Übungen zur Vorlesung
„Stochastik für Studierende der Informatik“

http://www.stochastik.uni-freiburg.de/lehre/SS-2016/VorStochInfoSS2016/InfoVorStochInfoSS2016

Sommersemester 2016, Blatt 6
Abgabetermin: 06.06.2016, zu Beginn der Vorlesung
(Bitte geben Sie auf jedem Lösungsblatt Ihren Namen und Ihre Übungsgruppe an)
Bitte nur maximal zu zweit abgeben!

Aufgabe 21 (Mittlere absolute Abweichung) (3 Punkte)
Wir definieren die mittlere absolute Abweichung einer reellwertigen Zufallsvariable X mit existierendem Erwartungswert $\mathbf{E}[X] = \mu_X$ durch

$$\mathbf{M}[X] := \mathbf{E}[|X - \mu_X|].$$

(a) Berechnen Sie für $Y \sim U([0,1])$ den Wert $\mathbf{M}[Y]$.

(b) Es sei $p = 1/2$, $n \in \mathbb{N}$ gerade und $Z \sim \mathrm{Bin}(n,p)$. Berechnen und vereinfachen Sie soweit wie möglich $\mathbf{M}[Z]$.

Aufgabe 22 (Varianzanalyse von FindMax) (5 Punkte)
Gegeben sei der Algorithmus FindMax aus Aufgabe 8 für den Vektor $(1,...,n)$. Berechnen Sie die Varianz der Anzahl an Überschreibungen $\max \leftarrow i$ bei zufälligem Input $\underline{X} = (\sigma(1),...,\sigma(n))$ für $\sigma \in \mathcal{S}_n$. Betrachten Sie dafür folgende Zufallsvariable für $2 \leq i \leq n$:

$$X_1 = 0, X_i(\sigma) := \begin{cases} 1, & \text{falls } \sigma(i) > \max(\sigma(1),...,\sigma(i-1)) \\ 0, & \text{sonst} \end{cases}.$$

(a) Zeigen Sie, dass für $2 \leq k \leq n$ und $x_i \in \{0,1\}$ für alle $1 \leq i \leq n$ folgendes gilt:

$$\left|\left\{\sigma : \bigcap_{i=1}^{k-1} \{X_i(\sigma) = x_i\}\right\}\right| = k \cdot \left|\left\{\sigma : \bigcap_{i=1}^{k-1} \{X_i = x_i\} \cap \{X_k = 1\}\right\}\right|.$$

Hinweis: Verwenden Sie eine Permutation, welche eine Permutation der rechten Menge sortiert und überlegen Sie, warum nun eines der k Elemente gestrichen werden kann, um eine Permutation zu erzeugen, die die linke Seite erfüllt.

(b) Zeigen Sie, dass die Zufallsvariablen $X_1,...,X_n$ stochastisch unabhängig sind.
Hinweis: Sie dürfen ohne Beweis verwenden, dass für $1 \leq k \leq n$

$$\begin{aligned}&\mathbf{P}(X_1 = x_1,...,X_{k-1} = x_{k-1}, X_k = 0)\\ &\quad = \mathbf{P}(X_1 = x_1,...,X_{k-1} = x_{k-1}) - \mathbf{P}(X_1 = x_1,...,X_{k-1} = x_{k-1}, X_k = 1)\end{aligned}$$

gilt.

(c) Berechnen Sie die Varianz für die Anzahl an Überschreibungen.

Aufgabe 23 (DNA-Sequenzanalyse) (5 Punkte)
Wir betrachten eine DNA-Sequenz bestehend aus den vier Aminosäuren Adenin (A), Cytosin (C), Guanin (G) und Thymin (T). Wir definieren $\mathcal{A} := \{A, C, G, T\}$. Einer zufälligen DNA-Sequenz $D_n = (D_1, ..., D_n)$ der Länge $n \in \mathbb{N}$ liegen die unabhängigen, $\mathcal{A}$-wertigen Zufallsvariablen $D_1, ..., D_n$ zugrunde, wobei $\mathbf{P}(D_i = \square) = p_\square$ mit $\square \in \mathcal{A}$. Wir wollen nun wissen wie oft man erwarten kann, dass eine Teilsequenz $d \in \mathcal{A}^m$ der DNA mit Länge $m < n$ in der Gesamtsequenz vorkommt. Sei dafür

$$S(d) := \sum_{i=0}^{n-m} Y_i, \qquad \text{wobei } Y_i := \mathbb{1}_{\{D_{i+1}=d_1,...,D_{i+m}=d_m\}}$$

die Anzahl an Teilsequenzwiederholungen in der Gesamtsequenz. Zeigen Sie:

(a) $\mathbf{E}[S(d)] = (n - m + 1)q(d)$, wobei $q(d) := P(D_m = d)$ ist.

(b) $\mathbf{Var}[S(d)] \leq (2m - 1)\mathbf{E}[S(d)]$.

Aufgabe 24 (Unkorreliert vs. unabhängig) (3 Punkte)
Es sei $X \sim \mathcal{N}(0, 1)$. Zeigen Sie, dass X und $Y := X^2$ unkorreliert, aber nicht unabhängig sind. (Insbesondere gilt also die Umkehrung aus Proposition 3.18 nicht!)