

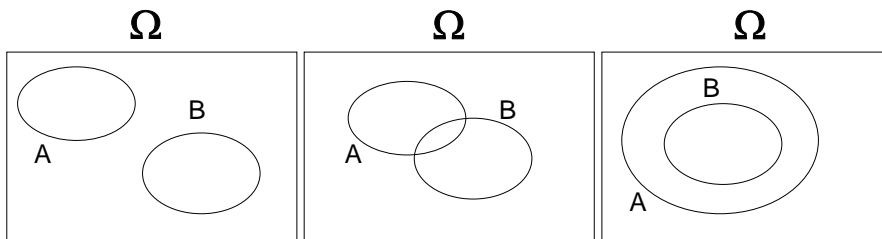
6 Mehrstufige zufällige Vorgänge – Lösungshinweise

Aufgabe 6.1: Begründen Sie, warum die stochastische Unabhängigkeit zweier Ereignisse bzw. zufälliger Vorgänge nur ein Modell der Realität darstellen kann.

Lösungsskizze Aufgabe 6.1 Die Variabilität statistischer Daten beschreibt das Phänomen, dass Datenerhebungen zu ein und demselben Merkmal zu unterschiedlichen Zeiten oder Orten im Allgemeinen zumindest im Detail unterschiedliche Ergebnisse erbringen. Selbst bei solchen zufälligen Vorgängen, die plausibel als stochastisch unabhängig *modelliert* werden können, etwa dem zweifachen Wurf eines Würfels, wird daher im Allgemeinen das Modell der stochastischen Unabhängigkeit nur annähernd, nicht aber exakt in den Daten erscheinen. Etwa wird für die Ereignisse „6₁“ und „6₂“ für das Fallen einer Sechs im ersten bzw. zweiten Wurf bei vielen Versuchen (n) des zweifachen Wurfs das Modell der stochastischen Unabhängigkeit ($P(6_1 \cap 6_2) = P(6_1) \cdot P(6_2)$) nur annähernd eine Entsprechung in den empirischen Daten haben: $h_n(6_1, 6_2) \approx h_n(6_1) \cdot h_n(6_2)$. Im Zusammenhang mit dem doppelten Münzwurf haben wir das im Buch auf S. 118 behandelt.

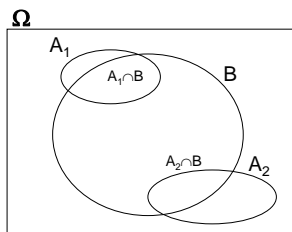
Aufgabe 6.2: Zeigen Sie, dass die drei Axiome der Wahrscheinlichkeitsrechnung für bedingte Wahrscheinlichkeiten gelten.

Lösungsskizze Aufgabe 6.2 Seien A und B zufällige Ereignisse mit $A, B \subseteq \Omega$ und $B \neq \emptyset$. Betrachtet werden soll das Ereignis $A|B$. Dann können für die Gültigkeit der ersten beiden Axiome folgende Szenarien betrachtet werden:



- Gilt $A \cap B = \emptyset$, so ist $P(A \cap B) = 0$ und damit $P(A|B) = \frac{P(A \cap B)}{P(B)} = 0$. Ist $P(A) \neq 0$, so wird deutlich, dass unvereinbare Ereignisse stets abhängig sind, da $P(A \cap B) = 0 \neq P(A) \cdot P(B) > 0$ gilt.
- Es gilt stets $P(A \cap B) \leq P(B)$. Für $A \cap B \neq \emptyset$ gilt damit $0 < P(A|B) = \frac{P(A \cap B)}{P(B)} < 1$. Im Falle $B \subseteq A$ (insbesondere $A = \Omega$) gilt $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

Für die Gültigkeit des dritten Axioms kann folgendes Szenario mit $A_1 \cap A_2 = \emptyset$ betrachtet werden:



Hier wird unmittelbar einsichtig, dass $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B)$ durch die Einschränkung von A auf B gilt.

Aufgabe 6.3: Untersuchen Sie anhand des Datensatzes zu den Studierenden der Hochschulen in Münster und in Freiburg, welche Merkmale ein verstecktes Indiz für ein anderes sein können. Kann z.B. das Beförderungsmittel ein Indiz für die Parteipräferenz oder den Erhalt von BAföG sein?

Lösungsskizze Aufgabe 6.3 Wir betrachten allein, ob das Verwenden des Fahrrads als Beförderungsmittel als Indiz für die Präferenz der Partei „Die Grünen“ gelten kann. Für den gesamten Datensatz ergibt sich:

	Fahrrad (F)	Sonstige (S_1)	Summe
Die Grünen (G)	116	104	220
Sonstige (S_2)	347	232	579
Summe	463	336	799

Aus den Daten ergibt sich:

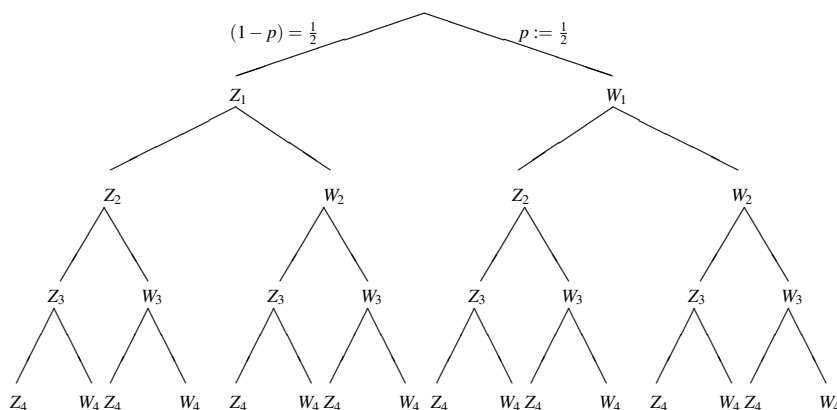
$$0,25 \approx h(G|F) = \frac{116}{336} < h(F) = \frac{220}{799} \approx 0,28$$

Demnach könnte das Beförderungsmittel Fahrrad höchstens ein negatives Indiz für die Präferenz für die Partei die Grünen sein. Da das Assoziationsmaß $A = h(G|F) - h(G|S_1) \approx -0,06$ klein ist, könnte man allerdings auch davon ausgehen, dass Beförderungsmittel und Parteipräferenz unabhängig voneinander sind.

Aufgabe 6.4: Eine Münze wird geworfen. X sei die Anzahl der Wappen (insgesamt), X_i die Anzahl der Wappen im i -ten Versuch. Begründen Sie, warum man die Teilvorgänge als stochastisch unabhängig modellieren kann. Bestimmen Sie die Wahrscheinlichkeitsverteilungen von X für den einfachen, den zweifachen, den dreifachen und den vierfachen Münzwurf. Zeichnen Sie einen Baum zu diesem zufälligen Vorgang.

Lösungsskizze Aufgabe 6.4 Die Einzelwürfe der Münze können plausibel als stochastisch unabhängige zufällige Vorgänge modelliert werden, da man andernfalls der Münze ein Art Gedächtnis zusprechen würde, aufgrund dessen das Ereignis *Wappen* (oder *Zahl*) einer der Ereignisse bei einem anderen Wurf beeinflussen würde.

Versieht man die Ereignisse W (Wappen) und Z (Zahl) mit einem Index für die Wurfnummer, so ergibt sich folgender Baum für vier Würfe, aus dem man die Bäume für $n = 1, 2, 3$ Würfe auslesen kann:



Aus dem Baum ergeben sich folgende Wahrscheinlichkeitsverteilungen:

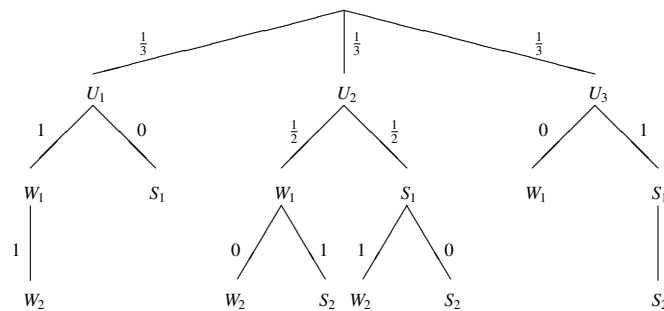
n	x	0	1	2	3	4
1	$P(X = x)$	$\frac{1}{2}$	$\frac{1}{2}$			
2	$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$		
3	$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	
4	$P(X = x)$	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

Aufgabe 6.5: Drei Urnen haben folgenden Inhalt. In der ersten Urne sind zwei weiße Kugeln, in der zweiten je eine schwarze und eine weiße, in der dritten zwei schwarze. Es wird eine Urne zufällig ausgewählt.

- Eine weiße Kugel wird gezogen (ohne diese zurückzulegen). Bestimmen Sie die Wahrscheinlichkeit dafür, dass die zweite Kugel schwarz sein wird.
- Im zweiten Zug ist eine schwarze Kugel gezogen worden. Bestimmen Sie die Wahrscheinlichkeit dafür, dass die erste Kugel weiß war.
- Bestimmen Sie die beiden Wahrscheinlichkeiten aus a) und b), wenn die Kugeln nach dem Ziehen zurückgelegt werden.

Stellen Sie das Problem mit Hilfe eines Baums und dem Einheitsquadrat dar.

Lösungsskizze Aufgabe 6.5 Unterscheidet man für mit einem Index einerseits die Urnen, andererseits die Kugelzüge, so ergibt sich für die Aufgabenteile a) und b) folgender Baum (bei dem an einigen Ästen Wahrscheinlichkeiten 0 oder 1 für unmögliche und sichere Ereignisse angetragen sind):



Mit Hilfe des Baums ergibt sich:

a)

$$P(S_2|W_1) = \frac{P(U_2 \cap W_1 \cap S_2)}{P(U_2 \cap W_1 \cap S_2) + P(U_1 \cap W_1 \cap W_2)} = \frac{\frac{1}{3} \cdot \frac{1}{2} \cdot 1}{\frac{1}{3} \cdot \frac{1}{2} \cdot 1 + \frac{1}{3} \cdot 1 \cdot 1} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{3}} = \frac{1}{3}$$

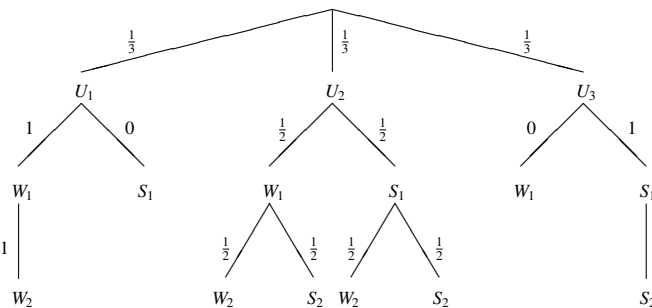
Im Baum wird also die Gesamtheit der Pfade durch die Bedingung (W_1) auf zwei Pfade eingeschränkt und auf der Basis des Teilbaums mit zwei Pfaden der Wahrscheinlichkeitsanteil des Pfades mit dem Ereignis S_2 bestimmt.

b)

$$P(W_1|S_2) = \frac{P(U_2 \cap W_1 \cap S_2)}{P(U_2 \cap W_1 \cap S_2) + P(U_3 \cap S_1 \cap S_2)} = \frac{\frac{1}{3} \cdot \frac{1}{2} \cdot 1}{\frac{1}{3} \cdot \frac{1}{2} \cdot 1 + \frac{1}{3} \cdot 1 \cdot 1} = \frac{\frac{1}{6}}{\frac{1}{6} + \frac{1}{3}} = \frac{1}{3}$$

Im Baum wird also die Gesamtheit der Pfade durch die Bedingung (S_2) auf zwei Pfade eingeschränkt und auf der Basis des Teilbaums mit zwei Pfaden der Wahrscheinlichkeitsanteil des Pfades mit dem Ereignis W_1 bestimmt.

Für Aufgabenteil c) ändert sich die Darstellung des zufälligen Vorgangs im Baum leicht (auf der dritten Ebene des Baums):



Mit Hilfe des Baums ergibt sich hier:

c)

$$\begin{aligned}
 P(S_2|W_1) &= \frac{P(U_2 \cap W_1 \cap S_2)}{P(U_2 \cap W_1 \cap S_2) + P(U_2 \cap W_1 \cap W_2) + P(U_1 \cap W_1 \cap W_2)} \\
 &= \frac{\frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot 1 \cdot 1} = \frac{\frac{1}{12}}{\frac{1}{12} + \frac{1}{12} + \frac{4}{12}} = \frac{1}{6}
 \end{aligned}$$

Im Baum wird also die Gesamtheit der Pfade durch die Bedingung (W_1) auf drei Pfade eingeschränkt und auf der Basis des Teilbaums mit drei Pfaden der Wahrscheinlichkeitsanteil des Pfades mit dem Ereignis S_2 bestimmt. Weiter ergibt sich

$$\begin{aligned}
 P(W_1|S_2) &= \frac{P(U_2 \cap W_1 \cap S_2)}{P(U_2 \cap W_1 \cap S_2) + P(U_2 \cap S_1 \cap S_2) + P(U_3 \cap S_1 \cap S_2)} \\
 &= \frac{\frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2}}{\frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{3} \cdot 1 \cdot 1} = \frac{\frac{1}{12}}{\frac{1}{12} + \frac{1}{12} + \frac{4}{12}} = \frac{1}{6}
 \end{aligned}$$

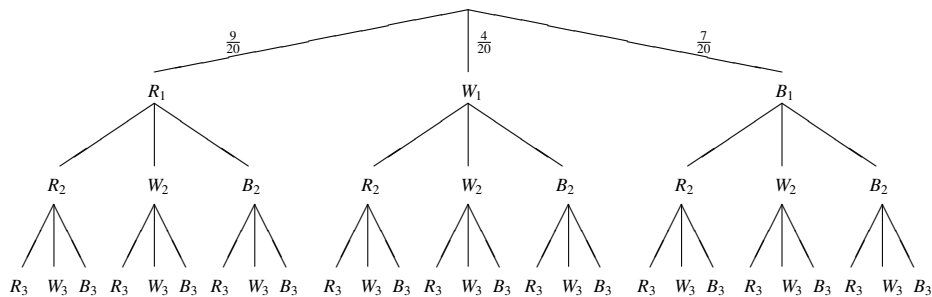
Im Baum wird also die Gesamtheit der Pfade durch die Bedingung (S_2) auf drei Pfade eingeschränkt und auf der Basis des Teilbaums mit drei Pfaden der Wahrscheinlichkeitsanteil des Pfades mit dem Ereignis W_1 bestimmt.

Alle Aufgabenteile lassen sich nicht mit dem Einheitsquadrat darstellen, da hier ein zufälliger Vorgang mit drei Teilvorgängen betrachtet wird. Hier hat die grafische Methode des Einheitsquadrats eine Grenze.

Aufgabe 6.6: In einer Urne sind 9 rote, 4 weiße und 7 blaue Kugeln. Es wird zufällig aus dieser Urne 3 Mal gezogen.

- Dabei werden die gezogenen Kugeln wieder zurückgelegt. Bestimmen Sie die Wahrscheinlichkeit für die Zugfolge: Zuerst eine weiße, dann eine blaue, dann eine rote Kugel. Zeichnen Sie einen Baum zu diesem Zufallsexperiment.
- Dabei werden die Kugeln nicht wieder zurückgelegt. Bestimmen Sie die Wahrscheinlichkeit für die Zugfolge in a).
- Bestimmen Sie für den Vorgang mit und ohne Zurücklegen die Wahrscheinlichkeit für das Ereignis A: Es werden zwei blaue und eine weiße Kugel gezogen. Zeichnen Sie einen Baum zu dieser Fragestellung.

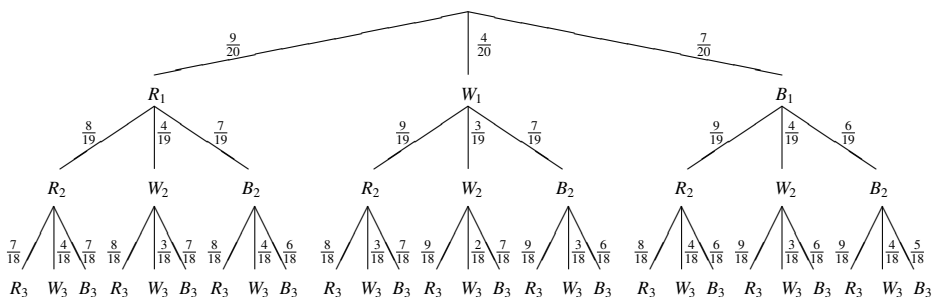
Lösungsskizze Aufgabe 6.6 Unterscheidet man für mit einem Index die Kugelzüge, so ergibt sich für Aufgabenteil a) folgender Baum, bei dem die bei jedem Zug gleichbleibenden Wahrscheinlichkeiten für die drei Farben nur an den Ästen der ersten Ebene des Baums eingezeichnet sind.



Mit Hilfe des Baums ergibt sich:

$$P(W_1 \cap B_2 \cap R_3) = \frac{4}{20} \cdot \frac{7}{20} \cdot \frac{9}{20} = \frac{252}{8000} = \frac{63}{2000} = 0,0315$$

Im Aufgabenteil b) müssen im Baum die sich auf jeder Ebene ändernden Wahrscheinlichkeiten beachtet werden:



Mit diesem Baum ergibt sich:

$$P(W_1 \cap B_2 \cap R_3) = \frac{4}{20} \cdot \frac{7}{19} \cdot \frac{9}{18} = \frac{252}{6840} = \frac{7}{190} \approx 0,0368$$

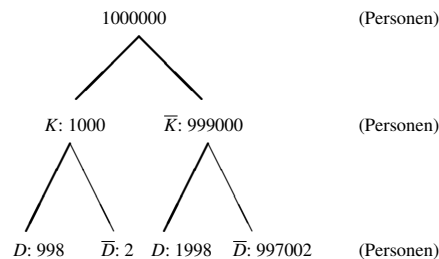
Für die Teilaufgabe c) können beide oben dargestellten Bäume verwendet werden. Es ergibt sich:

- Mit Zurücklegen: Jeder Pfad, der den Zug zweier blauer und einer weißer Kugel repräsentiert, hat die Wahrscheinlichkeit $\frac{7 \cdot 7 \cdot 4}{20^3} = \frac{196}{8000} = \frac{49}{2000} = 0,0245$. Es gibt drei solcher Pfade ($W_1 \cap B_2 \cap B_3$, $B_1 \cap W_2 \cap B_3$, $B_1 \cap B_2 \cap W_3$). Die gesuchte Wahrscheinlichkeit beträgt damit $3 \cdot 0,0245 = 0,0735$.
- Ohne Zurücklegen: Jeder Pfad, der den Zug zweier blauer und einer weißer Kugel repräsentiert, hat die Wahrscheinlichkeit $\frac{7 \cdot 6 \cdot 4}{20 \cdot 19 \cdot 18} = \frac{168}{6840} = \frac{7}{285} \approx 0,0246$. Es gibt drei solcher Pfade ($W_1 \cap B_2 \cap B_3$, $B_1 \cap W_2 \cap B_3$, $B_1 \cap B_2 \cap W_3$). Die gesuchte Wahrscheinlichkeit beträgt damit $3 \cdot \frac{7}{285} = \frac{7}{95} \approx 0,0737$.

Aufgabe 6.7: Gegeben sei der erste HIV-Test mit: K : krank (bzw. infiziert), \bar{K} : nicht krank (nicht infiziert), D : Test positiv und \bar{D} : Test negativ. Weiterhin sei gegeben: $P(K) = 0,001$, $P(D|K) = 0,998$ (Sensitivität) und $P(D|\bar{K}) = 0,002$ (Spezifität).

- Bestimmen Sie Wahrscheinlichkeit für $P(K|D)$.
- Was bedeutet im Sachkontext $P(K|\bar{D})$?
- Da das Ergebnis eines Tests (siehe a)) offensichtlich nicht ausreichend ist, um die Diagnose an einen Patienten zu geben, besteht der HIV-Test aus mehreren (unterschiedlichen) Tests. Bestimmen Sie die notwendige Spezifität des Tests insgesamt so, dass bei einer Sensitivität von $P(D|K) = 0,999$ die Wahrscheinlichkeit für $P(K|D)$ größer als 0,995 ist.
- $P(K)$ ist die offizielle Basisrate für Deutschland. In anderen Ländern, insbesondere in Teilen Afrikas, ist die Basisrate erheblich höher. Welche Auswirkung hat das auf den HIV-Test? Bestimmen Sie die Wahrscheinlichkeit $P(K|D)$ in Abhängigkeit von $P(K)$.
- Warum ist die Annahme $P(K) = 0,001$, also die Wahrscheinlichkeit, ohne es zu wissen und ohne Symptome mit HIV infiziert zu sein, im Allgemeinen keine sinnvolle Modellierung, wenn es um die Frage $P(K|D)$ für eine bestimmte Person geht?

Lösungsskizze Aufgabe 6.7 Wir verwenden zur Lösung des Aufgabenteils a) einen Baum mit absoluten Häufigkeiten (1.000.000 virtuelle Patienten):



Es ergibt sich damit:

$$P(K|D) = \frac{P(D|K) \cdot P(K)}{P(D|K) \cdot P(K) + P(D|\bar{K}) \cdot P(\bar{K})} = \frac{998}{998 + 1998} \approx 0,333$$

Die Wahrscheinlichkeit, trotz positiver Diagnose infiziert zu sein, würde also bei den gegebenen Modelldaten 33,3 Prozent betragen.

Im Gegensatz dazu gibt es allerdings auch die Wahrscheinlichkeit, infiziert zu sein, obwohl eine negative Diagnose erfolgt ist. Die Wahrscheinlichkeit für dieses Ereignis ist

$$P(K|\bar{D}) = \frac{P(\bar{D}|K) \cdot P(K)}{P(\bar{D}|K) \cdot P(K) + P(\bar{D}|\bar{K}) \cdot P(\bar{K})} = \frac{2}{2 + 997002} \approx 0,000002$$

Diese Wahrscheinlichkeit ist in diesem Modell also sehr gering.

Soll durch die Hintereinanderschaltung verschiedener Tests mit einer im Modell angenommenen Sensitivität von 99,9 Prozent eine Spezifität bestimmt werden, so dass $P(K|D) > 0,995$ ist, muss die Gleichung

$$P(K|D) = \frac{P(D|K) \cdot P(K)}{P(D|K) \cdot P(K) + P(D|\bar{K}) \cdot P(\bar{K})} > 0,995$$

gelöst werden.

$$\frac{P(D|K) \cdot P(K)}{P(D|K) \cdot P(K) + P(D|\bar{K}) \cdot P(\bar{K})} > 0,995 \Leftrightarrow P(D|\bar{K}) < \frac{0,005 \cdot P(D|K) \cdot P(K)}{0,995 \cdot P(\bar{K})} = 0,000005$$

Unter 100000 untersuchten Nicht-Infizierten, dürfen also nur 5 fälschlicherweise eine positive Diagnose erhalten.

In Ländern, in denen die Basisrate der HIV-Infektionen deutlich höher ist als in Deutschland, ist auch schon ein Test (das oben genannte Modell angenommen) recht sicher. In der Tabelle unten ist jeweils die Wahrscheinlichkeit $P(K|D)$ für drei Länder auf der Basis der offiziell angegebenen HIV-Basisraten angegeben (Stand 2010, Quelle UNAIDS):

Land	Infektionsrate	$P(K D)$ (im Modell)
Panama	0,01	0,834
Kenia	0,05	0,963
Botswana	0,25	0,994

Alle Rechnungen sind Modell-Rechnungen, die ein Individuum betreffen. Damit sind aber auch das Individuum und dessen Lebensumstände von Belang. Im Extremen: Hat ein Individuum die vergangenen 50 Jahre als Emerit auf einer einsamen Hallig im Wattenmeer verbracht, so wird $P(K) = 0,001$ wohl deutlich zu hoch angesetzt sein, sind dagegen im Lebenslauf eines Individuums alle Risikofaktoren für eine HIV-Infektion über Jahre deutlich vertreten, so ist $P(K) = 0,001$ (für Deutschland) eher zu niedrig angenommen.

Für die Berechnungen dieser Art Aufgaben stehen in den Zusatzmaterialien zu diesem Buch ein interaktiver Baum mit absoluten Häufigkeiten wie auch ein interaktives Einheitsquadrat zur Verfügung.