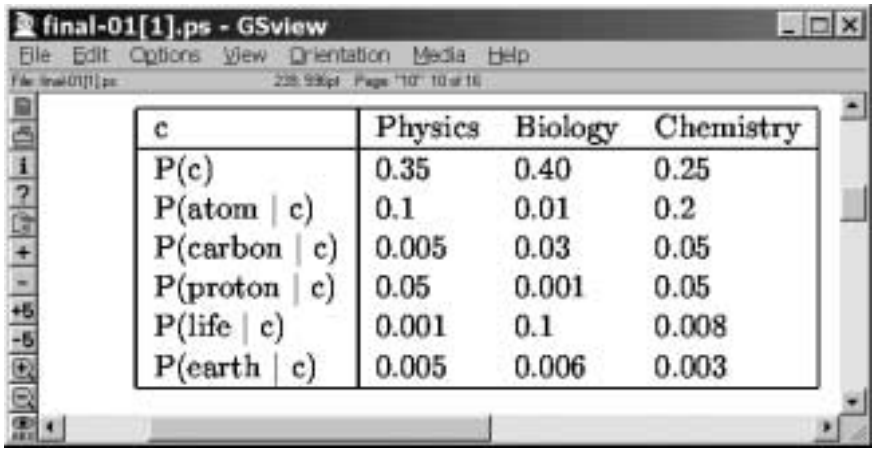
Qn 1. [10pt] [Mooney] Assume we want to classify science texts into three categories—physics, biology and chemistry. The following probabilities have been estimated from analyzing a corpus of pre-classified web-pages gathered from Yahoo.



c	Physics	Biology	Chemistry
P(c)	0.35	0.40	0.25
P(atom | c)	0.1	0.01	0.2
P(carbon | c)	0.005	0.03	0.05
P(proton | c)	0.05	0.001	0.05
P(life | c)	0.001	0.1	0.008
P(earth | c)	0.005	0.006	0.003

Assuming that the probability of each evidence word is independent of other word occurrences given the category of the text, compute the (posterior) probability for each of the possible categories each of the following short texts; and based on that, their most likely classification. Assume that the categories are disjoint and exhaustive (i.e., every text is either physics, or biology or chemistry and no text can be more than one). Assume that words are first stemmed to reduce them to their base form (atoms→atom) and ignore any words that are not in the table:

A: ***the carbon atom is the foundation of life on earth.***
B. the carbon atom contains 12 protons.