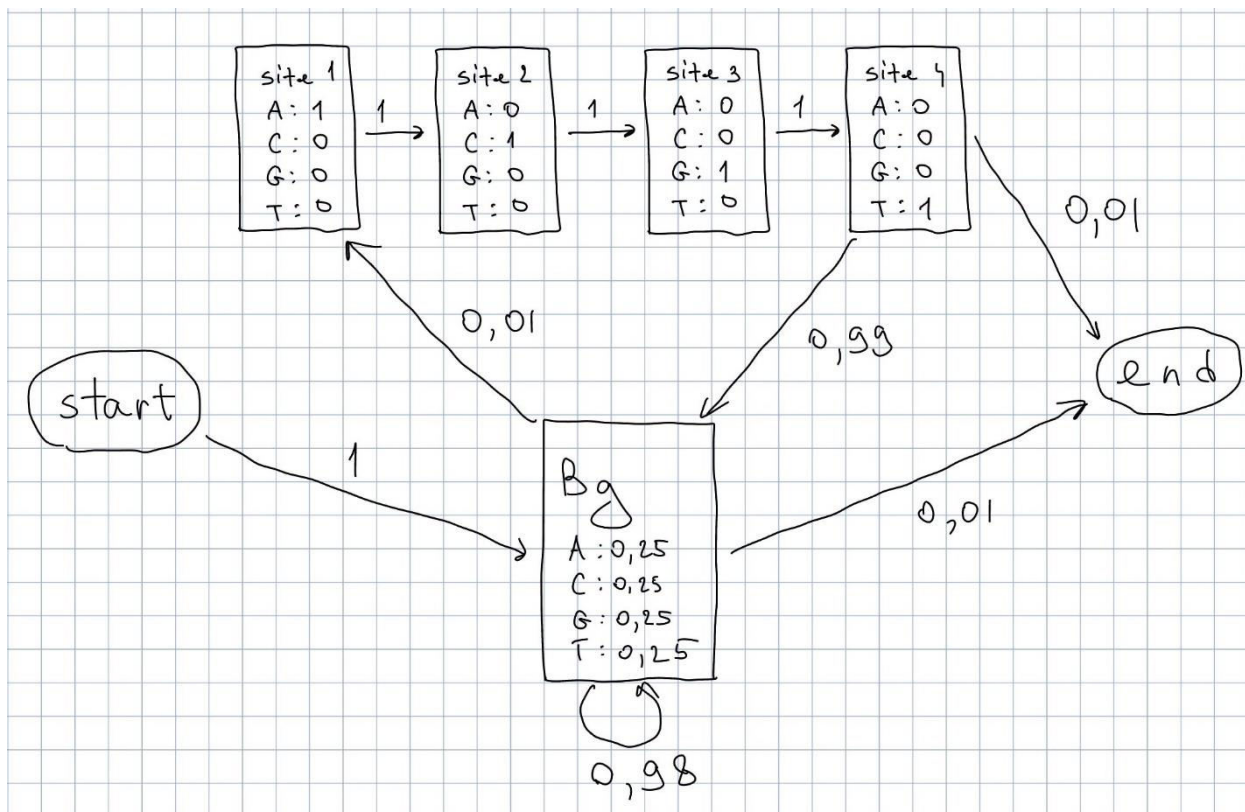


Автомат для модели:



Алгоритм Витерби:

Модель была протестирована на данной в ДЗ последовательности.

Модель с вероятностью перехода Background \rightarrow Site = 0.01 нашла все 3 сайта, и при этом не нашла ни одного ложного сайта (т.е. во всех остальных нуклеотидах состояние В – background)

Модель с вероятностью перехода Background \rightarrow Site = 0.001 не нашла ни одного сайта – всю последовательность разметила как Bg.

[illegible]

Forward-Backward:

Алгоритм был протестирован на данной в ДЗ последовательности для моделей со значениями $p_{site} = 0.01, 0.001, 0.05, 0.5$

Независимо от значения вероятности перехода $Bg \rightarrow Site$, вне сайта рестрикции правдоподобия нахождения модели в состоянии Bg всегда равны 1, а в состоянии Site равны 0.

В зависимости от значения p_{site} меняется “уверенность” модели в правдоподобии сайта. Чем выше p_{site} , тем больше правдоподобие нахождения модели в состоянии Site (и, соответственно меньше в Bg) в настоящем сайте рестрикции. При $p_{site} = 0.001$ в сайтах

рестрикции правдоподобие все же больше для состояния Bg. То есть порог p_{site} , при котором модель начинает быть “уверена” в сайте лежит где-то между 0.001 и 0.01.



Генерация последовательности моделью:

Генерация последовательности реализована в двух вариантах – с заданной длиной и со случайной длиной, зависящей от вероятностей перехода в состояние END.

Проводилась генерация последовательностей моделями с разной заданной вероятностью сайта, а затем оценивалась частота переходов BG → SITE на заданных последовательностях

Последовательности длины 1000:

```
Model with p_site: 0.01, p_site assessed from generated sequences: 0.01361914064208998 ± 0.003942830865226739
Model with p_site: 0.1, p_site assessed from generated sequences: 0.1055659522959163 ± 0.011549068778566846
Model with p_site: 0.5, p_site assessed from generated sequences: 0.5038984895501988 ± 0.027072931728099797
```

Последовательности случайной длины:

```
Model with p_site: 0.01, p_site assessed from generated sequences: 0.013380240750248453 ± 0.0033288125047799214
Model with p_site: 0.1, p_site assessed from generated sequences: 0.1044355694586851 ± 0.012330381683173618
Model with p_site: 0.5, p_site assessed from generated sequences: 0.5046027461292348 ± 0.027338335118119932
```

Распределение длин экспоненциальное, как и ожидалось:

Distribution of the lengths of generated sequences

