Hausübungsblatt 9

Abgabe: Freitag, 20.01.2012, **vor** der Vorlesung.

Aufgabe 1

Die DNA enthält bekanntlich die Nukleotide zu den Basen A (Adenin), G (Guanin), C (Cytosin) und T (Thymin).

a) Gegeben sei die DNA-Sequenz (mit Beachtung der Reihenfolge)
$A\ T\ C\ T\ A\ G\ C\ T\ G\ G\ C\ C.$
Bezeichne s_i die Base an Position i, $i = 1, \ldots, 12$. Zeigen Sie, dass die Anzahl der Index-Permutationen π von $\{1, \ldots, 12\}$, so dass die permutierte Sequenz $s_{\pi(i)}$ *nicht* mit der ursprünglichen Sequenz oben übereinstimmt, gegeben ist durch

$$\frac{12!}{2!\ 3!\ 4!\ 3!}$$

(3 Punkte)

b) Folgern Sie aus a), dass die Anzahl unterschiedlicher DNA-Sequenzen der Länge n, die genau n_A-mal A, n_G-mal G, n_C-mal C und n_T-mal T enthalten, so dass $n = n_A + n_G + n_C + n_T$ gegeben ist durch

$$\frac{n!}{n_A!\ n_G!\ n_C!\ n_T!}.$$

(2 Punkte)

c) Betrachten Sie erneut die DNA-Sequenz aus a). Wie groß ist die Wahrscheinlichkeit für diese Sequenz, wenn die Wahrscheinlichkeiten für A, G, C, T durch $\pi_A, \pi_C, \pi_G, \pi_T$ gegeben sind, $\pi_A + \pi_C + \pi_G + \pi_T = 1$, und die Positionen der Sequenz als unabhängig angenommen werden ? (2 Punkte)

d) Gegeben sei eine Urne, die N_l Kugeln mit der Nummer l enthält, $l = 1, \ldots, k$. Definiere $N = \sum_{l=1}^k N_l$ und $\pi_l = N_l/N$, $l = 1, \ldots, k$. Es wird n mal mit Zurücklegen aus der Urne gezogen. Betrachte die Zufallsvariablen

$$X_l: \text{ Anzahl gezogener Kugeln mit Nummer } l, \quad l = 1, \ldots, k.$$

Zeigen Sie, dass die gemeinsame Wahrscheinlichkeitsverteilung des Zufallsvektors $(X_1, \ldots, X_k)^\top$ gegeben ist durch

$$P(X_1 = n_1, \ldots, X_k = n_k) = \frac{n!}{\prod_{l=1}^k n_l!} \prod_{l=1}^k \pi_l^{n_l}, \quad n_1, \ldots, n_k \geq 0, \ \sum_{l=1}^k n_l = n.$$

Man sagt, $(X_1, \ldots, X_k)^\top$ ist *multinomialverteilt* mit Parameter $(\pi_1, \ldots, \pi_k)^\top$.
(4 Punkte).

Hinweis zu d): Berechnen Sie zunächst die Wahrscheinlichkeit, eine *bestimmte* Sequenz von Kugeln zu erhalten, so dass die Nummern $1, \ldots, k$, genau n_l-mal, $l = 1, \ldots, k$, vorkommen (analog zur Teilaufgabe c)). Multiplizieren Sie das Resultat mit der Anzahl *aller möglichen* Sequenzen, in denen die Nummern $1, \ldots, k$ genau n_l-mal, $l = 1, \ldots, k$, vorkommen (analog zu Teilaufgabe b)).

Aufgabe 2

Birthday paradox.

a) Gegeben seien N Personen. Zeigen Sie unter den Annahmen, dass alle Tage Wahrscheinlichkeit $1/365$ haben und dass die Geburtstage der Personen unabhängig identisch verteilt sind, dass

$$P(\text{ 'Es gibt mindestens zwei Personen haben, die am 1. Januar Geburtstag haben ' })$$
$$= \sum_{i=2}^{N} \binom{N}{i} \left(\frac{1}{365}\right)^{i} \left(\frac{364}{365}\right)^{N-i} = 1 - \left(\frac{364}{365}\right)^{N} - N\frac{1}{365}\left(\frac{364}{365}\right)^{N-1}.$$

b) Zeigen Sie, dass unter den Annahmen in a)

$$P(\text{ 'Es gibt einen Tag, an dem mindestens zwei Personen Geburstag haben' })$$
$$= 1 - \frac{\frac{365!}{(365-N)!}}{365^N}$$
$$= 1 - \prod_{i=1}^{N} \frac{365 - i + 1}{365} = 1 - \prod_{i=1}^{N}\left(1 - \frac{i-1}{365}\right) \geq 1 - \exp\left(-\frac{N(N-1)}{730}\right).$$

Werten Sie die Wahrscheinlichkeit in a) und die untere Schranke in b) für $N = 30$ numerisch aus (3 + 4 Punkte).

Hinweis zu b): Betrachten Sie das Komplementärereignis und benutzen Sie, dass für $x \geq 0$, $(1 - x) \leq \exp(-x)$ und dass $\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$.

Aufgabe 3

Eine Zufallsvariable X heisst geometrisch verteilt mit Parameter $\pi \in (0, 1]$, falls ihre Wahrscheinlichkeitsfunktion gegeben ist durch

$$P(X = k) = (1 - \pi)^{k-1}\pi, \quad k = 1, 2, \ldots$$

a) Bestimmen Sie die wahrscheinlichkeitserzeugende Funktion $G(s) = \mathbf{E}[s^X]$ von X für $s \in [0, 1]$. (2 Punkte)

b) Benutzen Sie das Ergebnis aus a), um den Erwartungswert und die Varianz von X zu bestimmen. (4 Punkte)