



Blatt 4

Aufgabe 4.1 In Aufgabe 3.1 haben wir uns mit der Häufigkeit von **nichtüberlap-penden** Trimeren in Binärsequenzen der Länge $N = 300$ beschäftigt. Berechnen Sie nun analytisch für jedes Trimer die Verteilung P_1 sowie den Erwartungswert und die Varianz von $N_{ijk}^{(1)}$ und D_{ijk} . Vergleichen Sie diese Ergebnisse mit denen der Simulation aus Aufgabe 3.1.

Aufgabe 4.2 In Aufgabe 3.2 haben wir uns mit der Häufigkeit von **überlappenden** Trimeren in Binärsequenzen beschäftigt. Berechnen Sie nun analytisch für jedes Trimer die Verteilung P_1 sowie den Erwartungswert und die Varianz von $N_{ijk}^{(1)}$ und D_{ijk} für Binärsequenzen der Länge $N = 3, 4, 5, 6$ (statt Binärsequenzen der Länge $N = 102$ wie in Aufgabe 3.2). Vergleichen Sie die 8 Verteilungen P_1 der Trimere untereinander und diskutieren Sie die Unterschiede. Wie stark variiert der Erwartungswert von Trimer zu Trimer? Wie stark variiert die Varianz von Trimer zu Trimer? Für welche Trimere ist die Varianz am größten/kleinsten? Stellen Sie eine Vermutung auf, wie der Erwartungswert und die Varianz von N abhängen könnten. Vergleichen Sie alle in Aufgabe 4.2 gewonnenen Ergebnisse mit denen aus Aufgabe 4.1.

Aufgabe 4.3 Beweisen Sie für beliebige diskrete Zufallsvariablen X_1, X_2, \dots, X_M und Y_1, Y_2, \dots, Y_N die Identitäten

$$\begin{aligned} \text{Cov}(X_1, X_2) &= E(X_1 \cdot X_2) - E(X_1) \cdot E(X_2) \\ E\left(\sum_{m=1}^M X_m\right) &= \sum_{m=1}^M E(X_m) \\ \text{Var}\left(\sum_{m=1}^M X_m\right) &= \sum_{m=1}^M \text{Var}(X_m) + 2 \sum_{m=1}^M \sum_{\ell=1}^{m-1} \text{Cov}(X_m, X_\ell) \\ \text{Cov}\left(\sum_{m=1}^M X_m, \sum_{n=1}^N Y_n\right) &= \sum_{m=1}^M \sum_{n=1}^N \text{Cov}(X_m, Y_n). \end{aligned}$$