디지털통신시스템설계

Digital Communication System Capstone Design

ICE4009-001

2 주차 실습과제

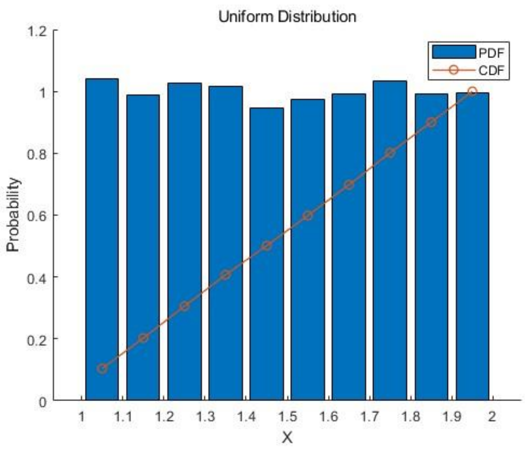
정보통신공학과

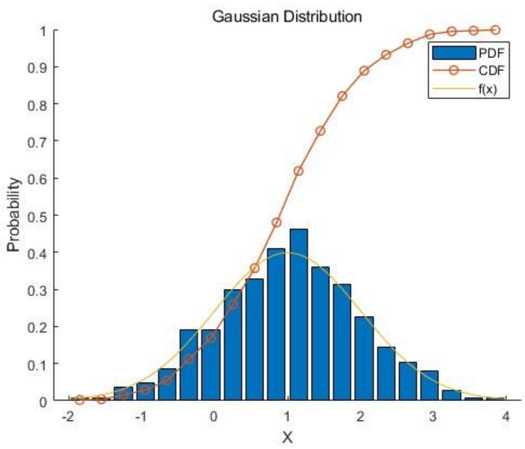
12191765

박승재

Introduction

PDF 는 Probability Density Function 으로, 연속확률분포에서 특정 구간의 확률을 나타내기 위한 함수이다. CDF 는 Cumulative Distribution Function 으로, 주어진 확률분포에서 확률변수가 특정 값보다 작거나 같은 확률을 나타낸다.





Uniform Distribution 은 모든 확률이 균일한 분포로, PDF 는 모든 구간에서 비슷한 Density 를 가지며 CDF 는 직선으로 나오게 된다.

Gaussian Distribution 은 정규분포라고도 불리며, PDF 는 가운데가 볼록한 형태가 되며 CDF 는 가운데에서 급격하게 증가하게 된다. 평균이 0 이고 표준편차가 1 인 정규분포 N(0,1)을 표준 정규분포(standard normal distribution)라고 한다.

`plot(X,Y)`는 X 값에 대한 Y 데이터의 2차원 선 플롯을 생성합니다.

- 선분으로 연결된 좌표의 집합을 플로팅하려면 X와 Y를 동일한 길이의 벡터로 지정하십시오.
- 동일한 좌표축에 여러 개의 좌표의 집합을 플로팅하려면 X와 Y 중 적어도 하나를 행렬로 지정하십시오.

`plot(X,Y,LineSpec)`은 지정된 선 스타일, 마커, 색을 사용하여 플롯을 만듭니다.

plot 은 주어진 X, Y 로 선 그래프를 그리는 함수이다. hold on 을 통해 여러 그래프를 하나의 figure 에 그릴 수 있다. plot 뒤에 그래프를 꾸미는 옵션을 지정할 수 있다.

LineSpec — 선 스타일, 마커, 색
string형 | 문자형 벡터

선 스타일, 마커, 색으로, 기호를 포함하는 string형 또는 문자형 벡터로 지정됩니다. 기호는 어떤 순서로 지정해도 좋습니다. 세 가지 특성(선 스타일, 마커, 색)을 모두 지정할 필요는 없습니다. 예를 들어 선 스타일을 생략하고 마커를 지정하면 플롯은 마커만 표시하고 선은 표시하지 않습니다.

예: "--or"은 원 마커로 표시된 빨간색 파선입니다.

선 스타일	설명	결과 선
"-"	실선	————
"--"	파선	- - - - -
":"	점선	··················

마커	설명	결과로 생성되는 마커
"o"	원	○
"+"	플러스 기호	+
"*"	별표	✳
"."	점	•
"x"	십자	×
"_"	가로선	—
"|"	세로선	|
"square"	정사각형	□
"diamond"	다이아몬드	◇
"^"	위쪽 방향 삼각형	△
"v"	아래쪽 방향 삼각형	▽
">"	오른쪽 방향 삼각형	▷
"<"	왼쪽 방향 삼각형	◁
"pentagram"	펜타그램	☆
"hexagram"	헥사그램	✡

색 이름	짧은 이름	RGB 3색	모양
"red"	"r"	[1 0 0]	
"green"	"g"	[0 1 0]	
"blue"	"b"	[0 0 1]	
"cyan"	"c"	[0 1 1]	
"magenta"	"m"	[1 0 1]	
"yellow"	"y"	[1 1 0]	
"black"	"k"	[0 0 0]	
"white"	"w"	[1 1 1]	

bar(y)는 y의 각 요소마다 막대가 하나씩 있는 막대 그래프를 만듭니다.

- 단일 막대 계열을 플로팅하려면 y를 m 길이 벡터로 지정하십시오. 막대는 x축을 따라 1에서 m까지 위치합니다.
- 여러 개의 막대 계열을 플로팅하려면 y를 계열마다 열을 하나씩 갖는 행렬로 지정하십시오.

---

bar(x,y)는 x로 지정된 위치에 막대를 그립니다.

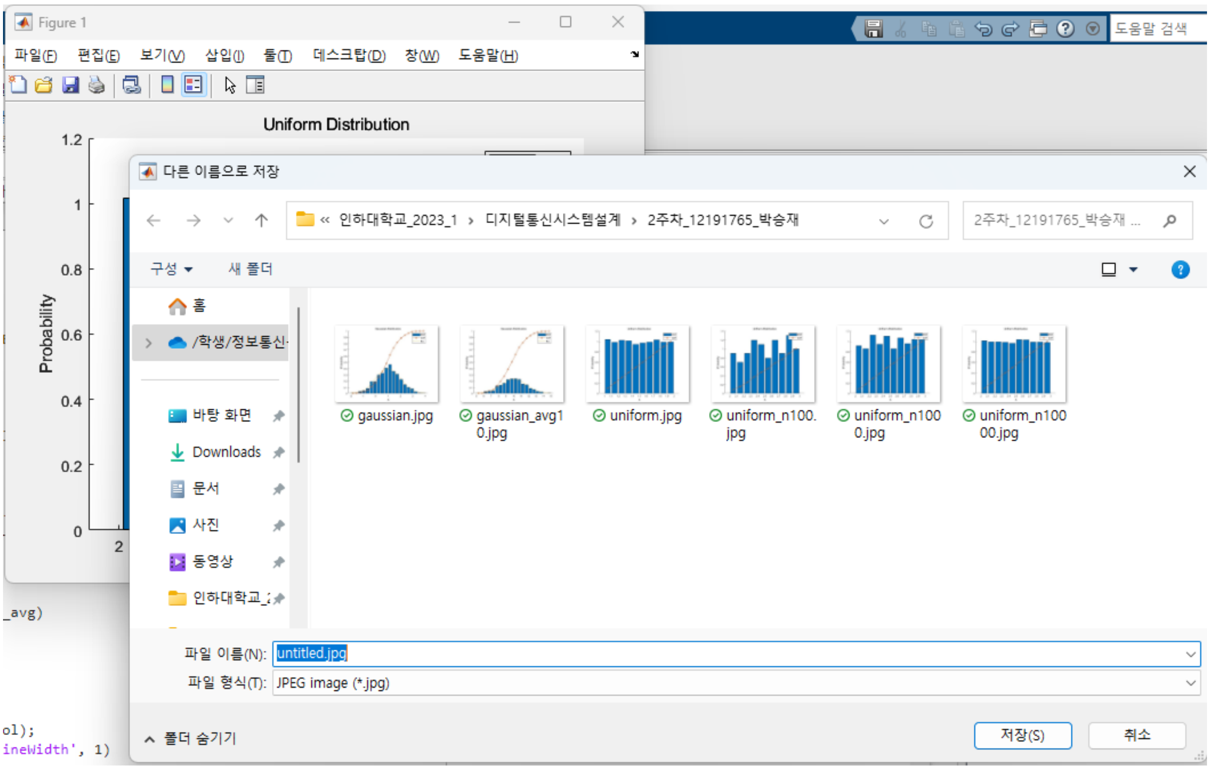
bar 는 막대그래프를 그리는 함수이다. plot 과 마찬가지로 인자로 X, Y 를 받는다.

설명

B = cumsum(A)는 크기가 1이 아닌 A의 첫 번째 배열 차원의 시작에서부터 계산하여 A의 누적합을 반환합니다.

- A가 벡터인 경우 cumsum(A)는 배열(A) 요소의 누적합을 계산하여 벡터로 반환합니다.
- A가 행렬인 경우 cumsum(A)는 A의 각 열별로 누적합을 계산하여 행렬로 반환합니다.

cumsum 은 배열 요소의 누적합을 계산하여 벡터로 반환한다.



Matlab 의 figure 에는 그래프를 jpg 이미지로 저장하는 기능이 있다. jpg 저장 기능을 이용하면 스크린샷을 찍는 것과 달리, 하얀 배경의 그래프를 얻을 수 있다.

Problem

1. [2, 3]의 범위를 가지는 Uniform 난수를 발생시키고 PDF, CDF 를 그리기

   ① 난수의 개수(N=100, 1000, 10000)에 따른 결과를 비교하여라(구간의 수 R=10)

   ② 발생시킨 난수의 평균을 계산하여라 (수식: $\mu = \frac{\sum_{i=1}^{N} X_i}{N}$)

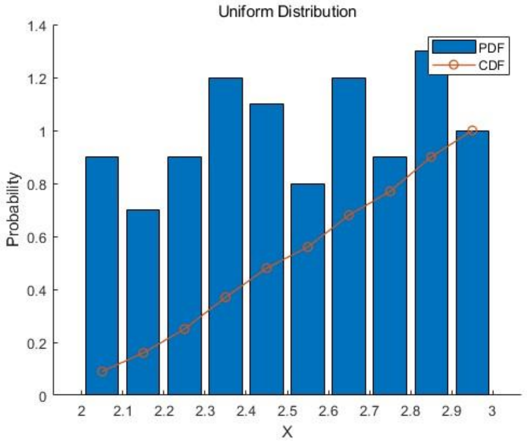
2. 평균 10, 분산 3 인 Gaussian 난수를 1000 개 발생시키고 PDF, CDF 를 그리기

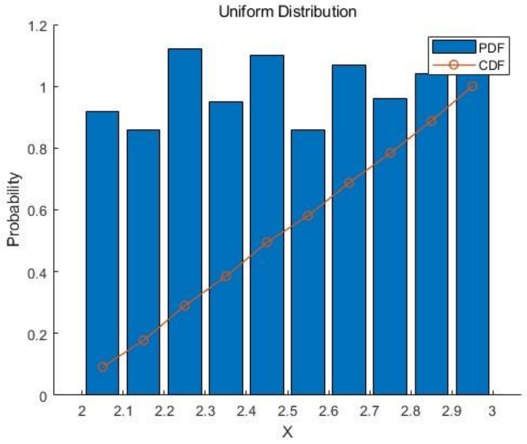
   ① 발생시킨 난수의 평균과 분산을 구하기 (수식: $\mu = \sum_{i=1}^{N} X_i / N,\ \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2$)

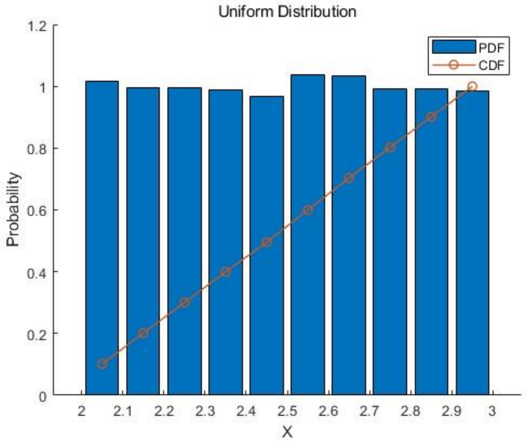
   ② Gaussian 분포의 수식을 이용한 그래프와 겹쳐 그려서 비교하여라

   (수식: $f(\mathrm{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$)

Result








```
>> uniform
avg: 2.508656
>> uniform
avg: 2.508506
>> uniform
avg: 2.498206
```

```
A = 2;
B = 3;
N = 10000;
R = 10;

U = A + rand(1, N) * (B - A);

% [M, X] = hist(U, R)
% resol = X(2) - X(1);
resol = (B - A) / R;
X1 = (A:resol:B - resol) + (resol / 2);
M1 = zeros(1, R);

for i = 1:N
    idx = floor((U(i) - A) / resol) + 1;
    M1(idx) = M1(idx) + 1;
end

U_avg = sum(U) / N;
```

```
fprintf('avg: %f\n', U_avg)

hold on;
PDF = M1 / N / resol;
bar(X1, PDF)

CDF = cumsum(PDF * resol);
plot(X1, CDF, '-o', 'LineWidth', 1)

title('Uniform Distribution')
xlabel('X')
ylabel('Probability')
legend('PDF', 'CDF')
hold off;
```

① 표본의 개수 N 을 증가시키면서 PDF 와 CDF 를 그려보았다. N 이 100 일 때보다 10000 일 때 각 막대의 높이가 비슷해지는 것을 확인할 수 있다. CDF 직선도 N 이 100 일 때는 완전한 직선이 아니지만, 10000 일 때는 거의 직선에 가까운 것을 확인할 수 있었다.

② N 의 크기와 상관없이 평균은 2.5 에 비슷하게 나오는 것을 확인할 수 있었다.

Matlab의 hist 함수 대신 for과 if로 직접 PDF를 만들었다. hist는 함수의 입력 인자 중 첫 인자인 행렬의 요소의 히스토그램 생성하는 함수이다.

[구문]

```
[M, X] = hist(A, R)
```

A : 입력 벡터 또는 행렬
R : 구간의 개수로 A의 요소를 R개의 구간으로 구분
M : 구간별 A의 요소의 개수
X : 구간의 중심 값

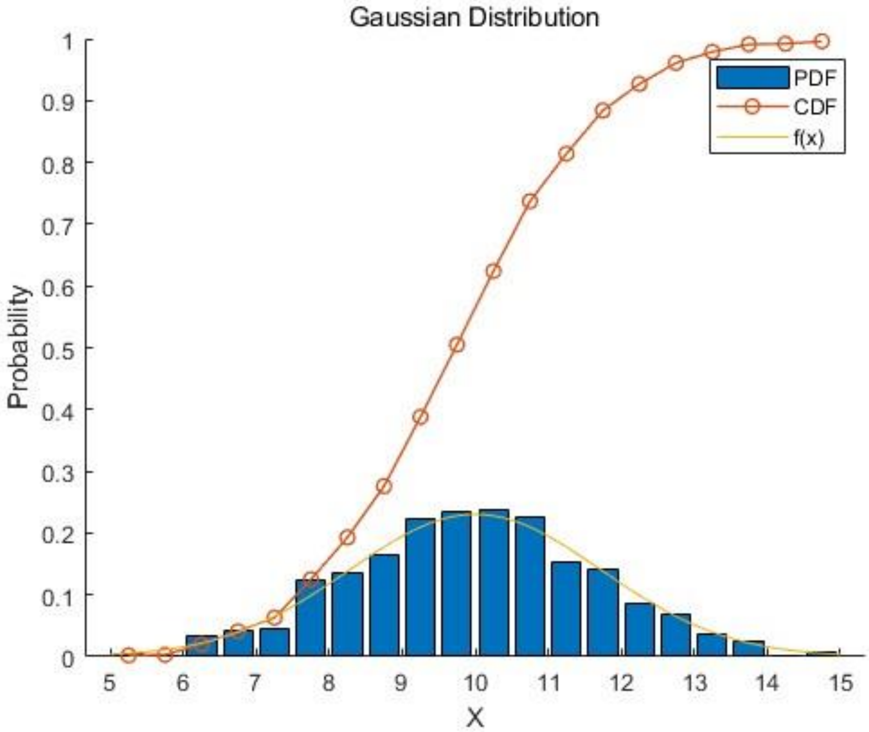
[hist 함수로 PDF 구하는 방법]

```
PDF = M/N/resol
resol = X(2)-X(1)
```

U(i)가 속한 구간의 위치 idx 를 알아내기 위해 U(i)에서 구간의 시작 값 A 를 빼고 구간 간격 resol 로 나누었다. Matlab 에서는 인덱스가 1 부터 시작하기에 마지막에 1 을 더하였다. 숫자가 속한 구간 M1(idx) 값을 1 증가시키면 hist 기능이 구현된다.

PDF 는 bar 를 통해 그렸고, CDF 는 cumsum 으로 값을 구하고 plot 을 이용해 그렸다. 두 그래프 모두 각 구간의 중심 값이 들어있는 X1 를 x 축으로 가진다.

PDF 와 CDF 를 동시에 보여주기 위해 hold 를 사용했다. title 과 xlabel 등으로 실습자료의 차트와 형태가 같도록 그래프를 꾸며주었다.




```
>> gaussian
avg: 10.063910, var: 2.988595
```

```
A = 5;
B = 15;
N = 1000;
avg = 10;
stdev = sqrt(3);
R = 20;

G = avg + randn(1, N) * stdev;

resol = (B - A) / R;
X1 = (A:resol:B - resol) + (resol / 2);
M1 = zeros(1, R);

for i = 1:N
    if G(i) >= A && G(i) < B
        idx = floor((G(i) - A) / resol) + 1;
        M1(idx) = M1(idx) + 1;
    end
end

G_avg = sum(G) / N;
G_var = sum((G - G_avg) .^ 2) / N;
G_stdev = sqrt(G_var);
fprintf('avg: %f, var: %f\n', G_avg, G_var)
```

```
x = A:0.1:B;
G_new = 1 / (stdev * sqrt(2 * pi)) * exp(-1/2 * ((x - avg) ./ stdev) .^ 2);

hold on;
PDF = M1 / N / resol;
bar(X1, PDF)

CDF = cumsum(PDF * resol);
plot(X1, CDF, '-o', 'LineWidth', 1)

plot(x, G_new)

title('Gaussian Distribution')
xlabel('X')
ylabel('Probability')
legend('PDF', 'CDF', 'f(x)')
hold off;
```

① 평균은 10.064, 분산은 2.988 이 나왔다. 문제에서 분산 3 을 주었기 때문에 `randn` 에 곱해지는 `stdev` 는 `sqrt(3)`가 들어가야 한다. 생성한 난수 G 의 분산 `G_var` 을 구할 때는 `.^ 2` 를 이용해 각 수를 제곱하고 `sum` 으로 합쳐주었다.

② 노란색 그래프가 $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ 이다. 파란색 막대그래프와 노란색 곡선이 거의 일치하게 그려지는 것을 확인할 수 있었다.

가우시안 분포를 갖는 난수를 만들기 위해 N(0, 1)인 `randn` 를 이용해 난수를 만들고 `avg` 를 더하고 `stdev` 를 곱해 `G` 를 만들었다. 문제에서 평균을 10 으로 주었기 때문에 그래프에서 가장 볼록한 부분은 10 이 되는 것을 확인할 수 있었다.

Uniform Distribution 과 달리 Gaussian Distribution 에서는 난수의 시작 값과 끝 값이 명확히 존재하지 않기에 CDF 가 1.0 까지 올라오는 적당한 값으로 A=5, B=15 를 잡았다. 평균을 중심으로 A, B 를 잡았고, CDF 가 1.0 까지 올라오는 것의 의미는 생성된 대부분의 난수가 그래프에 그려졌다는 뜻이다.

노란색 f(x)를 그리기 위해 시작 값부터 끝 값까지 0.1 간격으로 숫자를 만들어 x 값으로 잡고, $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$에 `A:0.1:B` 를 넣어 y 값 `G_new` 을 구했다. `G_new` 를 만들 때는 `x` 가 1×101 크기의 행렬임을 유의해 계산하도록 한다.

Conclusion

Uniform Distribution 과 Gaussian Distribution 을 가진 난수를 만들고, 각각 PDF 와 CDF 를 그려보았다. Uniform Distribution 의 PDF 는 N 이 커질수록 막대의 높이가 비슷해지며 이는 각 구간에서 생성된 난수가 들어갈 확률이 거의 비슷하다는 것을 의미한다. CDF 는 누적분포로 Uniform Distribution 에서는 직선의 형태를 띄게 된다.

Gaussian Distribution 의 PDF 는 가운데가 볼록한 정규분포 모양을 가진다. 그래프의 가장 볼록한 부분이 평균이며 평균에서 멀어질수록 막대의 크기가 줄어든다. CDF 는 평균에서 가장 급격하게 증가하며 이때의 값은 대략 0.5 로 나오게 된다. 평균에서 가장 급격하게 증가하는 이유는 대부분의 난수 값들이 평균에 가깝게 생성되기 때문에 CDF 그래프의 기울기가 가파르다. 생성된 난수의 분포가 올바른지 확인하기 위해 가우스 분포 함수 $f(\mathrm{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ 를 이용해 노란색 그래프를 그렸고, 파란색 막대와 노란색 곡선이 거의 일치함을 확인했다.