**확률과 통계**

**HW #1**

**담당교수 : 심동규**

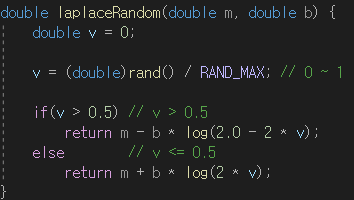
**제출일 : 2021. 03. 28.**

**학부 : 컴퓨터정보공학부**

**학번 : 2017202088**

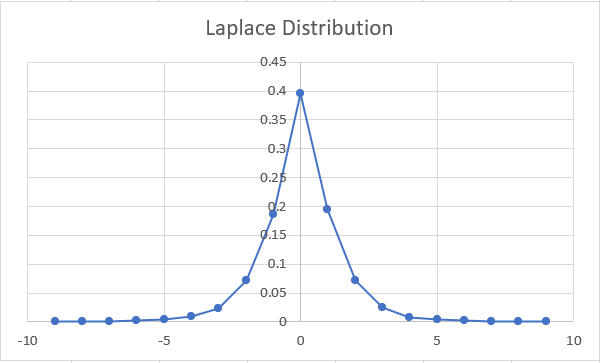
**이름 : 신해담**

1. **과제 설명**
   * Uniform distribution을 따르는 난수를 생성하는 rand함수를 이용하여 Laplace distribution을 따르는 난수를 생성하는 SW를 작성한다. 평균과 표준편차를 조정할 수 있어야 하며, 생성된 난수의 분포가 Laplace 분포인지 검증한다.
2. **접근 방법** 
   * 0~1 사이의 random variable X가 Uniform distribution을 따를 때, FX(x) = x이다. 이 때 X를 이용하여 Laplace distribution을 따르는 random variable Y를 얻고 싶다. 그러므로 X를 어떤 Transformation 함수 T에 넣으면 Y를 얻을 수 있고, 이 Y의 CDF는 다음과 같다.  
     FY(y) = exp() x ≤ μ  
      1 - exp() x > μ  
     Y = T(x)이므로, FY[y=T(x)] = FX(x)가 성립한다. 따라서 FY(y) = x이고 T(x)를 구하면 다음과 같다.  
     y = T(x) = μ + b ln(2x) x ≤   
      μ – b ln(2 \* (1 – x) x >

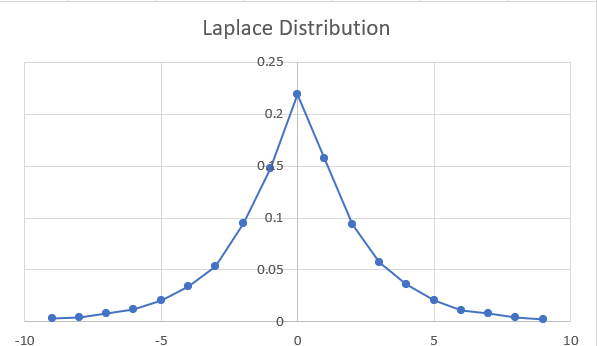


* + 위에서 구한 T(x)의 식을 이용하여 Laplace 분포를 따르는 난수 생성 함수를 작성하였다. laplaceRandom 함수는 μ와 b를 입력받아서 txt파일에 난수를 출력하도록 했다. X는 0~1 사이의 값이지만, rand() 함수로 생성되는 난수는 0에서 RAND\_MAX 사이의 int형 난수를 발생시키므로 생성된 난수를 RAND\_MAX로 나누어 0~1 사이의 무작위 값을 나타낼 수 있도록 했다.  
    if문에서는 생성한 0~1 사이의 난수가 0.5보다 클 때와 작을 때에 맞는 T(x)로 들어가 결과값을 return하도록 구현했다.

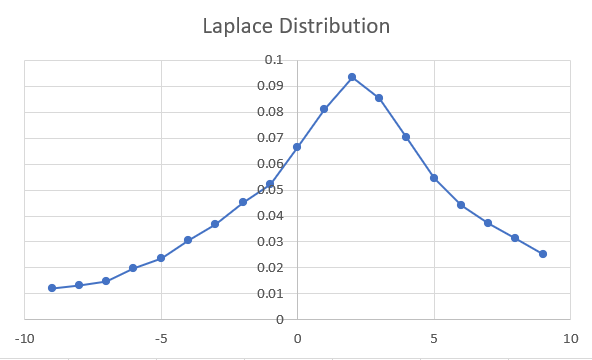
1. **실험 및 검증 결과**



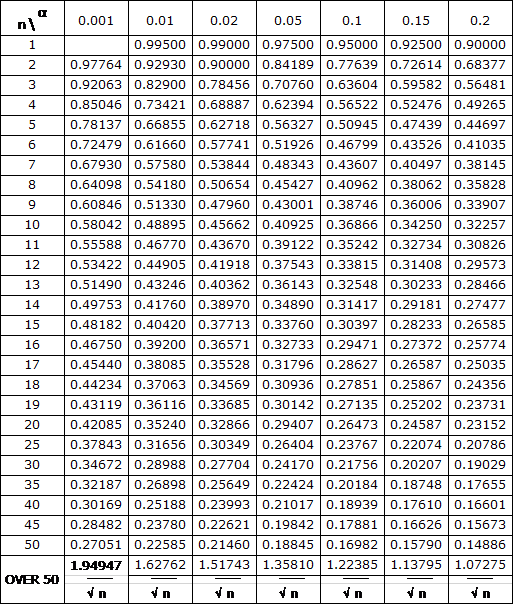
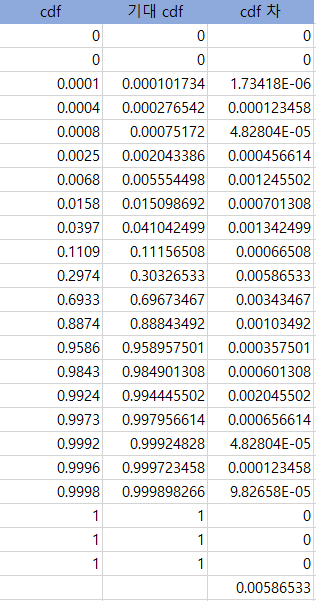
* + μ = 0, b = 1 에서의 분포



* + μ = 0, b = 2 에서의 분포  
    b = 1일 때보다 높이는 낮아지고 옆으로 퍼졌다.



* + μ = 2, b = 5 에서의 분포  
    중심이 μ = 2로 바뀌고 b = 2일 때보다 옆으로 더 퍼졌다.

* + 검증은 K-S test를 이용했다.  
    μ = 0, b = 1 에서의 난수 10000개를 생성하여  
    이론적 CDF – 생성한 난수 빈도의 CDF의 최댓값 D를 구하니 0.00586533이 나왔다. 이를 위의 표를 이용하여 분석하면  
    n = 10000이며 1.94947/sqrt(10000) = 0.0194947 > D 이므로, 생성한 난수의 분포가 Laplace 분포를 따른다고 할 수 있다.

1. **고찰**
   * Transformation 함수에 대한 이해 없이 과제를 진행하니 내가 입력해야 하는 값과 출력해야 하는 값, 무엇의 분포를 구해야 하는지조차 명확하지 않아 헤맸다. 결국 처음으로 돌아가서 강의를 다시 듣고 강의자료를 복습하여 이해한 다음에야 입력값을 어떻게 줄 건지, 출력값을 얻으려면 어떤 식을 세워야 하는지 인지할 수 있었다.
   * 생성된 난수 표본이 너무 작으면 오차가 커졌다. 반대로 평균값에서 크게 벗어난 구간에서 표본이 너무 적어 값이 틀어지는 경우가 발생해 표본 크기를 키우면 크기가 너무 커서 엑셀이 분석을 못하는 경우가 발생했다. 이런 상황을 유연하게 통제하기 위해 입력값으로 난수 생성 수를 넣어 상황에 따라 조절했다.