Lösungen zur Übung (6)

(1) (a) Einen $X-$ Wert nicht aus dem mittleren Drittel zu ziehen, hat offenbar Wahrscheinlichkeit 2/3, dies sechs mal hintereinander zu tun (unabhängig!), hat also Wahrscheinlichkeit $(2/3)^6 = 0.088$ (gerundet).

(b) Dabei zusätzlich je genau drei Exemplare des unteren und des oberen Drittels zu bekommen, hat die Wahrscheinlichkeit $\binom{6}{3}\left(\frac{1}{3}\right)^6 = 0.027$. Überlegung dazu: Man hat die Wahrscheinlichkeit $\left(\frac{1}{3}\right)^6$ dafür, diese Exemplare in einer bestimmten Reihenfolge von 'unteres Drittel', 'oberes Drittel' zu ziehen, und es gibt $\binom{6}{3}$ Möglichkeiten dafür, die Plätze, auf denen 'unteres Drittel' gezogen wird, auszuwählen. Eine andere Überlegung dazu: Wir fragen, welche Wahrscheinlichkeit man für 'je drei aus dem unteren und dem oberen Drittel' hat bedingt dadurch, dass kein Exemplar aus dem mittleren Drittel gezogen wird. Das ist offenbar $\binom{6}{3}\left(\frac{1}{2}\right)^6$. Diese Wahrscheinlichkeit ist nun zu multiplizieren mit der Wahrscheinlichkeit $\left(\frac{2}{3}\right)^6$ dafür, kein Exemplar aus dem mittleren Drittel zu ziehen, und wir haben dasselbe Resultat.

(2) Die Wahrscheinlichkeit einer korrekten Gruppenlösung ist

$$1 - \frac{19}{20} \cdot \frac{14}{15} \cdot \frac{9}{10} \cdot \frac{9}{10} \cdot \frac{5}{6} \cdot \frac{3}{4} \cdot \frac{2}{3} \approx 0.7.$$

Es ist wichtig, über die Verneinung zu gehen, weil wir dann Unabhängigkeit für die 'und'-Verbindung nutzen können, während man für die 'oder'-Verbindung des fraglichen Ereignisses nicht addieren darf. Beobachtung: Diese Wahrscheinlichkeit ist viel größer als jede der einzelnen. Verallgemeinerung und zugleich Verschärfung des Resultats: Wenn alle Gruppenmitglieder einzeln eine Wahrscheinlichkeit $\geq p > 0$ (p eine beliebig kleine feste Zahl) haben, die Aufgabe zu lösen, so kommt die Wahrscheinlichkeit der Gruppenlösung beliebig nahe an 1 heran, wenn die Gruppe nur groß genug ist. Denn die Wahrscheinlichkeit der Gruppenlösung ist dan mindestens $1 - (1-p)^n$, wenn n der Umfang der Gruppe ist, und mit $0 < 1 - p < 1$ geht $(1-p)^n$ mit $n \to \infty$ nach Null.

(3) Wir bezeichnen mit K das Ereignis 'krank', mit POS das Ereignis 'positiv getestet. Gefragt ist nach $P(K|POS)$, und das ist ein Fall für die Bayessche Formel:

$$P(K|POS) = \frac{P(POS|K)\,P(K)}{P(POS)}.$$

Beide Faktoren des Zählers sind nach den Angaben direkt bekannt:

$$\begin{aligned} P(POS|K) &= 0.99, \\ P(K) &= 0.01. \end{aligned}$$

Der Nenner ist nicht sofort bekannt, aber - typisch für diesen Kontext (!) - mittels der Formel von der totalen Wahrscheinlichkeit einfach zu ermitteln:

$$\begin{aligned} P(POS) &= P(POS|K) \cdot P(K) + P\left(POS|\overline{K}\right) \cdot P\left(\overline{K}\right) \\ &= 0.99 \cdot 0.01 + 0.01 \cdot 0.99. \end{aligned}$$

Damit haben wir

$$P(K|POS) = \frac{0.99 \cdot 0.01}{2 \cdot 0.99 \cdot 0.01} = \frac{1}{2}.$$

Wenn dagegen 50% der Population krank sind, so haben wir

$$P(K|POS) = \frac{0.99 \cdot 0.5}{0.99 \cdot 0.5 + 0.01 \cdot 0.5} = 0.99.$$

Dies ist tatsächlich genau das Ergebnis, das allgemein mit (schlechter) Intuition erwartet wird, aber nur in diesem (für die interessierenden Situationen völlig unrealistischen) Spezialfall wirklich stimmt.

(4) Pro Wurf erwarten wir $\frac{1}{6}$ Sechsen, also in 6 Würfen eine Sechs, und so benötigen wir im Mittel sechs Würfe, bis wir die erste Sechs geworfen haben. Dass man genau 10 Würfe benötigt, bedeutet: Zuerst wird neun mal keine Sechs gewürfelt, dann eine Sechs, und die Wahrscheinlichkeit dafür ist $\left(\frac{5}{6}\right)^9\left(\frac{1}{6}\right) = 0.03$. Es ist gar nicht so einfach, mit der Bestimmung der Wahrscheinlichkeiten für alle benötigten Wurfzahlen bis zur ersten Sechs den Erwartungswert formal auszurechnen, das ergäbe $\mu = \sum_{n=0}^{\infty}\left(\frac{5}{6}\right)^n \cdot \frac{1}{6}(n+1)$. Tatsächlich kommt da 6 heraus.

(5) Der benötigte Umfang ist jedenfalls so groß, dass Näherung der Variablen 'relative Häufigkeit' mit Normalverteilung in Ordnung ist. Wir haben $\mu = p$, $\sigma = \sqrt{p(1-p)}/\sqrt{n}$. Also ist der Fehler, der mit Wahrscheinlichkeit 0.99 eingehalten wird:

$$z_{0.995} \cdot \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq 2.58 \cdot \frac{1}{2\sqrt{n}}.$$

Nun ist die kleinste natürliche Zahl n derart anzusetzen, dass dies einen Wert von höchstens 0.01 bekommt. Dazu lösen wir die Gleichung

$$2.58 \cdot \frac{1}{2\sqrt{n}} = 0.01$$

und finden (exakt!) $n = 16641$. (Zufällig ist das bereits eine ganze Zahl. Allerdings ist ja auch der Wert 2.58 in Wahrheit etwas größer als für das Vertrauensintervall genau benötigt.) Dass $p = 1/2$ der schlimmste Fall ist, erkennt man daran, das $\sqrt{p(1-p)}$ den maximalen Wert bei $p = 1/2$ annimmt. Wenn man bereits wüsste, dass man eine relative Häufigkeit zu schätzen hat, die auf keinen Fall größer als 0.2 sein kann, so könnte man 0.2 für p einsetzen und ein deutlich günstigeres Resultat erzielen ($n = 10651$ würde dann reichen). Allerdings: Je kleiner Wahrscheinlichkeiten werden, desto genauer möchte man sie auch schätzen. Kurzum: Man braucht erstaunlich große Stichprobenumfänge, um relative Häufigkeiten mit erwünschter Genauigkeit zu schätzen.

(6) Hier wird die $t-$ Verteilung (mit 19 Freiheitsgraden) benötigt, da es sich um einen recht kleinen Stichprobenumfang und nur geschätzte Streuung handelt, also lautet das gesuchte Intervall

$$\begin{aligned} &20 \pm t_{0.995}^{19} \cdot 5, \text{ d.h.} \\ &20 \pm 2.86 \cdot 5 \text{ oder } [5.7; 34.3]. \end{aligned}$$

(7) Die Aussage ist falsch. Einfaches Gegenbeispiel: Das Quadrat (Ω entsprechend) teile man in zwei Hälften, sagen wir eine linke und eine rechte; die eine repräsentiert A, die andere $\overline{A}$. Das obere Drittel von A stelle $B \cap A$ dar, die oberen zwei Drittel von $\overline{A}$ stellen $B \cap \overline{A}$ dar. Dann ist offenbar $P(B|A) = 1/3$, $P(B) = 1/2$, also A und B abhängig. Nun sei C die obere und $\overline{C}$ die untere Hälfte des Quadrats. Dann ist $P(C|B) = \frac{5}{6}$, aber $P(C) = \frac{1}{2}$, somit sind auch B und C abhängig. Aber offenbar sind A und C unabhängig, $P(C|A) = P(C) = 1/2$. Das Ganze kann man auch leicht mit $\Omega = \{1, 2, 3, ..., 12\}$ darstellen, wobei man alle Zahlen mit gleicher Wahrscheinlichkeit 1/12 herauskommen lässt, so dass Wahrscheinlichkeit dasselbe ist wie relative Häufigkeit in Ω. Genau der graphischen Konstruktion entsprechend kann man ansetzen: A sei die Menge der geraden Zahlen in Ω. B sei die Menge $\{2, 4, 1, 3, 5\}$. C sei $\{1, 2, 3, 4, 5, 6\}$. Dann haben wir genau die oben aufgeführten Verhältnisse. **Beachten Sie, dass diese Befunde noch einmal zeigen, welchen gewaltigen Unterschied es zwischen statistischer Abhängigkeit auf der einen Seite und auf der andern Seite logischer oder auch kausaler Abhängigkeit gibt: Wenn aus einer Aussage** α eine Aussage β folgt und aus β folgt γ, dann folgt auch γ aus α. Wenn ein Zustand a einen Zustand b verursacht und b verursacht c, so verursacht auch a den Zustand c. Für statistische Abhängigkeit dagegen gilt das Analogon überhaupt nicht, A und B können sogar sehr stark abhängig sein, ebenso B und C sehr stark, während A und C völlig unabhängig voneinander sind. Beachten Sie, dass man aus der Konstruktion auch lernen kann, wie leicht das möglich ist (es handelt

sich also nicht um gewaltsam konstruierte und gewissermaßen 'pathologische' Beispiele, sondern um das ganz Normale) und dass die vorausgesetzten Abhängigkeiten beliebig stark gemacht werden können.