

Algorithmische Bioinformatik Übungsblatt 7

Ausgabe: 17. Dezember 2019 · Besprechung: 07. Januar 2020

Aufgabe 7.1 Bestimme die Verteilung des Minimums nach 5-maligem Würfeln mit einem 6-seitigen fairen Würfel. Gib formal den zugehörigen PAA an.

Aufgabe 7.2 Erinnere dich an die Medizinstudentin von Blatt 2 und die damals ungelöste Frage.

Eine Medizinstudentin schreibt eine Klausur, die 300 Minuten dauert und aus 300 Aufgaben besteht. Tatsächlich löst sie pro Minute genau eine Aufgabe. Dabei ändert sich ihr Zustand zwischen erschöpft, wach und hochkonzentriert; sie beginnt die Prüfung wach. Abhängig von ihrem Zustand beantwortet sie jede Prüfungsfrage mit einer gewissen Wahrscheinlichkeit richtig oder falsch. Ist sie erschöpft, liegt ihre Fehlerwahrscheinlichkeit bei 50%; ist sie wach, bei 10%; ist sie hochkonzentriert, bei nur 2%. Ihr Zustand kann sich von Minute zu Minute ändern, bleibt aber zu 90% gleich. Ist sie hochkonzentriert, ist sie in der nächsten Minute mit 10% Wahrscheinlichkeit nur noch wach. Ist sie wach, dann ist sie in der nächsten Minuten mit je 5% in einem der anderen Zustände. Ist sie erschöpft, dann ist sie in der nächsten Minute mit 10% Wahrscheinlichkeit wieder wach.

Berechne die Wahrscheinlichkeit, dass sie mindestens 200 der 300 Aufgaben richtig beantwortet, und gib den zugehörigen PAA an.

Aufgabe 7.3 Auf einer Kreuzfahrt geratet ihr in die Fänge von Piraten. Diese spielen ein grausames Würfelspiel, indem sie mit einem sechsseitigen Würfel, dessen Flächen mit LIVEDA beschriftet sind, spielen und die sich ergebende Sequenz notieren, bis die vier letzten Würfe (also das Suffix des sich entwickelnden Textes) entweder LIVE oder DEAD lautet. Entsprechend werdet ihr freigelassen oder gegessen (es sind Kannibalen-Piraten!). Das Würfelspiel dauert so lange, bis eine Entscheidung gefallen ist. Welche Überlebenschance habt ihr, wenn der Würfel fair ist und die Piraten nicht schummeln? Wie sieht eine mögliche PAA-Konstruktion aus, um die Überlebenschance zu berechnen? Wie muss der PAA modifiziert werden, um die Verteilung der Dauer des Spiels bis zu einer Entscheidung zu berechnen?

Aufgabe 7.4 Wir betrachten das Genom von *Corynebacterium diphtheriae* (Diphtherie-Erreger). Es steht auf der Webseite zur Vorlesung als .fasta-Datei zur Verfügung; der Original-Datenbankeintrag kann unter <http://www.ncbi.nlm.nih.gov/nuccore/38231477> gefunden werden.

Wir wollen ein Markov-Modell zweiter Ordnung (M2) aus dem Genom schätzen; dies ist äquivalent zum Zählen aller 3-Gramme (und geeigneter Normalisierung). Dafür können wir (entweder selbst Software schreiben oder) MoSDi (Motif Search and Discovery) verwenden: <https://bitbucket.org/tobiasmarschall/mosdi>. MoSDi ist ein Software-Paket und besteht aus mehreren Teilprogrammen.

Der Aufruf

```
mosdi-utils --help
```

erklärt die Benutzung. Der Aufruf

```
mosdi-utils count-qgrams -r sequence.fasta 3
```

liefert die gewünschten Zahlen symmetrisch bezüglich des reversen Komplements. Die resultierende Ausgabe kann in eine Datei umgeleitet und für weitere Schritte verwendet werden. Zusätzlich benötigen wir auch das entsprechende M0-Modell (1-Gramme), das sich analog erhalten lässt.

Aufgabe 7.5 Für zukünftige Bewerbungen und die Planung des Berufswegs gibt es nützliche Tipps von Heiko Mell über die VDI Nachrichten, und zwar unter <https://www.ingenieur.de/karriere/arbeitsleben/heiko-mell/>