 방향으로 연구 개발되어 왔다. 지금까지 알려진 생성기 중 본 논문에 사용할 생성기들을 소개하면 다음과 같다.

2.1 혼합 합동 생성기(Mixed Congruential Generator : MiCG)

혼합 합동 생성기의 기본 수식은 다음과 같다.

$$X_{i+1} = (aX_i + c) \pmod{m}, \qquad i = 0, 1, 2, \cdots, n \tag{1}$$

여기서 a는 승수(multiplier), c는 증분(increment), m은 법수(modulus)를 나타내는 상수 값이며, 주어진 초기값 X_0는 씨앗 값(seed value)이라 한다.
이 식을 이용하여 생성된 수는 $0 \le X_i \le m$의 정수이며, 이 수는 다음 식을 사용하여 자료로 활용될 단위 구간$[0, 1)$ 상의 난수로 변환할 수 있다.

$$R_i = \frac{X_i}{m} \tag{2}$$

식(1)로부터 모든 i에 대하여 $X_i < m$이 성립하며, 이는 수열 X_i가 기껏해야 m개의 구별되는 수 $(0 \sim m-1)$로 표현된다. 물론 하나의 주기 내에서 충분히 큰 구별되는 수를 보장받기 위해 가능한 한 큰 모듈러스를 선택하는 것이 바람직하다.

2.2 승산 합동 생성기(Multiplicative Congruential Generator : MuCG)

승산 합동 생성기의 수식 형태는 식(1)에서 $c = 0$인 경우로 다음 식과 같다.

$$X_{i+1} = (aX_i) \pmod{m} \tag{3}$$

승산 합동 생성기에는 IBM 360의 균일난수 생성기를 사용하였으며, 알고리즘은 다음 식과 같다. 이 생성기의 법수는 1 워드(32비트) 크기에서 부호 비트를 뺀 $2^{31} - 1$이다.

$$X_n = (7^5 X_n - 1) (\text{mod } 2^{31} - 1) = (16,807 X_{n-1}) (\text{mod } 2^{31} - 1) \tag{4}$$

$$U_n = X_n / (2^{31} - 1), \qquad n > 1, \quad X_0 > 0 \tag{5}$$

2.3 유니버셜 난수 생성기(Universal Random-number Generator : URG)

유니버셜 난수 생성기는 생성 초기값을 구성하기 위하여 식(6)과 같이 3개항의 곱을 사용하고, 식(7)의 모듈러스 169를 갖는 선형 합동 생성기를 이용한다.

$$y_n = (y_{n-3} \times y_{n-2} \times y_{n-1}) \pmod{179}, \quad (n = 4, 5, \cdots) \tag{6}$$

$$z_n = (53\, z_{n-1} + 1) (\text{mod } 169), \quad (n = 2, \cdots) \tag{7}$$

$$b_i = \begin{cases} 0, & \text{if } \{(y_i \times z_i) \bmod 64\} < 32 \\ 1, & \text{otherwise} \end{cases} \tag{8}$$

식(8)의 b_i를 생성하기 위하여 4개의 정수값 (y_1, y_2, y_3, z_1)이 설정되어야 하며, 97개의 초기값 $U(1), U(2), \cdots, U(97)$은 다음 식에서 구해진 값으로 소수부를 구성한다.

$$\begin{cases} U(1) & = 0.b_1\ b_2\ b_3 \cdots b_{24} \\ U(2) & = 0.b_{25}\ b_{26}\ \cdots b_{48} \\ & \vdots \\ U(97) & = 0.b_{2305}\ b_{2306}\ \cdots b_{2328} \end{cases} \tag{9}$$

최종적으로 난수열은 $U_n = X_n \cdot C_n$의 식을 이용하여 생성하며, 이 난수생성기의 주기는 약 $2^{144}(= 2.23 \times 10^{43})$이다.

2.4 역함수 합동 생성기(Inversive Congruential Generator : ICG)

m을 소수(prime number), $Z_m = \{x \mid x = a \bmod m\ ,\ a$는 정수$\}$라고 정의하고, $c \in Z_m$인 c에 대하여 $\bar{c}$를 다음과 같이 정의한다.

$$\bar{c} = \begin{cases} 0, & \text{if } c = 0 \\ c^{-1}, & \text{if } c \neq 0 \end{cases} \tag{10}$$

이 식의 값은 다음 식을 이용하여 구할 수 있다.

$$\bar{c} = c^{p-2} \pmod{p} \tag{11}$$

역함수 합동 생성기는 Eichenauer & Lehn(1986)에 의해 제안되었으며, 기본 수식은 다음과 같다.

$$X_{n+1} = (a\ \overline{X_n} + b) \pmod{m},\quad n \geq 0 \tag{12}$$

여기서 a는 승수(multiplier), b는 증분(increment), m은 소수 법수(prime modulus), X_0는 초기값(initial value)이다. 이 식으로 얻어진 수열 $(X_n)_{n\geq 0}$은 집합 $\{0, 1, \cdots, m-1\}$의 한 원소이다. 따라서 자료로 활용될 단위 구간 $[0, 1)$ 상의 난수는 식(2)를 이용하여 얻을 수 있다. 이 생성기의 주기는 m이며 긴 주기를 가지기 위해서는 법수를 크게 설정하는 것이 좋다.

2.5 양의 역함수 합동 생성기(Explicit Inversive Congruential Generator : EICG)

이 생성기는 Eichenauer-Herrmann(1993)에 의해 제안된 것으로 기본 수식은 다음과 같다.

$$X_n = \overline{a(n + n_0) + b} \pmod{m},\quad n \geq 0 \tag{13}$$

여기서 m은 소수 법수(prime modulus), $a(\in Z_m,\ a \neq 0)$는 승수(multiplier), $b(\in Z_m)$ 증분(increment), $n_0(\in Z_m)$는 초기값(initial value)이다. 이 식으로 얻어진 수열 $(X_n)_{n\geq 0}$은 집

합 {0, 1, ⋯, $m-1$}의 한 원소이다. 따라서 자료로 활용될 단위 구간 [0, 1) 상의 난수는 식(2)를 이용하여 얻을 수 있다.

3. 통계적 검정 방법

난수생성기에 의해 생성된 의사난수열이 $U(0,1)$상에서 진정한 난수가 되려면 통계적 검정을 통과하여야 한다. 이 장에서는 생성된 의사난수열의 일양성과 독립성 검정을 하는 방법들에 대하여 알아본다.

일양성을 판단하는 것은 발생된 난수들이 [0, 1) 구간에 균일하게 분포되어 있는지를 검정하는 것으로 K-S 검정, χ^2 검정을 시행한다. 일양성 검정을 위해 쓰이는 귀무가설 H_0는 다음과 같이 난수 R_i가 구간 [0,1)에서 균일분포를 따른다는 것을 가정한다.

$$\begin{cases} H_0 & : R_i \sim U(0,\ 1) \\ H_a & : R_i \sim U(0,\ 1)\text{이 아니다.} \end{cases} \tag{14}$$

독립성 조사 방법으로 알려진 테스트 중에서 poker 검정, run 검정, autocorrelation 검정을 시행 한다. 독립성을 조사하는데 쓰이는 가설은 다음과 같다.

$$\begin{cases} H_0 & : R_i \sim\text{독립적이다.} \\ H_a & : R_i \sim\text{독립적이 아니다.} \end{cases} \tag{15}$$

3.1 K-S(Kolmogorov-Smirnov) 검정

이 방법은 연속형 균일분포의 이론적 분포함수 $F_X(x) = x,\ 0 \le x \le 1$와 발생된 난수의 경험적 분포함수 $F_N(x)$를 비교한다. 발생된 난수의 경험적 분포함수는 발생된 난수를 $R_1,\ R_2, \cdots,\ R_N$이라 하고, 이들을 오름차순으로 정렬하여 구한다.

$$F_N(x) = \frac{N\text{개의 난수 중 } x \text{ 이하인 난수의 수}}{N} \tag{16}$$

다음은 K-S 방법으로 검정하는 과정이다.

첫째, 발생된 난수를 $R_{(1)} \le R_{(2)} \le \cdots \le R_{(i)} \le \cdots \le R_{(N)}$와 같이 오름차순으로 배열한다. 여기서 $R_{(i)}$는 i번째로 작은 난수를 표시한다.

둘째, D^+와 D^-를 다음과 같이 구한다.

$$\begin{cases} D^+ = \max\left\{\frac{i}{N} - R_{(i)}\right\}, & 1 \le i \le N \\ D^- = \max\left\{R_{(i)} - \frac{i-1}{N}\right\}, & 1 \le i \le N \end{cases} \tag{17}$$

셋째, K-S 검정 통계량 $D = \max(D^+, D^-)$를 구한다.

넷째, 정해진 유의수준 α와 자유도 N을 이용하여 임계값 $D_{(\alpha;N)}$를 구한다.

다섯째, 만일 $D > D_{(\alpha;N)}$인 경우 귀무가설 H_0를 채택하지 못하고, 일양성을 갖지 않는다고 결론을 내린다.

3.2 χ^2(Chi-square) 검정

K-S 검정 방법은 표본의 수가 적은 경우에도 사용이 가능하지만, Pearson(1900)에 의해 제안된 χ^2 검정은 표본의 수가 큰 경우 ($N \geq 50$)에 정확성이 높고 효과적이다.

χ^2 검정 과정을 다음과 같다.

첫째, 검정통계량 $\chi_0^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$을 구한다. 여기서 기대값 $E_i (= \frac{N}{n})$는 i번째 구간에 존재하는 이론적 기대치이며, O_i는 i번째 구간에 존재하는 표본 수를 나타낸다. 모든 구간은 등구간(等區間)이며 본 논문에는 10개의 구간 ($n = 10$) 으로 구분하였다.

둘째, 검정통계량 χ_0^2이 $\chi_0^2 > \chi^2(\alpha : n-1)$이면 귀무가설 H_0를 채택할 수 없다. 임계값 $\chi^2(\alpha : n-1)$은 자유도가 $n-1$인 카이제곱 분포에서 $100(1-\alpha)$ 백분위수이다.

3.3 poker 검정

각 난수에서 연속된 숫자(digit)들을 의미있는 그룹단위로 묶어 동일한 숫자가 몇 번 나타나는지 관찰해 보는 독립성 검정 방법이다.

먼저 각 그룹에 들어가는 관측도수와 기대도수를 구한다. 이 때 기대도수 E_r은 $n \times P_r$이며, 구성 요소 P_r은 각 그룹이 r개의 상이한 숫자로 구성될 확률식을 의미하며 다음 식에 의하여 구한다.

$$P_r = \frac{S(k,r) \times_d P_r}{d^k} \tag{18}$$

여기서 $S(k,r)$은 Stirling number로서 k개의 요소를 r개의 부분으로 분할하는 경우의 수이며, ${}_dP_r$은 d개(10진수 표현시 : $0 \sim 9$)의 숫자로 된 집합으로부터 r개를 택하는 순열의 수, 즉 ${}_dP_r = d(d-1)(d-2) \cdots (d-r+1)$이며, d^k은 $k-tuple$로 표현되는 그룹 수이다.

본 논문에서는 세자리 수($k = 3$)의 경우에 대하여 포카 검정을 실시하였으며, 그 과정은 다음과 같다.

첫째, 확률 P_r을 구한다. 다음은 Stirling number가 각각 $S(3,1)$, $S(3,2)$, $S(3,3)$인 경우 확률 값을 나타낸다.

- P(세 숫자가 모두 동일한 경우) : 0.01
- P(두 개의 숫자가 동일한 경우) : 0.27
- P(세 숫자가 모두 상이한 경우) : 0.72

둘째, 검정통계량 $\chi_0^2 = \sum_{r=1}^{k} \frac{(O_r - E_r)^2}{E_r}$ 을 구한다. 여기서 O_r ($r = 1,2,3$)은 각 범주에 들어가는 그룹 수를 나타낸다.

셋째, 주어진 유의수준과 자유도 $k-1$을 기준으로 임계값 $\chi^2(\alpha : k-1)$을 구한다.

넷째, $\chi_0^2 > \chi^2(\alpha : k-1)$이면 귀무가설을 채택할 수 없으며, 발생된 난수는 각 그룹마다 동일한 숫자가 반복적으로 나타난다고 결론을 내린다.

3.4 run 검정

run이란 연속된 같은 종류의 사건(event)을 말하며, 런의 길이(length)는 하나의 런에 있는 사건의 수를 말한다. 그리고 runs up은 점점 증가하는 추세의 숫자들의 연속집합을 뜻하며, runs down은 반대로 점점 감소하는 숫자들의 연속된 집합을 말한다. 다음 예에서 숫자 상단에 표시된 '+'는 다음 숫자의 증가 상태를 나타내고, '-'는 그 반대 경우이다.

$$\begin{array}{cccccc} - & + & + & - & - & \\ 0.65 & 0.15 & 0.47 & 0.73 & 0.26 & 0.18 \end{array}$$

마지막 숫자는 연속 숫자가 없으므로 표시되지 않는다. 여기서 런은 3개이고, 첫째 런의 길이는 1, 둘째와 셋째 런의 길이는 2이다. 따라서 N개의 난수가 있을 경우 런의 최대 개수는 $N-1$개, 최소 개수는 한 개가 존재할 수 있다.

런 검정 과정은 다음과 같다.

첫째, 런의 개수를 구한다.

둘째, 평균과 분산을 계산한다. (단 a는 런의 수, N은 난수의 개수이다.)

$$\text{평균 } \mu_a = \frac{2N-1}{3}, \qquad \text{분산 } \sigma_a^2 = \frac{16N-29}{90} \tag{19}$$

셋째, $N > 20$인 경우 a는 근사적으로 정규분포$(\mu_a,\ \sigma_a^2)$를 따른다. 표준정규분포를 사용하기 위하여 다음의 검정통계량을 만든다.

$$Z_0 = \frac{a-\mu_a}{\sigma_a} \quad \sim \quad N(0,\ 1) \tag{20}$$

넷째, 검정통계량이 $|Z_0| \le Z_{\alpha/2}$이면 발생된 난수들이 독립적이라는 결론을 내린다. 임계값 $Z_{\alpha/2}$은 정규분포에서 $100(1-\alpha)$백분위수이다.

3.5 autocorrelation 검정

자동상관관계 검정은 발생된 난수들 상호간에 어떤 관계가 존재하는지를 알아보는 방법으로 본 검정에 사용되는 모수 ρ_{im}은 i번째 난수에서 시작하여 매 m번째마다의 난수 즉 $R_i,\ R_{i+m},\ R_{i+2m},\ \cdots,\ R_{i+(M+1)m}$사이에 존재하는 자동상관관계를 나타낸다. 이때 M은 $i+(M+1)m \le N$을 만족시키는 최대 정수이며, N은 난수의 개수이다. 본 논문에서 $i=3,\ m=5$로 설정하였다.

자동상관관계 검정과정을 살펴보면 다음과 같다.

첫째, 상관관계의 유·무를 검정하므로 가설은 양측검정(two-tailed test)을 적용한다.

$$\begin{cases} H_0 : \rho_{im} = 0 \\ H_a : \rho_{im} \neq 0 \end{cases} \tag{21}$$

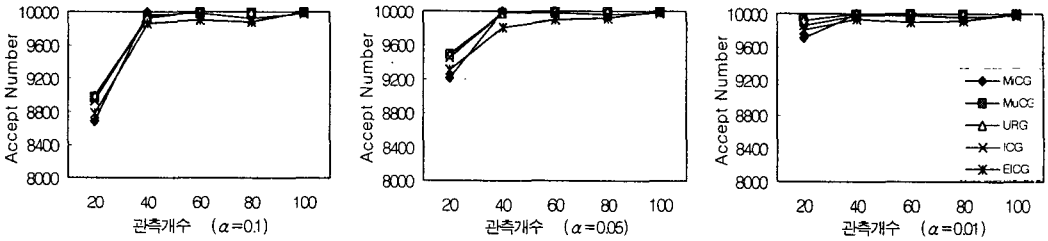
둘째, 검정통계량은 $Z_0 = \frac{\hat{\rho_{im}}}{\sigma_{\hat{\rho_{im}}}}$와 같다.

셋째, 검정통계량이 $|Z_0| \le Z_{\alpha/2}$이면 귀무가설을 채택하고, 검정대상인 $M+2$개의 난수가 모두 크거나 혹은 모두 작아지는 경향이 나타나지 않으며, 또한 매 m번째마다 교대로 큰 수, 작은 수가 나타나지 않는다고 결론을 내린다.

4. 검정 결과 분석

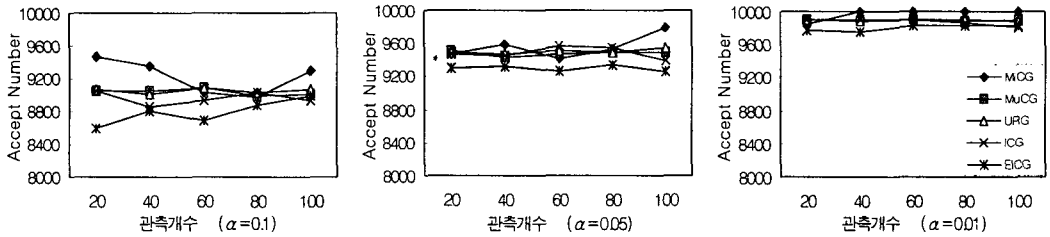
난수 생성기 MiCG, MuCG, URG, ICG, EICG로 생성한 난수가 일양성과 독립성을 만족하는지 검정하기 위해 유의수준 (α) 0.1, 0.05, 0.01에 대하여 획득한 자료로 그래프를 그려서 비교 분석하여 본 결과 다음과 같다.

4.1 K-S 검정 결과



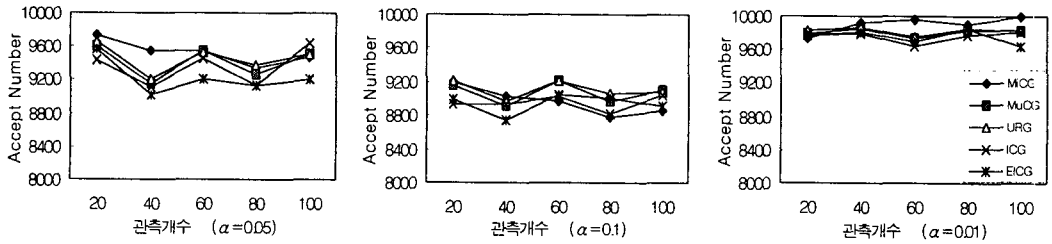
K-S 검정 결과 관측개수가 증가함에 따라 5가지 모두 유의수준(α) 0.1, 0.05, 0.01에 대하여 일양성을 가진다고 볼 수 있다.

4.2 χ^2 검정 결과



χ^2 검정 결과 유의수준(α) 0.1, 0.05, 0.01에 대하여 EICG를 제외한 다른 생성기 모두 일양성을 가진다고 볼 수 있다.

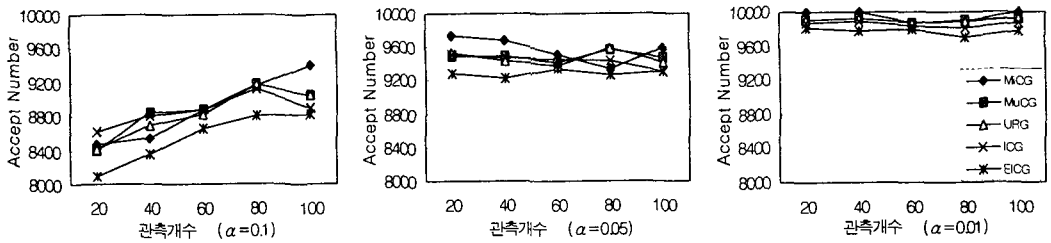
4.3 poker 검정 결과



포카 검정 결과 유의수준(α) 0.1, 0.05에서는 모든 생성기가 기복현상이 심하며, 0.01에

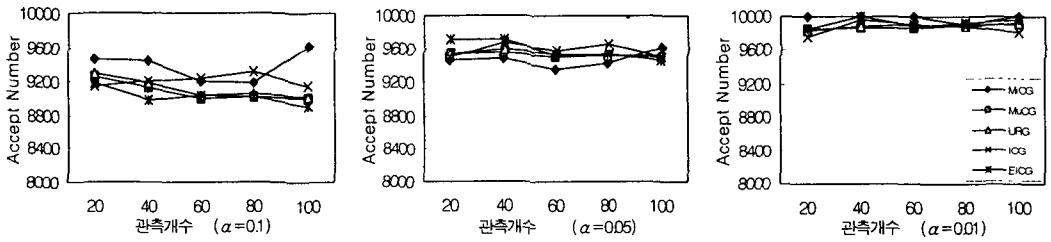
서는 기복이 완만하다. MiCG와 MuCG, URG는 독립성을 대체로 만족하고, ICG와 EICG는 독립성을 만족하지 못한다고 볼 수 있다.

4.4 run 검정 결과



런 검정 결과 유의수준 (α) 0.1에 대하여 관측개수가 증가함에 따라 독립성을 만족하게 되며, 0.05, 0.01에 대하여 EICG를 제외한 모든 생성기가 독립성을 만족한다고 볼 수 있다.

4.5 autocorrelation 검정 결과



자동상관관계 검정 결과 유의수준(α) 0.1, 0.05, 0.01에 대하여 모든 생성기가 독립성을 만족한다고 볼 수 있다.

5. 결 론

지금까지 고안되어 알려진 의사난수 생성기 중에서 혼합 합동 생성기(MiCG), 승산 합동 생성기(MuCG), 유니버셜 난수 생성기(URG), 역함수 합동 생성기(ICG), 양의 역함수 난수 생성기(EICG)에 대하여 20, 40, 60, 80, 100개의 자료를 생성하여 일양성과 독립성을 유의수준(α) 0.1, 0.05, 0.01 기준으로 10,000번의 시행 과정을 통하여 검정하였다.

그 결과 5가지 의사난수 생성기 중 EICG를 제외한 다른 발생기들은 대체로 일양성 검정과 독립성 검정을 통과하기 때문에 임의성을 지닌다고 볼 수 있으며, 모두 수식에 의한 발생기인 점을 들어 반복생성이 가능한 난수열을 만들 수 있다. 소개된 생성기 중에서는 테스트를 모두 만족시키고 2.23×10^{43}의 긴 주기를 가지고 있는 유니버셜 난수발생기를 사용하는 것이 바람직하다고 사료된다. 앞으로 연구 방향은 좀 더 다양한 형태(선형, 비선형)의 의사난수 생성기에 대한 연구가 이루어지기를 바란다.

참 고 문 헌

1. Fishman, G.(1978). *Principles of Discrete Event Simulation*, Wiley, New York.
2. George Marsaglia & Arif Zaman(1990). Toward a Universal Random Number Generator, *Statistics & Probability Letters*, 8, pp 35-39,
3. Karl Entacher(1998). A Collection of Selected Pseudorandom number Generators with Linear Structures, http://random.mat.sbg.ac.at
4. D. E. Knuth(1981). *The Art of Computer Programming : Seminumerical Algorithms*, vol. 1, Addison-Wesley, 2nd edition.
5. D. E. Knuth(1981). *The Art of Computer Programming : Seminumerical Algorithms*, vol. 2, Addison-Wesley, 2nd edition, pp 1-65.
6. Paul D. Coddington(1997). *Random Number Generators for Parallel Computers*, http:// www.npac.syr.edu
7. Peter Hellekalek, *Inversive Pseudorandom Number Generators*, http://random.mat.sbg.ac.at
8. Reuven Y. Rubinstein(1981). *Simulation and The Monte Carlo Method*, Wiley Series in Probability and Mathematical Statistics, March, 20-37.
9. Shu Tezuka, *Uniform Random Numbers: Theory And Practice,* Kluwer Academic Publisher
10. 김진광(1997). 의사난수발생기의 임의성 검정과 비교 분석, 영남대학교 대학원 통계학과 계 산학전공 석사논문.
11. 신종태(1987). 의사 난수열 생성 방법과 임의성에 관한 연구, 숭실대학교 대학원 전자계산 학과 석사논문.
12. 조영석 외 4 인(1992). 난수 발생기의 비교, 영남 통계학회 논문집, 3, No. 2, 75- 87.

Uniformity and Independency Tests of Pseudo-random Number Generators

Park Kyong Youl[4] · Kwon Gi-Chang [5] · Kwon Young Dam[6]

Abstract

We put the pseudo-random number generator into catagories like MiCG, MuCG, URG, ICG, EICG, and test uniformity and independency by 10,000 times through an empirical trial after selecting this random number generator. Here, from a fraction of data(20, 40, 60, 80, 100) with a significance level of 0.1, 0.05 and 0.01, we drive cumulative frequency with K-S, χ^2, poker, run, autocorrelation test. As a result from the uniformity and independency among five random number generators based on all these data, all random number generator except EICG passed uniformity and independency test, and the URG turn out to be excellent in periodicity.

[4]Department of Statistics, Yeungnam University, Kyongsan, Kyongbuk, 712-749, South Korea
[5]Full-time Instructor, Office Automation, Yechon Junior College, Yechon, Kyongbuk
[6]Professor, Department of Statistics, Yeungnam University, Kyongsan, Kyongbuk, 712-749, South Korea