# Musterlösungen zum Hausübungsblatt 9

## Aufgabe 1

**Lösung:**
**WARNUNG**: Die Aufgabe war nicht korrekt gestellt, insofern, als das angegebene Resultat nicht zur Aufgabenstellung passt.

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 $i$, $i = 1, \ldots, 12$.

**ursprüngliche Problemstellung**:
Bestimmen Sie die Anzahl der Index-Permutationen $\pi$ von $\{1, \ldots, 12\}$, so dass die permutierte Sequenz $(s_{\pi(i)})$ *nicht* mit der ursprünglichen Sequenz oben übereinstimmt.

**Lösung der ursprünglichen Problemstellung**:
Die Anzahl aller Index-Permutationen von $\{1, \ldots, 12\}$ ist 12!. Wir müssen davon alle Index-Permutationen abziehen, die die Sequenz unverändert lassen. Index-Permutationen, die die Sequenz unverändert lassen, tauschen $A$ gegen $A$, $T$ gegen $T$, $C$ gegen $C$ und $T$ gegen $T$. So lässt zum Beispiel die Permutation, die Index $1$ auf Index $4$ abbildet und Index $4$ auf Index $1$, die Sequenz unverändert, da sowohl an Position $1$ als auch an Position $4$ die Base $A$ steht. Die Anzahl Index-Permutationen, die die gegebene Sequenz unverändert lassen, ist gegeben durch

$$2!\ 3!\ 4!\ 3!$$

Dies ist das der Produkt der Permutationen von $\{1, \ldots, n_A\}$, $\{1, \ldots, n_G\}$, $\{1, \ldots, n_C\}$ und $\{1, \ldots, n_T\}$, wobei $n_A$, $n_G$, $n_C$ und $n_T$ die Häufigkeiten von $A$,$G$,$C$ und $T$ in der angegebenen Sequenz sind. Wir erhalten also das Ergebnis

$$12! - 2!\ 3!\ 4!\ 3!.$$

**Problemstellung, die zum angegeben Ergebnis passt**:

Wir betrachten Index-Permutationen von $\{1, \ldots, 12\}$. Zwei Index-Permutationen $\pi, \pi'$ seien äquivalent, falls die permutierten Sequenzen $(s_{\pi(i)})$ und $(s_{\pi'(i)})$ übereinstimmen. Zeigen Sie, dass die Anzahl der Äquivalenzklassen gegeben ist durch

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

**Begründung**:

Die Anzahl aller Permutationen von $\{1, \ldots, 12\}$ ist gegeben durch 12!. Nun ist

$$\begin{aligned} \text{Anzahl aller Permutationen } &= 12! \\ &= \text{Anzahl unterschiedlicher DNA-Sequenzen aus 2xA, 3xG, 4xC, 3xT} \\ &\times \text{Anzahl Index-Permutationen, die die eine solche Sequenz unverändert lassen.} \end{aligned}$$

Mit der gleichen Argumentation wie oben ist

$$\text{Anzahl Index-Permutationen, die eine solche Sequenz unverändert lassen} = 2!\ 3!\ 4!\ 3!$$

Die gesuchte Anzahl ergibt sich dann durch Umstellen der ersten Gleichung.
b) Dies ist die allgemeine Fassung des Problems aus a). Man kann der Argumentation aus a) kann folgen, indem man 12! durch $n!$ und die Häufigkeiten $2, 3, 4, 3$ der Basen $A, G, C, T$ durch $n_A$,$n_G$,$n_C$ und $n_T$ ersetzt.
c) Die gesuchte Wahrscheinlichkeit ergibt sich wegen der Unabhängigkeit als Produkt der Einzelwahrscheinlichkeiten

$$\prod_{i=1}^{12} \pi_{s_i} = \pi_A^{n_A} \pi_G^{n_G} \pi_C^{n_C} \pi_T^{n_T} = \pi_A^2 \pi_G^3 \pi_C^4 \pi_T^3.$$

d) Der angegebene Ausdruck für die Wahrscheinlichkeitsfunktion ergibt sich, indem man die Resultate aus den Aufgabenteilen b) und c) kombiniert und verallgemeinert. Die vier Basen lassen sich mit den Nummern $\{1, 2, 3, 4\}$ identifizieren, und eine DNA-Sequenz der Länge $n$ dementsprechend als eine Folge von diesen Nummern auf Kugeln, die mit Zurücklegen aus einer Urne gezogen werden. Die Verallgemeinerung entsteht durch Betrachtung von $k$ 'Kategorien' (im Beispiel der DNA-Sequenz ist $k = 4$). Die Wahrscheinlichkeit für eine *bestimmte* Sequenz von Kategorien mit Kategorienhäufigkeiten $n_1, \ldots, n_k$ ist

$$\prod_{l=1}^{k} \pi_l^{n_l}, \tag{1}$$

mit der gleichen Argumentation wie in Teil c). Um den angebenen Ausdruck für die Wahrscheinlichkeit zu erhalten, muss man die Wahrscheinlichkeit (1) mit der Anzahl aller Möglichkeiten multiplizieren, Sequenzen von Kategorien zu bilden, die die gleiche Kategorienhäufigkeiten aufweisen. Diese Anzahl ist

$$\frac{n!}{\prod_{l=1}^{k} n_l!}$$

mit der gleichen Argumentation wie in Teil b).

## Aufgabe 2

**Lösung:** a) Sei $X$ die Zufallsvariable 'Anzahl Personen, die am 1. Januar Geburstag' haben. Unter den gemachten Annahmen ist $X$ binomialverteilt mit Parametern $N$ und Wahrscheinlichkeit $\frac{1}{365}$. Die Wahrscheinlichkeit, dass *genau* $i$ Personen am 1. Januar Geburstag haben, ist also gegeben durch

$$P(X = i) = \binom{N}{i} \left(\frac{1}{365}\right)^i \left(\frac{364}{365}\right)^{N-i}.$$

Folglich ist die gesuchte Wahrscheinlichkeit

$$\sum_{i=2}^{N} P(X = i) = 1 - P(X = 0) - P(X = 1),$$

was den angegebenen Ausdrücken entspricht. Numerisches Auswerten der rechten Seite für $N = 30$ liefert approximativ die Wahrscheinlichkeit 0.003.
b) Das Komplementärereignis zum Ereignis, dessen Wahrscheinlichkeit zu berechnen ist, ist 'Alle Personen haben an verschiedenen Tagen Geburstag'. Unter den gemachten Annahmen beträgt die Wahrscheinlichkeit für das Komplementärereignis

$$\frac{365 \cdot 364 \ldots \cdot (365 - N + 1)}{365^N} = \frac{\frac{365!}{(365-N)!}}{365^N}.$$

Die gesuchte Wahrscheinlichkeit ist also gegeben durch

$$\begin{aligned} 1-\prod_{i=1}^{N}\frac{365-i+1}{365} &= 1-\prod_{i=1}^{N}\left(1-\frac{i-1}{365}\right) \\ &\geq 1-\prod_{i=1}^{N}\exp\left(-\frac{i-1}{365}\right) \\ &= 1-\exp\left(-\frac{1}{365}\sum_{i=1}^{N}(i-1)\right) \\ &= 1-\exp\left(-\frac{N(N-1)}{730}\right). \end{aligned}$$

Numerisches Auswerten der unteren Schranke liefert approximativ $0.696$.

## Aufgabe 3

**Lösung:** a) Die wahrscheinlichkeitserzeugende Funktion von $X$ erhält man gemäss

$$\begin{aligned} G(s) = \mathbf{E}[s^X] = \sum_{k=1}^{\infty} s^k(1-\pi)^{k-1}\pi &= \pi\, s\sum_{k=1}^{\infty}(s(1-\pi))^{k-1} \\ &= \pi s\sum_{k=0}^{\infty}(s(1-\pi))^k = \frac{\pi s}{1-s(1-\pi)}. \end{aligned}$$

Der letzte Ausdrück folgt aus der geometrischen Reihe. Man beachte dabei, dass $s(1-\pi) < 1$, da $s \in [0,1]$ und $(1-\pi) \in [0,1)$ nach Annahme, so dass die Konvergenz der Reihe sichergestellt ist.
b) Differenzieren von $G(s)$ bezüglich $s$ liefert unter Anwendung der Quotientenregel

$$G'(s) = \frac{(1-s(1-\pi))\cdot\pi - \pi s(\pi-1)}{(1-s(1-\pi))^2} = \frac{\pi}{(1-s(1-\pi))^2}.$$

Also ist $\mathbf{E}[X] = G'(1) = \frac{1}{\pi}$. Zur Berechnung der Varianz brauchen wir noch die zweite Ableitung.

$$G''(s) = \frac{2\pi(1-\pi)}{(1-s(1-\pi))^3}.$$

Es ist

$$G''(1) = \frac{2\pi(1-\pi)}{\pi^3} = \mathbf{E}[X(X-1)] = \mathbf{E}[X^2] - \mathbf{E}[X].$$

Auflösen nach $\mathbf{E}[X^2]$ liefert

$$\mathbf{E}[X^2] = \frac{2\pi-\pi^2}{\pi^3}$$

Man erhält schliesslich

$$V(X) = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{1-\pi}{\pi^2}.$$