# 고급 소프트웨어 실습1 1주차 보고서

전공: 컴퓨터공학 학년: 3학년 학번: 20171602 이름: 강지혁

1. 난수(Random Number)

   난수는 어떤 확률분포로부터 무작위로 추출이 된 수를 말한다. 난수는 난수 간 연관성이 없어 누구라도 그 다음에 나올 값을 확신할 수 없어야 한다. 난수들은 각각 추출될 동일한 확률을 가지며 다른 수의 선택에 영향을 주지 않아야 한다. 난수를 사람이 계속해서 만들어내기는 매우 어렵기 때문에 컴퓨터가 난수를 만들어 내는데, 이 수는 난수를 흉내내기 위해 알고리즘으로 생성되는 값이므로 유사 난수(psedo-random numbers)라고 부른다.

   Ripley(1990) 교수는 난수생성기에서 나온 난수는 다음 특성을 갖추어야 한다고 정의했다.

   1. 난수생성기에서 생성된 숫자는 거의 균일하게 분포해야 한다.
   2. 난수생성기에서 생성된 숫자는 상호 독립적이어야 한다.
   3. 난수생성기에서 동일한 숫자는 충분히 긴 기간 후에 생성되어야 한다.
   4. 씨앗(seed)이 주어지지 않는 경우, 난수생성기에서 생성된 숫자는 예측이 불가능해야 한다.

2. 난수 생성 방법

   A. 중앙제곱법

   $$X_{n+1} = (X_n)^2\text{의 가운데 a자리}$$

   중앙제곱법은 폰 노이만이 1949년에 고안한 유사 난수 생성법으로, seed라고 부르는 임의의 숫자를 제곱한 다음 이 숫자의 일부분을 가져와 새로운 난수를 생성하는 방법이다. 중앙제곱법에는 경우에 따라 같은 수가 반복될 수도 있고, 가운데 수가 0이 되는 경우에는 더 이상 난수를 만들어내지 못한다는 단점이 있다.

   B. Linear Congruential

   $$X_{n+1} = (aX_n + c) \bmod m$$

   (a = multiplier, c = increment, m = modulus)

   Linear Congruential은 위와 같은 재귀 관계식으로 정의되며 위 관계식은 난수들의 수열을 반환한다. Linear Congruential은 오늘날 컴퓨터 프로그램에 많이 사용되는데 대표적으로 C언어의 rand 함수가 Linear Congruential을 이용해 난수를 생성한다. Linear Congruential은 일정한 주기가 있고 주기에 따라서 숫자 나열이 반복된다는 단점이 있지만 알고리즘의 계산 속도가 충분히 빠르기 때문에 정교한 수준의 난수가 필요하지 않은 분야에서 많이 사용된다.

   C. 메르센 트위스터

메르센 트위스터는 1997년 마츠모토 마코토와 니시무라 다쿠지가 만든 난수생성기이다. 메르센 트위스터의 이름은 난수의 반복 주기가 메르센 소수인 데서 유래했다. 메르센 소수는 프랑스의 수학자 메르센이 정의한 수로, 2의 거듭제곱에서 1이 모자란 수들 중 소수인 수를 메르센 소수라 이름 붙였다. 이 알고리즘의 동작 원리는 다음과 같다.

1. seed를 사용하여 624 만큼의 길이를 가진 벡터를 생성한다. seed는 보통 하드웨어 노이즈나 오늘 날짜를 사용한다.
2. 이 벡터를 사용하여 624개의 유사 난수를 생성한다.
3. 이 벡터에 노이즈를 준 후 다시 2번을 반복한다. 이 과정을 twist한다고 한다.

메르센 트위스터는 기존의 난수생성기보다 주기가 2^19937 – 1으로 훨씬 더 길며 속도도 빠르다. 메르센 트위스터는 엑셀, MATLAB, PHP, Python, R, C++등 많은 프로그램에서 사용되고 있다.