

ASSIGNMENT-2

Submitted by- Vandit Sharma

Roll No.- 17EC10060

Topic- Frequency count

Aim-

- To plot the accuracies of the frequent count algorithms (Misra-Gries, Space Saving, and Count Min sketch) on various generated streams w.r.t. respective distribution entropy

Normal Distribution-

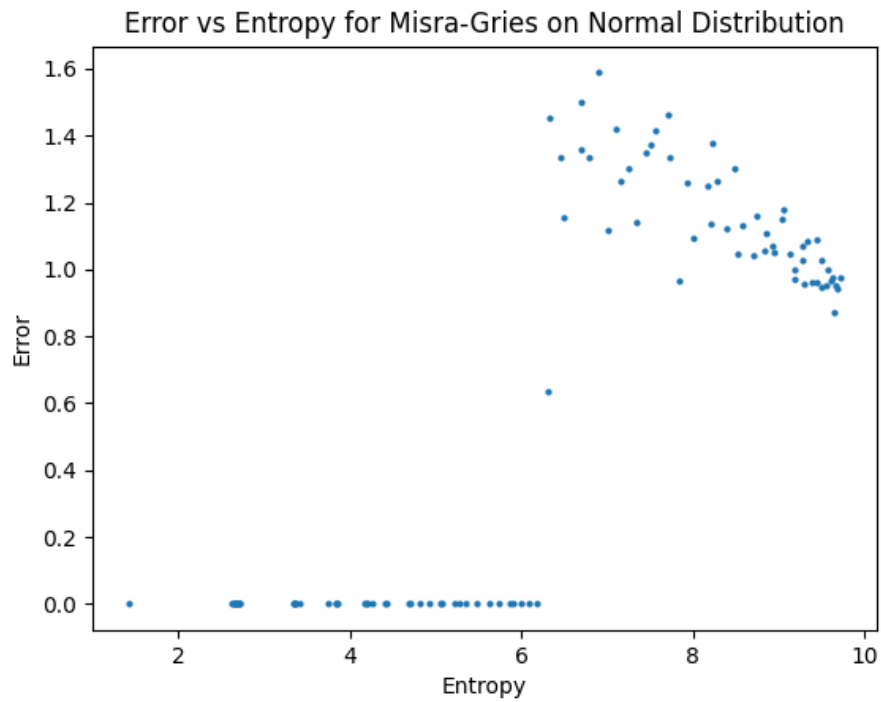
A) *Generation of Distribution*-

I used the random.normal function to generate the normal distribution, as shown below.

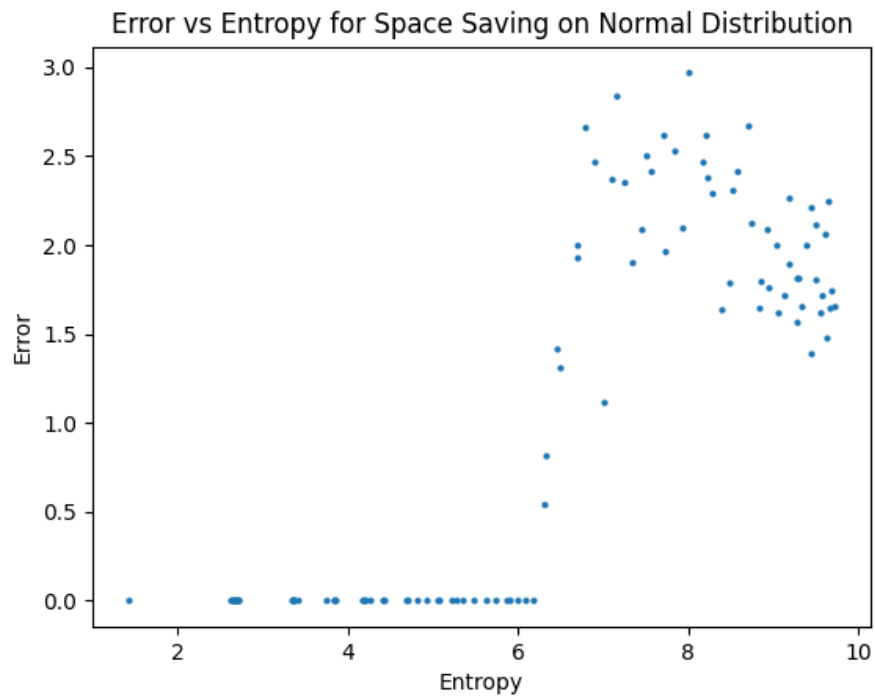
```
def gen_norm(stream_length):  
    norm = random.normal(0, 1, stream_length).astype(int)  
    return (norm)
```

For the purpose of plotting, we had to generate 100 such distributions by varying the standard deviation of the distribution. I used a non-linear logspace function (np.logspace(0.0, 3.0, num=100)) for varying the variance of the distribution, since the entropies of the generated distributions in this case were more uniformly spread, whereas in case of a linear function(range(1, 1001, 10)), the entropies were more towards the higher side, causing emptiness in the graph towards the lower range of entropies.

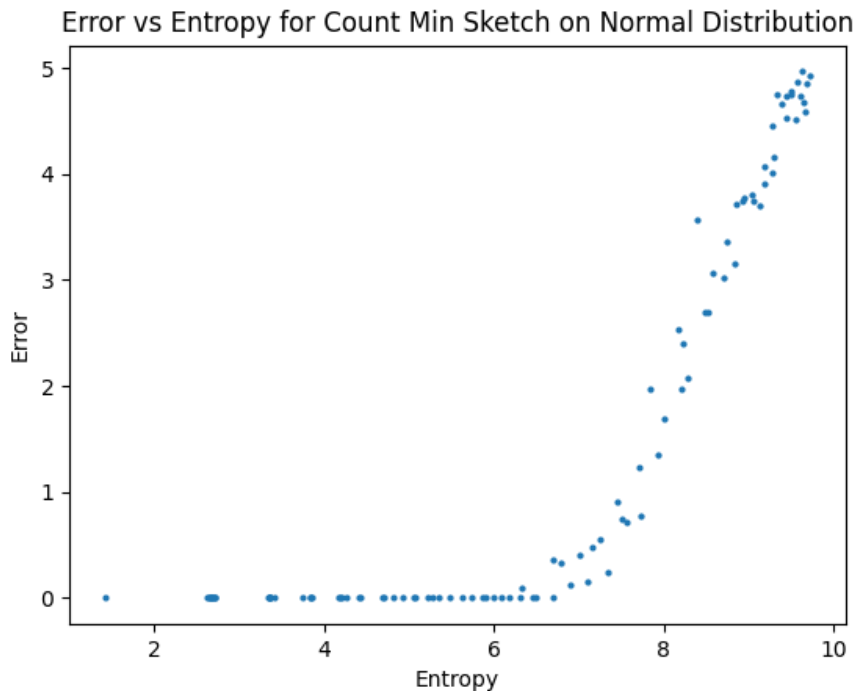
B) *Accuracy versus Entropy plots-*



Plot 1: Error vs Entropy for Misra-Gries on Normal Distribution



Plot 2: Error vs Entropy for Space Saving on Normal Distribution



Plot 3: Error vs Entropy for Count Min Sketch on Normal Distribution

Gamma Distribution-

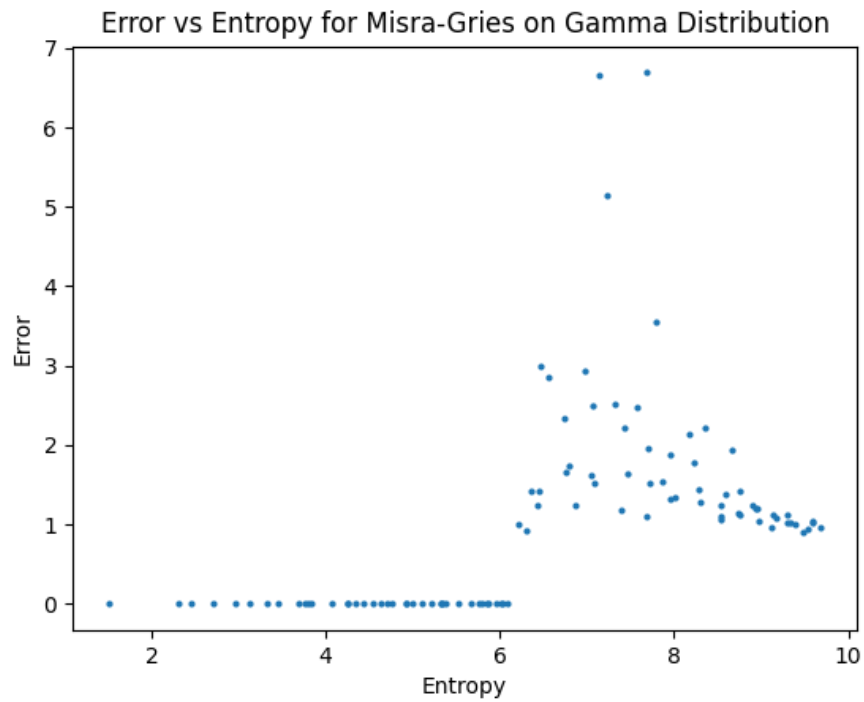
A) Generation of Distribution-

I used the `random.gamma` function to generate the normal distribution, as shown below.

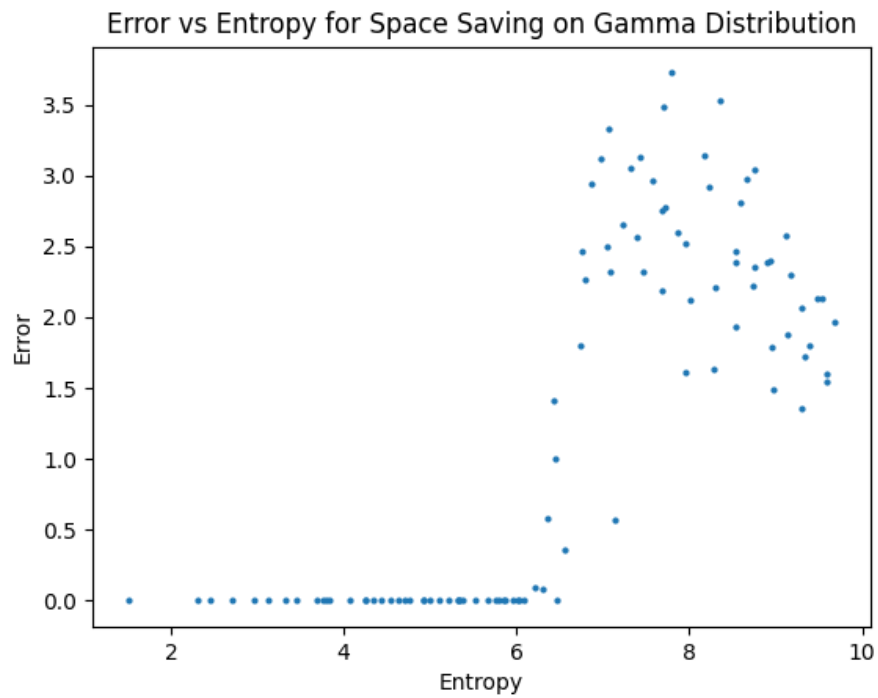
```
data sample gamma = random.gamma(i, j, stream length).astype(int)
```

For the purpose of plotting, we had to generate 100 such distributions by varying the standard deviation of the distribution. I used the non-linear `logspace` function (`np.logspace(0.0, 2.0, num=10)`) for varying both the shape and scale of the distribution, since the entropies of the generated distributions in this case were more uniformly spread, whereas in case of a linear function (`range(1, 101, 10)`), the entropies were more towards the higher side, causing emptiness in the graph towards the lower range of entropies.

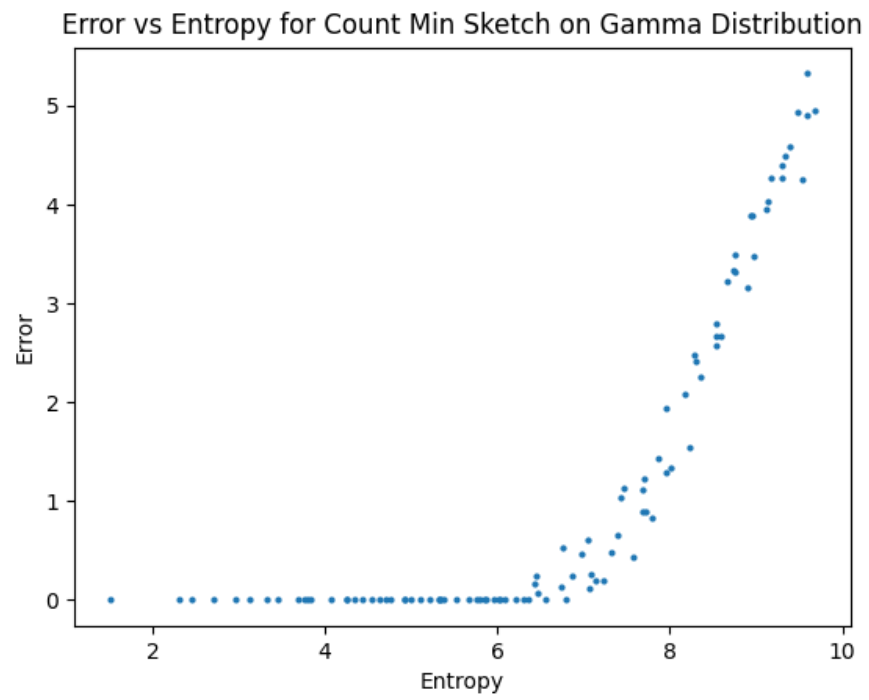
B) Accuracy versus Entropy plots-



Plot 4: Error vs Entropy for Misra-Gries on Gamma Distribution



Plot 5: Error vs Entropy for Space Saving on Gamma Distribution



Plot 6: Error vs Entropy for Count Min Sketch on Gamma Distribution