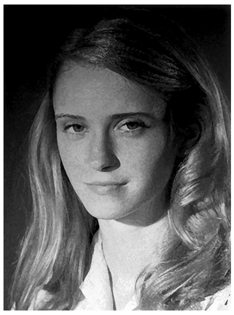
Aufgaben Sum of Us 2010

Eine Mathematikstudentin Emmy wird vermisst. In der Nacht zuvor wurde Sie noch bei einer Studentenfete am Rheinufer gesehen. Ihr Verschwinden ist rätselhaft. Emmy war eine sehr introvertierte Person, die nur noch wenig Motivation für ihr Studium aufbringen konnte. Sie ließ durchblicken, dass Sie auf Weltreise gehen wollte, ohne sich zuvor von Irgendjemandem zu verabschieden. Sie hatte kaum Kontakt zu ihren Eltern, und die Beziehung zu ihrem Freund David, ebenfalls ein Mathematikstudent, war derart angespannt, dass es niemanden überrascht hätte, wenn sie auch ihm eine solche Weltreise nicht angekündigt hätte. Einen Monat später wird jedoch eine Leiche im Rhein gefunden. Das Opfer ist nicht mehr zu erkennen. Eine DNA-Analyse zeigt jedoch, dass es sich bei dem Opfer mit großer Wahrscheinlichkeit um Emmy handelt.



- *In Aufgabe 1 wirst du diese Wahrscheinlichkeit bestimmen. Im folgenden kannst du davon ausgehen, dass es tatsächlich Emmys Leiche ist.*

Am Hals der Leiche hängt ein Gürtel, mit dem das Opfer offenbar erwürgt wurde. Die Ermittlungen führen zu der Vermutung, dass die Studentin während oder unmittelbar nach der oben genannten Fete ermordet wurde. Untersuchungen in ihrem begrenzten Freundeskreis führen zu vier Verdächtigen, die alle gemeinsam mit Emmy auf der Fete waren: David, Luitzen, Egbertus und Jan. David hätte ein Motiv, falls Emmy während oder unmittelbar nach der Fete die Beziehung mit ihm beendet hätte. Das Motiv von Luitzen, Egbertus und Jan könnte Abweisung oder Eifersucht sein. Unter den Mathematikstudenten ist bekannt, dass jeder der drei schon länger versucht Emmy für sich zu gewinnen. Auch während der Fete versuchten sie wieder mit Emmy zu flirten. Sie versuchten die Instabilität der Beziehung zwischen Emmy und David zu nutzen. Letzterer besitzt übrigens eine eindrucksvolle Figur sodass er wohl von keinem der anderen Verdächtigen überwältigt werden könnte. Emmy hingegen wäre ein leichtes Opfer für jeden der vier Verdächtigen.

Während der Ermittlungen zaubert David einen an ihn gerichteten Brief von Emmy hervor. Er ist auf den Vortag des Festes datiert. Emmy fordert David in dem Brief mit den liebevollsten Worten dazu auf, mit ihr am darauffolgenden Tag zur besagten Fete zu gehen. Der Brief ist getippt, jedoch handschriftlich mit „♡*Emmy*“ unterzeichnet. Die Ermittler vermuten jedoch, dass David den Brief gefälscht hat, um sich zu entlasten.

- *In Aufgabe 2 untersuchst du, ob die Unterschrift unter Emmys Brief gefälscht ist und damit der Brief nicht echt ist.*

Informationen verschiedener Augenzeugen, die die Verdächtigen während ihrer Heimwege gesehen haben, ermöglichen es, die Zahl der Verdächtigen durch logische Schlussfolgerungen zu reduzieren.

- *In Aufgabe 3 wirst du diese logische Analyse durchführen.*

Schließlich versuchen die Ermittler anhand von Daten, wie der Länge des Gürtels, herauszufinden, welcher der noch Verdächtigen am wahrscheinlichsten der Täter ist.

- *In Aufgabe 4 machst du dir die Technik der Bayes'schen Netze[1] zu eigen, um hiermit aus gegebenen Informationen die Wahrscheinlichkeit, mit der ein bestimmter Verdächtiger schuldig ist, zu ermitteln.*

Aus all diesen Indizien folgt schließlich, dass einer der Studenten höchstwahrscheinlich der Täter ist.

- *Schreibe unten auf dein Antwortblatt, wer der Meinung deiner Gruppe nach der Täter ist.*

Die Aufgaben 1 bis 4 können unabhängig voneinander bearbeitet werden!

Gib die Antworten zu den Aufgaben 1,2 und 4 als gekürzten Bruch an.

Schreibe alle Antworten auf das Antwortblatt (Seite 8)

1. Nach dem jüngsten Flugzeugunglück in Tripoli ist die Opferindentifizierung durch das NFI beispielsweise mithilfe solcher Netze erfolgt.

1. Ist Emmy die Leiche aus dem Rhein?

Im Rhein wird eine weibliche Leiche gefunden. Es gibt nur zwei Vermisste, v und d, die die Leiche sein könnten.

- v ist Emmy.
- d ist eine Obdachlose aus Köln, die von einem Tag auf den anderen verschwunden war.

Über die Obdachlose ist nichts bekannt, aber du darfst annehmen, dass sie keine Verwandte von Emmy war. Das verfügbare Beweismaterial (E, E_v) ist folgendes:

- E: Die Blutgruppe der Leiche ist AB.
- E_v: die Blutgruppen von Emmys Eltern sind bekannt, nämlich A (Vater) und B (Mutter).

H_v sei die Hypothese, dass Emmy die Leiche ist, und H_d die Hypothese, dass die Obdachlose die Leiche ist. Wir suchen die Wahrscheinlichkeit $\mathrm{P}(H_v \mid E \cap E_v)$, dass bei gegebenem Beweismaterial (E, E_v), die gefundene Leiche von Emmy ist. Wir nehmen an, dass $\mathrm{P}(H_v \mid E_v) = \mathrm{P}(H_d \mid E_v) = \frac{1}{2}$ ist; ohne die Information über die Blutgruppe der Leiche ist es gleichwahrscheinlich, dass die Leiche von Emmy oder der Obdachlosen ist.

Hintergrundinformation über Blutgruppen
Deine Blutgruppe wird durch die Kombination zweier Allele bestimmt, wovon eines von deiner Mutter und das andere von deinem Vater stammt. Allele sind die möglichen Ausprägungsformen eines Gens, das sich an einem bestimmten Ort im Erbmaterial DNA befindet. Die Kombination der beiden Allele bildet den *Genotyp* deiner Blutgruppe. Das Gen besitzt drei mögliche Allele: A, B und 0. Für den Genotyp der Kombination aus A und B schreiben wir kurz: AB usw. Aus jedem möglichen Genotyp ergibt sich nun eine bestimmte Blutgruppe, wie folgt:

Genotyp	*Blutgruppe*
AB	AB
AA	A
AO	A
BB	B
BO	B
OO	O

In der Bevölkerung treten die drei relevanten Allele mit den folgenden relativen Häufigkeiten p auf:

$$\begin{aligned} \mathrm{p}(\mathsf{O}) &= 2/3; \\ \mathrm{p}(\mathsf{A}) &= 1/4; \\ \mathrm{p}(\mathsf{B}) &= 1/12. \end{aligned}$$

Die Wahrscheinlichkeit P für einen bestimmten Genotyp wird gegeben durch („Hardy-Weinberg-Gleichgewicht“)

$$\begin{aligned} \mathrm{P}(\mathsf{OO}) &= \mathrm{p}(\mathsf{O})^2; \\ \mathrm{P}(\mathsf{OA}) &= 2\,\mathrm{p}(\mathsf{O})\,\mathrm{p}(\mathsf{A}), \end{aligned}$$

usw.

Aufgabe
(a) *Wie groß ist die Wahrscheinlichkeit* $\mathrm{P}(E \mid H_d \cap E_v)$*?*
(b) *Wie groß ist die Wahrscheinlichkeit* $\mathrm{P}(E \mid H_v \cap E_v)$*?*
(c) *Wie groß ist die Wahrscheinlichkeit* $\mathrm{P}(H_v \mid E \cap E_v)$*, dass Emmy die gefundene Leiche ist?*

2. Hat David den ♡-Brief von Emmy selbst geschrieben?

Echte Unterschriften – wie zum Beispiel handgeschriebene Namen – einer bestimmten Person enthalten immer ein gewisses Maß an Variation. Ein laienhafter Fälscher weiß dies jedoch nicht und imitiert daher eine gegebene Unterschrift – die in seinem Besitz ist – so genau wie möglich, um seine Fälschung echt aussehen zu lassen. Wenn eine verdächtige Unterschrift nun einer erwiesenermaßen echten Unterschrift zu sehr gleicht, so ist erstere wahrscheinlich gefälscht.

Die Ermittler verfügen über 100 verschiedene echte Unterschriften von Emmy. Außerdem haben sie bei einer Durchsuchung von Davids Wohnung einen älteren von Emmy unterzeichneten Brief an David gefunden, an dessen Echtheit ebenfalls nicht gezweifelt wird. Man möchte nun wissen, ob David die Unterschrift auf dem Liebesbrief kurz vor dem Fest von diesem älteren Brief kopiert hat. Dies wirst du untersuchen mithilfe einer Methode, die der Logiker und Philosoph Charles S. Pierce (1839 – 1914) im Jahr 1865 bei einem Erbschaftsstreit eingeführt hat (dieser ‚Howland Will Case' war der erste Rechtsfall in der Geschichte, bei dem mathematische Wahrscheinlichkeitsberechnungen eine Rolle gespielt haben).

Wir digitalisieren alle 101 verfügbaren, echten Unterschriften von Emmy, sowie die umstrittene Unterschrift. Wir überführen hierzu jede Unterschrift in eine Folge aus tausend Nullen und Einsen. Wir wählen aus jeder dieser digitalen Unterschriften eine Folge von 10 Stellen, die hinreichend weit voneinander entfernt sind, sodass wir annehmen können, dass die Nullen und Einsen an diesen Stellen unabhängig voneinander sind. (Man bedenke nämlich, dass die Art, einen Buchstaben zu schreiben, in der Regel von den Buchstaben unmittelbar davor oder danach beeinflusst werden kann. Für Buchstaben die weit voneinander entfernt liegen, ist dieser Einfluss allerdings vernachlässigbar.) Im Folgenden nennen wir eine solche Folge von Nullen und Einsen einfach „die Unterschrift" (obwohl es sich in Wahrheit nur um einen kleinen Teil der digitalisierten Unterschrift handelt). Auf diese Weise wird Emmys Unterschrift zu eine stochastische Größe $\mathbf{H}$, die als mögliche Werte h jede Folge aus 10 Nullen und Einsen annehmen kann (das sind insgesamt also $2^{10} = 1024$ Werte).

Die Unterschrift, die David möglicherweise gefälscht hat, stimmt an 9 von 10 Stellen mit der Unterschrift auf dem älteren Brief, der in Davids Besitz war, überein. Wie (un)wahrscheinlich ist diese Tatsache, wenn wir annehmen, dass David die Wahrheit sagt und der umstrittene Brief somit echt ist? Es ist vernünftig in dieser Situation anzunehmen, dass

- in einer gegebenen Unterschrift jede Stelle unabhängig von den anderen Stellen ist,
- die Wahrscheinlichkeit $\mathrm{P}_n(0)$, dass Emmy an der n-ten Stelle eine Null schreibt nicht vom Wert von n abhängt,
- verschiedene Unterschriften von Emmy voneinander unabhängig sind.

Da die Wahrscheinlichkeit $\mathrm{P}_n(0)$ unabhängig von n ist, nennen wir diese im Folgenden einfach P. Eine zweite relevante Wahrscheinlichkeit ist $\mathrm{P}^k_{i,j}$. Dies ist die Wahrscheinlichkeit, dass, wenn wir zwei verschiedene echte Unterschriften i und j vergleichen, an der k-ten Stelle dasselbe Symbol steht (also entweder in beiden Unterschriften eine 0 oder in beiden Unterschriften eine 1). Hierbei laufen die Indizes i und j von 1 bis einschließlich 100, und k von 1 bis einschließlich 10. Den Annahmen zufolge hängt die Wahrscheinlichkeit $\mathrm{P}^k_{i,j}$ nicht von i,j oder k ab (solange natürlich $i \neq j$ ist), so dass wir $\mathrm{P}^k_{i,j}$ im Folgenden einfach Q nennen.

Aufgaben

a) *In den 100 verfügbaren, echten Unterschriften von Emmy kommen genau 250 Nullen vor. Was ist auf Grund dieser Information eine vernünftiger Schätzwert für die Wahrscheinlichkeit* Q*?*

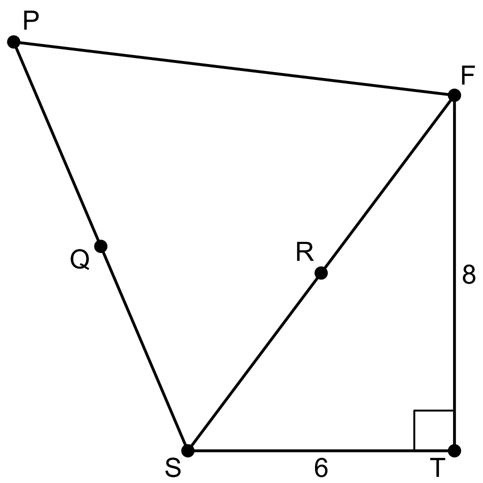
b) *Was ist – auf Grund des soeben gefundenen Wertes für* Q *– die Wahrscheinlichkeit, dass zwei gegebene (verschiedene) echte Unterschriften zufällig an 9 Stellen übereinstimmen?*

3. Wer hat ein Alibi?

Die zuvor erwähnte Fete begann um 20.30 Uhr. Alle vier Mordverdächtigen verließen die Fete zwischen 1.00 Uhr und 1.30 Uhr. Die Ermittlungen haben ergeben, dass der Mord zwischen 1.15 Uhr und 2.00 Uhr am Rheinufer, in unmittelbarer Nähe zum Veranstaltungsort der Festivität, verübt worden sein muss.

Über die vier Verdächtigen (Luitzen, Egbertus, Jan und David) ist bekannt, dass genau einer von ihnen zu Fuß nach Hause ging, einer von ihnen mit dem Fahrrad aufbrach, einer von ihnen den Bus nahm und einer von ihnen mit dem Auto wegfuhr. Der Fußgänger lief nach ein paar Bierchen mit einer Geschwindigkeit von 4 km/h, der Radfahrer fuhr 15 km/h, der Bus 56 km/h, und das Auto fuhr 60 km/h.

Jeder der vier Verdächtigen wurde von einem Zeugen zu einem bestimmten Zeitpunkt gesehen. Es handelt sich um die folgenden Zeitpunkte: 1.28 Uhr, 1.37 Uhr, 2.07 Uhr und 2.20 Uhr.



Auf obigem Plan sind die Orte P, Q, R und S abgetragen, an denen die Verdächtigen gesehen wurden. Nimm an, dass die Personen zu Fuß, mit dem Fahrrad und mit dem Auto jeweils die kürzest mögliche Route zu dem Ort, an dem sie gesehen wurden, genommen haben, der Bus jedoch die zweitkürzeste Route gefahren ist. Die Verdächtigen sind nach ihrem Aufbruch nicht mehr zum Fest zurückgekehrt.

Ferner gilt, dass der Abstand vom Fest F nach P gleich ist dem Abstand von P nach S und ebenfalls gleich dem Abstand von F nach S. Ort Q liegt genau in der Mitte zwischen P und S, und R liegt genau in der Mitte zwischen F und S. Der Winkel $\angle FTS$ ist ein rechter Winkel. Der Abstand von F nach T beträgt 8 km und der von T nach S beträgt 6 km.

Es folgen einige Fakten darüber, an welchen Orten die Verdächtigen gesehen wurden, zu welchen Zeitpunkten das geschah und mit welchem Fahrzeug der jeweilige unterwegs war.

- David wurde an Ort P gesehen.
- Der Fußgänger wurde um 2.20 Uhr bemerkt.
- Luitzen wurde nicht an Ort S gesehen.
- Egbertus wurde mit dem Auto gesehen, doch nicht um 1.28 Uhr.
- Der Fußgänger war nicht an Ort Q.
- An Ort P wurde niemand zu den Zeitpunkten 1.37 Uhr und 2.20 Uhr gesehen.
- An Ort S wurde jemand um 1.28 Uhr gesehen.
- Jan war nicht mit dem Fahrrad unterwegs.

Aufgabe
Schließe aus obiger Information, welche Verdächtigen ein wasserdichtes Alibi haben. Mit anderen Worten: Welche Verdächtigen können den Mord unmöglich begangen haben?

4. Wer hat Emmy erwürgt?

Wir nummerieren die vier des Mordes an Emmy Verdächtigen wie folgt: David ($i = 1$), Luitzen ($i = 2$), Egbertus ($i = 3$), und Jan ($i = 4$). Betrachte die folgenden Ereignisse, wobei $i = 1, 2, 3, 4$:

- H_i: Der Verdächtige i ist der Täter.
- V_i: Der Gürtel am Hals der Leiche stammt von einer Hose des Verdächtigen i.
- D: Der Gürtel am Hals der Leiche stammt von einer Hose des Täters.
- L: Die Länge des Gürtels (in Inch) am Hals der Leiche beträgt $l = 50$.
- T: Die Taillenumfänge (in Inch) der vier Verdächtigen betragen: $(t_1 = 38, t_2 = 31, t_3 = 33, t_4 = 30)$.

Wir haben also fünf Variablen $\mathbf{H}$, $\mathbf{V}$, $\mathbf{D}$, $\mathbf{L}$ und $\mathbf{T}$, wobei $\mathbf{D}$ nur zwei Werte nämlich $\mathbf{D} = D$ und $\mathbf{D} = \overline{D}$ annehmen kann. L steht für $\mathbf{L} = \ell$ und T ist das Ereignis, dass $\mathbf{T}$ den Wert $\mathbf{T} = (t_1, t_2, t_3, t_4)$ annimmt. Das Beweismaterial besteht aus $E = (L, T)$ und wir sind interessiert an den Wahrscheinlichkeiten $\mathrm{P}\left(H_i | E\right)$ (für $i = 1, 2, 3, 4$). Diese folgen aus Formel (9) des Vorbereitungsmaterials: Für $i = 1$ haben wir beispielsweise

$$\mathrm{P}\left(H_1 \mid E\right) = \frac{1}{1 + \frac{\mathrm{P}(E\cap H_2)}{\mathrm{P}(E\cap H_1)} + \frac{\mathrm{P}(E\cap H_3)}{\mathrm{P}(E\cap H_1)} + \frac{\mathrm{P}(E\cap H_4)}{\mathrm{P}(E\cap H_1)}}.$$

Nach den Regeln der Wahrscheinlichkeitsrechnung ist der rechte Term auszurechnen mithilfe von

$$\begin{aligned}
\mathrm{P}\left(E \cap H_i\right) &= \mathrm{P}\left(H_i \cap L \cap T \cap D \cap V_1\right) + \mathrm{P}\left(H_i \cap L \cap T \cap D \cap V_2\right) \\
&+ \mathrm{P}\left(H_i \cap L \cap T \cap D \cap V_3\right) + \mathrm{P}\left(H_i \cap L \cap T \cap D \cap V_4\right) \\
&+ \mathrm{P}\left(H_i \cap L \cap T \cap \overline{D} \cap V_1\right) + \mathrm{P}\left(H_i \cap L \cap T \cap \overline{D} \cap V_2\right) \\
&+ \mathrm{P}\left(H_i \cap L \cap T \cap \overline{D} \cap V_3\right) + \mathrm{P}\left(H_i \cap L \cap T \cap \overline{D} \cap V_4\right).
\end{aligned}$$

Das sieht kompliziert aus, weil die fünf Variablen voneinander abhängen. Wie geht es weiter?

Bedingte Wahrscheinlichkeiten
Wie du weißt – siehe Formel (4) aus dem Vorbereitungsmaterial – gilt für zwei Ereignisse

$$\mathrm{P}\left(A \cap B\right) = \mathrm{P}\left(A\right) \cdot \mathrm{P}\left(B \mid A\right).$$

Für drei Ereignisse gilt

$$\mathrm{P}\left(A \cap B \cap C\right) = \mathrm{P}\left(A\right) \cdot \mathrm{P}\left(B \mid A\right) \cdot \mathrm{P}\left(C \mid A \cap B\right).$$

Für viele Ereignisse $A, B, \ldots, Y, Z$ bekommst du

$$\mathrm{P}\left(A \cap \cdots \cap Z\right) = \mathrm{P}\left(A\right) \cdot \mathrm{P}\left(B \mid A\right) \cdot \mathrm{P}\left(C \mid A \cap B\right) \cdots \mathrm{P}\left(Z \mid A \cap \cdots \cap Y\right).$$

Ein Nachteil dieser allgemeinen Formel ist, dass der rechte Term von der Reihenfolge $A, \ldots, Z$, in der die Ereignisse aufgezählt wurden, abzuhängen *scheint*. Dass da also $\mathrm{P}\left(A\right)$ steht wird den tatsächlichen (Un)abhängigkeiten innerhalb des Problems nicht gerecht, da A sehr wohl von (einigen) anderen Ereignissen abhängen kann. Und da steht $\mathrm{P}\left(C | A \cap B\right)$, während C vielleicht durchaus von A und D, doch ausgerechnet von B nun gerade nicht abhängt. Kurzum, so wahr die Formel für $\mathrm{P}\left(A \cap \cdots \cap Z\right)$ auch ist, veranschaulicht sie uns kaum, wie die Wahrscheinlichkeit am besten ausgerechnet werden kann. Um die tatsächlichen (Un)abhängigkeiten innerhalb des Problems ins Spiel zu bringen, greift man auf die *Bayes'schen Netze* zurück.

Bayes'sche Netze
Ein Bayes'sches Netz ist ein Diagramm, in dem die Variablen $\mathbf{A}, \ldots, \mathbf{Z}$ eines bestimmten Problems durch Pfeile verbunden werden. Dabei ist es üblich, den Namen der Variablen nicht als $\mathbf{X}$ sondern als Ⓧ zu schreiben. Zeigt ein direkter Pfeil von Ⓧ nach Ⓨ, so wird $\mathbf{X}$ als Elternknoten von $\mathbf{Y}$ bezeichnet. Zu jeder Variablen $\mathbf{X}$ gehört also eine Menge $Eltern[\mathbf{X}]$, die aus den Eltern von $\mathbf{X}$ besteht. Beispiele:

(i) Für das Netz Ⓐ $\rightarrow$ Ⓑ gilt $Eltern[\mathbf{B}] = \{\mathbf{A}\}$, während $\mathbf{A}$ keine Eltern hat.
(ii) Beim pfeillosen Netz Ⓐ Ⓑ sind $\mathbf{A}$ und $\mathbf{B}$ elternlos (oder auch Waisen).

(iii) Im Netz $A \rightarrow C \leftarrow B$ gilt $Eltern[\mathbf{C}] = \{\mathbf{A}, \mathbf{B}\}$, während **A** und **B** Waisen sind.

Die Pfeile im Netz werden durch eine gegebene Wahrscheinlichkeitsfunktion P auf dem Ereignisraum bestimmt, der wiederum durch die Variablen $\mathbf{A}, \ldots, \mathbf{Z}$ bestimmt wird. Dieser Ereignisraum besteht aus allen Kombinationen der Form $(\mathbf{A} = a, \ldots, \mathbf{Z} = z)$, wobei a ein möglicher Wert für **A** ist, usw. Wir schreiben das Ereignis $\mathbf{A} = a$ wie üblich als A, usw. Ein korrektes Bayes'sches Netz für eine gegebene Wahrscheinlichkeitsfunktion P erfüllt für alle Werte $a, \ldots, z$ der Variablen, also für alle möglichen Ereignisse $A, \ldots, Z$ folgende Gleichung:

$$\mathrm{P}\left(A \cap \ldots \cap Z\right) = \mathrm{P}\left(A \mid Eltern[A]\right) \cdots \mathrm{P}\left(Z \mid Eltern[Z]\right). \qquad (*)$$

Die rechte Seite von $(*)$ ist das Produkt der Faktoren $\mathrm{P}\left(X \mid Eltern[X]\right)$, für jedes X, das in der Menge $A, \ldots, Z$ vorkommt. Zur Bestimmung dieser Faktoren gilt folgende Vorgehensweise:

(1) Bestimme zuerst $Eltern[\mathbf{X}]$; dabei handelt es sich um einen Teil der Menge der Variablen $\mathbf{A}, \ldots, \mathbf{Z}$ (oder leere Menge).
(2) Ersetze dann jede Variable **Y** aus der Menge $Eltern[\mathbf{X}]$ durch das entsprechende Ereignis Y (d.h. $\mathbf{Y} = y$). Die so erhaltene Menge heißt $Eltern[X]$.

Ist die Menge $Eltern[\mathbf{X}]$ leer, so wird $\mathrm{P}\left(X | Eltern[X]\right)$ schlichtweg als $\mathrm{P}\left(X\right)$ gelesen.

Der große Vorteil von $(*)$ gegenüber der vorherigen Formel für $\mathrm{P}\left(A \cap \cdots \cap Z\right)$ ist, dass alle Variablen auf dieselbe Weise behandelt werden, und dass alle Abhängigkeiten zum Ausdruck kommen.

Beispiele
Um diese Vorschrift zu veranschaulichen, kommen wir zurück auf die vorherigen Beispiele **(i) – (iii)**.

(i) Der linke Term von $(*)$ ist gleich $\mathrm{P}\left(A \cap B\right)$. Der rechte Term ist $\mathrm{P}\left(A \mid Eltern[A]\right) \cdot \mathrm{P}\left(B \mid Eltern[B]\right)$. Weil **A** keine Eltern hat, gilt $\mathrm{P}\left(A \mid Eltern[A]\right) = \mathrm{P}\left(A\right)$. Ferner hat **B** als Elternknoten **A**, so dass $\mathrm{P}\left(B \mid Eltern[B]\right) = \mathrm{P}\left(B \mid A\right)$. In diesem Fall liefert $(*)$ also die wohlbekannte Formel für die bedingte Wahrscheinlichkeit.
(ii) Nun gilt $\mathrm{P}\left(A \mid Eltern[A]\right) = \mathrm{P}\left(A\right)$ und $\mathrm{P}\left(A \mid Eltern[B]\right) = \mathrm{P}\left(B\right)$, da **A** und **B** beide keine Eltern haben. In diesem Fall liefert $(*)$ die Formel $\mathrm{P}\left(A \cap B\right) = \mathrm{P}\left(A\right) \cdot \mathrm{P}\left(B\right)$, und damit die Definition der Unabhängigkeit: siehe Gleichung (2) aus dem Vorbereitungsmaterial.
(iii) Nun lautet $(*)$

$$\mathrm{P}\left(A \cap B \cap C\right) = \mathrm{P}\left(A\right) \cdot \mathrm{P}\left(B\right) \cdot \mathrm{P}\left(C \mid A \cap B\right).$$

Die Variablen A und B haben schließlich wieder keine Eltern, während die beiden zusammen die Eltern von C bilden. Es gilt also $\mathrm{P}\left(C \mid Eltern[C]\right) = \mathrm{P}\left(C \mid A \cap B\right)$.

Das Bayes'sche Netz, dass die Situation im Mordfall Emmy beschreibt, sieht wie folgt aus:

$$\begin{tikzcd}
H \arrow[r] & V \arrow[d] & D \arrow[l] \\
 & L & T \arrow[l]
\end{tikzcd}$$

Die Wahrscheinlichkeiten $\mathrm{P}\left(E \cap H_i\right)$ können mithilfe des gegebenen Bayes'schen Netzes und $(*)$ ausgerechnet werden. Um das zu tun, muss man über eine ganze Reihe von (bedingten) Wahrscheinlichkeiten verfügen.

(a) Stelle erst für dich selbst eine Liste aller (bedingten) Wahrscheinlichkeiten auf, die benötigt werden, um alle Terme in den Wahrscheinlichkeiten $\mathrm{P}\left(E \cap H_i\right)$ auszurechnen.
(b) Überlege, durch logisches Nachdenken über die Bedeutung der Variablen, welche dieser (bedingten) Wahrscheinlichkeiten direkt bestimmt werden können.
(c) Gehe davon aus, dass der Täter seinen eigenen Gürtel benutzt hat.
(d) Gehe davon aus, dass die Verdächtigen zunächst alle im gleichen Maße verdächtig sind. Greife also nicht auf die vorherigen Aufgaben zurück!

(e) Das Verhältnis zwischen Gürtellänge und Taillenumfang ist wie folgt:

Taillenumfang t	$\mathrm{P}(t \mid \ell = 50)$	$\mathrm{P}(\ell = 50 \mid t)$
30	$1/10$	$11/100$
31	$1/10$	$1/10$
33	$1/4$	$7/40$
38	$1/3$	$2/3$

(f) Die Taillenumfänge der Verdächtigen kommen mit den folgenden Wahrscheinlichkeiten unter deutschen Studenten vor: $\mathrm{P}(30) = 1/11$, $\mathrm{P}(31) = 1/10$, $\mathrm{P}(33) = 1/7$, $\mathrm{P}(38) = 1/20$.

Aufgabe
Berechne auf Grundlage dieser Informationen die Wahrscheinlichkeiten $\mathrm{P}(H_1 \mid E), \ldots, \mathrm{P}(H_4 \mid E)$.

Lösungen zu „Sum of Us 2010“

Aufgabe 1:

a) $P(E|H_d \cap E_v) = P(E|H_d) = P(E) = 2 \cdot \frac{1}{4} \cdot \frac{1}{12} = \frac{1}{24}$.

b) Gesucht wird die Wahrscheinlichkeit für das Ereignis, dass Emmy die Blutgruppe AB hat, wenn ihr Vater die Blutgruppe A und ihre Mutter die Blutgruppe B haben. Dazu betrachten wir die möglichen Genotypen der Eltern (AA,A0,BB,B0). So ergibt sich:

$$P(E|H_v \cap E_v)$$

$$= \frac{\text{Wahrscheinlichkeit, dass die Tochter Blutgruppe AB hat}}{\text{Wahrscheinlichkeit für alle möglichen Blutgruppen der Tochter}}$$

$$= \frac{1 \cdot P(AA)P(BB) + \frac{1}{2} \cdot P(A0)P(BB) + \frac{1}{2} \cdot P(AA)P(B0) + \frac{1}{4} \cdot P(A0)P(B0)}{P(AA)P(BB) + P(A0)P(BB) + P(AA)P(B0) + P(A0)P(B0)}$$

$$= \frac{1 \cdot \frac{1}{16} \cdot \frac{1}{144} + \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{144} + \frac{1}{2} \cdot \frac{1}{16} \cdot \frac{1}{9} + \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{9}}{\frac{1}{16} \cdot \frac{1}{144} + \frac{1}{3} \cdot \frac{1}{144} + \frac{1}{16} \cdot \frac{1}{9} + \frac{1}{3} \cdot \frac{1}{9}} = \frac{99}{323}$$

c) Wir benutzen Formel 14 aus dem Vorbereitungsmaterial.

$$P(H_v|E \cap E_v) = \frac{1}{1 + \frac{P(E|H_v \cap E_v)}{P(E|H_d \cap E_v)}} = \frac{1}{1 + \frac{\frac{1}{24}}{\frac{99}{323}}} = \frac{1}{1 + \frac{323}{2376}} = \frac{2376}{2699}$$

Aufgabe 2:

a) Wir addieren die Wahrscheinlichkeiten für die Ereignisse, dass an der Stelle k bei zwei verschiedenen Unterschriften entweder in beiden Unterschriften eine 0 oder in beiden Unterschriften eine 1 steht:

$$\left(\frac{1}{4}\right)^2 + \left(\frac{3}{4}\right)^2 = \frac{5}{8}$$

b) Mit Hilfe der Binomialverteilung erhalten wir:

$$\binom{10}{9} \cdot \left(\frac{5}{8}\right)^9 \cdot \left(\frac{3}{8}\right)^1 = \frac{58593750}{1073741824}$$

Aufgabe 3:

Mit Hilfe der gegebenen Informationen von a) bis h) füllen wir die nachfolgende Tabelle aus:

Name	Ort	Zeitpunkt der Beobachtung	Verkehrsmittel
Luitzen	R_{10}	$(2:20)_7$	Fußgänger$_6$
Egbertus	Q_9	$(1:37)_8$	Auto$_2$
Jan	S_3	$(1:28)_4$	Bus$_{12}$
David	P_1	$(2:07)_5$	Fahrrad$_{11}$

Dabei gilt:
$a) \Rightarrow 1$
$d) \Rightarrow 2$
$c) \wedge d) \wedge g) \Rightarrow 3$
$g) \Rightarrow 4$
$f) \Rightarrow 5$
$b) \Rightarrow 6 \wedge 7 \wedge 8$
$e) \Rightarrow 9 \wedge 10$
$h) \Rightarrow 11 \wedge 12$

Als Abfahrtszeitpunkte ergeben sich:

- Luitzen verlässt die Feier um 1:05 Uhr.
- Egbertus verlässt die Feier um 1:22 Uhr.
- Jan verlässt die Feier um 1:13 Uhr.
- David verlässt die Feier um 1:27 Uhr.

Damit haben Luitzen und Jan ein wasserdichtes Alibi.

Aufgabe 4:

Mit Hilfe des Bayes'schen Netzes können wir den folgenden Ausdruck umschreiben:

$$P(H_i \cap L \cap T \cap D \cap V_j) = P(H_i) \cdot P(L|V_j \cap T) \cdot P(T) \cdot P(D) \cdot P(V_j|H_i \cap D)$$

Damit ist

$$P(E \cap H_i) \overset{P(D)+P(\overline{D})=1}{=} \sum_{j=1}^{4} P(H_i) \cdot P(L|V_j \cap T) \cdot P(T) \cdot P(V_j|H_i \cap D)$$

$$= P(T) \cdot P(H_i) \cdot \sum_{j=1}^{4} P(L|V_j \cap T) \cdot P(V_j|H_i \cap D) = P(T) \cdot P(H_i) \cdot P(L|V_i \cap T)$$

Somit ergibt sich zum Beispiel für

$$\frac{P(E\cap H_2)}{P(E\cap H_1)}=\frac{P(T)\cdot P(H_2)\cdot P(L|V_2\cap T)}{P(T)\cdot P(H_1)\cdot P(L|V_1\cap T)}=\frac{P(L|V_2\cap T)}{P(L|V_1\cap T)},$$

sodass wir insgesamt erhalten:

$$P(H_1|E)=\frac{1}{1+\frac{P(L|V_2\cap T)}{P(L|V_1\cap T)}+\frac{P(L|V_3\cap T)}{P(L|V_1\cap T)}+\frac{P(L|V_4\cap T)}{P(L|V_1\cap T)}}\overset{\text{siehe Tabelle}}{=}\frac{1}{1+\frac{\frac{1}{10}+\frac{7}{40}+\frac{11}{100}}{\frac{2}{3}}}=\frac{400}{631}$$

Analog folgt:

$$P(H_2|E)=\frac{60}{631}$$

$$P(H_3|E)=\frac{105}{631}$$

$$P(H_4|E)=\frac{66}{631}$$