수치해석 HW#4

2018008559 컴퓨터소프트웨어학부 신상윤

1. Introduction

1. 목적

 임의의 분포에서 data를 생성해보고, 샘플의 수가 많아질수록 어떤 변화가 있는지 논의한다.

2. Uniform distribution, Normal distribution

 if $X \sim U(0,1)$ then $X*(b-a) + a \sim U(0,1)$
 if $X \sim N(0,1)$ then $X*\sigma + \mu \sim N(\mu, \sigma^2)$

2. Process

```
int cnt[101] = {0};
int a = -3, b = 2;
int x;
for(int i=0; i<100; i++){
    double rdv = NR::ran1(x);
    double rdv_u = rdv*(b-a) + a;
    cout << rdv_u << "₩n";
    cnt[(int((rdv_u + 3)*100))/5]++;
}
for(int i=0; i<100; i++){
    cout << cnt[i] << "₩n";
}
```

각 구간에서 얼마만큼의 data가 생성되었는지를 count 해주는 cnt 배열을 만들고,
우리는 U(-3,2)가 궁금하므로 a = -3, b = 2로 주었다.
ran1(x)로 rdv = U(0,1)을 생성하고 새로운 변수 rdv_u를 선언하여
rdv*(b-a) + a를 대입했다. 이는 U(a,b)에서 나온 data와 같다.
출력을 통해서 값이 잘 생성되는지 확인하고, 구간을 100개로 나눴으므로 count도 그에 맞게 해주었다.

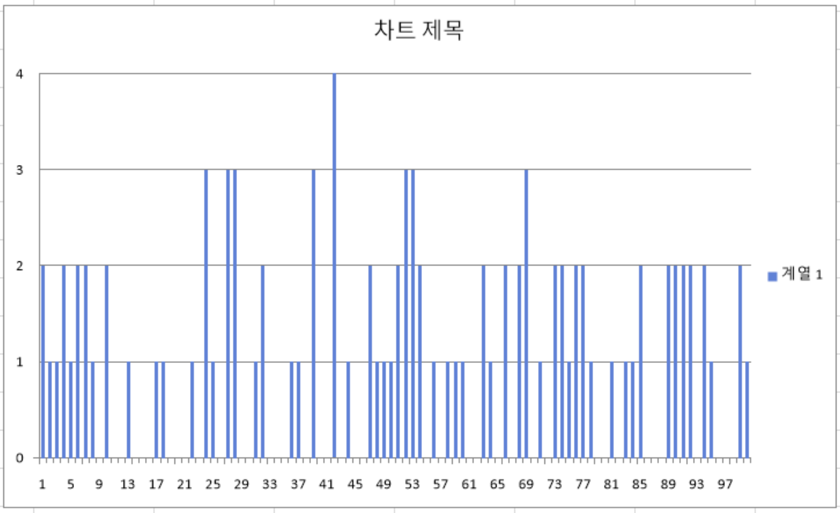
```cpp
int cnt1[101] = {0};
int L = 0;
int R = 0;
double m = 0.5, s = 1.5;
for(int i=0; i<100; i++){
    double rdv = NR::gasdev(x);
    double rdv_n = rdv*s + m;
    cout << rdv_n << "₩n";
    int idx = (int((rdv_n+6-m)*100))/12;
    if(idx < 0) L++;
    else if(idx > 99) R++;
    else cnt1[idx]++;
}
cout << L << "₩n";
for(int i=0; i<100; i++){
    cout << cnt1[i] << "₩n";
}
cout << R << "₩n";
```

각 구간에서 얼마만큼의 data가 생성되었는지를 count 해주는 cnt1 배열을 마찬가지로 만들고, 우리는 N(0.5, 1.5*1.5)가 궁금하므로 m = 0.5, s = 1.5로 주었다.
gasdev(x)로 rdv = N(0,1)을 생성하고 새로운 변수 rdv_n를 선언하여
rdv*s + m을 대입했다. 이는 N(m, s^2)에서 나온 data와 같다.
출력을 통해서 값이 잘 생성되는지 확인하고, 구간을 100개로 나눴으므로 count도 그에 맞게 해주었다.
또한 Uniform distribution과는 다르게 normal distribution은 data가 모든 실수에서 정의되므로 매우 작은 값을 포함해주는 구간 L, 매우 큰값을 포함해주는 구간 R을 추가로 선언해주었다.

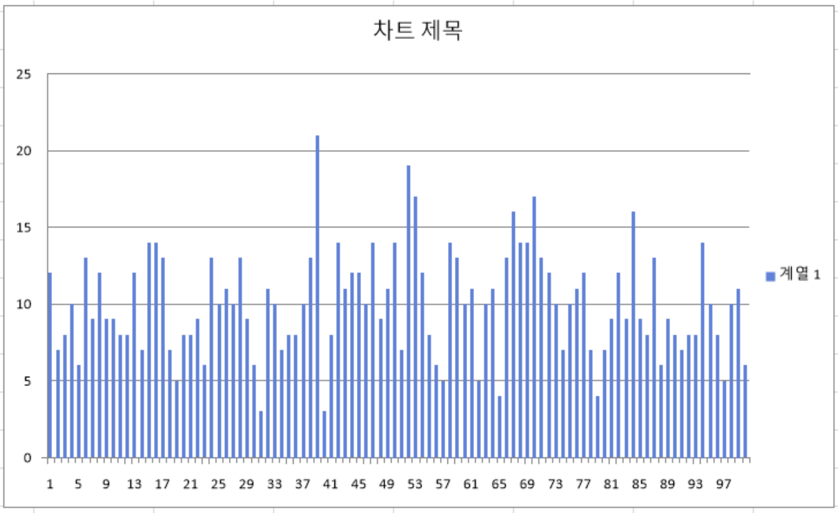
3. Result

3.1 U[-3,2]

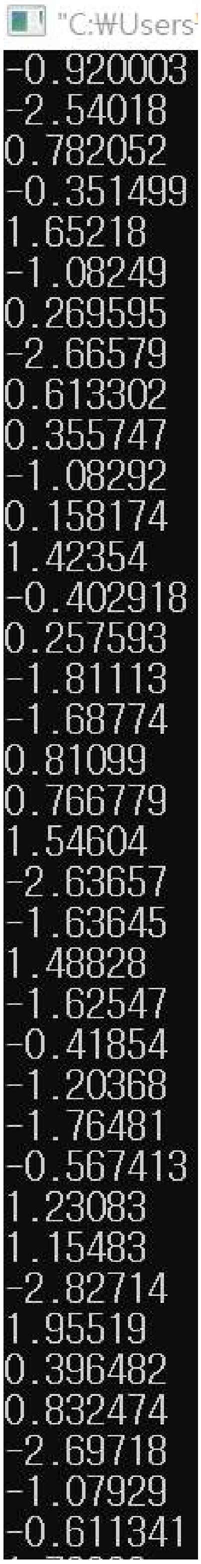
오른쪽과 같이 -3 ~ 2의 data가 잘 형성되었고,
각 data의 빈도를 각각 100, 1000, 10000, 100000번에서
막대그래프로 나타내었다.



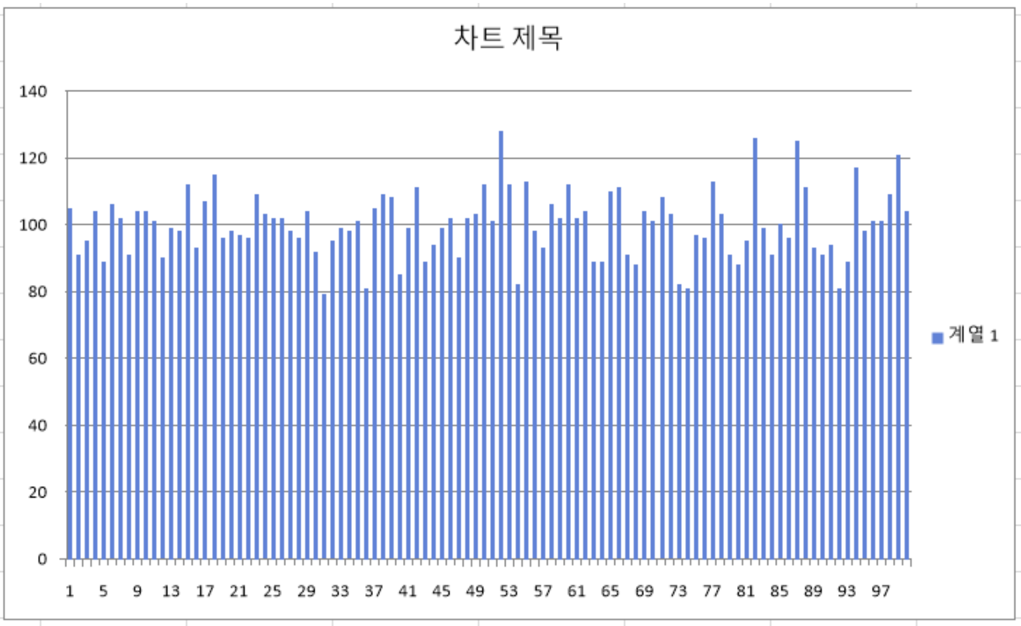
< n = 100 >



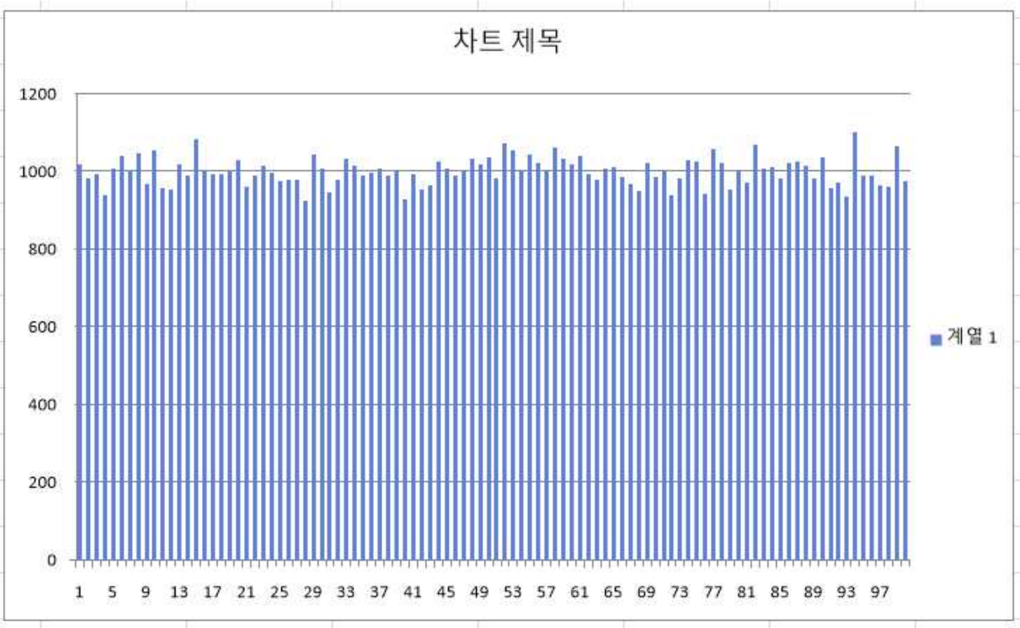
< n = 1000 >



들어 -2.77687532가 생성되었다면 이는 100개로 나눠진 구간 중 [-2.8, -2.75]에 속하고 이는 4번 그룹이다.



< n = 10000 >

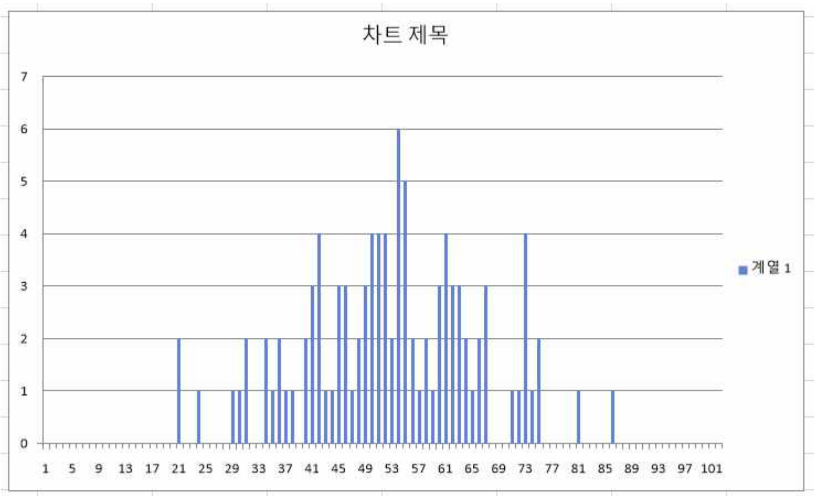


< n = 100000 >

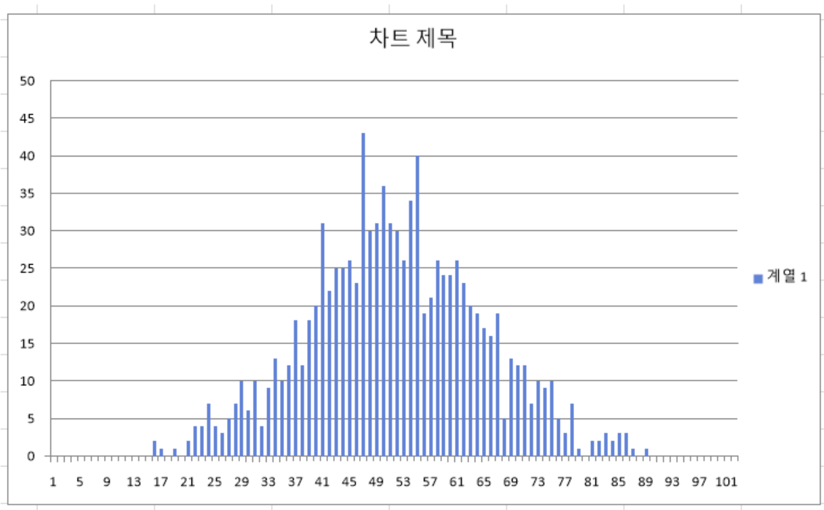
막대그래프의 세로축은 빈도를 나타내며 가로축은 data가 속해있는 그룹이다. 예를 들어 -2.77687532가 생성되었다면 이는 100개로 나눠진 구간 중 [-2.8, -2.75]에 속하고 이는 4번 그룹이다.

3.2 N[0.5, 2.25]

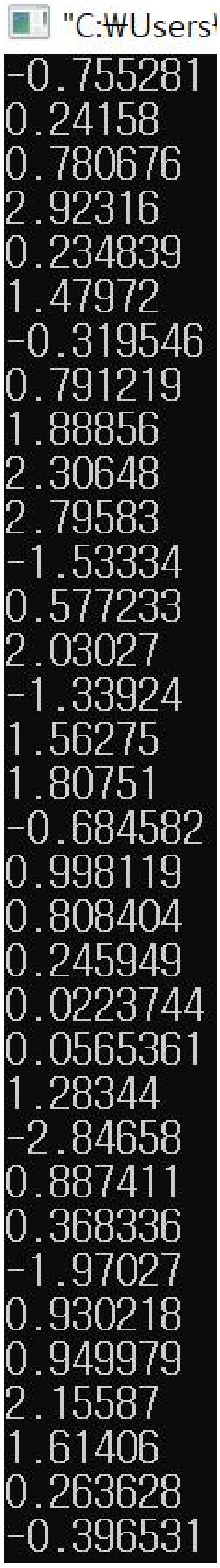
오른쪽과 같이 실수의 data가 잘 형성되었고,
각 data의 빈도를 각각 100, 1000, 10000, 100000번에서
막대그래프로 나타내었다.

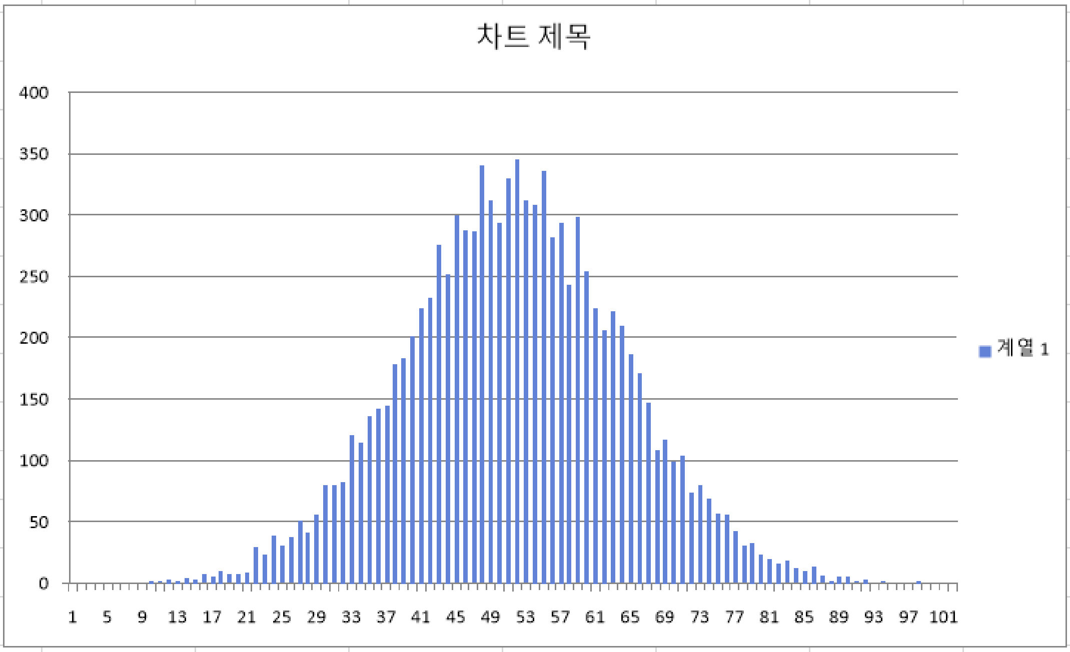


< n = 100 >

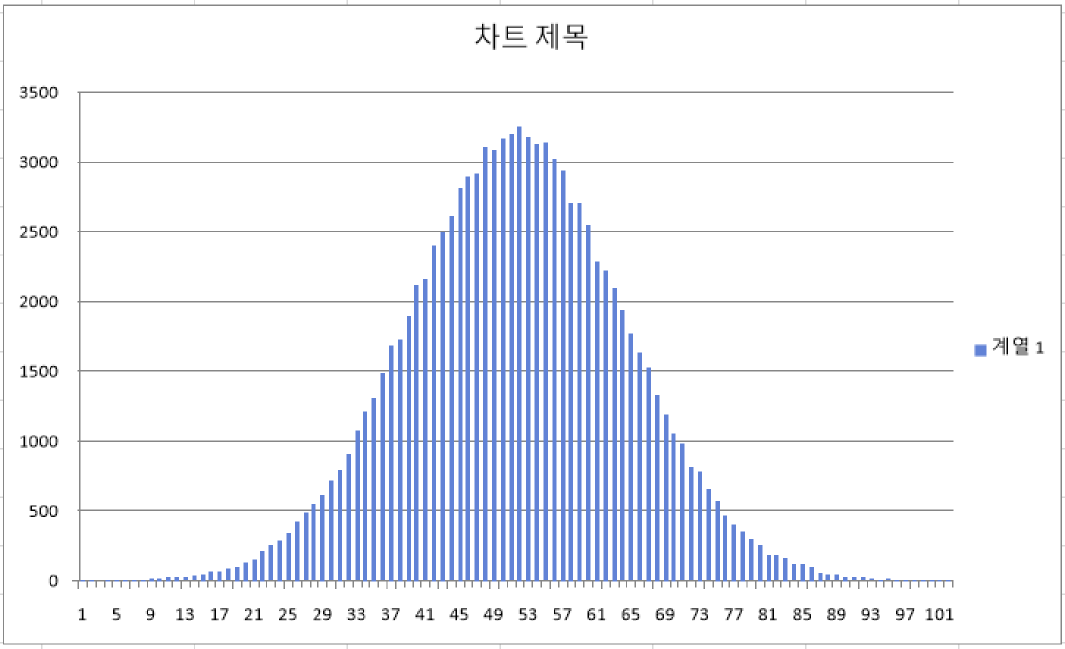


< n = 1000 >





< n = 10000 >



< n = 100000 >

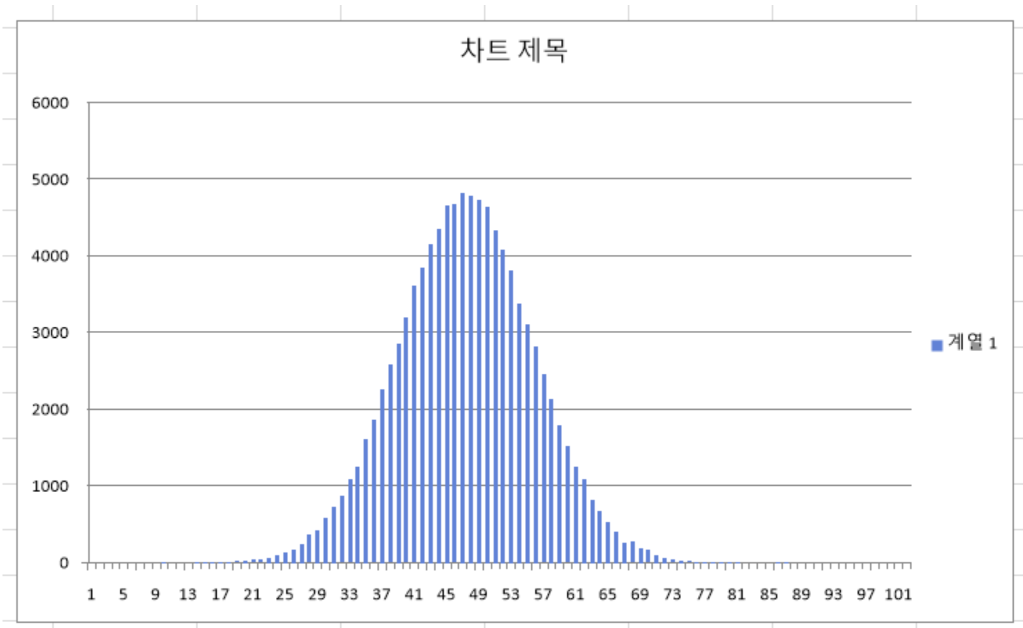
막대그래프의 세로축은 빈도를 나타내며 가로축은 data가 속해있는 그룹이다. 그룹은 [-6, 6]을 100의 구간으로 나눴고, 그보다 작은 값은 L, 큰 값은 R에 넣었다. 예를들어 -2.77687532가 생성되었다면 이는 100개로 나눠진 구간 중 [-2.88, -2.76]에 속하고 이는 26번 그룹이다.

4. Discussion

4.1 그래프의 모양 변화와 실제 분포 그래프와의 차이

실험횟수가 늘어나면서 점점 각 분포 그래프와 비슷해져 간다. 특히 n=100000일 때는 매끄러운 그래프가 보이며 각 분포 그래프와 값이 거의 일치한다. 수학적 확률분포와 비교해봤을 때 가장 큰 차이점은 수학적 확률분포는 연속적이고 어떤 값이 나타날 확률을 나타낸 것이기 때문에 예를 들어 100이 나타날 확률보다 101이 나타날 확률이 작다고 하면 수학적 확률분포에서는 연속적으로 나타내기 위해 매끄럽게 작아지지만 난수 발생 히스토그램에서는 101이 100보다 많이 일어날 수도 있고 99보다 많이 일어날 수도 있으므로 그래프가 매끄럽지 않다. 즉, 수학적 확률분포는 말 그대로 각각이 일어날 확률을 나타낸 것이므로 확률적으로 어떤거는 작고, 어떤거는 크다를 나타내지만 실제 실험에서는 실제 일어난 일이므로 꼭 확률대로 100번 시행하면 30번 나오는 것이 아닌 29번이 나올 수도 있고 10번도 나올 수 있다. 이는 n = 100일 때 실험결과로 확인할 수 있다.

4.2 N(0,1) 과 N(0.5, 2.25) 비교



위의 막대그래프는 표준 정규분포의 실험 결과인데, 크기가 달라서 알기 어려울 수 있지만, 가운데 부분이 더 크다는 것을 알 수 있다. 즉, 우리가 실험을 통해 얻은 N(0.5, 2.25)는 표준정규분포에서 가운데가 눌린 모양이라고 생각할 수 있다.
또한 미세하게 왼쪽에 있는 것을 확인할 수 있는데 이는 평균이 0.5를 기준으로 그래

프를 그렸기 때문이다. 결론적으로 N(0.5, 2.25)가 표준정규분포와 비교했을 때 중심축이 오른쪽으로 옮겨졌고, 분산이 증가했기 때문에 표준정규분포보다 눌린 모양이 나타남을 알 수 있다.

5. Reference

[1] http://numerical.recipes/book/book.html