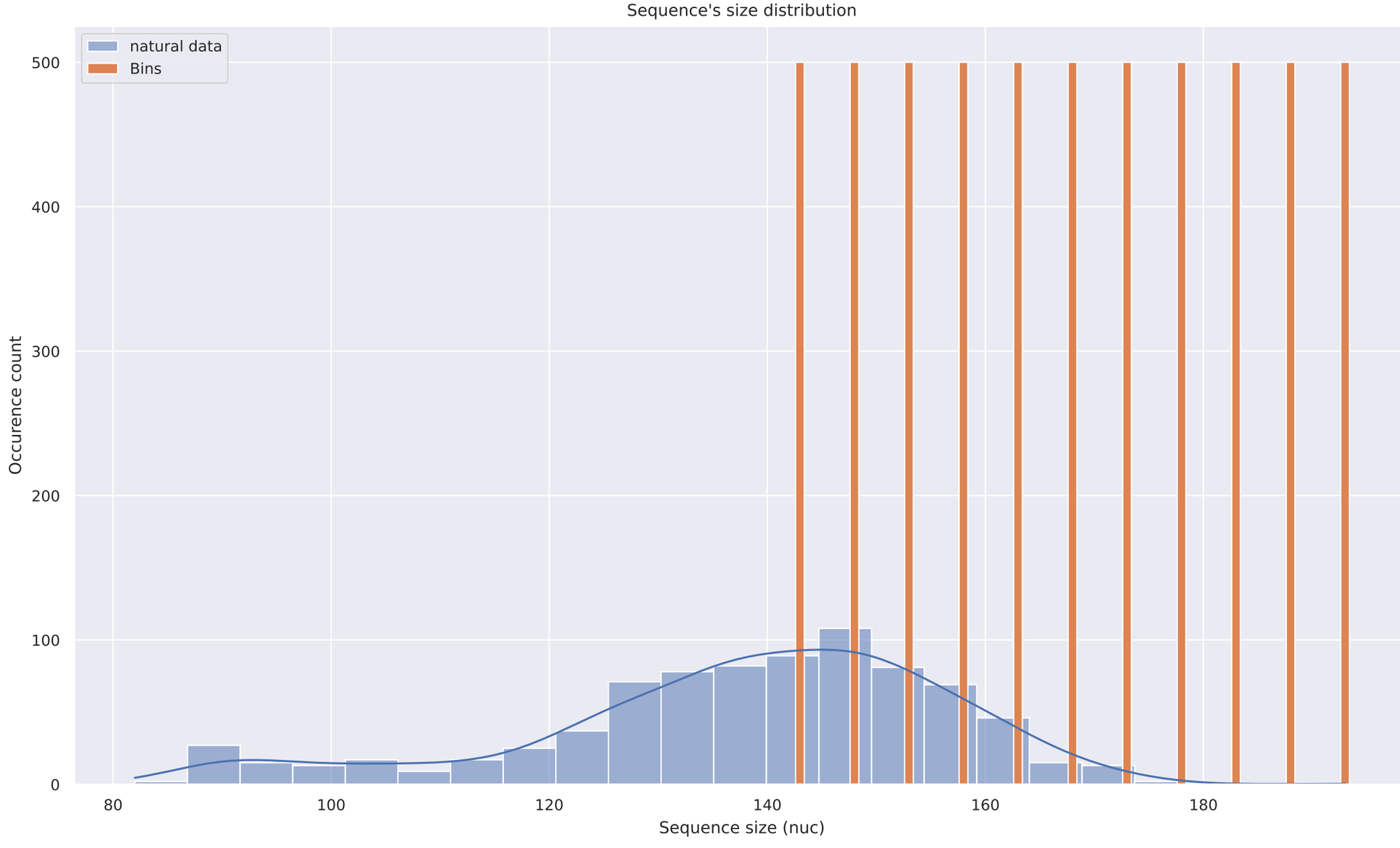
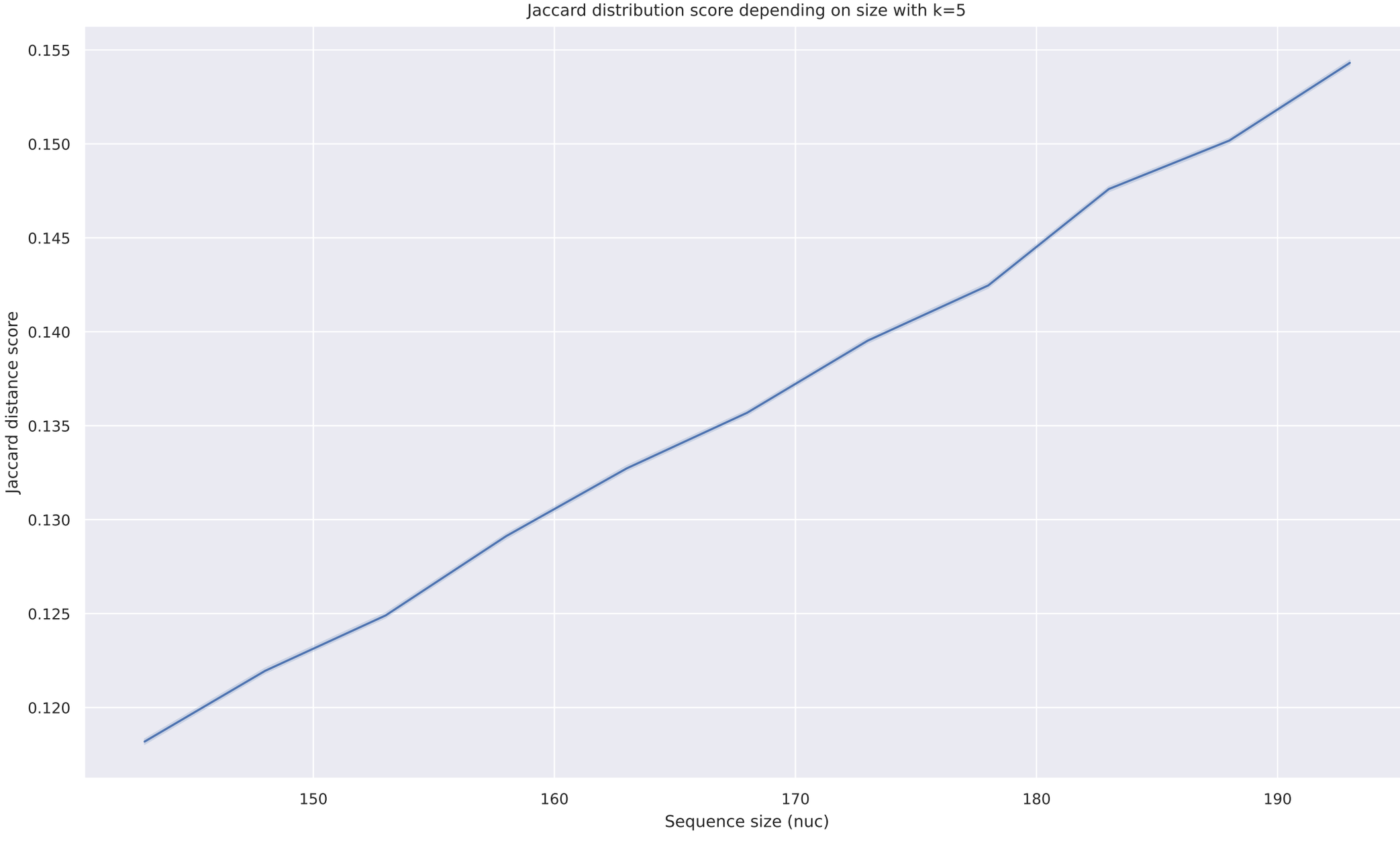
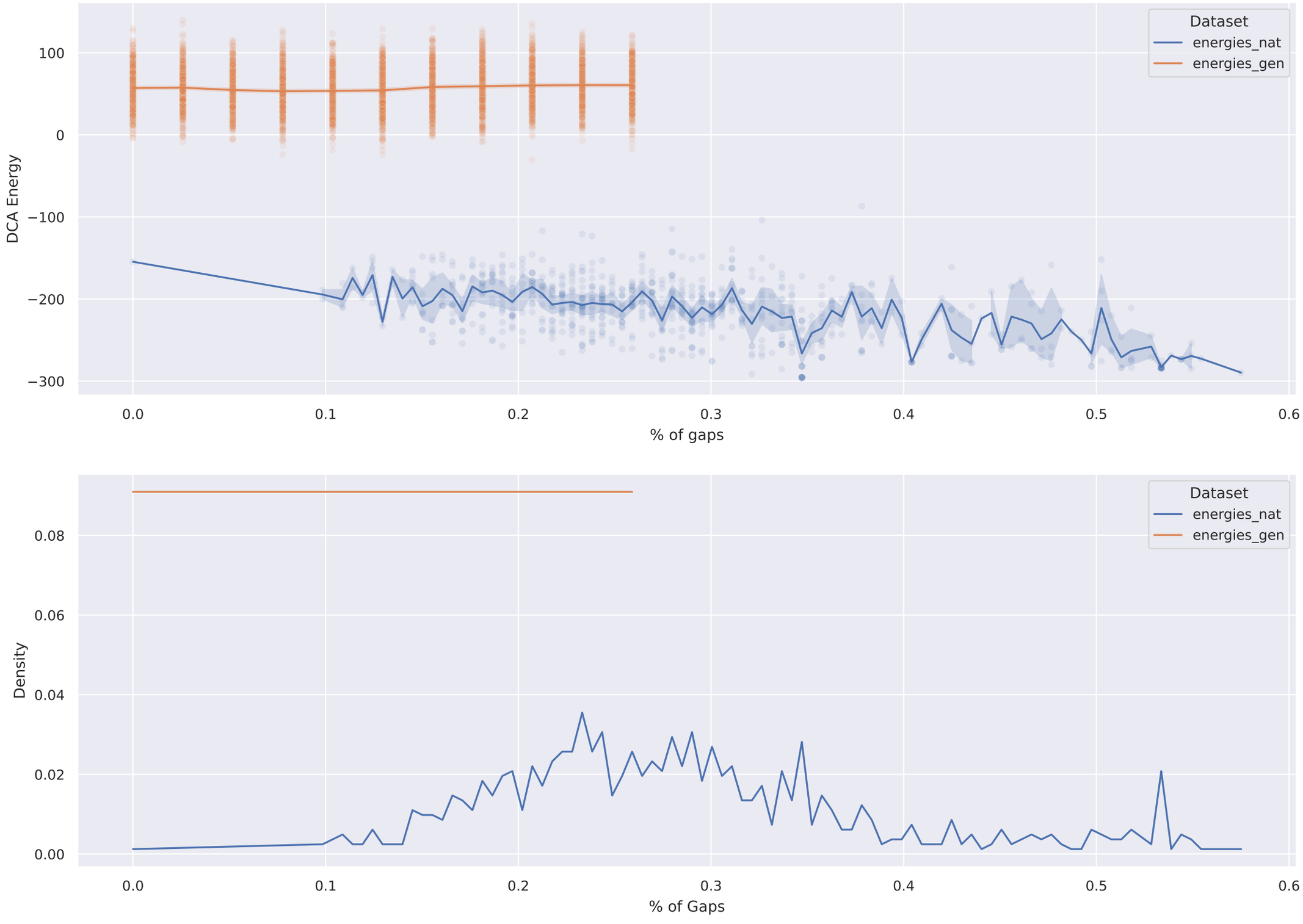
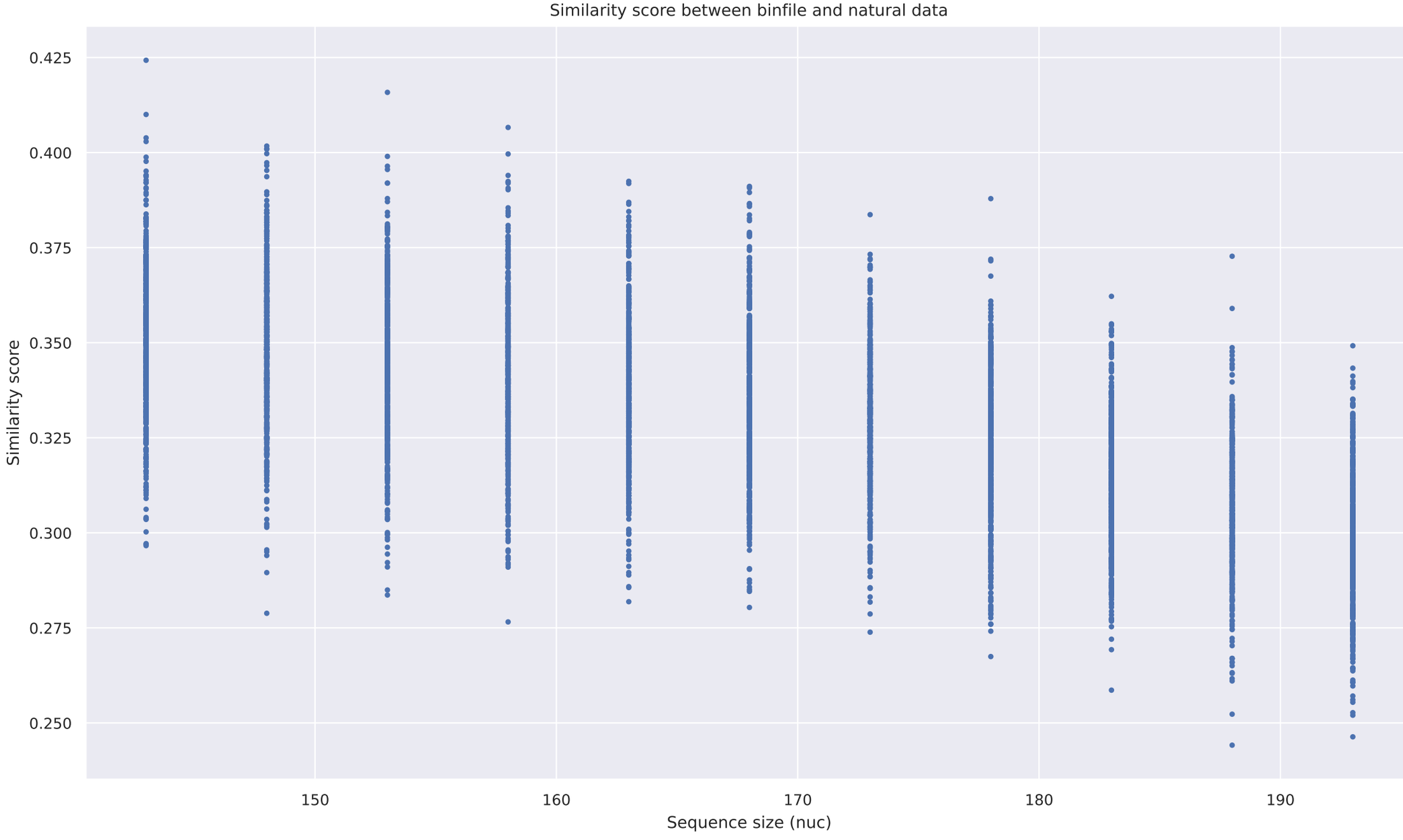

Sequence's size distribution
natural data
Bins
Occurence count
0
100
200
300
400
500
80
100
120
140
160
180
Sequence size (nuc)


Jaccard distribution score depending on size with k=5
Jaccard distance score
0.155
0.150
0.145
0.140
0.135
0.130
0.125
0.120
150
160
170
180
190
Sequence size (nuc)


Dataset
energies_nat
energies_gen
DCA Energy
100
0
−100
−200
−300
0.0
0.1
0.2
0.3
0.4
0.5
0.6
% of gaps
Dataset
energies_nat
energies_gen
Density
0.08
0.06
0.04
0.02
0.00
0.0
0.1
0.2
0.3
0.4
0.5
0.6
% of Gaps


Similarity score between binfile and natural data
Similarity score
0.425
0.400
0.375
0.350
0.325
0.300
0.275
0.250
150
160
170
180
190
Sequence size (nuc)