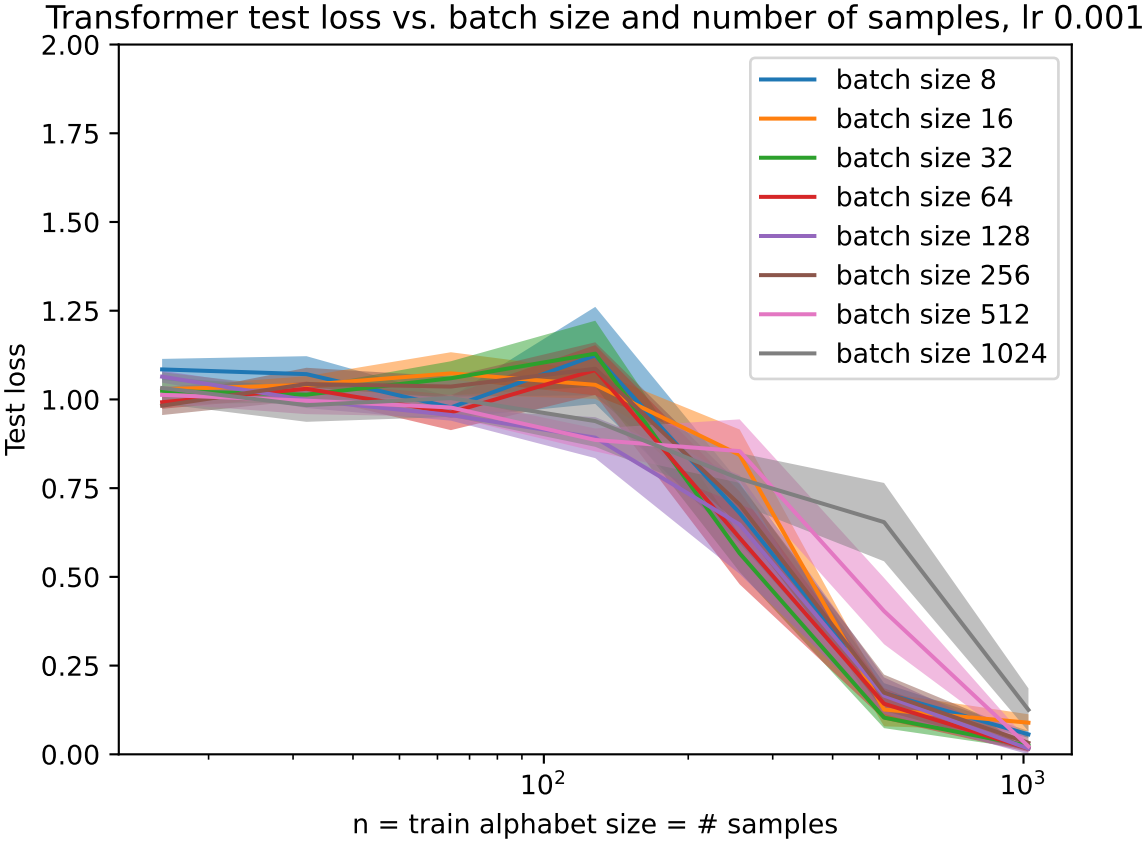
Transformer test loss vs. batch size and number of samples, lr 0.001
Test loss
n = train alphabet size = # samples
batch size 8
batch size 16
batch size 32
batch size 64
batch size 128
batch size 256
batch size 512
batch size 1024
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
10^2
10^3