HW # 1 Solution to part C

Let $D = \{A, C, G, T\}$. At any point in time a DNA segment ℓ of length m is an element of

$$D^m = \underbrace{D \times \cdots \times D}_{m \text{ times}} = \{\ell : \ell = (\ell_1, \ldots, \ell_m),\ \ell_i \in D\}$$

(m could be zero here, in which case D^m is just the empty string). Since insertions and deletions are allowed m may vary randomly. Let $S = \bigcup_{m=0}^{\infty} D^m$. This is the most general state space.

Note: An alternative representation of the state space is

$$S = \{\ell : \ell = (\ell_1, \ell_2, \ldots) \text{ and } \forall \ell\ \exists\ j(\ell) \geq 1 \text{ s.t. } \forall\ j \geq j(\ell)\ \ell_j = 0 \text{ and for } j < j(\ell)\ \ell_j \in D\}$$

In the above setting $\ell_j = 0$ means that the jth slot is empty.

The assumption above that $\ell_j \in D$ for $j < j(\ell)$ and $\ell_j = 0$ for $j \geq j(\ell)$ is essential. This assumption ensures that each DNA segment has a unique representation.