



Blatt 2

Aufgabe 2.1 Berechnen Sie von der geometrischen Verteilung und der Poisson Verteilung jeweils den Erwartungswert und die Varianz.

Aufgabe 2.2 $N = 300$ verschiedene Markergene wurden mit einer BAC Bibliothek aus $M = 5 \times 10^4$ BACs hybridisiert. Dabei stellte sich heraus, dass jedes Markergen mit genau einem BAC hybridisierte, dass aber nicht alle 300, sondern nur $L = 293$, BACs verschieden sind. Dies erscheint auf den ersten Blick überraschend wenig und könnte darauf hindeuten, dass die Gene nicht gleichmäßig über alle BACs verteilt sind. Um dies genauer zu untersuchen, führen wir Simulationen und Berechnungen unter den folgenden Modellannahmen durch:

Alle $M = 5 \times 10^4$ BACs sind gleich lang und überdecken das Genom vollständig und nichtüberlappend. Die Wahrscheinlichkeit, dass Markergen n auf BAC m liegt, ist gleich $1/M$ für alle $n = 1, 2, \dots, N$ und alle $m = 1, 2, \dots, M$.

- Führen Sie unter diesen Modellannahmen Simulationen durch und erstellen Sie ein Histogramm der Anzahl der hybridisierten BACs. Wie groß ist die Wahrscheinlichkeit, genau L hybridisierte BACs zu finden für $L = 293, 294, \dots, 300$? Wie groß ist die Wahrscheinlichkeit, $L = 293$ oder weniger hybridisierte BACs zu finden?
- Wie groß ist in der Simulation der Erwartungswert der hybridisierten BACs? Berechnen Sie den Erwartungswert analytisch, und vergleichen Sie das analytische mit dem simulierten Ergebnis.
- Berechnen Sie die Wahrscheinlichkeit analytisch, genau L hybridisierte BACs zu finden für $L = 300$ und $L = 299$. (Zusatzaufgabe: Berechnen Sie die Wahrscheinlichkeit analytisch, genau $L = 298$ hybridisierte BACs zu finden.)
- Approximieren Sie die Wahrscheinlichkeitsverteilung von $M - L$ nun durch eine Poisson Verteilung mit Parameter $\lambda = M \exp(-N/M)$. Berechnen Sie die Wahrscheinlichkeit, genau L hybridisierte BACs zu finden für $L = 293, 294, \dots, 300$? Vergleichen Sie diese approximierten Wahrscheinlichkeiten mit denen aus der Simulation.
- Vergleichen Sie graphisch die Poisson Verteilung mit dem aus den Simulationen gewonnenen Histogramm. Wie gut ist die Approximation?

- f) Was lässt sich aus diesen Simulationen und Berechnungen über die Verteilung der Markergene über die BACs aussagen?

Aufgabe 2.3 Nehmen Sie nun an, dass es zwei Klassen von BACs gibt. Die BACs aus Klasse 1 enthalten keine Gene. Die BACs aus Klasse 2 enthalten Gene, und die Gene sind gleichmäßig über alle BACs aus Klasse 2 verteilt. Unbekannt ist lediglich, wie viele BACs in Klasse 2 liegen. Bestimmen Sie nun die Anzahl der BACs in Klasse 2 so, dass der Erwartungswert von L mit dem experimentell beobachteten Wert von $L = 293$ übereinstimmt. Was lässt sich aus diesem Ergebnis über die Verteilung der Markergene über die BACs aussagen?