Frequency count demo

This assignment is for you to get a hands-on feel for different frequent count algorithms and their accuracy on different types of streams. We are going to measure the entropy for various randomly generated streams using their entropy, which is also a measure of "surprise" in the distribution of items in the stream.

Your task is 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.

Steps:

1. Generate different streams using the distributions: normal with different standard deviations, and discrete probability values generated by the gamma distribution. Generate 100 different streams for each distribution with a varying range of entropy values.

2. Run the stream frequent counting algorithms described above for each of the above streams and measure the difference in frequencies of the top 10% of items.

Report the above-described plots (6 total) for each type of stream (2 of them) and each streaming algorithm (3 of them). You can get the demo code here: https://colab.research.google.com/drive/1d3lrGWYcNuviGPfR0Za75SXTRJXYx7Oj?usp=sharing