

Assignment 1

Laplace distribution of

Random variable

|  |  |  |
| --- | --- | --- |
| 과목명 |  | 확률및통계 |
| 담당교수 |  | 심동규 교수님 |
| 학과 |  | 컴퓨터정보공학부 |
| 학년 |  | 3학년 |
| 학번 |  | 2019202009 |
| 이름 |  | 서여지 |
| 제출일 |  | 21.03.26(금) |



1. 과제설명

Uniform 분포를 갖는 random variable X를 Laplace 분포를 갖는 random variable Y로 변환하는 프로그램을 작성하고, random variable Y가 실제 Laplace분포인지 검증하는 과제이다. 프로그램은 평균과 표준편차를 조정할 수 있으며 그에 따라 생성되는 random variable Y의 분포가 적절하게 변화해야 한다.

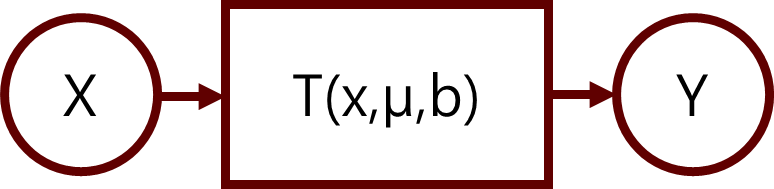
1. 접근방법

과제를 해결하기 위해 크게 세가지 부분으로 문제를 나누었다.

1. 실제 random variable을 변환하는 것에 필요한 T(x, μ, b)를 구하는 것
2. 생성한 분포를 시각화 하는 것
3. random variable Y가 Laplace분포임을 검증하는 것

아래에서 각 부분에 대해 설명한다.

1. T(x, μ, b)



FY[y=T(x)] = FX(x) //T(X)는 단조증가한다.

이때 x는 uniform분포이고, cdf는 x이다. 이것을 위 식에 대입하면

y = T(x) = FY-1(FX(x)) = FY-1(x) 이다.

따라서 uniform분포를 다른 분포로 변환하는 T(x)는 해당 분포의 cdf의 역함수이다.

Laplace분포의 cdf에 대해 역함수를 구하는 과정은 다음과 같다.

=

y = 1-0.5exp[(-x+ μ)/b]

exp[(-x+ μ)/b] = 2(1-y)

ln(2-2y) = [(-x+ μ)/b]

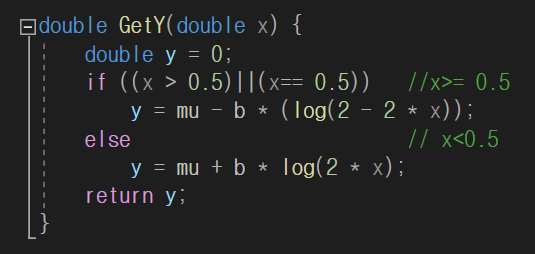
x = -bln(2-2y)+ μ

y = (1/2)exp[(x- μ)/b]

ln2y = [(x- μ)/b]

x = bln2y + μ

위 식을 이용하여 uniform 분포를 Laplace 분포로 변환할 수 있다. 프로그램을 작성할 때 C코드에 다음과 같이 작성되었다.

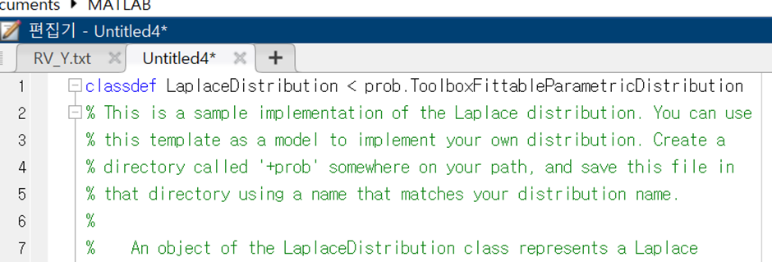


1. 분포시각화

분포시각화에 MATLAB을 이용하여 히스토그램을 확인하는 방법을 이용하였다. 작성한 프로그램을 이용하여 생성한 난수 파일을 readmatrix함수를 이용하여 불러오고, histogram함수를 이용해서 확인한다. 분포의 cdf는 distributionFitter명령어를 이용하여 분포피팅기에서 확인했다.

1. 분포 검증

random variable Y가 Laplace 분포임을 검증하기 위해 Kolmogorov-Smirnov test를 이용하였다. Laplace 분포의 데이터는 사용자 지정 분포 정의의 샘플에 저장된 것을 이용하였다. random variable Y의 cdf와 Laplace 분포의 cdf를 시각적으로 확인할 수 있고, MATLAB에서 지원하는 함수인 kstest함수를 이용하여 분포를 검증하였다.

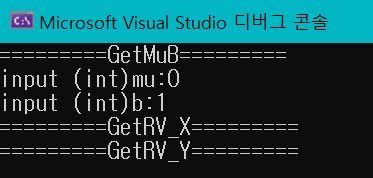


<Laplace 분포는 샘플로 제공된 것을 이용함>

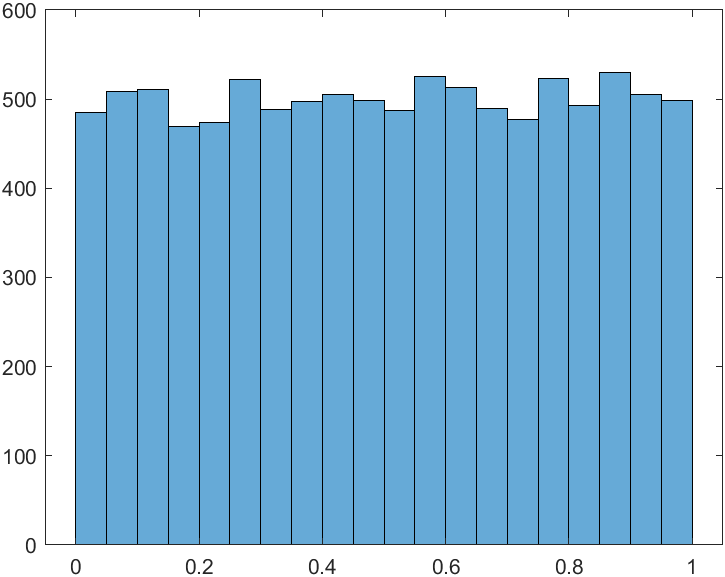
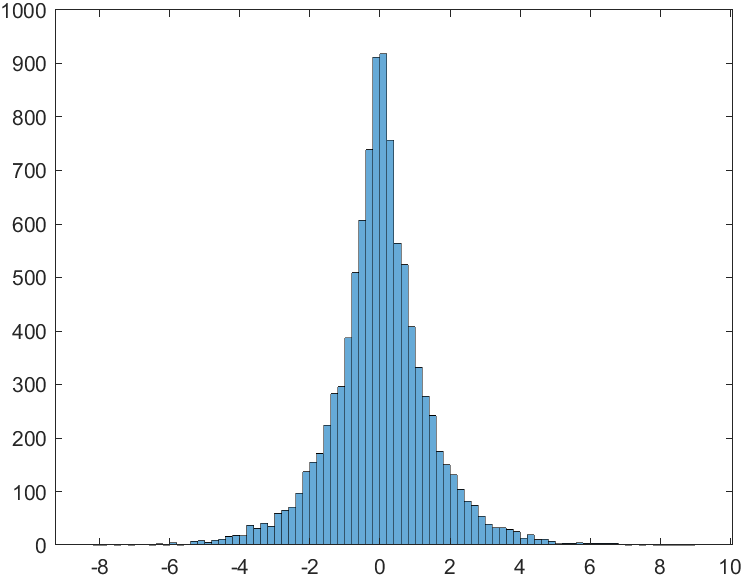
1. 실험 및 검증 결과

실험1. random variable X와 Y의 분포는 각각 uniform분포와 Laplace분포이다.

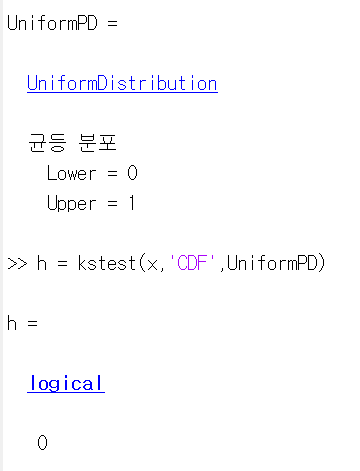
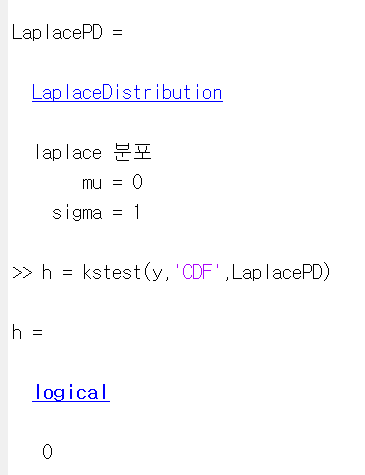
작성한 프로그램에 mu=0, b=1로 인자를 전달하고, 실행 결과로 얻은 RV\_X.txt와 RV\_Y.txt를 MATLAB의 histogram함수를 이용하여 확인한다. 또한 MATLAB에서 지원하는 K-S 검증법인 kstest함수를 이용하여 각각의 분포를 확인한다.



<프로그램 실행 화면>

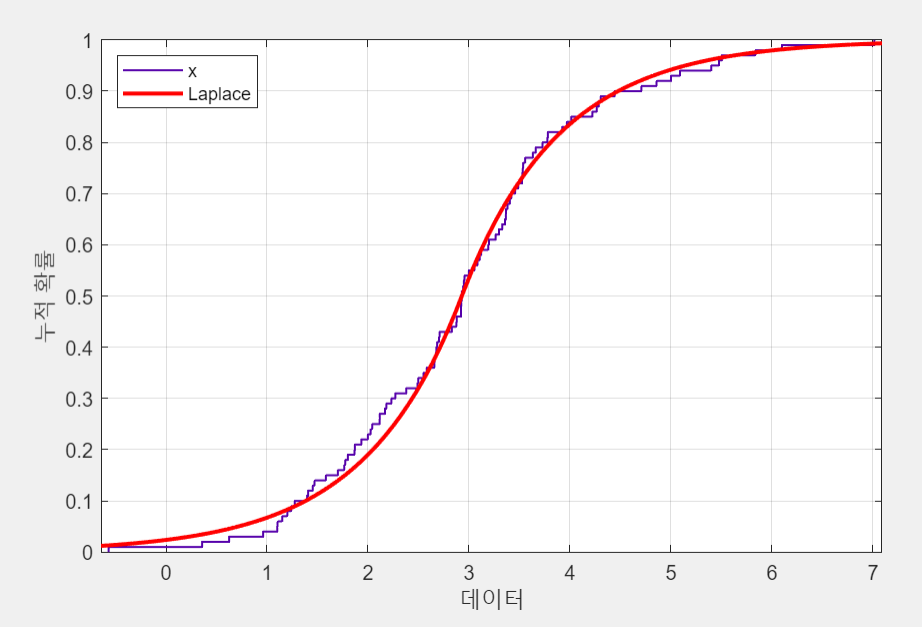
 

<random variable X> <random variable Y>

 ****

<random variable X> <random variable Y>

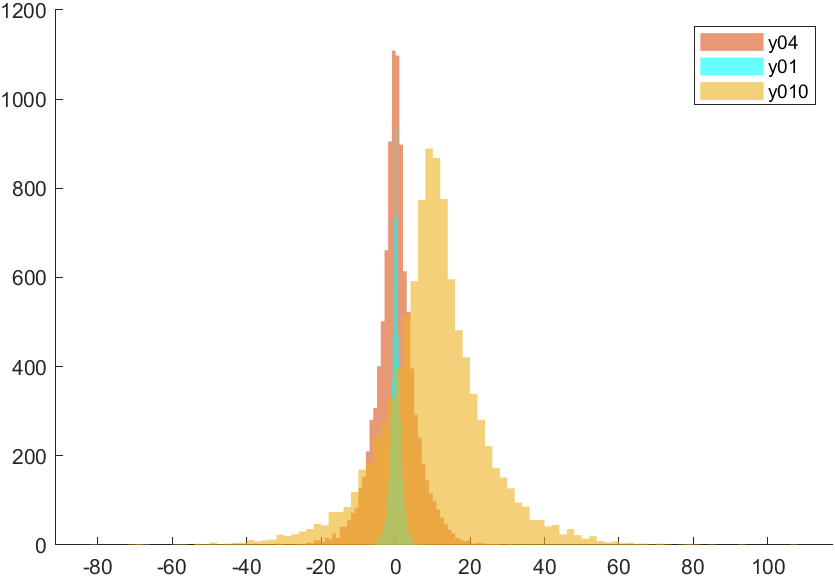
h = 0일 때 가설이 부정되지 않음(같은 분포)이므로 random variable X는 uniform 분포를, random variable Y는 Laplace 분포임을 검증할 수 있었다.



<random variable Y와 Laplace 분포 cdf 비교>

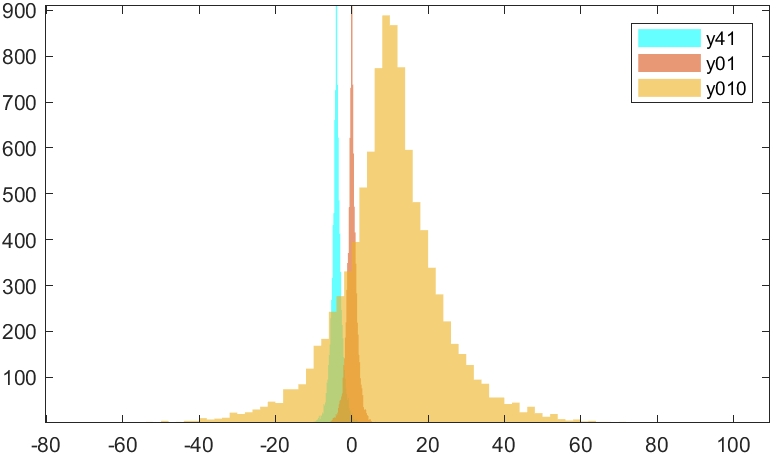
실험2. 프로그램에 입력한 평균과 표준편차에 따라 Y의 분포가 적절하게 변화한다.

작성한 프로그램에 인자를 변화시키며 여러 번 실행하고, 실행 결과로 얻은 여러 개의 RV\_Y파일을 MATLAB의 histogram함수를 이용하여 시각적으로 비교한다.



<T(x,0,1), T(x,0,4), T(x,10,10) 히스토그램 비교>

입력 값으로 전달한 표준편차의 크기가 커질수록 히스토그램의 형태가 넓게 퍼지는 것을 확인할 수 있다.



<T(x,0,1), T(x,4,1), T(x,10,10) 히스토그램 비교>

입력 값으로 전달한 평균의 크기에 따라 히스토그램의 위치가 이동하였다.

1. 고찰

rand함수는 uniform 분포를 따르는 난수를 발생시킨다는 것을 확인하였다. 이것은 0에서 RAND\_MAX까지의 수가 발생할 확률이 모두 동일하다는 의미이고, 무작위로 난수를 발생하는 용도에 적함함을 알 수 있다. 또한 rand함수에 대해 조사하며 사실 rand함수를 비롯한 몇 가지 함수가 생성하는 난수는 유사난수임을 알게 되었다. 정해진 알고리즘을 따라 동작하는 컴퓨터는 무작위의 난수를 발생시키기 어렵고, 그렇기 때문에 이미 정해진 난수표 등을 이용하여 난수에 근접한 유사난수를 생성한다.

프로그램을 작성하며 자주 사용한 함수인 rand()함수가 어떤 수의 발생확률이 모두 같을 것이라고 생각하고 자세한 분포를 확인하지 않았다. 과제를 진행하며 자주 사용하던 함수의 구조를 들여다보는 과정이 흥미로웠다. 이외에도 특정한 분포를 갖는 난수 발생 함수가 필요한 경우가 존재하며 기존의 rand함수를 이용하여 원하는 분포를 갖는 함수를 직접 정의할 수 있다는 것을 알 수 있었다.

1. 출처

rand() c++ reference, <https://www.cplusplus.com/reference/cstdlib/rand/>

kstest MATLAB mathworks, <https://kr.mathworks.com/help/stats/kstest.html>

makedist MATLAB mathworks, <https://kr.mathworks.com/help/stats/makedist.html>

distributionfitter MATLAB mathworks, <https://kr.mathworks.com/help/stats/distributionfitter-app.html>

Laplace distribution wikipedia <https://en.wikipedia.org/wiki/Laplace_distribution>