Project 3: Mutation

Juan, Patrick and John

October 28, 2019

1 Problem Formalization

Let r be one particular rule of a mutagen, then r can be represented as

$$r = p_0; p_1; p_2; p_3; p_4; p_5; p_6; p_7; p_8; p_9@a_0a_1a_2a_3a_4a_5a_6a_7a_8a_9, \tag{1}$$

where each $p_i \in \{\mathtt{a}, \mathtt{c}, \mathtt{g}, \mathtt{t}\}^*$ and $a_i \in \{\mathtt{a}, \mathtt{c}, \mathtt{g}, \mathtt{t}, \mathtt{0}, \mathtt{1}, \mathtt{2}, \mathtt{3}, \mathtt{4}, \mathtt{5}, \mathtt{6}, \mathtt{7}, \mathtt{8}, \mathtt{9}\}$.
Let R represent a random variable that takes rules as values. Similarly, let P_i and A_i be variables representing, respectively, individual elements of the pattern and action part of the rule. For simplicity let $P = \{P_i\}$, $A = \{A_i\}$ for $i = 0, \ldots, 9$. Lowercase letters p_i, a_i, p, a represent a particular pattern element, action element, a whole pattern and a whole action, respectively.
Consider a string $S = \{s_i\}_{i=0,\ldots,9}$, $s_i \in \{\mathtt{a}, \mathtt{c}, \mathtt{g}, \mathtt{t}\}$. If every s_i matches the corresponding pattern position p_i, the rule's action can be applied. Let M_i, $i = 0, \ldots, 9$ be indicator variables of these individual matches. Moreover, let M be an indicator for the event $\forall_i M_i = 1$, so that $M = 1$ if the string matches the pattern P and 0 otherwise.
It follow that, if $M = 1$, the original string S can be mutated into a string $T = \{t_i\}_{i=0,\ldots,9}$, $t_i \in \{\mathtt{a}, \mathtt{c}, \mathtt{g}, \mathtt{t}\}$ based on the rule's action A. If $M = 0$, then $T = S$.
Finally, let N be an indicator variable such that $N = 1$ if R is minimal and $N = 0$ otherwise.
Based on the process described above, the causal process by which S is transformed into T, can be represented with the graph depicted in Figure 1.

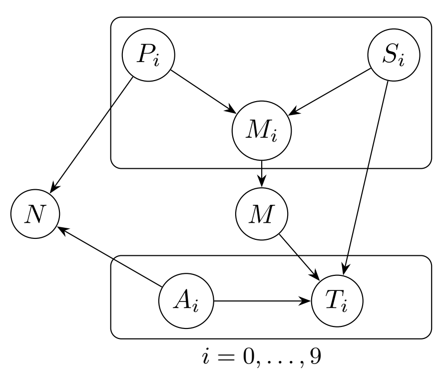

Figure 1: Graphical model representing the causal process by which a mutagen rule transforms a string S into a string T. Each box in the graph is repeated for $i = 0, \ldots, 9$, and arrows crossing boxes' boundaries should be considered as being from/to each variable with subscript i.

This is not only a curious picture, but it encodes important constraints of the process that are greatly helpful in solving the problem at hand, as we will see in the next section. Formally speaking, this is causal diagram can be though of as a Bayesian Network, and conditional independencies among variables can be read using the *d-separation* criterion [Koller and Friedman, 2009, Pearl, 2000].

References

[Koller and Friedman, 2009] Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*, volume 2009.

[Pearl, 2000] Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, NY, USA, 2nd edition.