Small Problem 5: Probabilistic Context-Free Grammar Sentence Completion

In this problem, you are given a probabilistic context free grammar and a prefix x of a string generated by that grammar. Your goal is to evaluate the conditional probability that the grammar will generate the complete string xy for a given suffix y: $P(xy|x, G)$.

Grammar: Note that this grammar produces strings of unbounded length, but produces a finite string with probability 1.

S $\rightarrow$ AB (0.25)
S $\rightarrow$ BC (0.2)
S $\rightarrow$ AC (0.4)
S $\rightarrow$ CA (0.15)
A $\rightarrow$ a (0.05)
A $\rightarrow$ b (0.3)
A $\rightarrow$ S (0.65)
B $\rightarrow$ b (0.5)
B $\rightarrow$ c (0.3)
B $\rightarrow$ d (0.2)
C $\rightarrow$ d (0.35)
C $\rightarrow$ e (0.1)
C $\rightarrow$ S (0.55)

Query 1: What is the conditional probability that a string beginning with “bd” terminates with the complete string “bdcb”: $P(y = \text{"bdcb"}|\text{prefix}(y) = \text{"bd"})$ where y is the complete string that is generated by S.

Query 2: What is $P(y = \text{"cddb"}|\text{prefix}(y) = \text{"cd"})$?

Metric: Square of the difference in negative log probability (“surprise”) between the true and the computed conditional probability. $\left(-\log P(xy|x) + \log \hat{P}(xy|x)\right)^2$.