

Statistics and Inferencing

Activity 02

Ali Muhammad Asad
aa07190

The purpose of this activity is to:

1. Consider a distribution. Generate data samples from the distribution using appropriate MATLAB function.
2. Visualize the data using histogram,
3. Normalize the histogram and compare it with the probability density function of considered distribution.

Modify the code given for Gamma Distribution and repeat for the following:

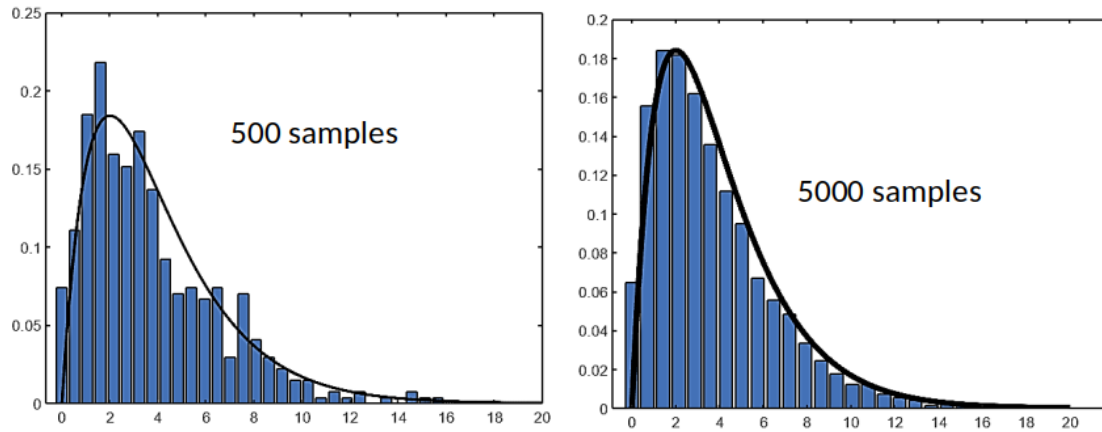
- (a) Uniform Distribution
- (b) Gaussian Distribution
- (c) Exponential Distribution
- (d) Rayleigh Distribution

Gamma Distribution

MATLAB Code:

```
1 >> alpha = 2; beta = 0.5;
2 >> xmax = 20; x = 0:0.01:xmax;
3 >> fx = beta^alpha*x.^(alpha-1).*exp(-beta*x)/gamma(alpha);
4 >> k = alpha; theta = 1/beta;
5 >> DATA = gamrnd(k, theta, 1, 5000);
6 >> BINS = 30;
7 >> [COUNTS, EDGES] = histcounts(DATA, BINS);
8 >> Ediff = diff(EDGES);
9 >> WIDTH = Ediff(1);
10 >> AREA = sum(COUNTS)*WIDTH;
11 >> figure; bar(EDGES(1:end-1), COUNTS/AREA);
12 >> hold on; plot(x, fx, 'k-', 'linewidth', 2)
```

Listing 1: Gamma Distribution



(a) Gamma Distribution with 500 Data Points

(b) Gamma Distribution with 5000 Data Points

The Gamma Distribution is a continuous probability distribution that is defined by two parameters: the shape parameter k or α , and the scale parameter θ where $\theta = \frac{1}{\beta}$. It's often used to model the time until an event occurs, such as the time until a radioactive particle decays or the time until a customer arrives at a service center.

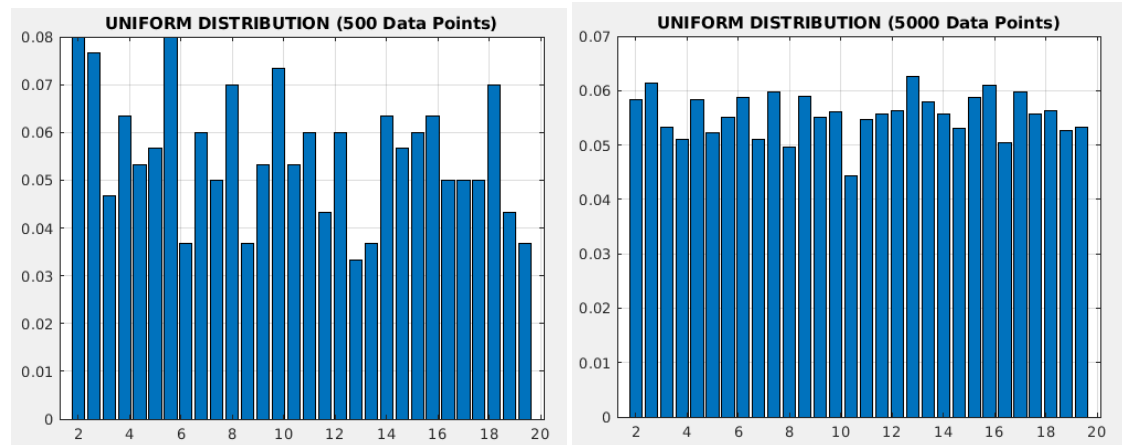
The shape of the Gamma Distribution can vary widely depending on the values of the shape and scale parameters. It can be right-skewed, left-skewed, or symmetrical. Here we can see that the Gamma Distribution is right-skewed, with a peak around 2, which was equal to the shape parameter α .

Uniform Distribution

MATLAB Code:

```
1 >> a = 2; b = 2; %Minimum and Maximum Values
2 >> samples = 500; %Change samples to 5000 for 5000 data points
3 >> DATA = a + (b - a)*rand(1, samples);
4 >> BINS = 30;
5 >> [COUNTS,EDGES] = histcounts(DATA,BINS);
6 >> Ediff = diff(EDGES);
7 >> WIDTH = Ediff(1);
8 >> AREA = sum(COUNTS)*WIDTH;
9 >> figure; bar(EDGES(1:end-1),COUNTS/AREA);
10 >> title('UNIFORM DISTRIBUTION (500 Data Points)'); grid on;
```

Listing 2: Uniform Distribution



The Uniform Distribution is a continuous probability distribution that is characterized by a constant probability density function (PDF) over a specified interval. It's defined by two parameters: the minimum value (a) and the maximum value (b). The PDF is flat within this interval and zero outside of it. The Uniform Distribution becomes more and more uniform as we increase the data samples, as can be clearly seen from the above figures which draws a comparison between the Uniform Distribution with 500 and 5000 data points.

In comparison with the Gamma Distribution, we can clearly see that the Uniform Distribution approaches a constant shape between its parameters a and b as the number of data points increases, whereas the Gamma Distribution approaches a peak value, and its shape can vary depending on its parameters.

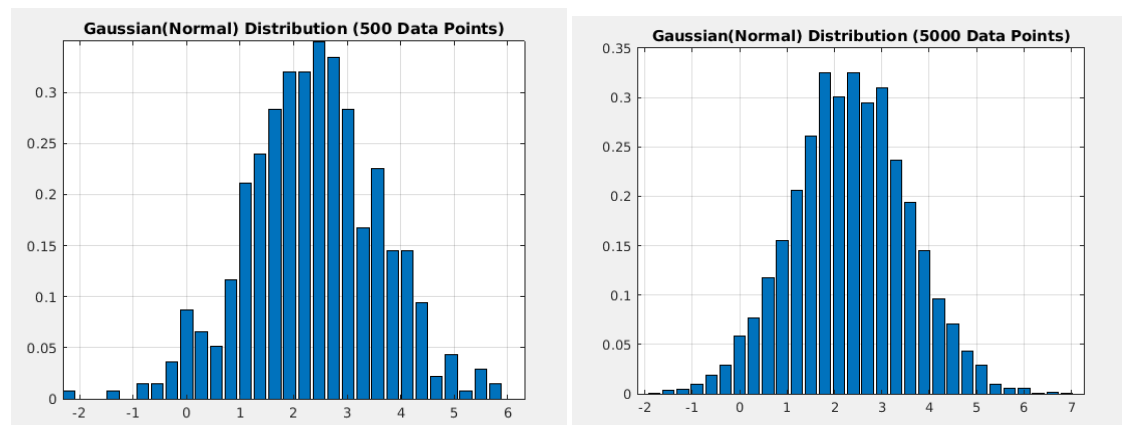
The Uniform Distribution has a fixed range $[a, b]$, which means all values within this range have equal probabilities. In contrast, the Gamma Distribution can have a variable range and is not constrained to a specific interval.

Gaussian (Normal) Distribution

MATLAB Code:

```
1 >> mu = 2.5; sigma = 1.2; %Mean and Standard Deviation
2 >> DATA = mu + sigma * randn(1, 500); %Change value of randn from 500 to
   5000 for data points
3 >> BINS = 30;
4 >> [COUNTS, EDGES] = histcounts(DATA, BINS);
5 >> Ediff=diff(EDGES);
6 >> WIDTH=Ediff(1); AREA = sum(COUNTS)*WIDTH;
7 >> figure; bar(EDGES(1:end-1), COUNTS/AREA);
8 >> title('Gaussian(Normal) Distribution (500 Data Points)');
9 >> grid on;
```

Listing 3: Gaussian (Normal) Distribution



The Gaussian Distribution is a continuous probability distribution characterized by a bell-shaped curve. It's defined by two parameters: the mean μ and the standard deviation σ . It's widely used to model data that tends to cluster around a central value with symmetrically decreasing probabilities as values move away from the mean - so the peak of the distribution occurs at its mean (expected) value. Unlike the Gamma Distribution which is not symmetrical and can take on various shapes or be skewed.

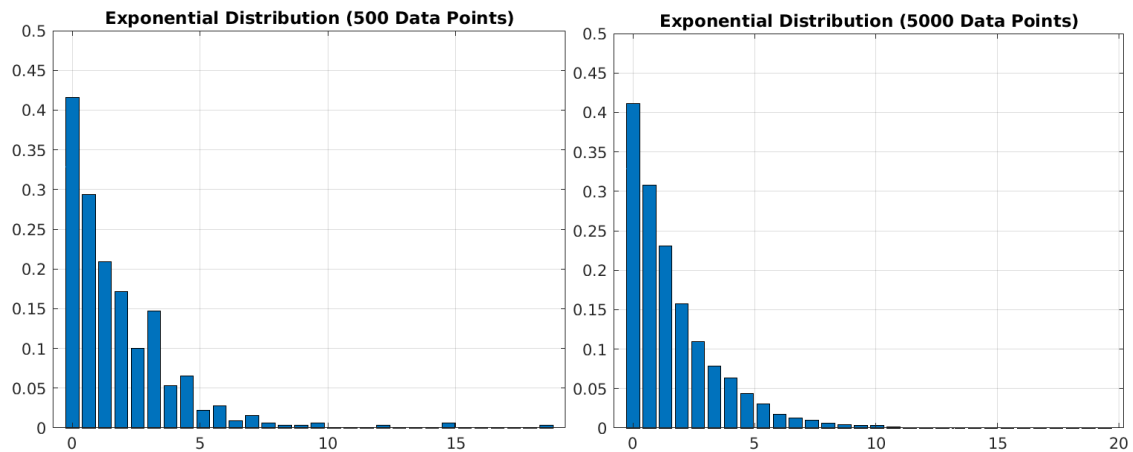
In addition, the shape of the Gaussian Distribution is primarily determined by the mean and standard deviation parameters. Changing these parameters shifts or scales the bell-shaped curve.

Exponential Distribution

MATLAB Code:

```
1 >> lambda = 0.5; %Rate Parameter
2 >> DATA = exprnd(1/lambda, 1, 500); %Change the value from 500 to 5000
   for 5000 data points
3 >> BINS = 30;
4 >> [COUNTS, EDGES] = histcounts(DATA, BINS);
5 >> Ediff=diff(EDGES);
6 >> WIDTH=Ediff(1);
7 >> AREA= sum(COUNTS)*WIDTH;
8 >> figure; bar(EDGES(1:end-1), COUNTS/AREA);
9 >> title('Exponential Distribution (500 Data Points)');
10 >> grid on;
```

Listing 4: Exponential Distribution



The Exponential Distribution is a continuous probability distribution that describes the time between events in a Poisson process, where events occur at a constant average rate λ . It's characterized by a single parameter, the rate parameter λ , which controls the rate of decay of the exponential curve. The mean of the Exponential Distribution is equal to $1/\lambda$, and the standard deviation is also equal to $1/\lambda$.

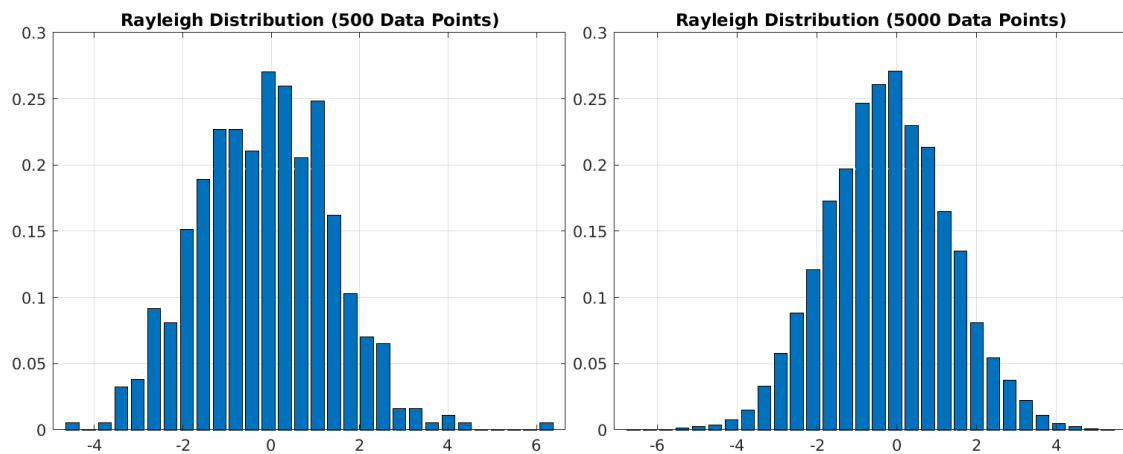
If the shape is considered, then the Exponential Distribution seems to follow a similar shape to the Gamma Distribution after its peak value. And indeed, the Exponential Distribution is a special case of the Gamma Distribution, where the shape parameter α is equal to 1. However, they both have different use cases.

Rayleigh Distribution

MATLAB Code:

```
1 >> sigma = 0.1; %Scale Parameter
2 >> DATA = sigma * randn(1, 5000); %Change value from 5000 to 500 for 500
   data points
3 >> BINS = 30;
4 >> [COUNTS,EDGES] = histcounts(DATA,BINS);
5 >> Ediff=diff(EDGES);
6 >> WIDTH=Ediff(1);
7 >> AREA=sum(COUNTS)*WIDTH;
8 >> figure; bar(EDGES(1:end-1), COUNTS/AREA);
9 >> title('Rayleigh Distribution (5000 Data Points)'); grid on;
```

Listing 5: Rayleigh Distribution



The Rayleigh Distribution is a continuous probability distribution, characterized by a single parameter σ , the scale parameter. It has a slightly right skewed, unimodal curve with a peak, commonly used to model the amplitude of signals in noise. In contrast, it looks more like the Gaussian Distribution, but it's not symmetrical, but its not completely skewed like the Gamma Distribution either.