



Blatt 1

Aufgabe 1.1

- Berechnen Sie für die diskrete und die kontinuierliche Gleichverteilung, $P(x) = \frac{1}{b-a}$, $x \in [a, b]$, jeweils Erwartungswert und Varianz.
- Generieren Sie verschiedene Listen mit kontinuierlich gleichverteilten Zufallszahlen, wobei die Anzahl der Elemente in den Listen und die Parameter der Gleichverteilung jeweils variieren. Stellen Sie die resultierenden Verteilungen in einem Histogramm dar und berechnen Sie die Mittelwerte der einzelnen Verteilungen. Vergleichen Sie diese mit den errechneten Erwartungswerten.

Aufgabe 1.2

Beweisen Sie für beliebige diskrete Zufallsvariablen X und Y

$$E[X + Y] = E[X] + E[Y]$$

Aufgabe 1.3 Wir betrachten ein Genom, das mehrere tausend Replikationen eines Sequenzstückes enthält, welches für 5.8 S rRNA codiert. Diese Sequenzen sind aufgrund evolutionärer Ereignisse unterschiedlich, einige haben auch ihre Funktionalität verloren, sind aber im Genom erhalten geblieben.

Wir interessieren uns für einen Teil dieser Sequenzen und wollen diese klonieren, um sie anschließend sequenzieren zu können. Zunächst nutzen wir PCR, da wir ein Primer-Paar kennen, das alle Sequenzen der 5.8 S rRNA codierenden Bereiche flankiert. Das PCR-Produkt wird mittels Gelelektrophorese bezüglich der Länge der Sequenzen getrennt, und im weiteren untersuchen wir eine der entstandenen Bande. Wir nehmen an, daß sie einen gewissen Anteil an funktionalen und nicht-funktionalen Sequenzen enthält. Diese stammen von a bzw. b Replikaten im Genom, die jeweils funktionale bzw. nicht-funktionale 5.8 S rRNA codieren.

Aus dieser Bande klonieren wir nun N (individuelle) Moleküle und erhalten so eine clone library mit N clonen. Diesen Prozeß modellieren wir als Ziehen einer Stichprobe **mit** Zurücklegen.

- Diskutieren Sie die geschilderte Modellierung.
- Wieviele Sequenzen für funktionale und nicht-funktionale 5.8 S rRNA haben wir im Mittel in unserer clone library?

- (c) Wie groß müssen wir N wählen, um mit Wahrscheinlichkeit ϕ mindestens je eine funktionale und nicht-funktionale 5.8 S rRNA in der clone library zu haben?
- (d) Führen sie für $a = 900$ und $b = 100$ Simulationen mit $N = 10, 100, 1000, 10000$ durch und diskutieren Sie die Ergebnisse insbesondere im Vergleich zu den theoretisch bestimmten Resultaten.

Hinweise zu praktischen Aufgaben (mehr Infos in der ersten Übung):

- Zur Lösung der praktischen Aufgaben, die im Verlauf des Semesters gestellt werden, eignen sich insbesondere Mathematik-Pakete wie *octave*, *Matlab* oder *R*. Darüber hinaus zulässig (aber nicht unbedingt empfehlenswert. . .) sind die Programmiersprachen C, C++ und Java, außerdem Perl, Python und Ruby.
- Falls gefordert, können gut dokumentierte Programme (Quellcode!) und Skripte per Email an `birgit.moeller@informatik.uni-halle.de` abgegeben werden. Wenn Programme und Skripte nicht intuitiv compilierbar/ausführbar sind, sollte eine kurze "Gebrauchsanleitung" mitgeliefert werden.
- Die Programme sollten grundsätzlich unter Linux/Unix lauffähig sein und insbesondere keine zusätzlichen Bibliotheken benötigen, die nicht in jeder Standardinstallation vorhanden sind.