Induktive Statistik

Übungsaufgaben mit Musterlösungen

Christian Reinboth

Inhaltsverzeichnis

Inhaltsverzeichnis 2
Vorwort und Einführung 3
Aufgabenblatt I: Venn-Diagramme und Pfaddiagramme 4
Aufgabenblatt II: Stochastisch unabhängige Ereignisse 11
Aufgabenblatt III: Rechnen mit Wahrscheinlichkeiten 15
Aufgabenblatt IV: Rechnen mit dem Satz von Bayes 18
Theorie-Einschub: Das Taxi-Problem und der Satz von Bayes 22
Aufgabenblatt V: Rechnen mit dem Satz von Bayes II 27
Aufgabenblatt VI: Hypergeometrische und Poisson-Verteilung 32
Aufgabenblatt VII: Poisson- und Bionomialverteilung 39
Aufgabenblatt VIII: Normal-, Gleich- und Exponentialverteilung 46
Aufgabenblatt IX: Normal-, Bionomial- und Poissonverteilung 51
Aufgabenblatt X: Konfidenzintervalle 56
Aufgabenblatt XI: Statistische Testverfahren 63
Aufgabenblatt XII: Alle Themenbereiche 70

Vorwort und Einführung

Während meines Studiums der Wirtschaftsinformatik an der Hochschule Harz hatte ich in den Jahren 2002 bis 2005 das große Glück, als Tutor für verschiedene Statistik-Vorlesungen von Dr. Walter Strube tätig werden zu dürfen. Zu Beginn meiner Tutorentätigkeit konnte ich eine über die Jahre von Tutor zu Tutor weiterentwickelte Aufgabensammlung in Form einer „Lose-Blatt-Sammlung“ von meiner direkten Vorgängerin, Frau Weinert, übernehmen. Um den Studierenden die Arbeit mit den Übungsaufgaben zu erleichtern, habe ich damals für sämtliche Aufgaben Musterlösungen und Erläuterungen verfasst und den Teilnehmern meiner Tutorien in Form von Handouts zur Verfügung gestellt.

In den vergangenen sieben Jahren bin ich immer wieder von Studierenden nach dieser Sammlung von Aufgaben mit Musterlösungen gefragt worden, so dass ich mich dazu entschlossen habe, sie noch einmal zu überarbeiten, mit einigen weiteren Erläuterungen zu versehen und über den GRIN-Verlag sowie verschiedene Webseiten als kostenfreien Download zur Verfügung zu stellen. Ich hoffe, dass sie so vielleicht auch für den einen oder anderen Studierenden außerhalb unserer Hochschule von Nutzen sein kann.

Ein wichtiger Hinweis: Obwohl sich in den Lösungen zu vielen Aufgaben Formeln und Hinweise zur Interpretation der Ergebnisse finden, ersetzt dieses Manuskript keinesfalls eine Einführung in die Theorie der induktiven Statistik. Wer also nicht nur Aufgaben lösen, sondern auch die theoretischen Grundlagen verstehen möchte, dem seien von meiner Seite die beiden Lehr- und Übungsbücher von Prof. Dr. Frank Lammers (ebenfalls Hochschule Harz) wärmstens empfohlen, die im GUC-Verlag Chemnitz erschienen sind.

Lammers, Frank: Statistik I: Deskriptive und explorative Statistik – Lehr- und Übungsbuch, GUC-Verlag der Gesellschaft für Unternehmensrechnung und Controlling, Chemnitz, 2003.

Lammers, Frank: Statistik II: Wahrscheinlichkeitsrechnung und Interferenzstatistik – Lehr und Übungsbuch, GUC-Verlag der Gesellschaft für Unternehmensrechnung und Controlling, Chemnitz, 2004.

Und noch ein weiterer Hinweis: Wie bereits erwähnt, basieren die Aufgaben in diesem Manuskript zum Großteil auf Aufgaben, die ich 2002 von meinen Tutorien-Vorgängern übernommen und teilweise überarbeitet und angepasst habe. Die Aufgabenideen stammen daher zum Großteil nicht von mir, die Musterlösungen sowie die begleitenden Hinweise zu den theoretischen Grundlagen entstammen dagegen meiner Feder.

Allen Studierenden wünsche ich bei der Bearbeitung der Übungsaufgaben einen maximalen Lernerfolg. Für Fragen zu einzelnen Aufgaben stehe ich per E-Mail unter creinboth@hs-harz.de grundsätzlich gerne zur Verfügung, bitte bei allen Anfragen aber zu bedenken, dass es durchaus einige Tage dauern kann, bis ich die Zeit für deren Beantwortung finde.

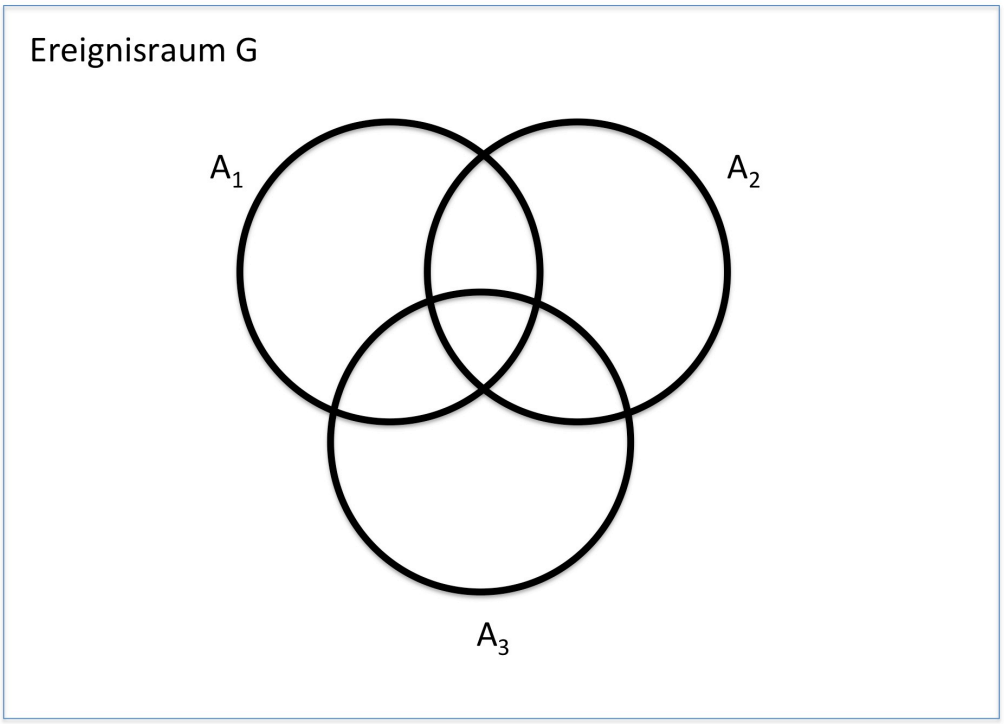
Wernigerode, den 03.11.2013

Christian Reinboth

Aufgabenblatt I: Venn-Diagramme und Pfaddiagramme

Aufgabe 1: Die Menge aller Studenten an einer kleinen Fachhochschule bildet den Ereignisraum G. Das Ereignis A_i beschreibt die Anzahl der Studenten, die einen der drei angebotenen Statistikkurse (mit i = 1,2,3) belegen. Wie sind vor diesem Hintergrund die nachfolgend aufgelisteten Ereignisse verbal und als Venn-Diagramm zu beschreiben?

(a) $A_1 \cup A_2 \cup A_3$
(b) $A_1 \cap A_2 \cap A_3$
(c) $(\overline{A_1 \cup A_2})$
(d) $(A_1 \cap A_2 \cap \overline{A_3})$
(e) $(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3})$



Aufgabe 2: Gegeben sind folgende Mengen und die Anzahl der darin enthaltenen Elemente:

N(G) = 1.000
$N(A_1) = 380$
$N(A_2) = 400$
$N(A_3) = 470$

$N\ (A_1 \cup A_2 \cup A_3) = 950$
$N\ (A_1 \cap A_2 \cap A_3) = 50$
$N\ (A_1 \cap A_2) = 80$
$N\ (A_1 \cap A_3) = 150$

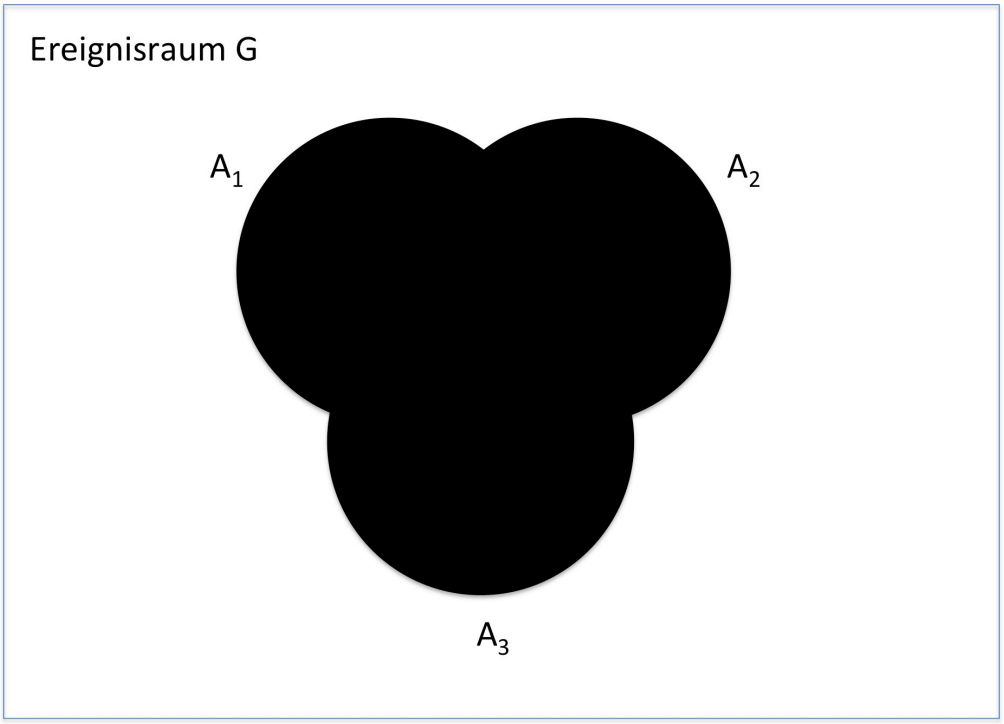
Stellen Sie diese Größen in einem Venn-Diagramm dar und vervollständigen Sie dieses.

Aufgabe 3: Statistikstudent C.R. bewirbt sich bei zwei Unternehmen – A und B – um einen Praktikumsplatz. Die Wahrscheinlichkeit des Bewerbungserfolgs schätzt er bei Firma A mit 0,5 und Firma B mit 0,6 ein. Weiterhin rechnet er mit einer Wahrscheinlichkeit von 0,3 von beiden Unternehmen gleichzeitig eine Zusage für einen Praktikumsplatz zu erhalten.

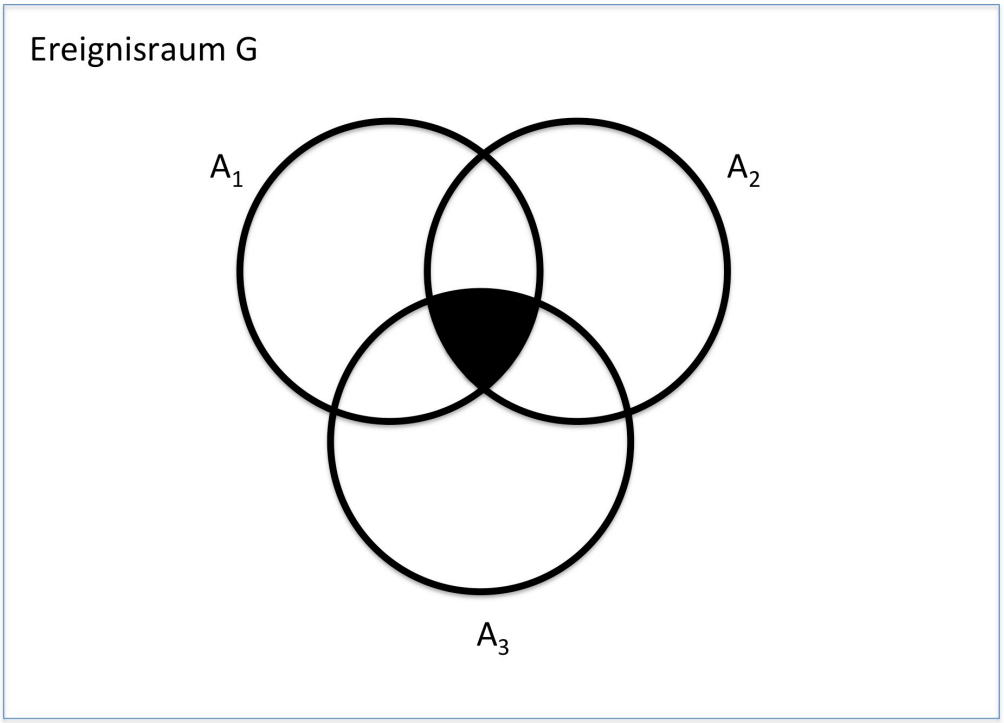
Wie groß ist die Wahrscheinlichkeit, von mindestens einer Firma eine Zusage zu erhalten?

Lösung zu Aufgabe 1:

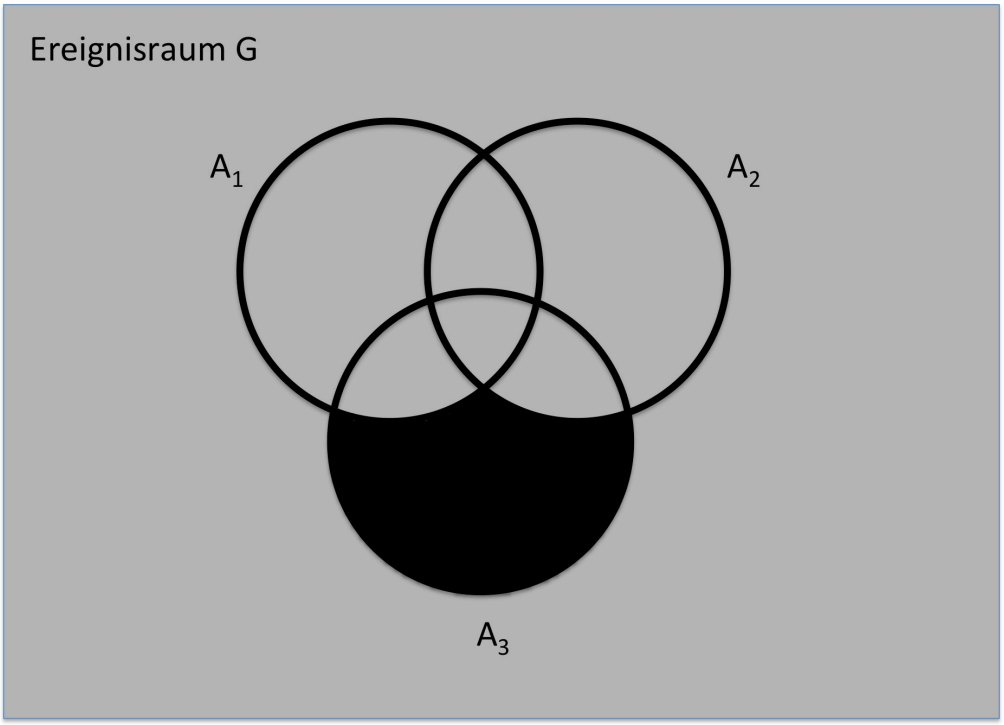
(a) $A_1 \cup A_2 \cup A_3$ = Studenten, die mindestens einen der drei Kurse belegen



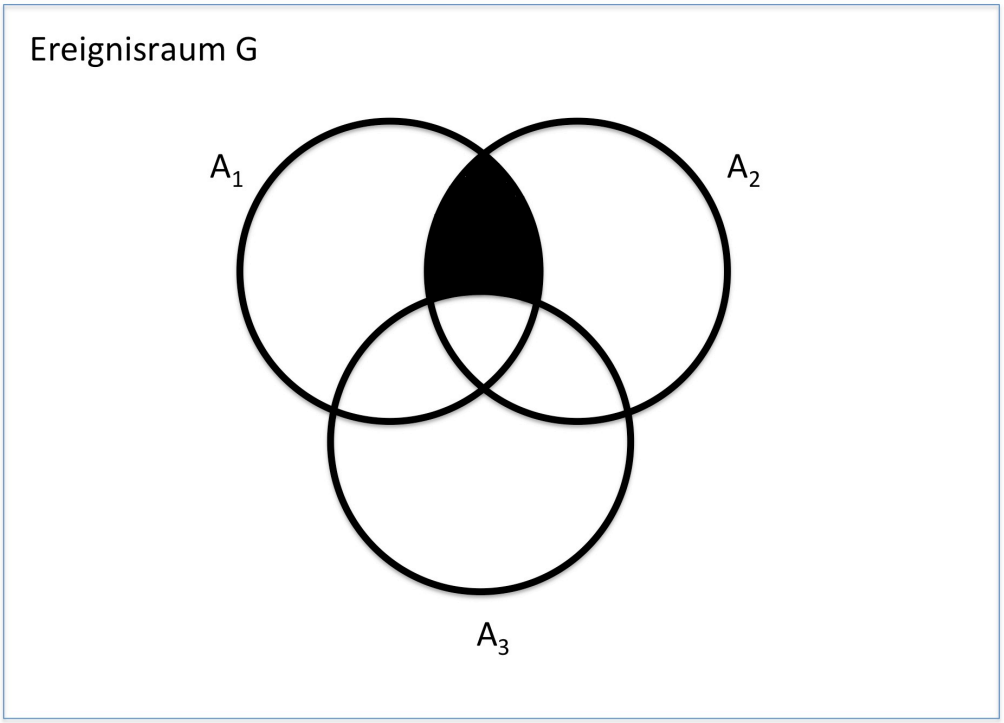
(b) $A_1 \cap A_2 \cap A_3$ = Studenten, die alle drei angebotenen Kurse belegen



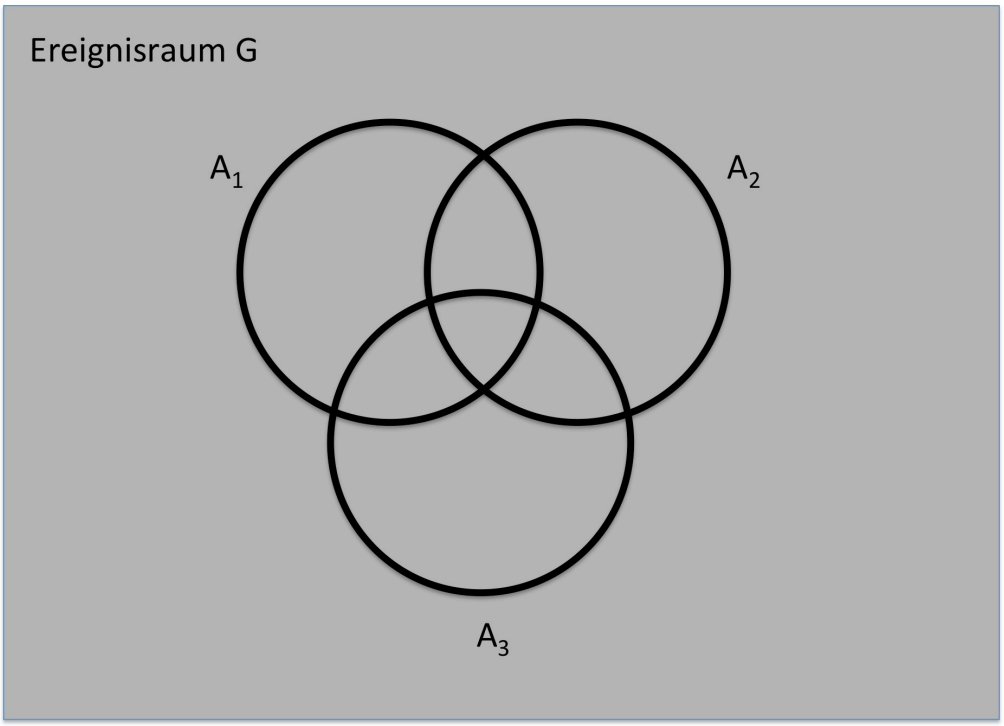
(c) $(\overline{A_1 \cup A_2})$ = Studenten, die weder Kurs 1 noch Kurs 2 belegen
(das Ausgrauen des Ereignisraums G zeigt dessen Einbeziehung)



(d) $(A_1 \cap A_2 \cap \overline{A_3})$ = Studenten, die Kurs 1 und 2, nicht aber Kurs 3 belegen



(e) $(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3})$ = Studenten, die keinen der drei Kurse belegen
(das Ausgrauen des Ereignisraums G zeigt dessen Einbeziehung)



Lösung zu Aufgabe 2:

Vier der noch fehlenden Größen sind vergleichsweise einfach zu ermitteln:

$N\ (A_1 \cap A_2 \cap \overline{A_3}) = \ 80 - 50 = 30$
$N\ (A_1 \cap \overline{A_2} \cap A_3) = \ 150 - 50 = 100$
$N\ (A_1 \cap \overline{A_2} \cap \overline{A_3}) = 380 - 30 - 50 \ - 100 = 200$
Für den Außenraum ergibt sich zudem: 1000 – 950 = 50

Ein wenig komplizierter ist die Kalkulation der verbliebenen Größen. Hier ist zunächst $N\ (A_2 \cap A_3)$ zu ermitteln, was über den Additionssatz für beliebige Ereignisse geschieht. Dieser Additionssatz lautet:

$$P\ (A_1 \cup A_2 \cup A_3) = P(A_1) + \ P(A_2) + P(A_3) - \ P\ (A_1 \cap A_2) - \ P\ (A_2 \cap A_3) - \ P(A_1 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$$

Gesucht wird $P\ (A_2 \cap A_3)$ (was zugleich $N\ (A_2 \cap A_3)$ entspricht). Da alle anderen Größen bekannt sind, ist der Additionssatz nach $P\ (A_2 \cap A_3)$ umzustellen und aufzulösen.

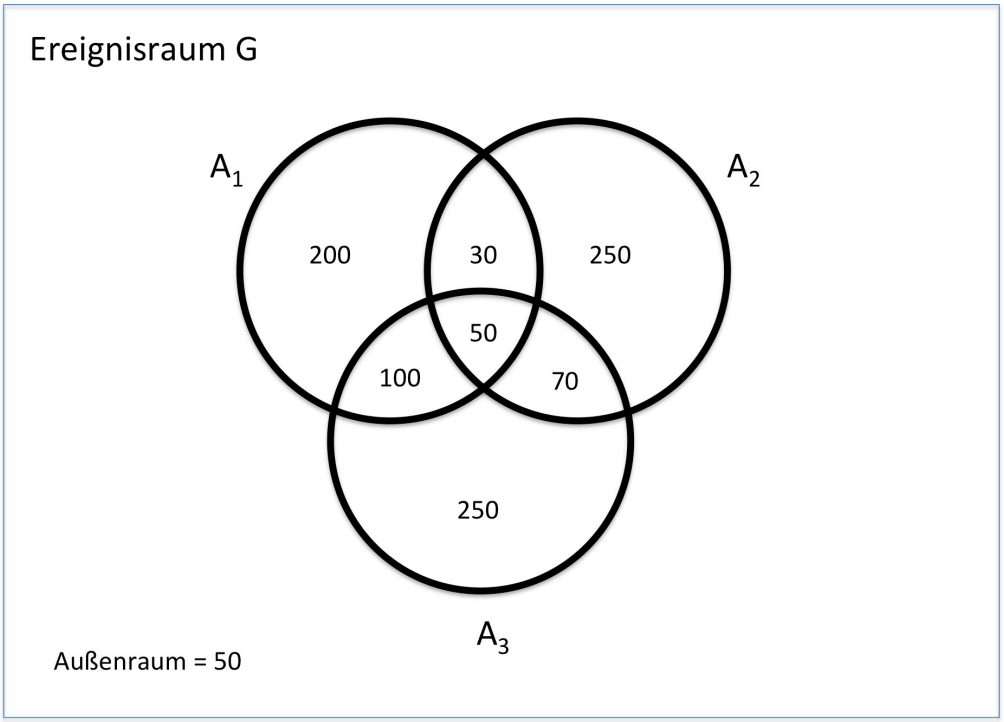
$$P(A_2 \cap A_3) = \ P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) + P(A_1 \cap A_2 \cap A_3) - P(A_1 \cup A_2 \cup A_3)$$

$$P(A_2 \cap A_3) = 380 + 400 + 470 - 80 - 150 + 50 - 950 = 120$$

Mit dieser Angabe lassen sich nun alle noch fehlenden Größen berechnen:

$N\ (\overline{A_1} \cap A_2 \cap A_3) = \ 120 - 50 = 70$
$N\ (\overline{A_1} \cap \overline{A_2} \cap A_3) = \ 470 - 100 - 50 - 70 = 250$
$N\ (\overline{A_1} \cap A_2 \cap \overline{A_3}) = \ 400 - 50 - 70 - 30 = 250$

Anschließend sind alle gefundenen und gegebenen Größen ins Venn-Diagramm einzutragen:



Zur Probe kann abschließend noch die Summe für N(G) ermittelt werden, die in der Aufgabenstellung mit 1.000 angegeben ist. Demnach müssten sich bei der Addition aller Einzelsummen aus dem Venn-Diagramm ebenfalls wieder 1.000 ergeben:

50 + 200 + 30 + 50 + 100 + 250 + 70 + 250 = 1.000 → stimmt!

Lösung zu Aufgabe 3: Hier bieten sich zwei Lösungsmöglichkeiten an, die nachfolgend kurz dargestellt werden sollen. Die erste Lösungsmöglichkeit besteht in der Berechnung über den Additionssatz für zwei beliebige Ereignisse (beide Unternehmen entscheiden schließlich unabhängig voneinander über das Praktikum).

Der Additionssatz lautet:

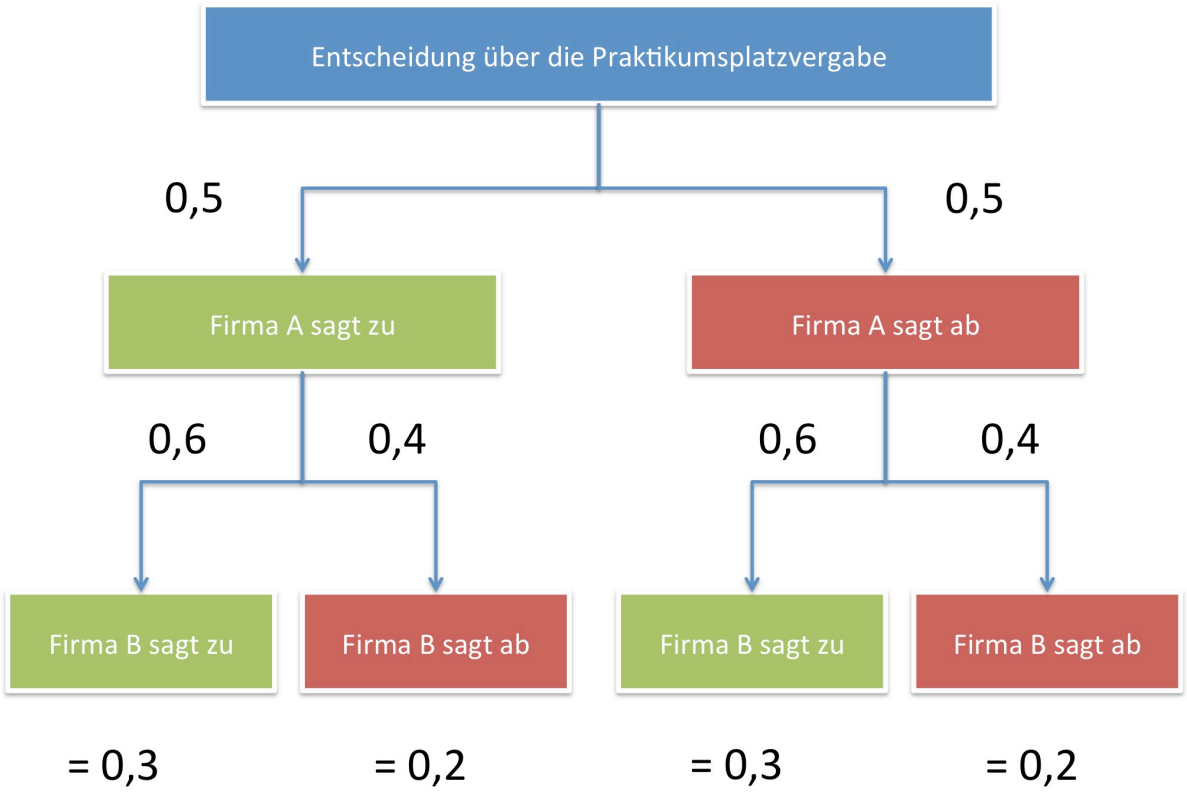
$P\ (A \cup B) = \ P(A) + P\ (B) - \ P\ (A \cap B)$

Wörtlich ausgedrückt ist also die Wahrscheinlichkeit für das Eintreten mindestens eines der beiden Ereignisse $(A \cup B)$ gleich der Summe der Wahrscheinlichkeiten des Eintretens der jeweiligen Ereignisse minus der Wahrscheinlichkeit des gleichzeitigen Eintretens $(A \cap B)$.

Setzt man die aus der Aufgabenstellung bekannten Werte ein, so ergibt sich folgendes Bild:

$$P(A \cup B) = 0{,}5 + 0{,}6 - 0{,}3 = 0{,}8$$

Die zweite Lösungsmöglichkeit besteht in der Erstellung eines Pfaddiagramms und der Berechnung der Wahrscheinlichkeit über die Pfadregeln. Diese besagen (verallgemeinert), dass entlang der Pfade zu multiplizieren ist, die Wahrscheinlichkeiten einzelner Pfade aber addiert werden. Auf das uns hier vorliegende Beispiel übertragen bedeutet dies, dass z.B. die Wahrscheinlichkeit für eine Zusage von beiden Firmen durch Multiplikation ermittelt wird (0,5 * 0,6 = 0,3), die Wahrscheinlichkeit für eine Zusage von Firma B dagegen durch die Addition der Wahrscheinlichkeiten beider Einzelpfade (0,3 + 0,3 = 0,6) – wobei die Wahrscheinlichkeiten der Einzelpfade wiederum durch Multiplikation kalkuliert wurden.



Da in der Aufgabenstellung nach der Wahrscheinlichkeit gefragt wird, dass mindestens eine Firma eine Zusage über einen Praktikumsplatz gibt, sind die Eintrittswahrscheinlichkeiten aller Pfade zu addieren, in denen es zu mindestens einer Zusage kommt (d.h. beide Firmen sagen zu, nur Firma A sagt zu, nur Firma B sagt zu). Es ergibt sich: $0{,}3 + 0{,}2 + 0{,}3 = 0{,}8$.

Die Wahrscheinlichkeit, dass Student R. eine Zusage für mindestens einen der beiden Praktikumsplätze erhält, liegt demnach bei 80%.

Aufgabenblatt II: Stochastisch unabhängige Ereignisse

Aufgabe 1: In der Bevölkerung Sachsen-Anhalts sind die Merkmale A (CDU-Wähler/in), B (Führerscheininhaber/in) und C (Raucher/in) mit den Wahrscheinlichkeiten P (A) = 0,3, P (B) = 02 und P (C) = 0,1 zu beobachten. Welchen Anteil haben demzufolge die Bürger/innen mit (1) keiner, (2) einer, (3) zwei und (4) drei dieser Eigenschaften an der Gesamtbevölkerung?

Aufgabe 2: Zwei Studenten versuchen unabhängig voneinander die gleiche Statistik-Aufgabe zu lösen, wobei jeder mit einer Lösungswahrscheinlichkeit von 0,6 arbeitet. Wie groß ist die Wahrscheinlichkeit dafür, dass wenigstens einer der Studenten das richtige Ergebnis findet?

Aufgabe 3: Bei der Produktion von zweifarbigem Keramikgeschirr treten Fehler in der einen Glasurfarbe bei 10% aller Stücke, Fehler in der zweiten Glasurfarbe dagegen bei 20% aller Stücke auf, wobei alle möglichen Fehler unabhängig voneinander auftreten. Während die fehlerfreien Stücke über den Fachhandel weiterverkauft werden, werden Stücke mit Fehlern in einer Glasurfarbe an Discountmärkte abgegeben. Stücke, bei denen Fehler in beiden Glasurfarben auftreten, werden dagegen als Abfall entsorgt.

Welche Stückzahlen lassen sich bei einem Produktionslos von 1.000 Stück jeweils für den Fachhandel, für den Discounthandel und für die Abfallentsorgung erwarten?

Lösung zu Aufgabe 1:

(1) Kein Merkmal: Die Berechnung erfolgt mittels des Multiplikationssatzes für stochastisch unabhängige Ereignisse, wobei in diesem Fall mit den Komplementärereignissen zu rechnen ist. Das Komplementärereignis eines Ereignisses E ist dasjenige Ereignis $\bar{E}$, welches eintritt, wenn E nicht eintritt. Da im ersten Teil der Aufgabe nach demjenigen Anteil der Bevölkerung gefragt wird, auf den genau keines der drei Merkmale zutrifft, müssen zur Lösungsfindung in diesem Fall nur die Komplementärereignisse berücksichtigt werden.

$$P\left(\bar{A} \cap \bar{B} \cap \bar{C}\right) = (0{,}7 * 0{,}8 * 0{,}9) = \ 0{,}504$$

Personen mit keinem der drei Merkmale machen demnach 50,4% der Bevölkerung aus.

(2) Ein Merkmal: Die Berechnung hier erfordert die Addition der Wahrscheinlichkeiten für die positive Verteilung eines Merkmals über die drei Merkmalsträger. Grundlage ist auch hier wieder der Multiplikationssatz für stochastisch unabhängige Ereignisse, wobei ebenfalls auf die Komplementärereignisse zurückgegriffen wird, da ja jeweils ein Merkmal auftreten und zwei Merkmale nicht auftreten sollen.

$$P(A \cap \bar{B} \cap \bar{C}) + \ P(\bar{A} \cap B \cap \bar{C}) + \ P(\bar{A} \cap \bar{B} \cap C)$$
$$= (0{,}3 * 0{,}8 * 0{,}9) + (0{,}7 * 0{,}2 * 0{,}9) + (0{,}7 * 0{,}8 * 0{,}1) = 0{,}398$$

Personen mit einem der drei Merkmale machen demnach 39,8% der Bevölkerung aus.

(3) Zwei Merkmale: Die Berechnung erfolgt analog zu (2), wobei der Unterschied darin besteht, dass anstatt zwei Komplementärereignissen pro Term eines zu berücksichtigen ist.

$$P(A \cap B \cap \overline{C}) + P(A \cap \bar{B} \cap C) + P(\bar{A} \cap B \cap C)$$
$$= (0{,}3 * 0{,}2 * 0{,}9) + (0{,}3 * 0{,}8 * 0{,}1) + (0{,}7 * 0{,}2 * 0{,}1) = 0{,}092$$

Personen mit zwei von drei Merkmalen machen demnach 9,2% der Bevölkerung aus.

(4) Drei Merkmale: Die Berechnung erfolgt analog zu (1), wobei der Unterschied darin besteht, dass nicht mit den Komplementärereignissen gerechnet werden muss, da nach dem Anteil der Bevölkerung gefragt wird, der alle drei Merkmale auf sich vereint.

$$P(A \cap B \cap C) = (0{,}3 * 0{,}2 * 0{,}1) = 0{,}006$$

Personen mit allen drei Merkmalen machen demnach 0,6% der Bevölkerung aus.

Probe: Nach Kolmogorov muss die Gesamtwahrscheinlichkeit stets 1 (=100%) betragen. Die Summe aller schnittmengenfreien Einzelwahrscheinlichkeiten muss daher ebenfalls 1 sein. Dieser Fall ist hier gegeben, da „kein Merkmal“, „ein Merkmal“, „zwei Merkmale“ sowie „alle drei Merkmale“ keine Schnittmengen aufweisen und den Ereignisraum G vollständig abdecken. Die Probe der zuvor angestellten Berechnungen lautet demnach:

$$0{,}504 + 0{,}398 + 0{,}092 + 0{,}006 = 1$$

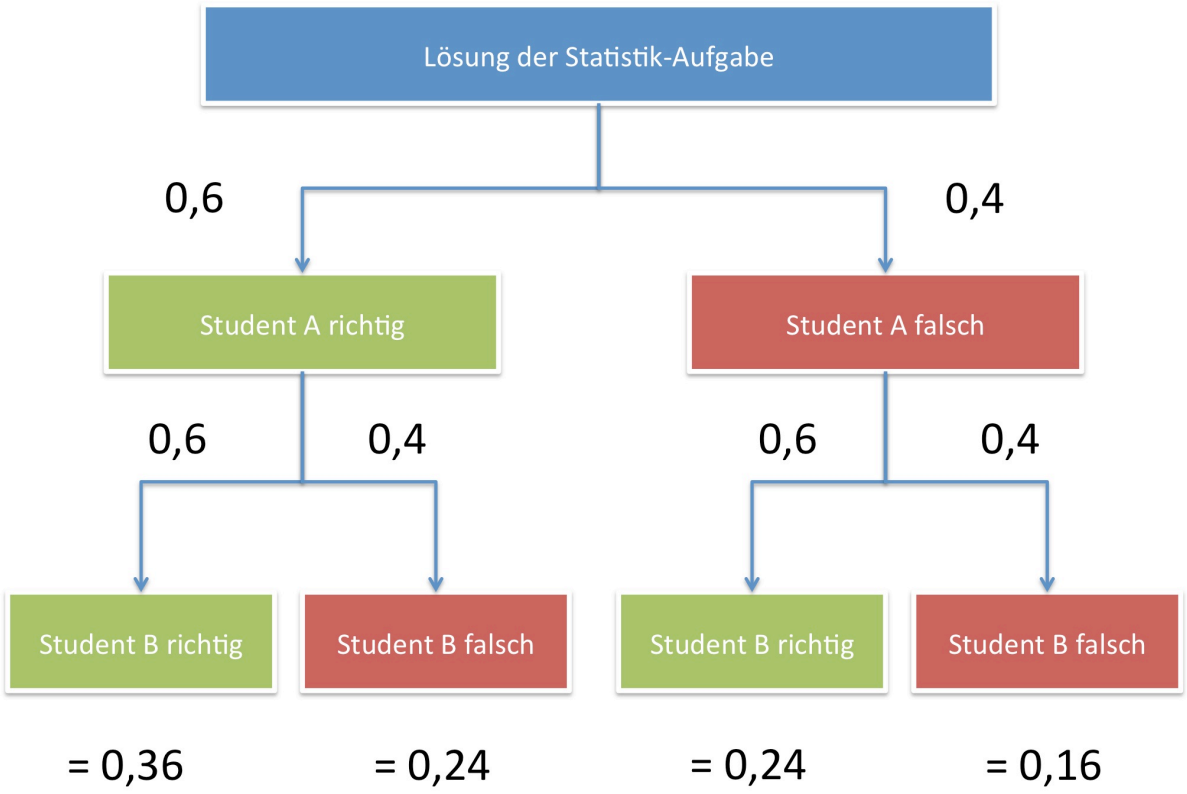
Lösung zu Aufgabe 2: Analog zu Aufgabe 3 des Arbeitsblatts I kann diese Aufgabe entweder über die Anlage eines Pfaddiagramms oder aber über die Anwendung des Additionssatzes für zwei beliebige Ereignisse (die beiden Studenten beeinflussen sich in ihrer Arbeit nicht gegenseitig) erfolgen.

Der Additionssatz lautet: $P\ (A \cup B) = P(A) + \ P(B) - \ P(A \cap B)$

Die beiden Wahrscheinlichkeiten P(A) und P(B) sind bereits durch die Aufgabenstellung vorgegeben (jeweils 0,6). Die Wahrscheinlichkeit für das gleichzeitige Eintreten beider Ereignisse $P(A \cap B)$ ergibt sich dagegen durch die Multiplikation von P(A) und P(B). Setzt man diese Werte nun in den Additionssatz ein, so erhält man:

$P(A \cup B) = \ 0{,}6 + 0{,}6 - (0{,}6 * 0{,}6) = \ 0{,}6 + 0{,}6 - 0{,}36 = 0{,}84$

Betrachten wir alternativ noch das Pfaddiagramm:



Um die Wahrscheinlichkeit für mindestens eine richtige Lösung zu ermitteln, sind die Einzelwahrscheinlichkeiten aller Pfade zu addieren, in denen mindestens einer der beiden Studenten die richtige Lösung findet (d.h. beide Studenten lösen die Aufgabe korrekt, nur Student A löst die Aufgabe korrekt, nur Student B löst die Aufgabe korrekt). Es ergibt sich:

$0{,}36 + 0{,}24 + 0{,}24 = 0{,}84$

Die Wahrscheinlichkeit, dass mindestens einer der beiden Studenten die gestellte Aufgabe korrekt lösen kann, liegt somit bei 84%.

Lösung zu Aufgabe 3: Die Anzahl der für den Fachhandel tauglichen Stücke lässt sich über den Multiplikationssatz für zwei stochastisch unabhängige Ereignisse ermitteln. Hierbei ist die Frage zu beantworten, mit welcher Wahrscheinlichkeit es bei einem beliebigen Stück zu keinerlei Fehlern (Komplementärereignisse) in der Färbung kommt, d.h. $P(\bar{A} \cap \bar{B})$.

$$P(\bar{A} \cap \bar{B}) = (0{,}9 * 0{,}8) = \ 0{,}72$$

Dies bedeutet, dass ein beliebiges Werkstück mit einer Wahrscheinlichkeit von 72% keine Fehler in beiden Glasurfarben aufweist. Bei einer Losgröße von 1.000 Stück ist demnach mit 720 fehlerfreien Werkstücken zu rechnen, die über den Fachhandel verkauft werden können.

Aufwändiger ist nun die Ermittlung der Anzahl von Werkstücken, die in den Discounthandel abgegeben werden müssen, da sie Fehler in einer Farbglasur (nicht aber in beiden) aufweisen.

Zu berechnen ist hierfür die Wahrscheinlichkeit für das Auftreten eines der beiden Ereignisse (Fehler in Glasur A oder Fehler in Glasur B). Dies geschieht in zwei Schritten. Zunächst wird mit dem Additionssatz für stochastisch unabhängige Ereignisse $P(A \cup B)$ berechnet um den Anteil der Werkstücke zu ermitteln, die einen Fehler entweder in Glasur A oder in Glasur B aufweisen. Stellt man sich die passend eingefärbte Fläche im Venn-Diagramm vor, so müsste von dieser nun allerdings noch die Fläche $P(A \cap B)$ subtrahiert werden, da die Werkstücke, bei denen Fehler in beiden Farbglasuren auftreten, nicht an den Discounter verkauft, sondern im Abfall entsorgt werden. Für die Berechnung von $P(A \cap B)$ ist analog zur Berechnung im ersten Lösungsschritt dieser Aufgabe der Multiplikationssatz für stochastisch unabhängige Ereignisse heranzuziehen.

$$P(A \cup B) - \ P(A \cap B) = \big((0{,}2 + 0{,}1) - (0{,}2 * 0{,}1)\big) - (0{,}2 * 0{,}1) = \ 0{,}26$$

Alternativ lassen sich natürlich auch die Wahrscheinlichkeiten für „Fehler in Glasur A, nicht aber in Glasur B“ und „Fehler in Glasur B, nicht aber in Glasur A“ berechnen und addieren.

$$P(A \cap \bar{B}) + P(\bar{A} \cap B) = (0{,}1 * 0{,}8) + (0{,}9 * 0{,}2) = \ 0{,}26$$

Bei einer Losgröße von 1.000 Stück ist demnach mit 260 Werkstücken zu rechnen, die einen Fehler in einer der beiden Glasuren aufweisen und die daher an die Discountmärkte gehen.

Die letzte noch offene Größe kann nun entweder analog zur ersten Berechnung über den Multiplikationssatz für zwei stochastisch unabhängige Ereignisse kalkuliert $P((A \cap B) = 0{,}2 * 0{,}1 = 0{,}02)$ oder aus den beiden bisherigen Ergebnissen berechnet werden: Da die Summe aller Teilwahrscheinlichkeiten nach Kolmogorov stets 1 bzw. 100% ergeben muss, lassen sich die beiden bereits berechneten Ergebnisse (0,72 und 0,26) auch von 1 abziehen, um den noch fehlenden Wert zu erhalten: 1 – 0,72 – 0,26 = 0,02. Bei einer Losgröße von 1.000 Stück ist demnach mit einem Ausschuss von 20 Werkstücken zu rechnen.

Aufgabenblatt III: Rechnen mit Wahrscheinlichkeiten

Aufgabe 1: Wie groß ist die Wahrscheinlichkeit, dass der Weihnachtsfeiertag in einem beliebigen Jahr entweder auf einen Samstag oder auf einen Sonntag fällt?

Aufgabe 2: Eine Kiste enthält insgesamt 20 Cremetuben und 10 Spraydosen, wobei 50% der Cremetuben und 20% der Spraydosen einen Defekt aufweisen. Wie groß ist die Wahrscheinlichkeit, bei der zufälligen Ziehung eines Gegenstands aus dieser Kiste

(a) eine Spraydose,
(b) eine Spraydose oder ein defektes Teil
(c) oder sowohl eine Spraydose als auch ein defektes Teil

zu erhalten?

Aufgabe 3: Zwei Fußballspieler schießen vom Elfmeterpunkt auf ein leeres Tor, wobei die Wahrscheinlichkeit für einen Treffer beim ersten Spieler 0,6 und beim zweiten Spieler 0,8 beträgt. Wie groß ist nun die Wahrscheinlichkeit dass

(a) keiner der beiden Spieler das Tor trifft,
(b) mindestens ein Tor bei beiden Schüssen erzielt wird
(c) oder beide Spieler erfolgreich sind und zwei Tore erzielt werden?

Lösung zu Aufgabe 1: Lösungsgrundlage ist die klassische Wahrscheinlichkeitsdefinition:

$$P(A) = \frac{Anzahl\ der\ günstigen\ Elementarereignisse}{Anzahl\ der\ möglichen\ Elementarereignisse}$$

Im vorliegenden Fall sind insgesamt sieben Ereignisse denkbar (der Weihnachtsfeiertag könnte auf einen Montag, Dienstag, Mittwoch, Donnerstag, Freitag, Samstag oder Sonntag fallen), von denen jedoch nur zwei (Samstag und Sonntag) für das Ergebnis relevant sind.

Es ergibt sich daher: $\frac{2\ (Samstag\ oder\ Sonntag)}{7\ (alle\ Wochentage)} = \frac{2}{7} = 0{,}2857$

Die gesuchte Wahrscheinlichkeit liegt demnach bei 28,57%.

Hintergrund: Die klassische Wahrscheinlichkeitsdefinition nach Laplace. Der klassische Wahrscheinlichkeitsbegriff geht auf den französischen Mathematiker Pierre-Simon Laplace (1749 – 1827) zurück. Besteht ein Zufallsvorgang A (wie etwa das Werfen eines Würfels) aus endlich vielen Elementarereignissen, ist die Wahrscheinlichkeit P(A) wie folgt definiert:

$$P(A) = \frac{Anzahl\ der\ günstigen\ Elementarereignisse}{Anzahl\ der\ möglichen\ Elementarereignisse}$$

Übertragen auf das Würfelbeispiel existieren sechs mögliche Elementarereignisse (nämlich die Zahlen 1,2,3,4,5 und 6). Die Wahrscheinlichkeit, bei einem einzelnen Wurf eine gerade Zahl zu würfeln, berechnet sich demnach wie folgt:

$$P(A) = \frac{3\ (günstige\ Elementarereignisse{:}2,4,6)}{6\ (alle\ Elementarereignisse{:}1,2,3,4,5,6)} = \frac{3}{6} = \frac{1}{2} = 0{,}5$$

Lösung zu Aufgabe 2: Für die Lösungsfindung sind folgende Vorüberlegungen hilfreich:

- In der Kiste befinden sich insgesamt 30 Gegenstände
- In der Kiste befinden sich 10 defekte Cremetuben
- In der Kiste befinden sich 2 defekte Spraydosen

Für die Lösung aller drei Teilaufgaben wird auf die klassische Wahrscheinlichkeitsdefinition nach Laplace zurückgegriffen (Hinweise hierzu siehe weiter oben auf dieser Seite).

(a) Wahrscheinlichkeit für das Ziehen einer Spraydose:

$$\frac{10\ Spraydosen}{30\ Gegenstände} = \frac{10}{30} = \frac{1}{3} = 33{,}33\%$$

(b) Wahrscheinlichkeit für das Ziehen einer Spraydose oder eines defekten Teils:

$$\frac{10\ Spraydosen + 10\ defekte\ Cremetuben}{30\ Gegenstände} = \frac{20}{30} = \frac{2}{3} = 66{,}66\%$$

(zu beachten ist, dass defekte Spraydosen nicht "doppelt gezählt" werden dürfen)

(c) Wahrscheinlichkeit für das Ziehen einer Spraydose sowie eines defekten Teils:
(gesucht ist hier also die Wahrscheinlichkeit des Ziehens einer defekten Spraydose)

$$\frac{2\ defekte\ Spraydosen}{30\ Gegenstände} = \frac{2}{30} = 6{,}66\%$$

Lösung zu Aufgabe 3:

(a) Die Wahrscheinlichkeit für keinen Tortreffer durch beide Spieler wird anhand des Multiplikationssatzes für zwei stochastisch unabhängige Ereignisse (da beide Spieler voneinander unabhängig auf das Tor schießen) sowie unter Berücksichtigung der Komplementärereignisse (also der Wahrscheinlichkeiten für "kein Tor", in der Aufgabenstellung gegeben sind die Wahrscheinlichkeiten für "Tor") ermittelt.

$$P(\bar{A} \cap \bar{B}) = (0{,}4 * 0{,}2) = 0{,}08$$

Die Wahrscheinlichkeit für zwei Fehlschüsse hintereinander liegt also bei 8%.

(b) Die Wahrscheinlichkeit für mindestens einen Tortreffer lässt sich auf verschiedene Wege ermitteln. Am schnellsten geht es, wenn man sich vor Augen führt, dass von den vier insgesamt möglichen Ereignissen (beide Spieler schießen ein Tor, nur Spieler A schießt ein Tor, nur Spieler B schießt ein Tor und beide Spieler schießen daneben) nur ein einziges Ereignis (beide Spieler schießen daneben) nicht das Kriterium "mindestens ein Treffer" erfüllt. Berechnet man also die Wahrscheinlichkeit für das Eintreten dieses Ereignisses und zieht das Ergebnis von der Gesamtwahrscheinlichkeit 1 ab, bleibt die geforderte Wahrscheinlichkeit für mindestens einen Tortreffer übrig.

$$1 - P(\bar{A} \cap \bar{B}) = 1 - (0{,}4 * 0{,}2) = \ 1 - 0{,}08 = 0{,}92$$

Die Wahrscheinlichkeit für mindestens einen Tortreffer liegt somit bei 92% und entspricht der Komplementärwahrscheinlichkeit zu (a) (zwei Fehlschüsse hintereinander). Bei Problemen mit dem Verständnis dieser Vorgehensweise kann die Erstellung des Pfaddiagramms helfen.

(c) Die Wahrscheinlichkeit für zwei Treffer hintereinander lässt sich analog zu (a) (d.h. der Wahrscheinlichkeit für zwei Fehlschüsse hintereinander) vergleichsweise einfach anhand des Multiplikationssatzes für stochastisch unabhängige Ereignisse berechnen:

$$P(A \cap B) = (0{,}6 * 0{,}8) = 0{,}48$$

Die Wahrscheinlichkeit für zwei aufeinanderfolgende Treffer liegt demnach bei 48%.

Aufgabenblatt IV: Rechnen mit dem Satz von Bayes

Aufgabe 1: Ein Unternehmen stellt auf diesen drei Maschinen Spitzgussteile her:

Maschine	A	B	C
Kapazität	60%	25%	15%
Ausschussquote	9%	12%	4%

(a) Mit welcher Wahrscheinlichkeit handelt es sich bei einem zufällig aus der laufenden Produktion entnommenen Spritzgussteil um Ausschuss?
(b) Mit welcher Wahrscheinlichkeit stammt ein als fehlerhaft identifiziertes Spritzgussteil von Maschine A / Maschine B / Maschine C?

Aufgabe 2: Ein Discountmarkt bezieht seine Einkaufstüten von zwei verschiedenen Lieferanten, wobei 70% der Tüten von Lieferant A und 30% der Tüten von Lieferant B stammen. Bei einer Normalbeladung liegt die Reißwahrscheinlichkeit bei einer Tüte von Lieferant A bei 10%, bei einer Tüte von Lieferant B dagegen bei 20%.

(a) Bestimmen Sie die Wahrscheinlichkeit für das Reißen einer beliebigen Tüte.
(b) Berechnen Sie die Wahrscheinlichkeit, mit der eine bereits gerissene Tüte von Lieferant B stammt.

Aufgabe 3: Die Wahrscheinlichkeit eines Studenten, sein Studium erfolgreich in der Regelstudienzeit abzuschließen, liegt bei 0,7. Gelingt ein erfolgreicher Abschluss in der Regelstudienzeit, so stehen die Chancen auf eine Einstellung innerhalb des ersten Monats nach der Exmatrikulation bei 0,8. Gelingt der Abschluss dagegen nicht, fallen die Chancen mit nur 0,1 sehr viel geringer aus. Wie groß aber ist nun die Wahrscheinlichkeit, dass es überhaupt zu einer Einstellung des Studenten im Monat nach der Exmatrikulation kommt?

Aufgabe 4: Ein Unternehmen produziert Radios in drei Betriebsstätten. Betrieb B1 produziert am Tag 1.500 Radios mit einem Ausschuss von 15 Stück, Betrieb B2 3.500 Stück mit einem Ausschuss von 175 Stück und Betrieb B3 5.000 Stück, darunter 100 fehlerhafte Exemplare. Wie groß ist nun die Wahrscheinlichkeit, dass

(a) ein zufällig aus der Produktion entnommenes Radio defekt ist?
(b) ein als defekt identifiziertes Radio von Betrieb B2 stammt?
(c) ein als defekt identifiziertes Radio von Betrieb B3 stammt?

Lösung zu Aufgabe 1:

(a) Mit welcher Wahrscheinlichkeit handelt es sich bei einem zufällig aus der laufenden Produktion entnommenen Spritzgussteil um Ausschuss?

Zur Beantwortung dieser Aufgabe kann wieder einmal auf die klassische Definition der Wahrscheinlichkeit nach Laplace zurückgegriffen werden. Angenommen, im Rahmen eines Produktionslots würden genau 100 Bauteile unter Einsatz aller drei Maschinen hergestellt. In diesem Fall würden 60 Stück auf Maschine A, 25 Stück auf Maschine B und 15 Stück auf Maschine C produziert. Unter den 60 auf Maschine A produzierten Bauteilen wären 5,4 defekte Bauteile zu erwarten (Ausschussquote 9%), während Maschine B 3 defekte Bauteile und Maschine C 0,6 defekte Bauteile auswerfen würde. In Summe sind dies $5{,}4 + 3 + 0{,}6 = 9$ defekte bei 100 insgesamt hergestellten Bauteilen. Unter Berücksichtigung der klassischen Wahrscheinlichkeitsdefinition nach Laplace lässt sich somit folgende Berechnung anstellen:

$$P(A) = \frac{Anzahl\ der\ günstigen\ Elementarereignisse}{Anzahl\ der\ möglichen\ Elementarereignisse}$$

$$P(defektes\ Bauteil) = \frac{9\ defekte\ Bauteile}{100\ hergestellte\ Bauteile} = \frac{9}{100} = 0{,}09$$

Die Wahrscheinlichkeit für die zufällige Ziehung eines defekten Bauteils liegt somit bei 9%. Die Kalkulation lässt sich natürlich auch mit einer beliebigen anderen Losgröße durchführen.

(b) Mit welcher Wahrscheinlichkeit stammt ein bereits als fehlerhaft identifiziertes Spritzgussteil von Maschine A / Maschine B / Maschine C?

Zur Lösung dieser Aufgabe wird der Satz von Bayes benötigt: Der Satz von Bayes spielt immer dann eine Rolle, wenn nach der Wahrscheinlichkeit für ein Ereignis B unter der Bedingung gefragt wird, dass zuvor ein anderes Ereignis A eingetreten ist. In solchen Fällen lässt sich die Wahrscheinlichkeit für das Eintreten von B nicht einfach nur an B festmachen, vielmehr ist bei der Berechnung der Eintrittswahrscheinlichkeit von B zu berücksichtigen ist, mit welcher Wahrscheinlichkeit A eingetreten ist. Ein klassisches Beispiel für die Anwendung des Satz von Bayes ist das sogenannte „Taxi-Problem“, das im Theorie-Einschub zwischen den Aufgabenblättern IV und V betrachtet wird.

Zunächst aber zur Aufgabe: Zur Berechnung der Wahrscheinlichkeit, dass ein fehlerhaftes Spritzgussteil von Maschine A stammt, sind die folgenden Überlegungen anzustellen. Die Wahrscheinlichkeit für das Auftreten eines defekten Teils haben wir bereits in Aufgabenteil (a) berechnet: Sie liegt bei 0,09 und bildet den Nenner des Bruchs. Im Zähler wird nun die Wahrscheinlichkeit für eine Produktion auf Maschine A (0,6) mit der Wahrscheinlichkeit der Ausschussproduktion auf Maschine A (0,09) multipliziert. Analog wird auch bei der Berechnung der beiden anderen gesuchten Wahrscheinlichkeiten verfahren.

Wahrscheinlichkeit für Maschine A: $P(A_1/E) = \frac{(0{,}6*0{,}09)}{0{,}09} = 0{,}6 = 60\%$

Wahrscheinlichkeit für Maschine B: $P(A_2/E) = \frac{(0{,}25*0{,}12)}{0{,}09} = 0{,}33 = 33\%$

Wahrscheinlichkeit für Maschine C: $P(A_3/E) = \frac{(0{,}15*0{,}04)}{0{,}09} = 0{,}06 = 6\%$

Probe: Die Summe aller drei Wahrscheinlichkeiten muss insgesamt 100% ergeben.

Lösung zu Aufgabe 2:

(a) Bestimmen Sie die Wahrscheinlichkeit für das Reißen einer beliebigen Tüte.

Die Berechnung erfolgt über den Additionssatz für zwei sich ausschließende Ereignisse in Kombination mit dem Multiplikationssatz für zwei stochastisch unabhängige Ereignisse:

$$P(A \cap B) = P(A) + P(B) = (0{,}7 * 0{,}1) + (0{,}3 * 0{,}2) = 0{,}13$$

Die Wahrscheinlichkeit für das Reißen einer beliebigen Tüte liegt also bei 13%.

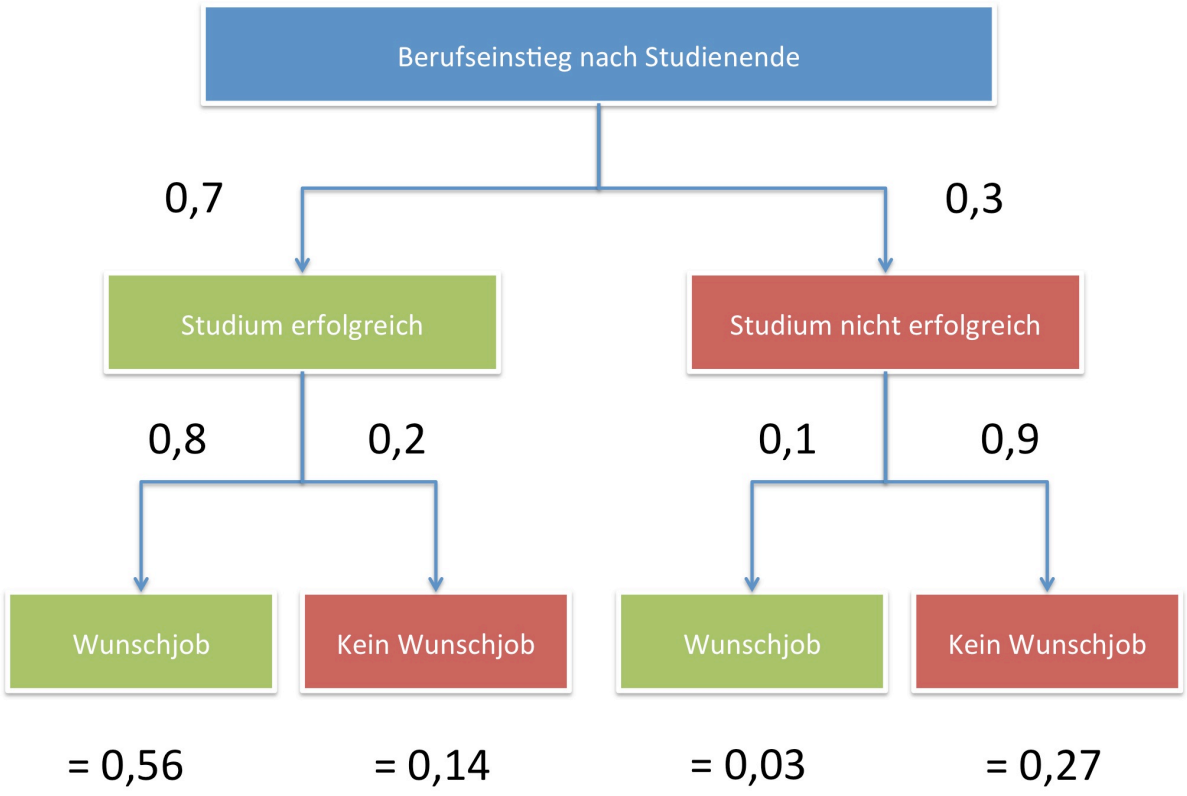
(b) Berechnen Sie die Wahrscheinlichkeit, mit der eine bereits gerissene Tüte von Lieferant B stammt.

Für die Lösung dieser Aufgabe wird erneut auf den Satz von Bayes zurückgegriffen. Die hier benötigte Wahrscheinlichkeit für das Reißens einer Tüte wurde bereits in (a) berechnet.

$$P(A/E) = \frac{(0{,}3 * 0{,}2)}{0{,}13} = 0{,}4615$$

Ist eine Tüte gerissen, so liegt demnach die Wahrscheinlichkeit dafür, dass diese Tüte von Lieferant B hergestellt wurde, bei 46,15%.

Lösung zu Aufgabe 3: Für diese Aufgabe empfiehlt sich die Erstellung des Pfaddiagramms:



Aus diesem Pfaddiagramm lassen sich nun alle gewünschten Wahrscheinlichkeiten ablesen. Konkret gefragt wurde nach der Wahrscheinlichkeit, dass der Student den gewünschten Job überhaupt erhält. Diese ist durch die Addition der beiden entsprechenden Pfade (Studium erfolgreich > Wunschjob sowie Studium nicht erfolgreich > Wunschjob) zu ermitteln:

$$0{,}56 + 0{,}03 = 0{,}59$$

Die Wahrscheinlichkeit für einen erfolgreichen Jobeinstieg liegt damit bei 59%.

Lösung zu Aufgabe 4: Da in der Aufgabenstellung bereits absolute Werte genannt werden, lässt sich die Aufgabe leicht mit Hilfe der klassischen Wahrscheinlichkeitsdefinition lösen.

$$Wahrscheinlichkeit = \frac{Anzahl\ der\ für\ das\ Ergebnis\ interessanten\ Ereignisse}{Anzahl\ der\ insgesamt\ möglichen\ Ereignisse}$$

Die Wahrscheinlichkeit, dass ein zufällig entnommenes Radio defekt ist, beträgt somit:

$$\frac{Summe\ der\ defekten\ Radios}{Summe\ der\ produzierten\ Radios} = \frac{(15+175+100)}{(1.500+3.500+5.000)} = \frac{290}{10.000} = 0{,}029 = 2{,}9\%$$

Die Wahrscheinlichkeit, dass ein defektes Radio aus Betriebsstätte B2 stammt liegt bei:

$$\frac{Defekte\ Radios\ aus\ Betriebsstätte\ B2}{Summe\ der\ defekten\ Radios\ insgesamt} = \frac{175}{290} = 0{,}609 = 60{,}9\%$$

Die Wahrscheinlichkeit, dass ein defektes Radio aus Betriebsstätte B3 stammt liegt bei:

$$\frac{Defekte\ Radios\ aus\ Betriebsstätte\ B3}{Summe\ der\ defekten\ Radios\ insgesamt} = \frac{100}{290} = 0{,}344 = 34{,}4\%$$

Theorie-Einschub: Das Taxi-Problem und der Satz von Bayes

(basierend auf einem von mir im Jahr 2008 verfassten Artikel für das Portal ScienceBlogs.de; das Taxi-Problem selbst wurde meines Wissens nach erstmals von Arthur Engel formuliert)

Das Ausgangsszenario: Grüne und blaue Taxen

Die Ausgangssituation des sogenannten Taxi-Problems gestaltet sich wie folgt: In einer nicht näher bezeichneten Stadt existieren zwei konkurrierende Taxi-Unternehmen. Die Taxen des ersten Unternehmens, „Green Cab", sind grün, die Taxen des zweiten Unternehmens, „Blue Cab", dagegen blau. Der Marktanteil von „Green Cab" liegt bei 85%, womit (da ja kein drittes Unternehmen existiert) 15% Marktanteil für „Blue Cab" verbleiben.

Nun kommt es zu einem Verkehrsunfall mit Fahrerflucht, für den es leider nur einen einzigen Zeugen gibt. Dieser Zeuge gibt an, der Fahrer eines blauen Taxis habe den Unfall verursacht, und sei anschließend mit dem Fahrzeug geflüchtet. Andere Spuren gibt es nicht, so dass die ermittelnde Behörde sich auf die Aussagen des Zeugen verlassen muss. Der antwortet zwar ehrlich und nach bestem Wissen und Gewissen – aber wie groß ist die Wahrscheinlichkeit, dass seine Schilderung des Unfallhergangs trotzdem fehlerhaft ist?

Dass er ein Taxi gesehen hat, steht außer Frage, da sich auf allen Taxen in unserer Stadt die Aufschrift „TAXI" in riesigen Lettern findet. Da der Zeuge sich in diesem Punkt also kaum irren kann und zudem keinen Grund für eine Falschaussage hat, arbeitet der flüchtige Fahrer also mit Sicherheit für eines der beiden Taxi-Unternehmen. Aber was ist mit der Farbe des Wagens? Der Zeuge ist sich zwar sicher, ein blaues Taxi erkannt zu haben, könnte sich aber auch irren, denn immerhin war es schon ziemlich dämmrig…

Die ermittelnde Behörde führt daher einen Test mit dem Zeugen durch, in dessen Rahmen er Farbbeispiele vorgeführt bekommt, an die er sich später erinnern soll. Zur Freude der Ermittler zeigt sich, dass der Zeuge über ein sehr gutes Farbgedächtnis verfügt, denn es gelingt ihm, 80% der Farbbeispiele richtig zuzuordnen. Lässt sich daraus nun der Schluss ziehen, dass auch die Beschreibung des Unfallwagens mit 80%iger Sicherheit korrekt ist?

Warum ist diese Annahme falsch?

Die meisten Menschen würden dieser Aussage wohl auf Anhieb zustimmen. Es klingt ja auch logisch: Wenn der Zeuge einen Wagen mit x-beliebiger Farbe sieht, dann liegt die Wahrscheinlichkeit dafür, dass er sich später korrekt an die Farbe erinnern kann, genau bei 80%. Demnach müsste das Unfalltaxi doch ebenfalls mit 80%iger Wahrscheinlichkeit blau gewesen sein – oder etwa nicht?

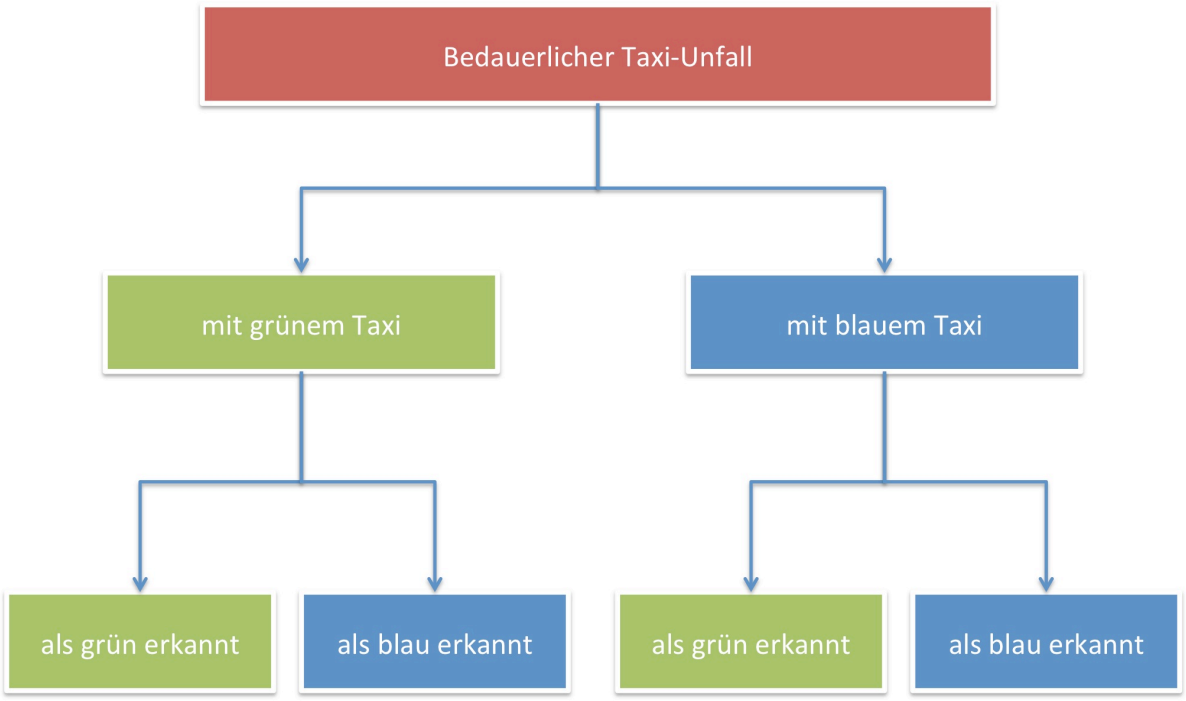
Tatsächlich ist dem nicht so, denn man vernachlässigt bei der Schlussfolgerung viel zu leicht den Umstand, dass in der Stadt sehr viel mehr grüne als blaue Taxen unterwegs sind. Nehmen wir einfach einmal an, in der Stadt sind insgesamt nur 100 Taxen unterwegs (es ist eine sehr kleine Stadt). Von diesen 100 Taxen muss eines in den Unfall verwickelt gewesen sein – und aus den jeweiligen Marktanteilen von 85% bzw. 15% ergibt sich, dass 85 grüne und 15 blaue Taxen die Straßen der Stadt „bevölkern".

Wie wir wissen, ist sich der Zeuge sicher, ein blaues Taxi gesehen zu haben. Dass er mit 80%iger Wahrscheinlichkeit richtig liegt bedeutet natürlich auch, dass er in 20% aller Fälle die Farbe nicht korrekt zuordnen kann. Er würde demnach von den 85 grünen Taxen 68 korrekt als grün erkennen und 17 fälschlicherweise als blau. Umgekehrt würde er von den 15 blauen Taxen 12 korrekt als blau erkennen und 3 fälschlicherweise als grün. Diese Umstände gilt es bei der die Berechnung der Wahrscheinlichkeit für ein blaues Unfalltaxi zu beachten.

Letztendlich gibt es vier Möglichkeiten für den Verlauf der ganzen Geschichte:

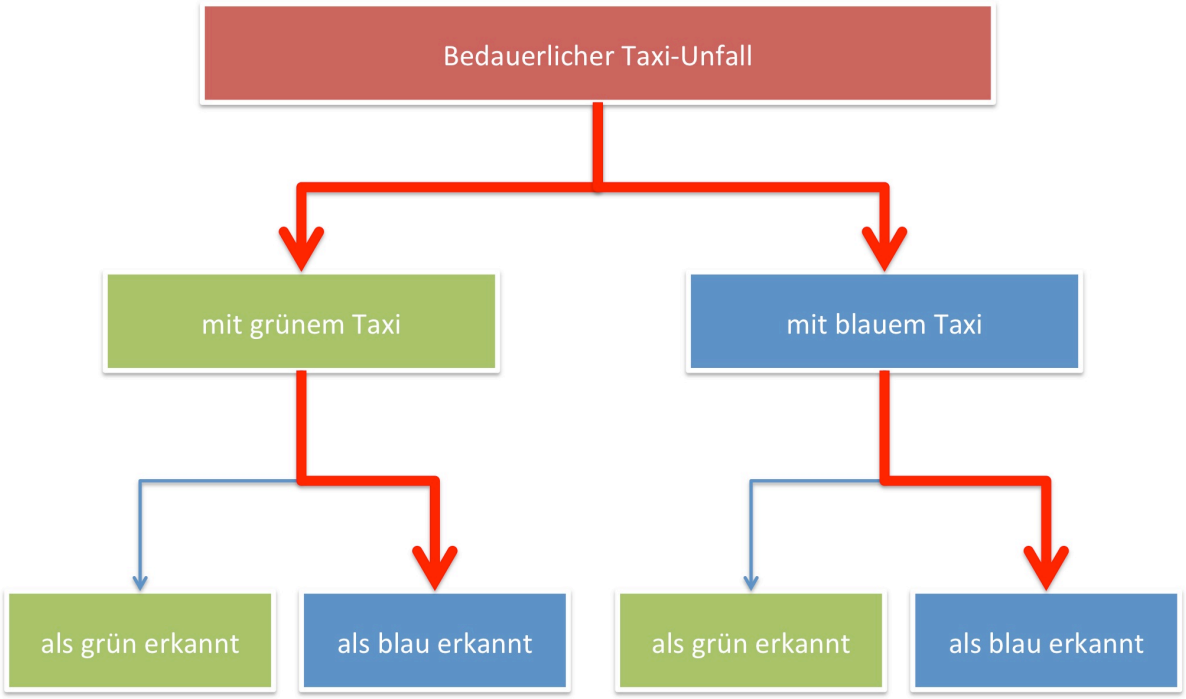
1. Das Unfalltaxi war grün und wird als grün erkannt
2. Das Unfalltaxi war grün und wird als blau erkannt
3. Das Unfalltaxi war blau und wird als blau erkannt
4. Das Unfalltaxi war blau und wird als grün erkannt

In Form eines Pfaddiagramms stellt sich die Situation demnach wie folgt dar:



Da die Zeugenaussage (das Unfalltaxi war angeblich blau) bereits vorliegt, interessieren uns zwei der vier Ereignisketten nicht weiter. Konzentrieren wir uns also auf die beiden anderen:

1. Das Unfalltaxi war grün und wird als blau erkannt
2. Das Unfalltaxi war blau und wird als blau erkannt



Wie man sieht, besteht jede dieser Ereignisketten aus zwei Einzelereignissen, nämlich

1. Taxi der Farbe xy ist in den Unfall verwickelt und
2. Farbe xy des Taxis wird richtig / falsch erkannt

Die Wahrscheinlichkeit für das richtige Erkennen der Farbe ist uns aus dem Test bekannt, sie liegt bei 80%. Die Wahrscheinlichkeit des vorausgehenden Ereignisses („ein Taxi der Farbe xy ist in den Unfall verwickelt") ergibt sich aus den Marktanteilen der beiden Taxi-Unternehmen. Wenn man nun noch beachtet, dass die Irrtumswahrscheinlichkeit des Zeugen bei 20% liegt (wenn er in 80% aller Fälle die Farbe richtig erkennt), dann kann man den beiden verbleibenden Ereignisketten leicht die Wahrscheinlichkeiten zuordnen:

- Das Unfalltaxi war grün (85%) und wird als blau erkannt (20%)
- Das Unfalltaxi war blau (15%) und wird als blau erkannt (80%)

Wie groß ist nun aber die Wahrscheinlichkeit dafür, dass das Unfalltaxi tatsächlich blau gewesen ist, wie der Zeuge behauptet? In der Statistik verwenden wir für Berechnungen mit sogenannten "bedingten Wahrscheinlichkeiten" den Satz von Bayes. Dabei wird die Wahrscheinlichkeit des zu untersuchenden Ereignisses (blaues Taxi wird als blau erkannt) ins Verhältnis zur Summe der bedingten Wahrscheinlichkeiten aller Ereignisketten gesetzt (blaues Taxi wird als blau erkannt und grünes Taxi wird als blau erkannt):

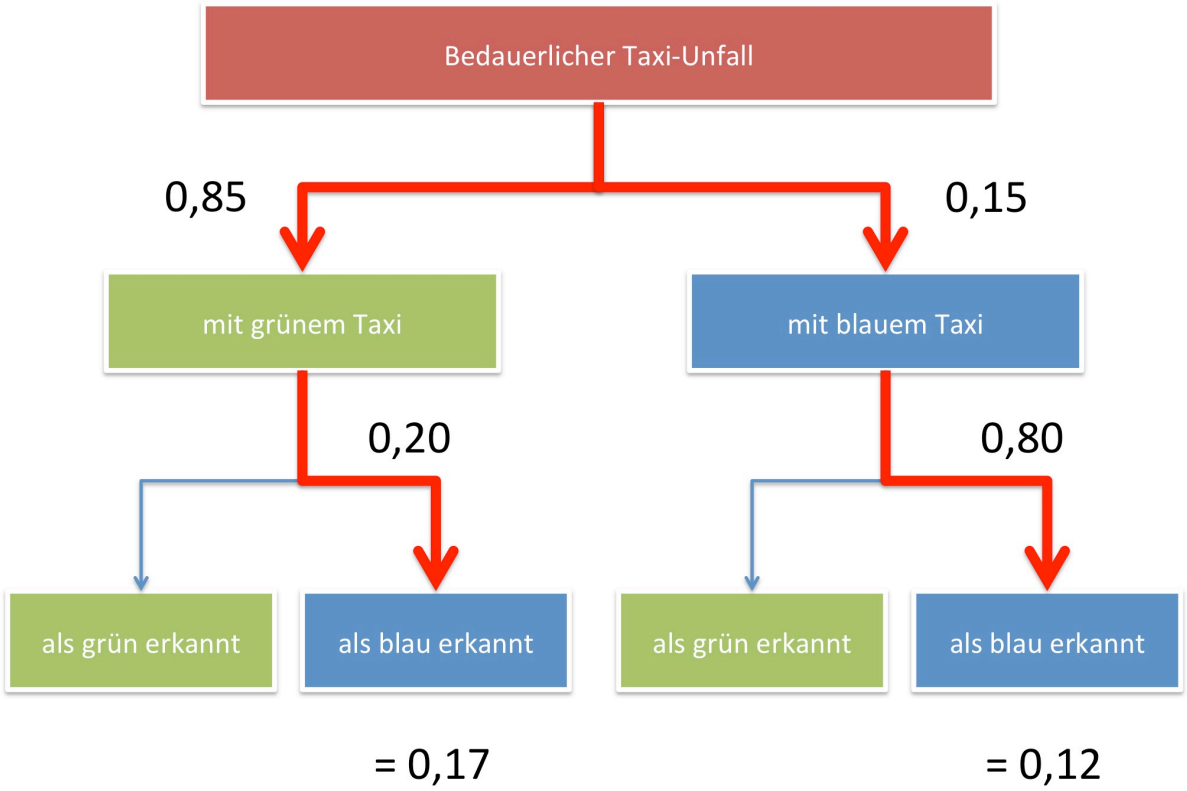
$$P(Tb \cup Zb) = \frac{P(Zb \cup Tb) * P(Tb)}{P(Zb \cup Tb) * P(Tb) + P(Zb \cup Tg) * P(Tg)} = \frac{0{,}80 * 0{,}15}{0{,}80 * 0{,}15 + 0{,}20 * 0{,}85}$$

= 0,41

Erklärung zu den Abkürzungen:

Tb = Taxi ist blau
Tg = Taxi ist grün
Zb = Zeuge erkennt Taxi als blau
Zg = Zeuge erkennt Taxi als grün

Die Wahrscheinlichkeit dafür, dass das Taxi tatsächlich blau gewesen ist, liegt also gerade einmal bei 41%. Ein verblüffendes Ergebnis – obwohl der Zeuge eine 80%ige Sicherheit bei der Identifikation von Farben an den Tag legt, liegt die Wahrscheinlichkeit dafür, dass er im Recht ist, unterhalb der 50%-Grenze. Was nichts anderes bedeutet, als dass man bessere Chancen hat, die Farbe des Unfalltaxis zu ermitteln, wenn man einfach eine Münze wirft…



Mathematische Spielerei oder relevantes Problem?

Auch wenn es auf den ersten Blick wie eine Spielerei erscheint, ist das hier kurz dargestellte Problem sogar im höchsten Maße praxisrelevant – unter anderem natürlich auch für „echte" Ermittlungen oder Gerichtsverfahren, in denen Zeugenaussagen eine Rolle spielen. Es begegnet uns aber auch anderswo – nämlich überall dort, wo bedingte Wahrscheinlichkeiten nicht richtig wahrgenommen werden. Wie im Falle des Taxi-Problems wird viel zu häufig eine Wahrscheinlichkeit (der Zeuge liegt in 80% aller Fälle richtig) korrekt interpretiert, eine andere (nur 15% aller Taxen sind blau) jedoch einfach „ausgeblendet".

Das Problem taucht etwa bei flächendeckenden HIV-Tests auf, die aus gutem Grund noch nirgendwo eingeführt wurden. Zwar sind die Tests an sich sehr ergebnissicher (mehr noch als der Zeuge im Taxi-Fall), der Anteil der HIV-Infizierten in der Bevölkerung ist jedoch – zum Glück – äußerst gering (sehr viel geringer als der Anteil der blauen Taxen). Unter diesen Umständen würde flächendeckender Test zu einer so großen Menge an Fehldiagnosen führen, dass er einfach keinen Sinn macht – trotz der hohen Güte des eigentlichen Testverfahrens.

Auch in der Terror-Abwehr spielen bedingte Wahrscheinlichkeiten eine große Rolle. Ein Beispiel mit Fehlalarmen in spanischen Hotels habe ich vor Jahren mal mit Studierenden meines SPSS-Kurses berechnet. In diesem Szenario gab es einen Terrorismus-Experten, der sogar mit 95%iger Sicherheit eine echte von einer falschen ETA-Bombendrohung unterscheiden konnte – eine wesentlich bessere „Trefferquote" als unser Zeuge im Taxi-Fall. Trotzdem veranlasst dieser Experte eine überflüssige Hotel-Evakuierung nach der anderen.

Warum? Natürlich deshalb, weil in unserem Szenario (wie eben auch im wahren Leben) auf einen echten ETA-Terroristen knapp hundert Anrufer kommen, die eine Bombendrohung nur so zum Spaß durchgeben. Echte Drohungen und „Spaßdrohungen" stehen also in einem sehr viel krasserem Missverhältnis zueinander, als grüne und blaue Taxen – und das führt dann auch bei einer sehr hohen Sicherheit von 95% zu so vielen „false positives", dass es auf Außenstehende fast wie ein Ratespiel des Experten wirkt.

Nicht richtig wahrgenommene bedingte Wahrscheinlichkeiten finden wir im Alltag an jeder Ecke. Wer sich für die Thematik begeistern kann, kann sich ja spaßeshalber einmal darüber Gedanken machen, welchen Zusammenhang es zwischen diesen Stichwörtern und dem Problem der falschen Wahrnehmung von bedingten Wahrscheinlichkeiten geben könnte:

- „Cold Reading"
- Ausländerkriminalität
- Horoskop-Vorhersagen
- DNA-Tests vor Gericht
- BSE-Massentests bei Kühen
- Gewalttätige Computerspieler
- National Terrorism Advisory Level

Aufgabenblatt V: Rechnen mit dem Satz von Bayes II

Aufgabe 1: Herr Vorsicht hat in seiner Wohnung eine neue Alarmanlage einbauen lassen. Diese löst im Falle eines Einbruchs mit einer Wahrscheinlichkeit von 0,9 Alarm aus. Die Wahrscheinlichkeit für einen Fehlalarm beträgt dagegen 0,01. Die Wahrscheinlichkeit, dass im Wohnviertel von Herrn Vorsicht überhaupt eingebrochen wird, liegt bei 0,001.

Die Alarmanlage von Herrn Vorsicht hat soeben einen Alarm ausgelöst.

(a) Wie groß ist die Wahrscheinlichkeit für dieses Ereignis?
(b) Mit welcher Wahrscheinlichkeit handelt es sich um einen echten Einbruch?

Aufgabe 2: Ein Biologe studiert zwei Kreuzungen, von denen eine im Mittel bei 25% aller Versuche, die andere bei lediglich 5% aller Versuche gelingt.

(a) Wie viele geglückte Versuche sind jeweils bei 12 Versuchen zu erwarten und wie groß ist die Varianz?
(b) Wie groß ist die Wahrscheinlichkeit, dass bei 10 sukzessiven Versuchen jeweils (1) zwei geglückte Kreuzungen, (2) mehr als vier geglückte Kreuzungen und (3) maximal 8 geglückte Kreuzungen zu beobachten sind?

Aufgabe 3: Die Fußballmannschaften des FC Sorge und des FC Elend spielen um den Oberharz-Cup, wobei die Mannschaft des FC Sorge Spiele mit einer Wahrscheinlichkeit von 0,6 gewinnt. Wie hoch ist die Wahrscheinlichkeit, dass der FC Sorge

(a) höchstens 3 Spiele gewinnt?
(b) mindestens 3 Spiele gewinnt?
(c) alle Spiele gewinnt?

Lösung zu Aufgabe 1:

(a) Wie groß ist die Wahrscheinlichkeit für die Auslösung der Alarmanlage?

Zu berechnen ist hier sowohl die Wahrscheinlichkeit für die Auslösung der Alarmanlage bei einem echten Einbruch (Einbruchswahrscheinlichkeit: 0,001; Auslösesicherheit: 0,9) sowie die Wahrscheinlichkeit für einen Fehlalarm (das Komplementärereignis des Einbruchs in Kombination mit der Fehlauslösewahrscheinlichkeit). Beide Wahrscheinlichkeiten sind nach dem Additionssatz für zwei stochastisch unabhängige Ereignisse zu addieren:

$$P(A \vee B) = P(A) + P(B) = (0{,}001 * 0{,}9) + (0{,}999 * 0{,}01) = 0{,}0108$$

Die Wahrscheinlichkeit für das Auslösen der Alarmanlage liegt demnach bei 1,08%.

(b) Mit welcher Wahrscheinlichkeit handelt es sich um einen echten Einbruch?

Zur Lösung dieser Aufgabe wird erneut auf den Satz von Bayes zurückgegriffen:

$$P(A/E) = \frac{(0{,}001 * 0{,}9)}{0{,}01089} = 0{,}08262$$

Die Wahrscheinlichkeit für einen echten Einbruch liegt demnach bei 8,262%. Dieses Ergebnis verdeutlicht – ebenso wie das Taxi-Problem – erneut die Relevanz bedingter Wahrscheinlichkeiten für die Wahrnehmung von Alltagsphänomenen. Da die Alarmanlage selbst sehr sicher ist (Auslösesicherheit 90%, Fehlalarmwahrscheinlichkeit 1%), würden viele Menschen die Wahrscheinlichkeit eines echten Einbruchs im Falle einer Alarmauslösung sehr viel höher als gerade einmal 8,2% einschätzen. Dies liegt daran, dass das äußerst geringe Einbruchsrisiko von 0,1% einfach „ausgeblendet“ wird.

Lösung zu Aufgabe 2: Bei der beschriebenen Verteilung handelt es sich um eine Binomialverteilung. Diese ist durch die folgenden drei Eigenschaften gekennzeichnet:

I. Es existieren nur zwei mögliche Ausprägungen, d.h. die Verteilung ist diskret.
II. Die Erfolgs-/Misserfolgs-Wahrscheinlichkeiten sind konstant und ändern sich auch bei der Wiederholung von Versuchen nicht.
III. Die einzelnen Versuche sind nicht voneinander abhängig, d.h. es handelt sich um ein sogenanntes Modell mit Zurücklegen (MMZ).

Ein klassisches Binomialverteilungsbeispiel ist der Münzwurf: Es existieren nur zwei mögliche Ausprägungen (Kopf oder Zahl), die Wahrscheinlichkeiten für Kopf oder Zahl (0,5) ändern sich auch bei wiederholten Münzwürfen nicht (vorausgesetzt, es handelt sich um eine sogenannte “faire” Münze) und die einzelnen Münzwürfe sind nicht voneinander abhängig.

Als Hilfsmittel sind bei derartigen Aufgaben üblicherweise zwei Verteilungstabellen zulässig: Eine für die Wahrscheinlichkeitsfunktion (für die Realisation genau einer Zufallsvariablen) und eine für die Verteilungsfunktion (für die Wahrscheinlichkeit bis zu einer bestimmten Ausprägung). Auf beide Tabellen wird nachfolgend zurückgegriffen. Wer die Tabellen nicht vorliegen hat, findet sie auf zahlreichen Webseiten in Form von PDF-Downloads.

(a) Wie viele geglückte Versuche sind jeweils bei 12 Versuchen zu erwarten und wie groß ist die Varianz?

Hier ist sowohl nach dem Erwartungswert als auch nach der Varianz gefragt.

Berechnung des Erwartungswerts: $E(X) = n * p$
Berechnung der Varianz: $Var(X) = n * p * (1 - p)$

Gegeben sind:

n = Anzahl der Versuche (hier: 12)
p = Erfolgs-Wahrscheinlichkeit (hier: 0,25 für Versuch 1 und 0,05 für Versuch 2)

Berechnungen für Versuch 1:

$E(X) = 12 * 0{,}25 = 3$
$Var(X) = 12 * 0{,}25 * (1 - 0{,}25) = 2{,}25$

Berechnungen für Versuch 2:

$E(X) = 12 * 0{,}05 = 0{,}6$
$Var(X) = 12 * 0{,}05 * (1 - 0{,}05) = 0{,}57$

Bei 12 Versuchen pro Versuchstyp sind für den ersten Versuch 3 geglückte Versuche mit einer Varianz von 2,25 Versuchen und für den zweiten Versuch 0,6 geglückte Versuche mit einer Varianz von 0,57 Versuchen zu erwarten.

(b) Wie groß ist die Wahrscheinlichkeit, dass bei 10 sukzessiven Versuchen jeweils (1) zwei geglückte Kreuzungen, (2) mehr als vier geglückte Kreuzungen und (3) maximal 8 geglückte Kreuzungen zu beobachten sind?

(1) Genau zwei geglückte Kreuzungen > Rückgriff auf die Wahrscheinlichkeitsfunktion

Gegeben sind:

n = 10
x = 2
$p_1 = 0{,}25$
$p_2 = 0{,}05$

Das Ergebnis lässt sich aus der Tabelle der Wahrscheinlichkeitsfunktion ablesen:

Versuch 1: $fB(x/n\,;p) = fB(2/10;\,0{,}25) = 0{,}2816$
Versuch 2: $fB(x/n\,;p) = fB(2/10;\,0{,}05) = 0{,}0746$

Zum Ablesen muss man in den meisten Tabellen zunächst den Wert für n in der linken Spalte suchen (10), anschließend sucht man den Wert für x (2) in der Spalte rechts daneben. Hat man erst die richtige Zeile gefunden, sucht man in der oberen Spalte nach dem Wert für p (0,25) und findet im Schnittpunkt dieser Spalte und der zuvor gefundenen Zeile den Ergebniswert (0,2816). Ist die Wahrscheinlichkeit (p) größer als 0,5, muss das Ergebnis über eine Komplementärrechnung gefunden werden, die üblicherweise in der Tabelle angegeben ist.

(2) Mehr als vier geglückte Kreuzungen > Rückgriff auf die Verteilungsfunktion

Die Verteilungsfunktion liefert den Wert für die Wahrscheinlichkeit bis zu einer bestimmten Ausprägung, d.h. bei x=4 erhält man die Wahrscheinlichkeit dafür, dass es zu höchstens vier geglückten Kreuzungen kommt. Gesucht ist hier aber die Wahrscheinlichkeit für mehr als vier geglückte Kreuzungen. Diese errechnet man, indem man zunächst die Wahrscheinlichkeit für höchstens vier geglückte Kreuzungen berechnet und anschließend durch Subtraktion von 1 das Komplementärereignis bildet.

Gegeben sind:

n = 10
x = 4
$p_1 = 0{,}25$
$p_2 = 0{,}05$

Versuch 1: $FB(x/n\,;p) = FB(4/10;0{,}025) = 0{,}9219 \rightarrow 1 - 0{,}9219 = 0{,}0781$
Versuch 2: $FB(x/n\,;p) = FB(4/10;0{,}05) = 0{,}9999 \rightarrow 1 - 0{,}9999 = 0{,}0001$

(3) Höchstens acht geglückte Kreuzungen > Rückgriff auf die Verteilungsfunktion

Im Gegensatz zu (2) wird hier auf die Berechnung des Komplementärereignisses verzichtet.

Gegeben sind:

n = 10
x = 8
$p_1 = 0{,}25$
$p_2 = 0{,}05$

Versuch 1: $W(X \leq 8) = FB(x/n\,;p) = FB(8/10;0{,}25) = 1$
Versuch 2: $W(X \leq 8) = FB(x/n\,;p) = FB(8/10;0{,}05) = 1$

Lösung zu Aufgabe 3:

(a) Wahrscheinlichkeit für den FC Sorge, höchstens 3 Spiele zu gewinnen.

Da die Wahrscheinlichkeit hier größer als 0,5 ausfällt, gilt für den gesuchten Wert die komplementäre Beziehung: $FB(x/n\,;p) = 1 - FB(n - x - 1/n\,;1 - p)$. Für die Ermittlung des Ergebnisses ist auf die Verteilungsfunktion zurückzugreifen.

Gegeben sind:

n = 5
x = 3
p = 0,6

$$FB(x/n\,;p) = 1 - FB(n - x - 1/n\,;1 - p) = 1 - FB(1/5;0{,}4) = 1 - 0{,}337 = 0{,}663$$

Die Wahrscheinlichkeit, dass der FC Sorge höchstens 3 Spiele gewinnt, liegt bei 66,3%.

(b) Wahrscheinlichkeit für den FC Sorge, mindestens 3 Spiele zu gewinnen.

Für diese Aufgabe ist erneut auf die Verteilungsfunktion zurückzugreifen. Da nach mindestens 3 gewonnenen Spielen gefragt wird, die Verteilungsfunktion aber das „höchstens"-Ergebnis liefert, ist das gleichzeitig eintretende Ereignis (es werden höchstens 2 Spiele verloren, mit x = 2 und p = 0,4) zu berechnen.

Gegeben sind:

n = 5
x = 2
p = 0,4

$FB(x/n\,;p) = FB(2/5;0{,}4) = 0{,}6826$

Die Wahrscheinlichkeit, dass der FC Sorge mindestens 3 Spiele gewinnt, liegt bei 68,26%.

(c) Wahrscheinlichkeit für den FC Sorge, sämtliche 5 Spiele zu gewinnen.

Hier kann der Einfachheit halber mit der Multiplikationspfadregel gerechnet werden, da es nur einen Pfad im Pfaddiagramm gibt, auf dem der FC Sorge alle Spiele gewinnt. Dieser berechnet sich wie folgt:

$0{,}6^5 = 0{,}6 * 0{,}6 * 0{,}6 * 0{,}6 * 0{,}6 = 0{,}07776$

Die Wahrscheinlichkeit, dass der FC Sorge alle Spiele gewinnt, liegt demnach bei 7,7%.

Alternativ kann auch mit der Wahrscheinlichkeitsfunktion (nicht mit der Verteilungsfunktion, da ein bestimmtes Ereignis gefragt ist) gerechnet werden. Da jedoch eine Wahrscheinlichkeit größer 0,5 vorliegt, müsste jedoch noch eine weitere Umrechnung vorgenommen werden. Diesen zusätzlichen Rechenschritt kann man sich somit durch das Ausweichen auf die Multiplikationspfadregel ersparen.

Aufgabenblatt VI: Hypergeometrische und Poisson-Verteilung

Aufgabe 1: In einer Kiste befinden sich 5 süße und 4 bittere Apfelsinen, wobei der Unterschied zwischen beiden Sorten visuell nicht zu erkennen ist. Es werden nun nacheinander 5 Apfelsinen entnommen, so dass 4 in der Kiste zurückbleiben.

Wie groß ist die Wahrscheinlichkeit

(a) 3 bittere und 2 süße,
(b) 4 bittere und 1 süße,
(c) mindestens 1 bittere
(d) oder mindestens 1 süße Apfelsine zu entnehmen?

Wie lauten die Erfahrungswerte und Varianzen für süß und bitter?

Aufgabe 2: Untersuchungen in einer Telefonzentrale ergaben, dass die dort pro Minute eingehenden Anrufe mit 2,5 Gesprächen pro Minute poissonverteilt sind. Wie hoch ist die Wahrscheinlichkeit, dass in einer bestimmten Minute

(a) gar kein Anruf erfolgt,
(b) höchstes zwei Anrufe erfolgen
(c) oder mindestens ein Anruf erfolgt?

Aufgabe 3: Zum Wernigeröder Rathausfest bietet ein Bäckermeister ein “Rathausbrot” an und schließt sich zur Steigerung des Verkaufserfolgs mit einem Gastronomen zusammen. Mit dem Kauf eines Brotes hat jeder Kunde die Chance, ein Frühstück im bekannten Restaurant “Harzhof” zu gewinnen. Gewinner sind dabei diejenigen Kunden, deren Brot auf der Unterseite mit einer bestimmten Markierung versehen ist. Jeweils eines von 50 Broten weist diese Gewinnmarkierung auf. Am Nachmittag des ersten Verkaufstages liegen noch genau 100 Brote im Regal des Bäckermeisters. Ein Kunde setzt auf sein Glück und wählt zufällig 5 Brote aus. Wie groß ist nun die Wahrscheinlichkeit,

(a) kein Frühstück zu gewinnen,
(b) genau ein Frühstück zu gewinnen
(c) oder mindestens ein Frühstück zu gewinnen?

Lösung zu Aufgabe 1: In dieser Aufgabe ist mit einer hypergeometrischen Verteilung zu rechnen. Derartige Verteilungen zeichnen sich durch folgende drei Eigenschaften aus:

I. Es existieren nur zwei mögliche Ausprägungen, d.h. die Verteilung ist diskret.
II. Es handelt sich um ein Modell ohne Zurücklegen (MOZ), d.h. die einzelnen Versuche beeinflussen sich gegenseitig, da die Wahrscheinlichkeiten eines Versuchs vom jeweils vorangegangenen Versuch abhängen (z.B. Lotto).
III. Es gilt die folgende Approximationsregel: Wenn der Umfang der Stichprobe im Verhältnis zum Umfang der Grundgesamtheit sehr klein ist, darf in die Binomialverteilung approximiert werden (bei $n/N < 0{,}05$).

Wie für die Binomialverteilung existiert auch für die hypergeometrische Verteilung eine Tabelle mit den Ergebniswerten der Wahrscheinlichkeitsfunktion sowie eine Tabelle mit den Ergebniswerten der Verteilungsfunktion. Beide Tabellen werden für die nachfolgenden Aufgabenlösungen benötigt und lassen sich bei Nichtvorliegen im Internet finden.

Bevor man mit der Lösung von Aufgabenteil (a) beginnt, muss noch geprüft werden, ob sich die hypergeometrische Verteilung eventuell in eine Binomialverteilung approximieren lässt, mit der sich einfacher (aber ungenauer) rechnen ließe. Der Umstand, dass eine „Stichprobe" von 5 Apfelsinen aus einer Grundgesamtheit von gerade einmal 9 Apfelsinen bereits einen sehr großen Teil der Grundgesamtheit abdeckt, spricht gegen eine solche Approximation. Dieser Eindruck bestätigt sich bei der rechnerischen Prüfung der Approximationsregel:

$\frac{n}{N} = \frac{5}{9} > 0{,}05$ → Es kann nicht approximiert werden.

(a) Wahrscheinlichkeit für die Entnahme von 3 bitteren und 2 süßen Apfelsinen.

Hier ist auf die tabellierte Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung zurückzugreifen, da nach genau einem Wert gefragt wird.

Gegeben sind:

N = 9 (Grundgesamtheit)
M = 4 (Anzahl der „richtigen" Elemente insgesamt; hier 4, da 4 bittere Apfelsinen)
n = 5 (Stichprobengröße)
x = 3 (Anzahl der „richtigen" Elemente in der Stichprobe)

Aus der tabellierten Wahrscheinlichkeitsfunktion lässt sich entnehmen:

$$fH(x\ /\ N; n; M) = fH(3/9; 5; 4) = 0{,}3175$$

Alternativ kann der gesuchte Wert natürlich auch manuell berechnet werden. Dies geschieht mit Hilfe der Formel für die Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung:

$$fH(x\ /\ N; n; M) = \frac{\binom{M}{x} * \binom{N-M}{n-x}}{\binom{N}{n}}$$

$$fH(2/9;5;5) = \frac{\binom{4}{3} * \binom{9-4}{5-3}}{\binom{9}{5}} = 0{,}3175$$

Die Wahrscheinlichkeit für die Entnahme von 3 bitteren und 2 süßen Apfelsinen liegt somit bei 31,75%.

(b) Wahrscheinlichkeit für die Entnahme von 4 bitteren und 1 süßen Apfelsine.

Auch hier ist auf die Tabelle der Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung zurückzugreifen, da nach genau einer Wertkombination gefragt wird.

Gegeben sind:

N = 9
M = 4
n = 5
x = 4 (hier: 4 bittere Apfelsinen, alle anderen Werte bleiben gleich)

Aus der tabellierten Wahrscheinlichkeitsfunktion ergibt sich:

$$fH(x\ /\ N;n;M) = fH(4/9;5;4) = 0{,}0397$$

Auch dieser Wert lässt sich natürlich mit der oben genannten Formel berechnen. Die Wahrscheinlichkeit für die Entnahme von 4 bitteren und einer süßen Apfelsine liegt damit bei 3,97%.

(c) Wahrscheinlichkeit für die Entnahme von mindestens einer bitteren Apfelsine.

Hier verwendet man nun die tabellierte Verteilungsfunktion, da nach einem Bereich von Werten (1,2,3,4 oder 5 bittere Apfelsinen) gefragt wird. Dieser Wert lässt sich allerdings nicht direkt aus der Tabelle ablesen, da diese sozusagen die Werte für „höchstens“ enthält, die Aufgabenstellung jedoch nach „mindestens“ fragt. Aus diesem Grund berechnet man das Parallelereignis. Dieses ist in der vorliegenden Aufgabe wie folgt zu definieren:

Es werden höchstens 4 süße Apfelsinen gezogen.

Gegeben sind:

N = 9 (der Umfang der Grundgesamtheit bleibt natürlich gleich)
M = 5 (da nun nach den süßen Apfelsinen gefragt wird)
n = 5 (auch der Stichprobenumfang ändert sich nicht)
x = 4 (wird mindestens 1 bittere gezogen, kommen höchstens 4 süße in die Stichprobe)

Der Tabelle für die hypergeometrische Verteilung lässt sich nun entnehmen:

$$FH(x\ /\ N;n;M) = FH(4/9;5;5) = 0{,}9921$$

Alternativ ist hier (neben der direkten Rechnung mit der Formel der Verteilungsfunktion der hypergeometrischen Verteilung) noch folgender Lösungsweg gangbar: Da das komplementäre Ereignis zu „mindestens eine bittere Apfelsine“ „gar keine bittere Apfelsine“ ist, kann dieser

Einzelpfad im Pfaddiagramm berechnet und das Ergebnis von 1 subtrahiert werden, um die Wahrscheinlichkeit für die Ziehung von höchstens 4 süßen Apfelsinen zu errechnen.

$$P(X = 0) = \frac{5}{9} * \frac{4}{8} * \frac{3}{7} * \frac{2}{6} * \frac{1}{5} = \frac{1}{126} = 0{,}00793 \rightarrow 1 - 0{,}00793 = 0{,}9921$$

Die Wahrscheinlichkeit für die Entnahme von mindestens einer bitteren Apfelsine liegt demnach bei 99,2%.

Bei der Aufstellung dieser Kalkulation ist zu berücksichtigen, dass es sich um ein Modell ohne Zurücklegen handelt, weshalb sich die Wahrscheinlichkeit für das Ziehen gar keiner bitteren Apfelsine Zug um Zug verändert. Liegen sie beim ersten Griff in die Kiste noch bei 5/9, da sich unter den 9 vorhandenen Apfelsinen 5 süße befinden, so liegt die Wahrscheinlichkeit, nach der Ziehung von 4 süßen Apfelsinen mit dem letzten Griff in die Kiste die noch letzte verbleibende süße Apfelsine zu erhaschen, lediglich noch bei 1/5.

(d) Wahrscheinlichkeit für die Entnahme von mindestens einer süßen Apfelsine.

Hier handelt es sich im Prinzip um die gleiche Aufgabe wie in (c), weshalb auch hier wieder auf die Tabelle der Verteilungsfunktion zurückgegriffen wird. Analog zu (c) ist zunächst das Parallelereignis „in bitteren Apfelsinen“ zu bestimmen, welches mit dem abgefragten Ereignis „in süßen Apfelsinen“ einhergeht. Wenn bei fünf Griffen in die Kiste mindestens eine süße Apfelsine gezogen werden soll, können höchstens vier bittere Apfelsinen gezogen werden.

Das Ereignis lautet also: Es werden höchstens 4 bittere Apfelsinen gezogen.

Gegeben sind:

N = 9
M = 4
n = 5
x = 4

Der Tabelle für die hypergeometrische Verteilung ist zu entnehmen:

$$FH(x/N; n; M) = FH(4/9; 5; 4) = 1$$

Die Wahrscheinlichkeit für die Ziehung von mindestens einer süßen Apfelsine liegt also bei 100%. Dieses Ergebnis überrascht wenig und hätte auch logisch hergeleitet werden können: Wenn sich nur vier bittere Apfelsinen in der Kiste befinden, man aber fünfmal zieht (ohne Zurücklegen) muss zwangsläufig mindestens eine süße Apfelsine gezogen werden.

(e) Berechnen Sie die Erwartungswerte und die Varianzen für süße und bittere Apfelsinen.

Der Erwartungswert einer hypergeometrischen Verteilung berechnet sich wie folgt:

$$E(X) = n * \frac{M}{N}$$

Die Varianz einer hypergeometrischen Verteilung berechnet sich wie folgt:

$$VAR(X) = n * \frac{M}{N} * \frac{N-M}{N} * \frac{N-n}{N-1}$$

Wir berechnen zunächst Erwartungswert und Varianz für die süßen Apfelsinen (5 von 9).

$$E(X) = 5 * \frac{5}{9} = 2{,}777777$$

$$VAR(X) = 5 * \frac{5}{9} * \frac{9-5}{9} * \frac{9-5}{9-1} = 5 * \frac{5}{9} * \frac{4}{9} * \frac{4}{8} = 0{,}61728$$

Anschließend werden Erwartungswert und Varianz für bittere Apfelsinen berechnet (4 von 9).

$$E(X) = 5 * \frac{4}{9} = 2{,}222222$$

$$VAR(X) = 5 * \frac{4}{9} * \frac{9-4}{9} * \frac{9-5}{9-1} = 5 * \frac{4}{9} * \frac{5}{9} * \frac{4}{8} = 0{,}61728$$

Lösung zu Aufgabe 2: Auch für die Poisson-Verteilung (die auch als Verteilung seltener Ereignisse bezeichnet wird) existieren Tabellen für die Wahrscheinlichkeitsfunktion und die Verteilungsfunktion. Beide Tabellen werden nachfolgend benötigt.

(a) Wahrscheinlichkeit dafür, dass gar kein Anruf erfolgt.

Da nach einem bestimmten Wert (0 Anrufe) gefragt wird, greifen wir hier auf die Wahrscheinlichkeitsfunktion zurück. Gegeben sind die beiden folgenden Werte:

μ = 2,5 (Erwartungswert)
x = 0 (gesuchter Wert für X – in diesem Fall 0 Anrufe)

Aus der Tabelle für die Poisson-Verteilung ergibt sich:

$$fPo(x/\mu) = fPo(0/2{,}5) = 0{,}0821$$

(b) Wahrscheinlichkeit dafür, dass höchstens zwei Anrufe erfolgen.

Im Gegensatz zu Aufgabenteil (a) wird hier nicht nach einem bestimmten Wert (genau 0 Anrufe), sondern nach einer „höchstens"-Angabe (höchstens 2 Anrufe) gefragt, weshalb auf die Tabelle der Verteilungsfunktion zurückzugreifen ist. Gegeben sind dabei:

μ = 2,5 (Erwartungswert)
x = 2 (Gesuchter Bereichswert für X – in diesem Fall höchstens 2 Anrufe)

Aus der Tabelle für die Poisson-Verteilung ergibt sich:

$$FPo = (x \ / \ \mu) = FPo(2/2{,}5) = 0{,}5438$$

(c) Wahrscheinlichkeit dafür, dass mindestens ein Anruf erfolgt.

Hier wird analog zu Aufgabenteil (b) nicht nach einem bestimmten Wert (genau 0 Anrufe), sondern nach einen „höchstens"-Bezug gefragt – formuliert allerdings in einer „mindestens"-Beziehung, die zunächst noch umgestellt werden muss. Dazu folgende Überlegung: Das Komplementärereignis von „mindestens ein Anruf" lautet „gar kein Anruf". Berechnet man dieses und zieht das Ergebnis von 1 ab, erhält man die Lösung für Aufgabe (c).

Da wir nun also das Komplementärereignis suchen, ist wieder nach einem ganz bestimmten Wert (genau 0 Anrufe) gefragt. Exakt diesen Wert hatten wir jedoch in Aufgabenteil (a) bereits ermittelt – die Wahrscheinlichkeit für gar keinen Anruf liegt bei 0,0821 oder 8,21%. Subtrahiert man diesen Wert nun von 1, erhält man 0,9179 und damit die Lösung für diesen Aufgabenteil. Die Wahrscheinlichkeit, dass während einer bestimmten Minute mindestens 1 Anruf erfolgt, liegt also bei 91,79%.

Lösung zu Aufgabe 3: Hier handelt es sich erkennbar um ein Modell ohne Zurücklegen (die einmal verkauften Brote stehen nicht erneut zur Auswahl) und damit eine hypergeometrische Verteilung. Zunächst ist daher zu überprüfen, ob die Verteilung in eine Binomialverteilung approximiert werden kann. Die Bedingung für eine solche Approximation lautet:

$\frac{n}{N} \leq 0{,}05 \rightarrow \frac{5}{100} = 0{,}05 \rightarrow$ Eine Approximation in die Binomialverteilung ist möglich.

Der p-Wert für die Bionomialverteilung berechnet sich als $\frac{M}{N} = \frac{2(Gewinnbrote\ aus\ 100)}{100} = 0{,}02$.

(a) Wahrscheinlichkeit dafür, gar kein Frühstück zu gewinnen.

Da hier nach genau einem Wert (0 Gewinne) gefragt wird, greifen wir nachfolgend auf die Tabelle der Wahrscheinlichkeitsfunktion der Binomialverteilung zurück.

Gegeben sind:

n = 5
x = 0
p = 0,02 (entspricht wie oben gezeigt dem M/N der hypergeometrischen Verteilung)

Aus der entsprechenden Tabelle lässt sich nun ablesen:

$fB(x/n\,;p) = fB(0/5;0{,}02) = 0{,}9030$

Die Wahrscheinlichkeit, kein Frühstück zu gewinnen, liegt demnach bei 90,3%.

(b) Wahrscheinlichkeit dafür, genau ein Frühstück zu gewinnen.

Da wir zuvor in die Binomialverteilung approximiert haben, bleiben wir nun auch in dieser und greifen auch weiterhin auf die Tabelle der Wahrscheinlichkeitsfunktion zurück, da analog zu Aufgabenteil (a) nach genau einem Wert (genau ein Gewinn) gefragt wird).

Gegeben sind:

n = 5
x = 1
p = 0,02

Aus der entsprechenden Tabelle lässt sich nun ablesen:

$$fB(x/n\,;p) = fB(1/5;0{,}02) = 0{,}0922$$

Die Wahrscheinlichkeit, genau ein Frühstück zu gewinnen, liegt demnach bei 9,22%.

(c) Wahrscheinlichkeit dafür, mindestens ein Frühstück zu gewinnen.

Der einfachste Weg zur Lösung dieses Aufgabenteils besteht in der Berechnung des Komplementärereignisses (der Wahrscheinlichkeit dafür, gar kein Frühstück zu gewinnen) und die Subtraktion dieser Wahrscheinlichkeit von 1. Da diese Wahrscheinlichkeit bereits in Aufgabenteil (a) berechnet wurde (0,9030), können wir einfach weiterrechnen:

$$1 - 0{,}9030 = 0{,}097$$

Die Wahrscheinlichkeit, mindestens ein Frühstück zu gewinnen, liegt also bei 9,7%.

Aufgabenblatt VII: Poisson- und Bionomialverteilung

Aufgabe 1: An einem gewöhnlichen Sonntag sind an einer Wernigeröder Tankstelle zwischen 7:30 und 8:00 Uhr im Durchschnitt 2 Kunden zu bedienen. Wie groß ist die Wahrscheinlichkeit, dass während dieser Zeitspanne 3 Kunden zu bedienen sind?

Aufgabe 2: Ein Schmuckgeschäft führt eine gefragte Markenarmbanduhr und hat von dieser 15 Exemplare auf Lager. Ein Hehler bietet dem Besitzer des Schmuckgeschäfts Fälschungen zu einem günstigen Preis an, von denen der Juwelier 5 erwirbt und zufällig mit den echten Markenuhren vermischt.

Ein Kunde kauft genau 5 Uhren. Wie groß ist nun die Wahrscheinlichkeit

(a) dass er nur Originale erhält?
(b) dass er mindestens 2 Fälschungen erhält?

Aufgabe 3: Reederei A stellt bei einer Untersuchung fest, dass lediglich 50 der 500 Frachtschiffe im Besitz der Reederei voll ausgelastet fahren.

(a) Wie lautet der Anteilswert der voll ausgelasteten Frachtschiffe?
(b) Wie viel voll ausgelastete Frachtschiffe würde man bei einer Stichprobe im Umfang n=10 und bei einer Ziehung ohne Zurücklegen zu finden erwarten?

Reederei A wird von der größeren Reederei B aufgekauft, die nun insgesamt 2.000 Schiffe besitzt. Vor der Übernahme waren 350 Schiffe von Reederei B voll ausgelastet.

(c) Wie viel Prozent der Frachtflotte sind nach der Übernahme voll ausgelastet?
(d) Wie hoch ist die Wahrscheinlichkeit, dass man bei einer zufälligen Auswahl (ohne Zurücklegen) von 10 Schiffen höchstens 8 unterausgelastete Schiffe erhält?
(e) Wie viel voll ausgelastete Frachtschiffe sind höchstens in einer Stichprobe im Umfang von n=10 mit einer Wahrscheinlichkeit von 0,95 enthalten?
(f) Wie groß ist die Wahrscheinlichkeit dafür, in einer Stichprobe von n=20 Frachtschiffen mindestens 6 und höchstens 11 voll ausgelastete Frachtschiffe zu finden?
(g) Im Durchschnitt laufen am Tag 2 Frachtschiffe der Reederei im Hamburger Hafen ein. Bestimmen Sie die Wahrscheinlichkeit dafür, dass an einem beliebigen Tag mehr als 3 Frachtschiffe der Reederei einlaufen.

Aufgabe 4: Dass derzeit am häufigsten bei Zahnfüllungen verwendete Material ist nach wie vor Amalgam. Es befindet sich in 96% aller plombierten Zähne. Die verbleibenden 4% sind dagegen mit Gold plombiert.

(a) In einem Büro mit 25 Angestellten werden 6 Personen auf Zahnfüllungen untersucht. Wie groß ist die Wahrscheinlichkeit, dass 5 Angestellte eine Amalgamfüllung besitzen – gesetzt den Fall, dass jeder der Angestellten über eine Zahnplombe verfügt?
(b) Welche Verteilung wäre zu wählen, wenn diese Untersuchung bei einer Stichprobe von 5 Personen in einem Betrieb mit insgesamt 150 Angestellten durchgeführt würde?
(c) Wie hoch ist bei einer Untersuchung von 120 plombierten Personen die Wahrscheinlichkeit, dass genau 7 Personen eine Goldfüllung besitzen?

Lösung zu Aufgabe 1: Bei dieser Aufgabe liegt eine typische Poisson-Verteilung vor, da die Wahrscheinlichkeit des Eintretens eines Ereignisses zu einem ganz bestimmten Zeitpunkt äußerst gering ist. Der Erwartungswert für den betrachteten Zeitraum liegt bei 2 (Kunden), gesucht ist nun also die Wahrscheinlichkeit dafür, dass die Zufallsvariable X den Wert 3 annimmt. Das Ergebnis findet sich in der Tabelle der Wahrscheinlichkeitsfunktion:

$$P(X = x) = fP(3; 2) = 0{,}1804$$

Die Wahrscheinlichkeit dafür, dass an einem Montag zwischen 7:30 und 8:00 Uhr drei Kunden zu bedienen sind, beträgt demnach 18,04%.

Lösung zu Aufgabe 2: Da ein Modell ohne Zurücklegen vorliegt (eine einmal verkaufte Uhr steht nicht erneut zur Auswahl), rechnen wir hier mit der hypergeometrischen Verteilung.

(a) Wahrscheinlichkeit dafür, nur Originale zu kaufen.

Da hier nach einem bestimmten Wert (genau 5 Originale) gefragt wird, ist auf die Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung zurückzugreifen.

Gegeben sind:

N = 20
M = 15
n = 5
x = 5

Wegen der großen Grundgesamtheit von N=20 findet sich das Ergebnis in den meisten Tabellen nicht mehr und ist daher manuell mit Hilfe der folgenden Formel zu berechnen:

$$fH(x/N\,; n; M) = \frac{\binom{M}{x} * \binom{N-M}{n-x}}{\binom{N}{n}}$$

$$fH(5/20; 5; 15) = \frac{\binom{15}{5} * \binom{20-15}{5-5}}{\binom{20}{5}} = 0{,}19369$$

Die Wahrscheinlichkeit für den ausschließlichen Kauf von Originalen liegt also bei 19,36%.

(b) Wahrscheinlichkeit dafür, mindestens zwei Fälschungen zu kaufen.

Im Gegensatz zu Aufgabenteil (a) lässt sich dieser Aufgabenteil nicht mehr mit der Wahrscheinlichkeitsfunktion, sondern nur mit der Verteilungsfunktion lösen. Zuvor muss allerdings die „mindestens-Formulierung“ des Aufgabentextes noch in eine „höchstens-Formulierung“ umgewandelt werden: Wenn mindestens zwei Fälschungen gekauft werden, werden höchstens drei Originale gekauft.

Gegeben sind:

N = 20
M = 15

n = 5
x = 3

Auch hier muss wegen der großen Grundgesamtheit auf die Formel zurückgegriffen werden:

$$FH(x/N;n;M) = \sum_{k=0}^{xk} \frac{\binom{M}{x} * \binom{N-M}{n-x}}{\binom{N}{n}}$$

$$FH(5/20;3;15) = \sum_{k=0}^{xk} \frac{\binom{15}{5} * \binom{20-15}{5-3}}{\binom{20}{5}} = 0{,}36609$$

Die Wahrscheinlichkeit für den Kauf mindestens zweier Fälschungen liegt also bei 36,6%.

Lösung zu Aufgabe 3: Die Aufgabenstellung weist bereits darauf hin, dass hier ein Modell ohne Zurücklegen vorliegt. Es ist also mit der hypergeometrischen Verteilung zu rechnen.

(a) Berechnung des Anteilswertes der voll ausgelasteten Frachtschiffe:

$$p = \frac{50}{500} = 0{,}1$$

Vor der Fusion waren 10% der Frachtschiffe von Reederei A voll ausgelastet.

(b) Berechnung des Erwartungswerts der hypergeometrischen Verteilung:

$$E(X) = n * \frac{M}{N} = 10 * \frac{50}{500} = 1$$

Bei einer Ziehung von zehn Schiffen ist ein voll ausgelastetes Frachtschiff zu erwarten.

(c) Anteil der voll ausgelasteten Frachtschiffe nach der Fusion:

$$p = \frac{50 + 350}{2000} = \frac{400}{2000} = 0{,}2$$

Nach der Fusion sind 20% der Frachtschiffe der neuen Reederei voll ausgelastet.

(d) Wahrscheinlichkeit, bei einer Auswahl von 10 Schiffen höchstens 8 nicht voll ausgelastete Schiffe zu finden.

Hier ist zunächst zu prüfen, ob eine Approximation in die Binomialverteilung möglich ist.

$\frac{10}{2000} = 0{,}005$ → Eine Approximation ist zulässig.

Das p der Binomialverteilung errechnet sich in diesem Falle als: $\frac{400}{2000} = 0{,}2$

Für die Lösung dieser Aufgabe ist neben p auch q als Gegenwert erforderlich, da ein „Erfolg“ in diesem Fall kein ausgelastetes, sondern ein nicht ausgelastetes Schiff ist. Es gilt daher:

$$q = 1 - p = 1 - 0{,}2 = 0{,}8$$

q wird nachfolgend analog zu p verwendet.

Gegeben sind demnach:

n = 10
x = 8
p = 0,8 (q)

Die Lösung kann nun der Tabelle für die Verteilungsfunktion („höchstens“) der Binomialverteilung entnommen werden. Da p > 0,5 ist, ist allerdings noch die Umbildungsvorschrift zu beachten:

$$FB(x/n\,;p) = 1 - FB(n - x - 1/n\,;1 - p) = 1 - FB(10 - 8 - 1/10;1 - 0{,}8)$$
$$= 1 - FB(1/10;0{,}2) = 1 - 0{,}3758 = 0{,}6242$$

Die Wahrscheinlichkeit, in einer zufälligen Auswahl von 8 Schiffen (nach der Fusion) höchstens 8 nicht voll ausgelastete Schiffe zu finden, liegt demnach bei 62,42%.

(e) Wie viel voll ausgelastete Frachtschiffe sind mit einer Wahrscheinlichkeit von 0,95 höchstens in einer Stichprobe des Umfangs n = 10 zu erwarten?

Gegeben sind:

n = 10
p = 0,2 (dieser Aufgabenteil fragt wieder nach voll ausgelasteten Frachtschiffen)
FB = 0,95

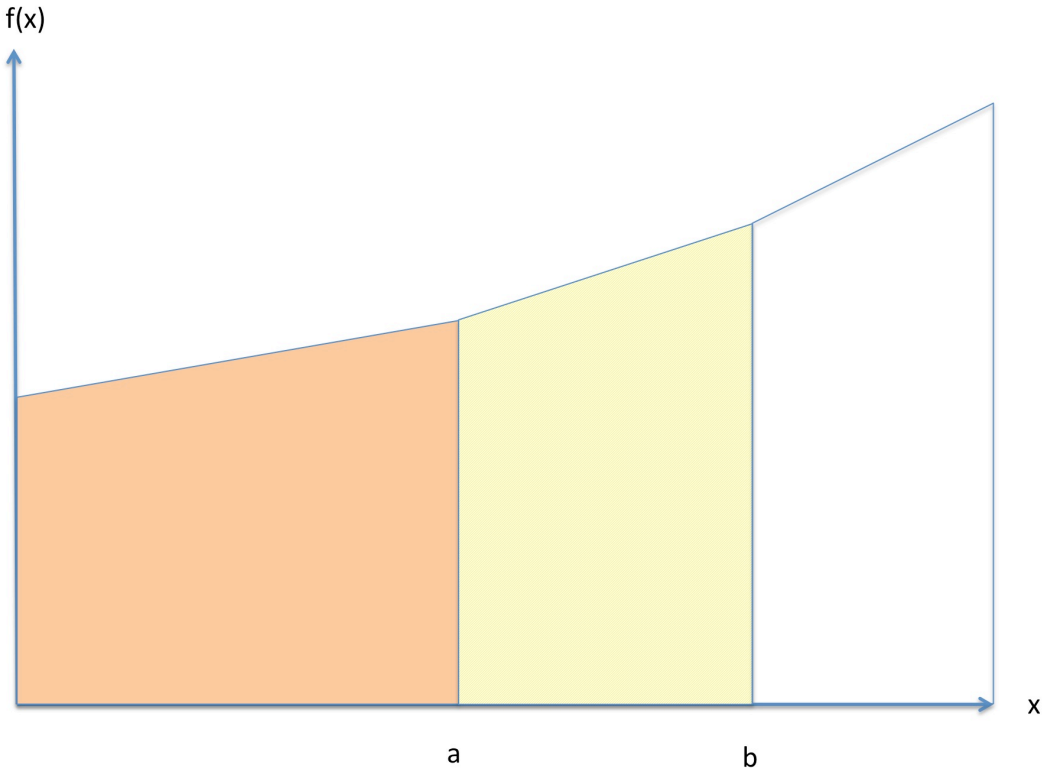
Nun bildet man die Rechenvorschrift für die Verteilungsfunktion der Binomialverteilung:

$$FB(x/n\,;p) = 0{,}95 \rightarrow FB(x/10;0{,}2) = 0{,}95$$

Über die Tabelle lässt sich für x der Wert 3 ermitteln (bei x = 4 wird die Wahrscheinlichkeit von 0,95 bereits überschritten). Bei der gegebenen Wahrscheinlichkeit sind demnach höchstens 3 voll ausgelastete Frachtschiffe in einer Stichprobe n = 10 zu erwarten.

(f) Wahrscheinlichkeit, in einer Stichprobe von n=20 Frachtschiffen mindestens 6 und höchstens 11 voll ausgelastete Schiffe zu finden.

Diese Fragestellung liefert uns den komplexesten Teil der gesamten Aufgabe. Das Problem besteht darin, dass sich die Wahrscheinlichkeit nicht direkt berechnen, sondern nur über eine Subtraktion ermitteln lässt. Die nachfolgende Abbildung illustriert die Vorgehensweise.



Zunächst wird die Wahrscheinlichkeit dafür berechnet, dass sich in einer Stichprobe des Umfangs n = 20 höchstens 11 voll ausgelastete Frachtschiffe befinden (Integral unter der Verteilungskurve bis zum Punkt b). Anschließend wird die Wahrscheinlichkeit dafür berechnet, dass sich in einer Stichprobe des Umfangs n = 20 höchstens 5 voll ausgelastete Frachtschiffe befinden (Integral unter der Verteilungskurve bis zum Punkt a). Zieht man nun die zweite errechnete Fläche von der ersten errechneten Fläche ab, bleibt exakt die „Zwischenfläche“ übrig, die für die Beantwortung der Fragestellung benötigt wird.

Wir berechnen also zunächst $P(X \leq 11)$ und subtrahieren dann $P(X \leq 5)$ (entsprechend $P(X < 6)$, da die 6 ja im Feld vorkommen soll („mindestens 6 voll ausgelastete Frachtschiffe“)). Demnach gilt: $P(6 \leq X \leq 11) = P(X \leq 11) - P(X \leq 5)$.

Gegeben sind:

$x_1 = 5$
$x_2 = 11$
$n = 20$
$p = 0{,}2$

Aus der Tabelle für die Verteilungsfunktion der Binomialverteilung (in die man bei der vorliegenden Größe von Stichprobe und Grundgesamtheit approximieren kann), lässt sich nun ablesen:

$$FB(x_1/n\,;p) = FB(5/20\,;0{,}2) = 0{,}8042$$
$$FB(x_2/n\,;p) = FB(11/20\,;0{,}2) = 0{,}9999$$

$$P(6 \leq X \leq 11) = 0{,}9999 - 0{,}8042 = 0{,}1957$$

Die Wahrscheinlichkeit, in einer Stichprobe von n = 20 Frachtschiffen mindestens 6 und höchstens 11 voll ausgelastete Schiffe zu finden, liegt demnach bei 19,57%.

(g) Wahrscheinlichkeit für das Einlaufen von mehr als 3 Frachtschiffen an einem Tag.

Bei diesem Aufgabenteil ist analog zu Aufgabe (1) mit der Poisson-Verteilung zu rechnen, da es sich offensichtlich um seltene Ereignisse handelt.

Gegeben sind:

μ = 2
X = 3

Zu berechnen ist das Komplementärereignis, d.h. $1 - P(X \leq 3)$:

$1 - FP(3/2) = 1 - 0{,}8571 = 0{,}1429$ (abzulesen in der Verteilungsfunktion)

Die Wahrscheinlichkeit auf das Einlaufen von mehr als 3 Schiffen liegt also bei 14,29%.

Lösung zu Aufgabe 4: Hier ist zunächst festzustellen, dass ein Modell ohne Zurücklegen (jeder Angestellte wird maximal einmal untersucht) und damit eine hypergeometrische Verteilung vorliegt.

(a) Wahrscheinlichkeit dafür, dass 5 von 6 Angestellten eine Amalgamfüllung besitzen.

Stichprobengröße und Größe der Grundgesamtheit lassen eine Approximation in die Bionomialverteilung nicht zu, weshalb mit der Wahrscheinlichkeitsfunktion der hypergeometrischen Verteilung gerechnet werden muss.

Gegeben sind:

N = 25
M = 24
n = 6
x = 5

$$fH(x/N\,;n;M) = \frac{\binom{M}{x} * \binom{N-M}{n-x}}{\binom{N}{n}}$$

$$fH(5/25\,;6;24) = \frac{\binom{24}{5} * \binom{25-24}{6-5}}{\binom{25}{6}} = 0{,}24$$

Die Wahrscheinlichkeit, dass 5 der 6 untersuchten Angestellten eine Amalgamfüllung besitzen, liegt demnach bei 24%.

(b) Welche Verteilung wäre zu wählen, wenn die Untersuchung bei einer Stichprobe von 5 Personen in einem Betrieb mit 150 Angestellten durchgeführt würde?

Auch in diesem Fall liegt natürlich eine hypergeometrische Verteilung vor, aus der heraus aber eventuell in die Binomialverteilung approximiert werden kann. Dies ist zu prüfen:

$\frac{5}{150} = 0{,}033333$ → Eine Approximation in die Binomialverteilung ist möglich.

Der p-Wert dieser Verteilung beträgt: $p = \frac{144}{150} = 0{,}96$

(c) Wahrscheinlichkeit dafür, dass 7 von 120 Personen eine Goldfüllung besitzen.

Da Goldfüllungen im Kontext dieser Aufgabenstellung offenkundig sehr selten sind, kann mit der Poisson-Verteilung (für seltene Ereignisse) gerechnet werden.

Gegeben sind:

μ = 4,8 (= 0,04 * 120)
x = 7

Alternativ zum Ablesen aus einer Tabelle der Wahrscheinlichkeitsfunktion für die Poisson-Verteilung (die den hier untersuchten Wertebereich meist nicht mehr umfassen) kann das Ergebnis auch mit Hilfe der entsprechenden Formel errechnet werden:

$$fP(7/4\,;8) = \frac{(4{,}8^7 * e^{(-4{,}8)})}{7!} = 0{,}09586$$

Die gesuchte Wahrscheinlichkeit liegt demnach bei 9,586%.

Aufgabenblatt VIII: Normal-, Gleich- und Exponentialverteilung

Aufgabe 1: Am Bahnhof von Wernigerode fährt exakt alle 20 Minuten ein Zug in Richtung Halberstadt ab. Wie groß ist die Wahrscheinlichkeit dafür, dass ein Fahrgast

(a) länger als 15 Minuten
(b) oder weniger als 10 Minuten wartet?
(c) Wie lauten Erwartungswert und Varianz der Verteilung?

Aufgabe 2: Die Zeit zwischen dem Eintreffen jeweils zweier PKW an einer Ampel sei mit einem Erwartungswert von 0,25 Minuten exponentialverteilt.

(a) Wie groß ist die Wahrscheinlichkeit dafür, dass die Zeit zwischen dem Eintreffen zweier PKW an dieser Ampel höchstens eine halbe Minute beträgt?
(b) Wie groß ist die Wahrscheinlichkeit dafür, dass die Zeit zwischen dem Eintreffen zweier PKW an dieser Ampel zwischen 0,2 und 0,3 Minuten liegt?

Aufgabe 3: Wie eine Untersuchung der Lebensdauer von Autoreifen ergab, beträgt der Mittelwert 36.000 Fahrt-Kilometer bei einer Abweichung von 4.800 Kilometern. Die Verteilung der Lebensdauer kann durch eine Normalverteilung approximiert werden.

Wie viele Reifen sind demnach in einem Produktionslos von 20.000 Stück zu erwarten

(a) deren Lebensdauer unter 26.400 km liegt,
(b) deren Lebensdauer über 48.000 km liegt
(c) oder deren Lebensdauer zwischen 23.636 und 48.364 km liegt?

Bei welcher Grenze liegt die Lebensdauer der

(d) 25% schlechtesten Reifen
(e) und der 10% besten Reifen?

Lösung zu Aufgabe 1: Da die Chancen auf einen eintreffenden Zug in jeder Minute gleich groß sind („exakt alle 20 Minuten"), liegt hier eine Gleichverteilung vor.

(a) Wahrscheinlichkeit einer Wartedauer von mehr als 15 Minuten.

Gegeben sind:

a = 0 (Untergrenze)
b = 20 (Obergrenze)
x = 15

Die Chance, dass ein Zug eintrifft, beträgt in jeder Minute $p = \frac{1}{20}$. Wenn die Wartedauer mehr als 15 Minuten betragen soll, so ist das Komplementräereignis einer Wartedauer von weniger als 16 Minuten ($p \leq 15$) zu berechnen. Das Ergebnis wird dann von 1 abgezogen, um die Lösung für Aufgabenteil (a) zu errechnen.

$$FG(x/a\,;b) = \frac{(x-a)}{(b-a)}$$

$$FG(15/0;20) = \frac{(15-0)}{(20-0)} = \frac{15}{20} = 0{,}75 \rightarrow 1 - 0{,}75 = 0{,}25$$

Die Wahrscheinlichkeit einer Wartedauer von mehr als 15 Minuten liegt also bei 25%.

(b) Wahrscheinlichkeit einer Wartedauer von weniger als 10 Minuten.

Diese Aufgabenstellung lässt sich analog zu Aufgabenteil (a) leicht lösen:

$$FG(x/a\,;b) = \frac{(x-a)}{(b-a)}$$

$$FG(10/0\,;20) = \frac{(10-0)}{(20-0)} = \frac{10}{20} = 0{,}5$$

Die Wahrscheinlichkeit einer Wartedauer von weniger als 10 Minuten liegt also bei 50%.

(c) Berechnung von Erwartungswert und Varianz der Verteilung.

Der Erwartungswert einer Gleichverteilung berechnet sich wie folgt:

$E(X) = \frac{(a+b)}{2}$ mit a; b = Unter- bzw. Obergrenze

Bezogen auf das konkrete Beispiel des Aufgabenteils ergibt sich:

$$E(X) = \frac{(0+20)}{2} = \frac{20}{2} = 10$$

Die Varianz einer Gleichverteilung berechnet sich wie folgt:

$$VAR(X) = \frac{(b-a)^2}{12}$$

Bezogen auf das konkrete Beispiel des Aufgabenteils ergibt sich:

$$VAR(X) = \frac{(20-0)^2}{12} = \frac{400}{12} = 33{,}3333$$

Lösung zu Aufgabe 2:

(a) Wahrscheinlichkeit einer Zwischenzeit von höchstens 0,5 Minuten.

Die Aufgabenstellung („höchstens“) impliziert bereits, dass hier mit der Verteilungsfunktion der ebenfalls in der Aufgabenstellung benannten Exponentialverteilung gerechnet werden muss. Hierfür ist zunächst λ aus dem Erwartungswert (0,25 Minuten) zu berechnen.

$$E(X) = \frac{1}{\lambda} = 0{,}25$$

$$\lambda = \frac{1}{0{,}25} = 4$$

Anschließend setzt man die Werte in die Formel für die Verteilungsfunktion ein:

$$FE(x/\lambda) = 1 - e^{(-\lambda * x)}$$
$$FE(0{,}5/4) = 1 - e^{(-4*0{,}5)} = 0{,}8647$$

Die Wahrscheinlichkeit dafür, dass die Zeit zwischen zwei PKW-Ankünften höchstens eine halbe Minute beträgt, liegt somit bei 86,47%.

(b) Wahrscheinlichkeit einer Zwischenzeit zwischen 0,2 und 0,3 Minuten.

$$FE(0{,}2 \leq x \leq 0{,}3) = FE(x = 0{,}3) - FE(x = 0{,}2)$$
$$FE(0{,}3/4) = 1 - e^{(-4*0{,}3)} = 0{,}699$$
$$FE(0{,}2/4) = 1 - e^{(-4*0{,}2)} = 0{,}551$$
$$0{,}699 - 0{,}551 = 0{,}148$$

Die Wahrscheinlichkeit dafür, dass die Zeit zwischen zwei PKW-Ankünften höchstens zwischen 0,2 und 0,3 Minuten liegt, beträgt somit 14,8%.

Lösung zu Aufgabe 3:

(a) Anzahl von Reifen mit einer Lebensdauer kleiner als 26.400 km (bei 20.000 Reifen).

Gegeben sind:

μ = 36.000 (Erwartungswert)
σ = 4.800 (Varianz)

Die Normalverteilung kann mittels der sogenannten Z-Transformation in eine Standardnormalverteilung transformiert werden, damit die entsprechende Tabelle verwendet werden kann. Dies geschieht über die folgende Formel:

$$Z = \frac{(x - \mu)}{\sigma}$$

Der gesuchte Wert für die Lebensdauer lässt sich nun aus der Verteilungsfunktion (es liegt wieder eine „höchstens"-Beziehung – höchstens 26.400 km – vor) ablesen:

$$FN(Z) = FN\left(\frac{(x - \mu)}{\sigma}\right) = FN\left(\frac{26400 - 36000}{4800}\right) = FN(-2) = 1 - FN(2) = 1 - 0{,}9772 = 0{,}0228$$

In einem Produktionslos von 20.000 Reifen sind demnach 0,0228 * 20.000 = 456 Reifen mit einer Lebensdauer von höchstens 26.400 km zu erwarten.

(b) Anzahl von Reifen mit einer Lebensdauer von mehr als 48.000 km (bei 20.000 Reifen).

Um diese Aufgabe zu lösen, müssen wir das Komplementärereignis (Anteil von Reifen mit einer Lebensdauer von höchstens 48.000 km) berechnen und das Ergebnis von 1 subtrahieren. Da es sich um eine stetige Verteilung handelt, ist mit genau 0 Reifen zu rechnen, die eine Lebensdauer von exakt 48.000 km aufweisen, weshalb die Position des Grenzelements für die Berechnung keine Rolle spielt.

$$1 - FN\left(\frac{(48000 - 36000)}{4800}\right) = 1 - FN(2{,}5) = 1 - 0{,}9938 = 0{,}0062$$

In einem Produktionslos von 20.000 Reifen sind demnach 0,0062 * 20.000 = 124 Reifen mit einer Lebensdauer von mehr als 48.000 km zu erwarten.

(c) Anzahl von Reifen mit einer Lebensdauer zwischen 23.636 km und 48.364 km (bei 20.000 Reifen).

Um diese Aufgabe zu lösen, müssen wir zunächst den Wert der Verteilungsfunktion der Standardnormalverteilung für 48.364 km berechnen und anschließend den entsprechenden Wert für 23.636 km subtrahieren:

$$FN\left(\frac{(48364 - 36000)}{4800}\right) - FN\left(\frac{(23636 - 36000)}{4800}\right) = FN(2{,}5758) - FN(-2{,}5758)$$
$$= FN(2{,}5758) - \big(1 - FN(2{,}5758)\big) = 0{,}995 - 0{,}005 = 0{,}9$$

In einem Produktionslos von 20.000 Reifen sind demnach 0,9 * 20.000= 19.800 Reifen mit einer Lebensdauer zwischen 23.636 km und 48.364 km zu erwarten.

(d) Bei welcher Grenze liegt die Lebensdauer der 25% schlechtesten Reifen?

Zur Lösung dieser Aufgabe müssen wir die bereits bekannte Formel für die Z-Transformation

$$Z = \frac{(x - \mu)}{\sigma}$$

nach x (Grenze) umstellen. Da μ und σ über die Aufgabenstellung gegeben sind, fehlt zur Auflösung der Gleichung noch der Wert für Z. Weil Z = 0,25 in den meisten Tabellierungen nicht enthalten ist, sucht man Z = 0,75 und kehrt das Vorzeichen um (Komplementärereignis). Für Z = 0,75 findet sich ein Wert von 0,7502 bei 0,67, also rechnet man mit Z = -0,67 weiter.

Nach x umgestellt lautet die Z-Transformationsformel: $x = Z * \sigma + \mu$
Eingesetzt ergibt sich: $x = -0{,}67 * 4800 + 36000 = 32784$

Die Lebensdauer-Grenze der 25% schlechtesten Reifen liegt demnach bei 32.784 km.

(e) Bei welcher Grenze liegt die Lebensdauer der 10% besten Reifen?

Hier kann man analog zu Aufgabenteil (d) vorgehen. Für Z = 0,9 findet sich in der Tabelle der Standardnormalverteilung der Wert 1,28, weshalb man zu folgender Kalkulation gelangt:

$$x = 1{,}28 * 4800 + 36000 = 42144$$

Die Lebensdauer-Grenze der 10% besten Reifen liegt demnach bei 42.144 km.

Aufgabenblatt IX: Normal-, Bionomial- und Poissonverteilung

Aufgabe 1: Bei der Herstellung von Werbegeschenken gibt es produktionsbedingt 10% Ausschuss. Wie groß ist die Wahrscheinlichkeit dafür, dass in einer Stichprobe von 50 zufällig ausgewählten Werbegeschenken

(a) genau die Hälfte,
(b) höchstens die Hälfte
(c) oder zwischen 5 und 10 Werbegeschenken defekt sind?

Nach einer Modernisierung der Produktionsanlagen liegt die neue Ausschussquote nun bei 5%. Diese soll anhand einer größeren Stichprobe im Umfang von 500 zufällig ausgewählten Werbegeschenken belegt werden. Wie groß ist hier die Wahrscheinlichkeit

(d) mindestens 100 defekte
(e) oder höchstens 25 defekte Werbegeschenke zu ziehen?

Als Werbegeschenk für besondere Kunden werden Füllfederhalter mit Goldfeder und Diamantspitze hergestellt. Der Ausschussanteil bei diesen Werbegeschenken liegt bei 25%. Berechnen Sie die Wahrscheinlichkeit, dass von 10 hergestellten Füllfederhaltern

(f) genau 1
(g) und zwischen 1 und 9 defekt sind.

Aufgabe 2: Der Alkoholgehalt eines Bieres der Marke “Auerhahntee” liegt gleichmäßig verteilt zwischen 3,8% und 4,4%. Eine Bierflasche wird zufällig der laufenden Produktion entnommen und auf ihren Alkoholgehalt getestet. Wie groß ist die Wahrscheinlichkeit, dass

(a) der Alkoholgehalt höchstens bei 4,0% liegt?
(b) Der Alkoholgehalt exakt 4,4% beträgt?

Eine Veränderung im Brauverfahren führt zu einem normalverhalten Alkoholanteil mit dem Erwartungswert 4,0 und einer Varianz von 0,01.

(c) Wie groß ist bei einer zufälligen Stichprobe nun die Wahrscheinlichkeit für einen Alkoholgehalt von mindestens 3,8%?

Lösung zu Aufgabe 1:

(a) Genau die Hälfte der Werbegeschenke in der Stichprobe erweist sich als Ausschuss.

Da lediglich zwei Ausprägungen existieren (defekt oder nicht defekt) und sich die einzelnen Zufallsexperimente nicht gegenseitig beeinflussen (da alle 50 Einheiten getestet werden und die größere Grundgesamtheit unbekannt ist), rechnen wir mit der Binomialverteilung.

Gegeben sind:

n = 50
x = 25
p = 0,1

Über die Tabelle für die Wahrscheinlichkeitsfunktion erhalten wir:

$$fB = (x/n\,;p) = fB(25/50\,;0{,}1) = 0$$

Die Wahrscheinlichkeit für genau 50% defekte Werbegeschenke liegt also bei 0%.

(b) Höchstens die Hälfte der Werbegeschenke in der Stichprobe erweist sich als Ausschuss.

Wir rechnen mit der Binomialverteilung weiter, werfen nun aber anstatt in die Tabelle der Wahrscheinlichkeitsfunktion wie in Aufgabenteil (a) einen Blick in die Tabelle der Verteilungsfunktion, da nach einem Wertebereich („höchstens“) gefragt wird.

Gegeben sind:

n = 50
x = 25
p = 0,1

Über die Tabelle für die Verteilungsfunktion erhalten wir:

$$FB(x/n\,;p) = FB(25/50\,;0{,}1) = 1$$

Die Wahrscheinlichkeit für höchstens 50% defekte Werbegeschenke liegt also bei 100%.

(c) Wahrscheinlichkeit für 5 bis 10 defekte Werbegeschenke in der Stichprobe.

Hier berechnen wir den Wert der Verteilungsfunktion für höchstens 10 defekte Werbegeschenke und ziehen dann den entsprechenden Wert derselben Funktion für höchstens 4 defekte Werbegeschenke ab.

$$FB(10/50\,;0{,}1) - FB(4/50\,;0{,}1) = 0{,}9906 - 0{,}4312 = 0{,}5594$$

Die Wahrscheinlichkeit für 5 bis 10 defekte Werbegeschenke liegt also bei 55,94%.

(d) Wahrscheinlichkeit für mindestens 100 defekte Produkte nach der Umstellung.

Da man bei einem Stichprobenumfang von mehr als 100 und einer Wahrscheinlichkeit von unter 5% aus der Binomialverteilung in die Poisson-Verteilung approximiert, werden die noch folgenden Aufgabenteile (d) und (e) in dieser Verteilung berechnet. Für (a) lässt sich gemäß der Approximationsbedingungen aus der Poisson-Verteilung weiter in die Normalverteilung approximieren. Diese wird wie üblich in die Standardnormalverteilung transformiert, wozu man sich der bereits bekannten Z-Transformationsformel bedient:

$$Z = \frac{(x - \mu)}{\sigma}$$

Erwartungswert und Varianz von 25 lassen sich der Aufgabenstellung entnehmen (5% defekte Werbegeschenke von 500 hergestellten = 25). Für x setzen wir in diesem Fall 400 ein, da wir „mindestens 100 defekte Werbegeschenke“ nicht direkt berechnen können, sondern die Fragestellung zuvor in das Komplementärereignis „höchstens 400 nicht defekte Werbegeschenke“ umformulieren müssen.

Es ergibt sich:

$$Z = \frac{(400 - 25)}{25} = \frac{375}{25} = 15$$

Eingesetzt in die Verteilungsfunktion der Standardnormalverteilung erhält man für Z=15 den Wert 1 (alle Werte oberhalb von Z=4 sind 1). Da wir das Komplementärereignis berechnet haben, muss das Ergebnis von 1 subtrahiert werden, d.h. 1-1=0. Dies bedeutet, dass die Wahrscheinlichkeit für mindestens 100 defekte Werbegeschenke in der Stichprobe bei 0% liegt.

(e) Wahrscheinlichkeit für höchstens 25 defekte Werbegeschenke nach der Umstellung.

Prinzipiell geht man hier wie bei Aufgabenteil (a) vor – zunächst wird approximiert, anschließend setzt man die entsprechenden Werte in die Transformationsformel ein:

$$Z = \frac{(x - \mu)}{\sigma}$$

$$Z = \frac{(475 - 475)}{475} = 0$$

Für Z=0 findet sich in der Tabelle der Wert 0,5. Dies bedeutet, dass die Wahrscheinlichkeit für höchstens 25 defekte Werbegeschenke in der Stichprobe bei 50% liegt.

(f) Genau 1 Füllfederhalter ist defekt.

Da der Wert für n hier wieder sehr klein ausfällt, lässt sich die Aufgabe durch den Rückgriff auf die Wahrscheinlichkeitsfunktion der Binomialverteilung lösen.

Gegeben sind:

n = 10
x = 1

p = 0,25

Aus der Tabelle für die Wahrscheinlichkeitsfunktion der Binomialverteilung lesen wir ab:

$$fB(x/n\,;p) = fB(1/10\,;0{,}25) = 0{,}1877$$

Die Wahrscheinlichkeit dafür, dass genau 1 von 10 produzierten Edel-Füllfederhaltern defekt ist, liegt demnach bei 18,77%.

(b) Zwischen 1 und 9 Füllfederhaltern sind defekt.

Hier berechnet man zunächst den Wert der Verteilungsfunktion der Binomialverteilung für 9 defekte Füllfederhalter und subtrahiert anschließend den entsprechenden Wert derselben Verteilung für 0 defekte Füllfederhalter (die 1 gehört mit in das betrachtete Feld).

Gegeben sind:

n = 10
$x_1 = 0$
$x_2 = 9$
p = 0,25

Aus der Tabelle für die Verteilungsfunktion der Binomialverteilung lesen wir ab:

$$FB(x_2/n\,;p) - FB(x_1/n\,;p) = FB(9/10\,;0{,}25) - FB(0/10\,;0{,}25) = 1 - 0{,}0536 = 0{,}9464$$

Die Wahrscheinlichkeit dafür, dass zwischen 1 und 9 von 10 produzierten Edel-Füllfederhaltern defekt sind, liegt demnach bei 94,64%.

Lösung zu Aufgabe 2:

(a) Es sind höchstens 4,0% Alkohol im Bier.

Für die Berechnung der „höchstens"-Wahrscheinlichkeit benötigen wir die Tabelle der Verteilungsfunktion der Gleichverteilung (die direkt in der Aufgabenstellung benannt wird):

$$FG(x/a\,;b) = \frac{(x-a)}{(b-a)} = FG(4{,}0/3{,}8\,;4{,}4) = \frac{(4{,}0-3{,}8)}{(4{,}4-3{,}8)} = \frac{0{,}2}{0{,}6} = 0{,}3333$$

Die gesuchte Wahrscheinlichkeit liegt somit bei 33,33%.

(b) Es sind genau 4,4% Alkohol im Bier.

In dieser Aufgabe ist nach der Wahrscheinlichkeit für das Auftreten einer ganz bestimmten Ausprägung in einer stetigen Verteilung gefragt. Dieser liegt jedoch stets bei 0 (Integral über einem Punkt). Die gesuchte Wahrscheinlichkeit liegt somit bei 0%.

(c) Es sind (nach der Umstellung) mindestens 3,8% Alkohol im Bier.

Da wir uns nunmehr in der Normalverteilung befinden, müssen wir zunächst mit Hilfe der bekannten Z-Transformationsformel in die Standardnormalverteilung transformieren:

$$Z = \frac{(x - \mu)}{\sigma}$$

$$Z = \frac{(3{,}8 - 4{,}0)}{\sqrt{0{,}01}} = \frac{-0{,}2}{0{,}1} = -2$$

Das Ergebnis kann nun mit Hilfe der Verteilungsfunktion der Standardnormalverteilung gefunden werden. Achtung: Es ist das Komplementärereignis zu berechnen („mindestens“).

$$1 - FS(-2) = 1 - \big(1 - FS(2)\big) = 1 - 0{,}0228 = 0{,}9772$$

Die Wahrscheinlichkeit für einen Alkoholgehalt von mindestens 3,8% liegt nach der Umstellung also bei 97,72%.

Aufgabenblatt X: Konfidenzintervalle

Aufgabe 1: Die Geschäftsleitung eines großen Kaufhauses möchte feststellen, in welchem Maß die Angestellten von den angebotenen Personalrabatten Gebrauch machen. Frühere Befragungen haben ergeben, dass die Rabattleistungen des Wahrenhauses an einzelne Angestellte relativ großen Schwankungen unterliegen, die Standardabweichung jedoch beinahe konstant war und bei 10 Euro im Monat lag.

(a) Eine neuerliche Befragung von 100 Angestellten ergab einen durchschnittlichen Rabattvorteil von 20 Euro im Monat. Geben Sie ein 95%-Konfidenzintervall um den Erwartungswert an.
(b) Berechnen Sie nun das 99%-Konfidenzintervall für die Standardabweichung. Wie verändert sich das Intervall bei einer Vergrößerung der Sicherheit?
(c) Wie verändert sich das 95%-Konfidenzintervall für die Standardabweichung wenn eine Befragung von 400 Angestellten denselben durchschnittlichen Rabattvorteil von 20 Euro im Monat ergibt?
(d) Wie groß muss der Stichprobenumfang sein, wenn das 95%-Konfidenzintervall so beschaffen sein soll, dass keine Abweichungen von mehr als +/- 2 vom Erwartungswert auftreten?
(e) Angenommen, es würden lediglich 20 Angestellte befragt, wobei sich $x = 20$ und $s = 10$ bei unbekannter Standardabweichung ergeben. Errechnen Sie das 95%-Konfidenzintervall für den Erwartungswert unter der Annahme einer normalverteilten Grundgesamtheit.
(f) In einer anderen Filiale ergab sich bei einer Befragung von 100 Angestellten ein durchschnittlicher Rabattvorteil von 30 Euro pro Monat bei einer Standardabweichung von 10 Euro. Kann man sagen, dass die Angestellten dieser Filiale die Rabattvorteile in größerem Maße nutzen? Berechnen Sie ein 99%-Konfidenzintervall für die Differenz der Mittelwerte μ_1 und μ_2.

Aufgabe 2: Ein Fabrikant nimmt die Produktion einer bestimmten, leicht verderblichen Konserve auf. Die mittlere Frischhaltedauer (Erwartungswert) und die Standardabweichung sind ihm nicht bekannt. Er weiß aber, dass die Verteilung der Frischhaltedauer gut durch eine Normalverteilung approximiert werden kann. Die Standardabweichung der Stichprobe ist $s=3$.

Welche Grenzen schließen die Standardabweichung bei 1% Irrtumswahrscheinlichkeit und 21 Beobachtungen ein?

Aufgabe 3: Aus Fragebögen, die zur Studienrückmeldung auszufüllen waren, wurden zufällig 2.500 Bögen ausgewählt. Dabei zeigte sich unter anderem, dass 1.500 Studenten überwiegend aus Mitteln der Eltern und 625 überwiegend durch staatliche Fördermittel unterhalten werden.

(a) Für die Anteilswerte der beiden Finanzierungstypen sind die Konfidenzintervalle mit einer Irrtumswahrscheinlichkeit von 1% zu ermitteln.
(b) Berechnen Sie die Konfidenzintervalle für eine Sicherheitswahrscheinlichkeit von 99%, wenn nur 200 Fragebögen mit 100 überwiegend von den Eltern unterhaltenen und 42 überwiegend staatlich geförderten Studenten herausgegriffen werden.
(c) Wie viele Bögen muss man herausgreifen, damit die Konfidenzintervalle für beliebige p bei einem Sicherheitsgrad von 95% eine Breite von 0,04 haben?

Lösung für Aufgabe 1:

(a) Berechnung des 95%-Konfidenzintervalls um μ.

Gegeben sind:

$\bar{x} = 20$
n = 100
α = 0,05 (= 1 – 0,95)
σ = 10

Die Bildungsvorschrift für das Konfidenzintervall lautet:

$$P\left(\bar{x} - z_{\left(1-\frac{\alpha}{2}\right)} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\left(1-\frac{\alpha}{2}\right)} * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Die beiden z-Werte lassen sich der Tabelle der Z-Verteilung entnehmen. Da es sich um ein symmetrisches Konfidenzintervall handelt, wird der gleiche Wert auf beiden Seiten der Ungleichung eingesetzt.

$$z_{\left(1-\frac{0,05}{2}\right)} = z_{(0,975)} = 1,96$$

Die z-Werte werden anschließend in die Bildungsvorschrift eingesetzt:

$$P(20 - 1,96 * 1 \leq \mu \leq 20 + 1,96 * 1) = 0,95$$
$$P(18,04 \leq \mu \leq 21,96) = 0,95$$

Mit einer Sicherheit von 95% liegt der wahre Erwartungswert der Grundgesamtheit also zwischen 18,04 und 21,96.

(b) Berechnung des 99%-Konfidenzintervalls für μ.

Diese Aufgabe ist analog zu Aufgabenteil (a) zu lösen, aufgrund des geänderten α muss im Grunde lediglich der z-Wert neu berechnet werden:

$$z_{\left(1-\frac{0,01}{2}\right)} = z_{(0,995)} = 2,576$$

Eingesetzt in die Bildungsvorschrift ergibt sich:

$$P(20 - 2,576 * 1 \leq \mu \leq 20 + 2,576 * 1) = 0,99$$
$$P(17,424 \leq \mu \leq 22,576) = 0,99$$

Wir sehen, dass das 99%-Konfidenzintervall mit einer Breite von 5,152 breiter als das zuvor berechnete 95%-Konfidenzintervall mit einer Breite von 3,92 ist. Dies liegt an der erhöhten Sicherheit, dass sich der reale Wert auch tatsächlich im Intervall befindet. Grundsätzlich gilt also, dass Konfidenzintervalle mit zunehmender Sicherheit breiter werden bzw. sich mit abnehmender Sicherheit verkleinern.

(c) Veränderung des 95%-Konfidenzintervalls durch neue 400-Personen-Befragung.

Im Vergleich zu Aufgabenteil (a) verbreitert sich hier lediglich die Basis für die Berechnungen – der Stichprobenumfang n. Alle anderen Werte bleiben gleich.

Wir können das neue Konfidenzintervall also wie folgt aufstellen:

$$P(20 - 1{,}96 * 0{,}5 \leq \mu \leq 20 + 1{,}96 * 0{,}5) = 0{,}95$$
$$P(19{,}02 \leq \mu \leq 20{,}98) = 0{,}95$$

Im Vergleich zum im Aufgabenteil (a) erstellten Konfidenzintervall lässt sich feststellen, dass die Breite des Konfidenzintervalls mit 1,96 verringert wurde. Dies passiert immer, wenn der Stichprobenumfang vergrößert wird, da eine größere Stichprobe auch eine größere Sicherheit bei der Vorhersage gestattet.

Im Gegensatz zur nachträglichen Manipulation des Konfidenzwertes α ist die Änderung des Stichprobenumfangs im übrigen eine legitime Maßnahme um zu breit geratene Konfidenzintervalle zu verringern. Dabei muss die Untersuchung natürlich neu durchgeführt, d.h. es muss eine neue Zufallsstichprobe gezogen werden.

(d) Größe des Stichprobenumfangs bei 95%-Konfidenzintervall um μ mit Umfang +/-2.

Hier muss mit dem Genauigkeitsmaß (absoluter Fehler) gerechnet werden.

Die benötigte Formel lautet:

$\Delta\mu = z * (\frac{\sigma}{\sqrt{n}})$ → Da nach n gesucht wird, muss die Gleichung quadriert werden.
$\Delta\mu^2 = z^2 * \frac{\sigma^2}{n}$ → Im nächsten Schritt ist die Gleichung nach n umzustellen.
$n = \frac{z^2 * \sigma^2}{\Delta\mu^2}$ → Nun sind die Werte aus der Aufgabenstellung einzusetzen.

$$n = \frac{(1{,}96^2 * 10^2)}{2^2} = \frac{384{,}16}{4} = 96{,}04$$

Es müssen demnach mindestens 97 Elemente in der Stichprobe sein, da 96 Elemente noch zu wenig wären und 96,04 kein brauchbares Ergebnis ist.

(e) 95%-Konfidenzintervall bei einer Befragung von 20 Personen und unbekanntem σ.

Wenn die Standardabweichung der Grundgesamtheit σ nicht bekannt ist, gilt für das Konfidenzintervall eine alternative Bildungsvorschrift, die die Standardabweichung der Stichprobe s miteinbezieht und auf der t-Verteilung basiert.

$$P\left(\bar{x} - t_{\left(1-\frac{\alpha}{2};n-1\right)} * \frac{s}{\sqrt{n-1}} \leq \mu \leq \bar{x} + t_{\left(1-\frac{\alpha}{2};n-1\right)} * \frac{s}{\sqrt{n-1}}\right) = 0{,}95$$

Gegeben sind:

$\bar{x} = 20$
n = 20

s = 10
α =0,05 (= 1 – 0,95)

Der Wert für t ist der Tabelle für die t-Verteilung / Student-Verteilung zu entnehmen.

$$t_{(1-\frac{\alpha}{2};n-1)} = t_{(0,975;19)} = 2,093$$

Mit diesem Wert wird nun die Bildungsvorschrift des Konfidenzintervalls komplettiert.

$$P(20 - 2,093 * 2,294 \leq \mu \leq 20 + 2,093 * 2,294) = 0,95$$
$$P(15,1986 \leq \mu \leq 24,8013) = 0,95$$

(f) Nutzen die Angestellten der zweiten Filiale den Rabattvorteil in größerem Maße?

Da hier die Standardabweichung der Grundgesamtheit σ wieder gegeben ist, kann auf die Bildungsvorschrift für das Konfidenzintervall um μ aus Aufgabenteil (a) zurückgegriffen werden. Zu beachten ist, dass mit zwei x-Werten zu rechnen ist, da das Konfidenzintervall für die Differenz(!) beider Mittelwerte benötigt wird.

Gegeben sind:

$\overline{x_1} = 30$
$\overline{x_2} = 20$
x = 20 – 30
n = 100
α = 0,01 (= 1 – 0,99)
σ = 10

$$P\left(\bar{x} - z_{\left(1-\frac{\alpha}{2}\right)} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{\left(1-\frac{\alpha}{2}\right)} * \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Die beiden z-Werte muss man nun aus der Tabelle der Z-Verteilung ablesen. Da es sich um ein symmetrisches Konfidenzintervall handelt, gibt es nur einen einzigen Wert, der auf beiden Seiten der Ungleichung eingesetzt wird.

$$z_{(1-\frac{0,01}{2})} = z_{(0,995)} = 2,576$$

Der Term $\frac{\sigma}{\sqrt{n}}$ muss noch umgerechnet werden, da in der Aufgabe zwei Standardabweichungen angegeben sind:

$$\sigma^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{100}{100} + \frac{100}{100} = 1 + 1 = 2 \rightarrow \sigma = \sqrt{2}$$

Die gefundenen Werte werden nun in die Bildungsvorschrift eingesetzt:

$$P((20 - 30) - 2,576 * \sqrt{2} \leq \mu_1 - \mu_2 \leq \left((20 - 30) + 2,567 * \sqrt{2}\right) = 0,99$$
$$P(-13,643 \leq \mu \leq -6,357) = 0,99$$

Dies bedeutet, dass der Erwartungswert für die Rabattnutzung bei der zweiten Filiale um 6,357 bis 13,643 größer ist als bei der ersten Filiale. Man kann daher feststellen, dass die Angestellten der zweiten Filiale die Rabattvorteile tatsächlich in größerem Maße nutzen.

Lösung für Aufgabe 2: Da für das Konfidenzintervall um σ keine Bildungsvorschrift existiert, ist zur Lösung dieser Aufgabe das Konfidenzintervall um σ^2 zu bilden, wobei aus der gesamten Ungleichung zum Schluss die Wurzel gezogen werden muss.

Gegeben sind:

$\bar{x} = 29$
s = 3
n = 21
α = 0,01

Vor der Berechnung des Konfidenzintervalls müssen nun noch die Chi^2-Werte aus der entsprechenden Tabelle abgelesen werden:

$$\chi^2_{(1-\frac{\alpha}{2};n-1)} = \chi^2_{(0,995;20)} = 39,997$$
$$\chi^2_{(\frac{\alpha}{2};n-1)} = \chi^2_{(0,005;20)} = 7,434$$

Eingesetzt in die Bildungsvorschrift des Konfidenzintervalls ergibt sich:

$$P\left(\frac{(21-1)*3^2}{39,997} \leq \sigma^2 \leq \frac{(21-1)*3^2}{7,434}\right) = 0,99$$
$$P(4,5 \leq \sigma^2 \leq 24,213) = 0,99$$

Nun muss noch die Wurzel gezogen werden:

$$P(2,1213 \leq \sigma \leq 4,9207) = 0,99$$

Lösung für Aufgabe 3:

(a) Konfidenzintervalle mit 1% Irrtumswahrscheinlichkeit für beide Anteilswerte.

Da hier ein Modell ohne Zurückliegen vorliegt, handelt es sich um eine hypergeometrische Verteilung, für die zunächst geprüft werden muss, ob in die Binomialverteilung approximiert werden darf.

$$\frac{n}{N} = \frac{1500}{2500} > 0,05$$

$$\frac{n}{N} = \frac{625}{2500} > 0,05$$

Eine Approximation in die Binomialverteilung ist nicht möglich. Denkbar wäre allerdings noch eine Approximation in die Normalverteilung. Hierfür sind zwei Bedingungen zu prüfen:

$$\hat{p} = \frac{x}{n}$$

$$\widehat{p_1} = \frac{1500}{2500} = 0{,}6$$

$$\widehat{p_2} = \frac{625}{2500} = 025$$

Die für eine Approximation zu erfüllende Bedingung lautet:

$$n * \hat{p} * (1 - \hat{p}) \geq 9$$

Dies wird nun für beide Finanzierungstypen überprüft:

$$2500 * 0{,}6 * 0{,}4 = 600 > 9 \rightarrow \text{OK}$$
$$2500 * 0{,}25 * 0{,}75 = 468{,}75 > 9 \rightarrow \text{OK}$$

Es kann in die Normalverteilung approximiert werden. Die hier als $\hat{p}$ berechneten Werte werden im Rahmen dieser Approximation zu den p-Werten der Normalverteilung.

Gegeben sind also:

$p_1 = 0{,}6$
$p_2 = 0{,}25$
$n = 2500$
$\alpha = 0{,}01 (= 1 - 0{,}99)$

Das Konfidenzintervall um p wird nach folgender Formel gebildet:

$$P\left(\hat{p} - z_{\left(1-\frac{\alpha}{2}\right)} * \sqrt{\frac{\hat{p} * (1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\left(1-\frac{\alpha}{2}\right)} * \sqrt{\frac{\hat{p} * (1-\hat{p})}{n}}\right) = 1 - \alpha$$

Der Wert für z ist der Tabelle für die Normalverteilung (Quantile) zu entnehmen:

$$z_{\left(1-\frac{\alpha}{2}\right)} = z_{(0{,}995)} = 2{,}5758$$

Nun kann man das Intervall für die elterliche Versorgung bilden:

$$P(0{,}6 - 2{,}5758 * 0{,}00979 \leq p \leq 0{,}6 + 2{,}5758 * 0{,}00979) = 0{,}99$$
$$P(0{,}5748 \leq p \leq 0{,}6252) = 0{,}99$$

Auch das Konfidenzintervall für die staatliche Versorgung lässt sich aufstellen:

$$P(0{,}25 - 2{,}5758 * 0{,}00866 \leq p \leq 0{,}25 + 2{,}5758 * 0{,}00866) = 0{,}99$$
$$P(0{,}22769 \leq p \leq 0{,}27239) = 0{,}99$$

(b) Berechnung neuer Konfidenzintervalle unter geänderten Umständen.

Gegeben sind nun:

$P_1 = 0{,}5$ (100/200)
$P_2 = 0{,}21$ (42/200)
$n = 2500$
$\alpha = 0{,}01 (= 1 - 0{,}99)$

Da alle anderen Werte gleichbleiben, kann sofort mit der Aufstellung der beiden Konfidenzintervalle begonnen werden. Zunächst das Intervall für die elterliche Unterstützung:

$$P(0{,}5 - 2{,}5758 * 0{,}03535 \leq p \leq 0{,}5 + 2{,}5758 * 0{,}03535) = 0{,}99$$
$$P(0{,}4089 \leq p \leq 0{,}5911) = 0{,}99$$

Anschließend folgt die Bildung des Intervalls für die staatliche Unterstützung:

$$P(0{,}21 - 2{,}5758 * 0{,}0288 \leq p \leq 0{,}21 + 2{,}5758 * 0{,}0288) = 0{,}99$$
$$P(0{,}1358 \leq p \leq 0{,}2842) = 0{,}99$$

(c) Wie viele Bögen müssen herausgegriffen werden, damit Konfidenzintervalle für beliebige p bei einen Sicherheitsgrad von 95% eine Breite von 0,04 haben?

Bei einer Breite des Konfidenzintervalls von 0,04 beträgt der absolute Fehler 0,02 (=0,04/2). Für die Berechnung weiterhin benötigt wird der Wert für z aus der Normalverteilung.

$$z_{(1-\frac{\alpha}{2})} = z_{(0{,}975)} = 1{,}96$$

Nun kann die Anzahl der herauszugreifenden Bögen über die Formel für das Genauigkeitsmaß ermittelt werden, die nach n umzustellen ist.

$\Delta\mu = z * (\frac{\sigma}{\sqrt{n}})$ → Da nach n gesucht wird, muss die Gleichung quadriert werden.
$\Delta\mu^2 = z^2 * \frac{\sigma^2}{n}$ → Im nächsten Schritt ist die Gleichung nach n umzustellen.
$n = \frac{z^2 * \sigma^2}{\Delta\mu^2}$ → Nun sind die Werte aus der Aufgabenstellung einzusetzen.

$$n = \frac{(1{,}96^2 * 0{,}5^2)}{0{,}02^2} = \frac{0{,}9604}{0{,}0004} = 2401$$

Es müssen daher genau 2401 Bögen herausgegriffen werden.

Aufgabenblatt XI: Statistische Testverfahren

Aufgabe 1: Bauer Mänz behauptet, dass seine Kartoffeln ein Durchschnittsgewicht von mindestens 100 g haben. Über die Streuung der Gewichte gibt er keine Auskunft. Bauer Meier kauft 36 Kartoffeln deren Durchschnittsgewicht 97 g bei einer Standardabweichung von 12 g beträgt. Kann Bauer Meier auf Basis dieser Stichprobe die Behauptung von Bauer Mänz bei einer Irrtumswahrscheinlichkeit von 5% wiederlegen?

Aufgabe 2: Ein Schütze trifft in 40% aller Fälle ins Schwarze, wenn er in Normalform ist. Vor einem Wettbewerb gibt er zur Kontrolle seiner Form 100 Schüsse ab, von denen 85 das Ziel treffen.

(a) Prüfen Sie bei einer Irrtumswahrscheinlichkeit von 5%, ob die Trefferquote vom Ergebnis bei Normalform signifikant nach oben abweicht (signifikant besser ist).
(b) Der gleiche Schütze gibt am nächsten Tag 200 Schüsse ab, von denen 90 ins Schwarze treffen. Testen Sie, ob dieses Ergebnis vom Ergebnis bei Normalform signifikant abweicht (α=0,05).

Aufgabe 3: Die Sollstärke von Unterrichtsgruppen an der Hochschule Harz liegt bei 30 Studenten. Eine Beobachtung von Gruppen zu 20 verschiedenen Zeitpunkten im Jahr 1995 ergab eine Gruppenstärke von durchschnittlich 35 Studenten bei einer Varianz von 16 Studenten^2(!).

(a) Überprüfen Sie die Behauptung, dass sich die Gruppenstärke an der Hochschule gegenüber der Sollstärke nicht erhöht hat (α=0,05).
(b) Im Jahr 1996 wird die Erhebung zu 22 Zeitpunkten wiederholt. Dabei wird eine durchschnittliche Gruppenstärke von 40 und eine Standardabweichung von 5 Studenten festgestellt. Hat sich die Gruppenstärke nun im Bezug auf das Vorjahr verändert? Unterstellen Sie für die Berechnung eine Normalverteilung der Gruppenstärke und eine über die Zeit unveränderte Varianz.

Lösung zu Aufgabe 1: Um eine Behauptung zu überprüfen, die auf einer oder mehreren Stichproben basiert, bedient man sich eines sogenannten statistischen Testverfahrens. Für alle derartigen Testverfahren gelten die folgenden Verfahrensregeln:

> Schritt I: Aufstellung der Parametermenge, der Nullhypothese und der Alternativhypothese sowie Festlegung des Signifikanzniveaus.
>
> Schritt II: Festlegung einer geeigneten Prüfgröße und Bestimmung der Testverteilung bei Gültigkeit der Nullhypothese.
>
> Schritt III: Bestimmung des kritischen Bereichs.
>
> Schritt IV: Berechnung des Wertes der Prüfgröße.
>
> Schritt V: Entscheidung und Interpretation.

Wir werden nachfolgend alle fünf Schritte für das geforderte Testverfahren durchlaufen.

Schritt I: Aufstellung der Parametermenge, der Nullhypothese und der Alternativhypothese sowie Festlegung des Signifikanzniveaus.

$H0: \bar{x} \geq 100$
$HA: \bar{x} < 100$
$\alpha = 0{,}05$
$t(0{,}995) = 2{,}576$

Schritt II: Festlegung einer geeigneten Prüfgröße und Bestimmung der Testverteilung bei Gültigkeit der Nullhypothese.

$T = \frac{(\bar{X}-\mu)}{(\frac{s}{\sqrt{n}})}$ ist t (n-1)-verteilt

Schritt III: Bestimmung des kritischen Bereichs.

Der Test ist linksseitig, es gibt daher einen Wert, der nicht unterschritten werden darf. Dieser Wert ist der Standardnormalverteilung zu entnehmen.

Schritt IV: Berechnung des Wertes der Prüfgröße.

$$T = \frac{(\bar{X}-\mu)}{(\frac{s}{\sqrt{n}})} = \frac{(97-100)}{(\frac{12}{\sqrt{36}})} = \frac{-3}{2} = -1{,}5$$

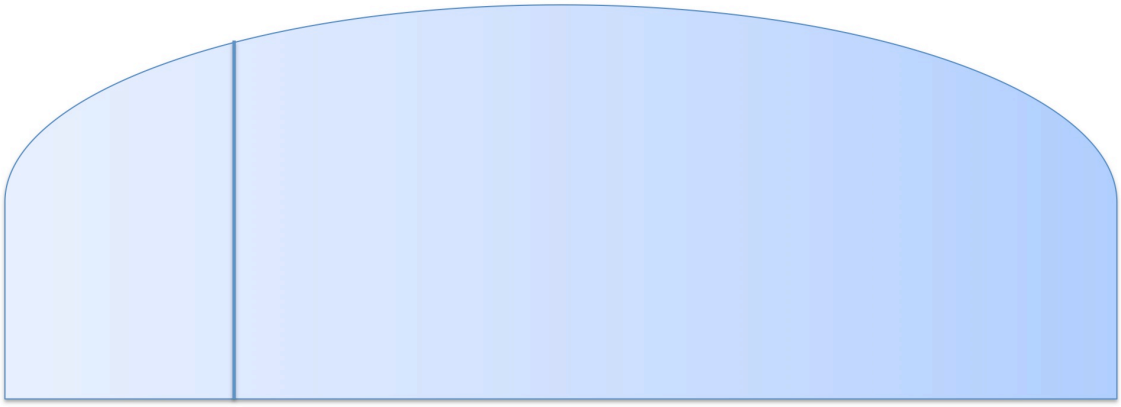
Schritt V: Entscheidung und Interpretation.

Die Nullhypothese ist in diesem Fall abzulehnen.

Lösung für Aufgabe 2:

(a) Weicht die Trefferquote vom Ergebnis bei Normalverteilung signifikant nach oben ab?

Hier handelt es sich um einen Einstichprobentest mit rechtsseitigem Ablehnungsbereich.



Rechtsseitiger Verwerfungsbereich

Zunächst werden Null- und Alternativhypothese aufgestellt und das Signifikanzniveaz festgelegt:

$H0: p = 0{,}4$
$HA: p > 0{,}4$
$\alpha = 0{,}05$

Da es sich um einen einseitigen Test handelt, wird α nicht durch 2 dividiert. Im nächsten Schritt wird der kritische Wert der Standardnormalverteilung entnommen:

$Z(0{,}95) = 1{,}6449$

Sollte die im nächsten Schritt zu berechnende Prüfgröße diesen Wert übersteigen, wäre die Nullhypothese abzulehnen. Zur Berechnung der Prüfgröße wird folgende Formel verwendet:

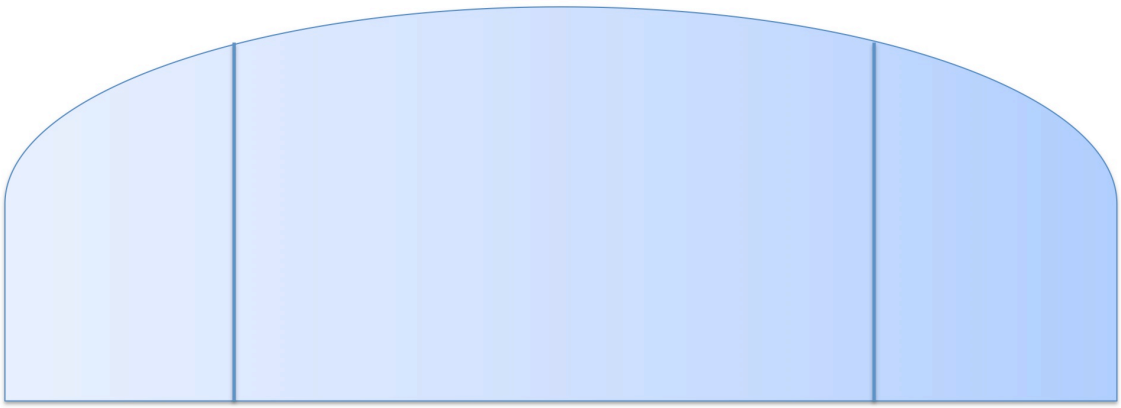
$$Z = \frac{(\hat{p} - p_0)}{\sqrt{\frac{(p_0 * (1 - p_0))}{n}}}$$

$$Z = \frac{(0{,}85 - 0{,}4)}{\sqrt{\frac{(0{,}4 * 0{,}6)}{100}}} = 9{,}18$$

Da die Prüfgröße den kritischen Wert eindeutig übersteigt, ist die Nullhypothese in diesem Fall abzulehnen.

(b) Weicht das zweite Ergebnis signifikant vom Ergebnis bei Normalform ab?

Auch hier ist ein Einstichprobentest gefordert, allerdings ist im Gegensatz zu Aufgabenteil (a) der Ablehnungsbereich diesmal beidseitig.



Beidseitiger Verwerfungsbereich

Zunächst müssen Null- und Alternativhypothese aufgestellt und das Signifikanzniveau festgelegt werden.

$H0: p = 0{,}4$
$HA: p \neq 0{,}4$
$\alpha = 0{,}05$

Der kritische Wert ist der Standardnormalverteilung zu entnehmen:

$Z(0{,}95) = 1{,}69$

Sollte die im nächsten Schritt zu berechnende Prüfgröße nun diesen Wert oder dessen negative Gegenwert (beidseitiger Test) über- bzw. unterschreiten, muss die Nullhypothese verworfen werden. Zur Berechnung der Prüfgröße ist die bekannte Formel zu verwenden:

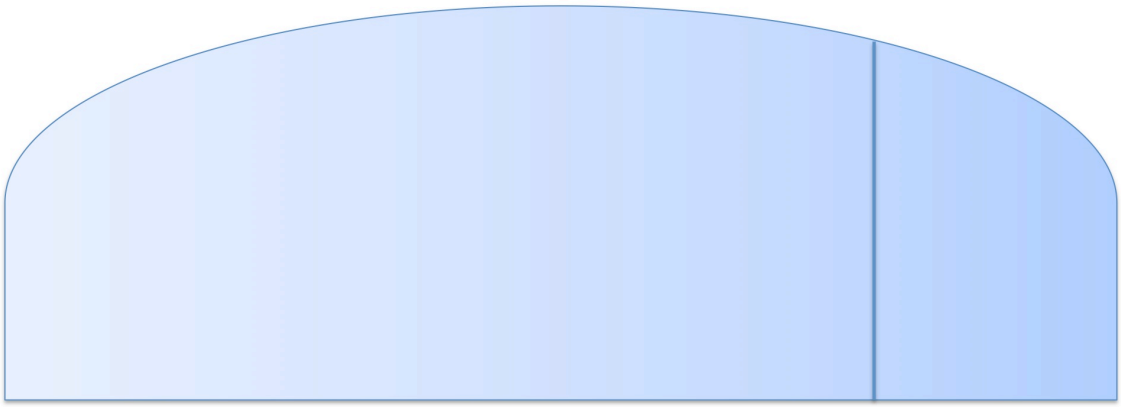
$$Z = \frac{(\hat{p} - p_0)}{\sqrt{\frac{(p_0 * (1 - p_0))}{n}}}$$

$$Z = \frac{(0{,}45 - 0{,}4)}{\sqrt{\frac{(0{,}4 * 0{,}6)}{100}}} = 1{,}4433$$

Da die Prüfgröße klar zwischen den beiden kritischen Werten liegt, ist die Nullhypothese in diesem Fall nicht abzulehnen. Die Trefferquote des Schützen weicht daher nicht signifikant von seinem Ergebnis bei Normalform ab.

Lösung zu Aufgabe 3: Hier ist ein weiterer Einstichprobentest gefordert, der jedoch bei einer unbekannten Varianz angewandt wird (die Varianz in der Aufgabenstellung bezieht sich auf die Varianz in der Stichprobe, nicht auf die Varianz der Grundgesamtheit). Auch hier sind zunächst Null- und Alternativhypothese aufzustellen und das Signifikanzniveau festzulegen.

$H0: \mu = 30$
$HA: \mu > 30$
$\alpha = 0{,}05$



Linksseitiger Verwerfungsbereich

Der kritische Wert ist in diesem Fall der t-Verteilung zu entnehmen, wobei die Stichprobe 19 Freiheitsgrade zulässt (20-1). Es ergibt sich ein kritischer Wert von 1,729. Die 0,05 werden nicht geteilt, da es sich um einen linksoffenen und rechtsgeschlossenen Test handelt. Die Nullhypothese ist zu verwerfen, wenn die Prüfgröße den kritischen Wert überschreitet.

Die Prüfgröße berechnet sich nach folgender Formel:

$$T = \frac{(\bar{X} - \mu)}{(\frac{s}{\sqrt{n}})} = \frac{(35 - 30)}{(\frac{4}{\sqrt{20}})} = 5{,}59$$

Hinweis zur Berechnung: Die Standardabweichung ist die Wurzel der Varianz.

Da der Wert der Prüfgröße den kritischen Wert klar übersteigt, ist die Nullhypothese in diesem Fall zu verwerfen. Wir können daher vermuten, dass sich die durchschnittliche Gruppenstärke in der Tat signifikant erhöht hat.

(b) Weist die Befragung im Folgejahr auf eine signifikant veränderte Gruppenstärke hin?

Da für diese Aufgabe die Werte zweier Stichproben miteinander verglichen werden müssen, greifen wir an dieser Stelle erstmalig auf einen Zweistichprobentest zurück. Auf für diesen ist im ersten Schritt eine Null- sowie eine Alternativhypothese aufzustellen sowie das Signifikanzniveau festzulegen.

$H0: \mu_1 = \mu_2$
$HA: \mu_1 \neq \mu_2$
$\alpha = 0{,}05$

Da es sich um einen zweiseitigen Test handelt, existiert sowohl ein linksseitiger wie auch ein rechtsseitiger Ablehnungsbereich. Der kritische Wert ist der t-Verteilung zu entnehmen, wofür zunächst die Anzahl der Freiheitsgerade zu bestimmen ist. Die Anzahl der Freiheitsgrade einer Stichprobe ergibt sich (bei bekanntem Mittelwert) aus dem Stichprobenumfang minus 1. Da in diesem Fall mit zwei Stichproben und zwei Mittelwerten gerechnet wird, müssen wir zunächst die Stichprobengrößen addieren und dann um die Anzahl der Mittelwerte verringern, d.h. 20 + 22 – 2 = 40.

Wir entnehmen also der tabellierten T-Verteilung den Zahlenwert bei 40 Freiheitsgeraden und 1 – α = 0,975 (1 - 0,05 / 2 = 0,975). Dieser Wert lautet 2,021. Die im nächsten Schritt zu berechnendet Prüfgröße führt demnach zu einer Ablehnung der Nullhypothese, wenn sie größer als 2,021 oder kleiner als -2,021 ausfällt.

Zur Berechnung der Prüfgröße ist folgende Formel zu verwenden:

$$T = \frac{(\overline{X_1} - \overline{X_2})}{(s * \sqrt{\frac{(n_1 + n_2)}{(n_1 * n_2)}})}$$

mit $s = \sqrt{\frac{((n_1-1)*s_1^2+(n_2-1)*s_2^2)}{(n_{1+}n_2-2)}}$

Setzt man die entsprechenden Werte aus der Aufgabenstellung ein, so erhält man:

$$T = \frac{(35-40)}{4{,}552 * \sqrt{\frac{42}{440}}} = -3{,}55$$

mit $s = \sqrt{\frac{(19*16+21*25)}{40}} = 4{,}552$

Da die Prüfgröße mit -3,55 eindeutig unter dem unteren kritischen Wert liegt, ist die Nullhypothese zu verwerfen. Zur Gruppenstärke kann damit ausgesagt werden, dass diese sich signifikant verändert zu haben scheint. Dabei deutet zwar alles auf eine weitere Erhöhung hin – allerdings lässt sich bei einem zweiseitigen Test keine eindeutige Aussage über die Art einer signifikanten Veränderung treffen.

Aufgabenblatt XII: Alle Themenbereiche

Aufgabe 1: Wie groß ist die Wahrscheinlichkeit, beim dreimaligen Werfen einer Münze genau dreimal "Wappen" zu erhalten?

Aufgabe 2: Auf zwei verschiedenen Bändern – B1 und B2 – wird der gleiche Fernsehröhrentyp gefertigt. B1 liefert 20%, B2 liefert 80% der Produktion. Der Ausschuss beträgt jeweils 10% bzw. 5%. Aus der Gesamtproduktion wird eine Röhre zufällig ausgewählt. Wie groß ist die Wahrscheinlichkeit dafür, dass

(a) die Röhre auf B1 gefertigt wurde, wenn sie einwandfrei ist?
(b) Die Röhre auf B2 gefertigt wurde, wenn sie defekt ist?

Aufgabe 3: Ein Medikament enthält einen Wirkstoff, die seine beste Wirksamkeit bei der Einnahme von 2,6 mg bis 9,5 mg zeigt. Der Gehalt des Wirkstoffs pro Pille ist normalverteilt mit einem Erwartungswert von 5 mg und einer Standardabweichung von 3 mg. Mit welcher Wahrscheinlichkeit hat eine zufällig ausgewählte Pille nun die beste Wirksamkeit?

Aufgabe 4: Ein Meinungsforschungsinstitut beabsichtigt, den prozentualen Stimmenanteil, den eine der beiden großen Parteien in einer bevorstehenden Wahl erreichen wird, zu prognostizieren. Wie viel zufällig ausgewählte Wahlbeteiligte müssen mindestens befragt werden, um ein Konfidenzintervall (α=0,01) zu erhalten, dessen Länge höchstens 2% beträgt?

Lösung zu Aufgabe 1: Beim Werfen einer Münze handelt es sich um ein typisches Zufallsexperiment, da die einzelnen Experimente (Münzwürfe) voneinander unabhängig sind und die beiden möglichen Ergebnisse (Wappen und Zahl) nach jedem Wurf wieder "zurückgelegt" werden, so dass es möglich ist, mehrfach Wappen und Zahl zu werfen. Es liegt demnach eine Binomialverteilung mit bekannter Wahrscheinlichkeit p = 0,5 vor. Der hier gesuchte Wert lässt sich der Tabelle der Wahrscheinlichkeitsfunktion entnehmen:

$$fB(x/n\,;p) = fB(3/5\,;0{,}5) = 0{,}3125$$

Die Wahrscheinlichkeit für einen dreifachen Wappenwurf liegt demnach bei 31,25%.

Lösung zu Aufgabe 2:

(a) Die Röhre ist einwandfrei und wurde auf B1 gefertigt.

Bei der Berechnung dieser Aufgabe muss der Satz von Bayes Anwendung finden. Benötigt werden auch die Komplementärereignisse:

$$\frac{(0{,}2 * 0{,}9)}{(0{,}2 * 0{,}9) + (0{,}8 * 0{,}95)} = 0{,}1941$$

Die Wahrscheinlichkeit dafür, dass eine einwandfreie Röhre auf B1 gefertigt wurde, liegt demnach bei 19,41%.

(b) Die Röhre ist defekt und wurde auf B2 gefertigt.

Die Berechnung dieser Aufgabe erfolgt analog zu Aufgabenteil (a):

$$\frac{(0{,}8 * 0{,}05)}{(0{,}2 * 0{,}1) + (0{,}8 * 0{,}05)} = 0{,}6666$$

Die Wahrscheinlichkeit, dass eine defekte Röhre auf B2 gefertigt wurde, liegt demnach bei 66,66%.

Lösung zu Aufgabe 3: Bei dieser Aufgabe muss berechnet werden, mit welcher Wahrscheinlichkeit der Wirkstoffgehalt in einer zufällig ausgewählten Pille zwischen 2,6 mg und 9,5 mg liegt. Hierfür transfomieren wir zunächst von der Normalverteilung in die Standardnormalverteilung (keine Approximation!) mittels der Z-Transformationsformel:

$$Z = \frac{(x - \mu)}{\sigma}$$

Da mit zwei Z-Werten zu rechnen ist, ergeben sich auch zwei Z-Variablen, nämlich Z_1= -0,8 und Z_2=1,5. Wie bei der Integralrechnung muss nun der Wert der unteren Grenze (der Z-Wert von 2,6) vom Wert der oberen Grenze (dem Z-Wert von 9,5) subtrahiert werden, damit nur noch die Fläche – die Wahrscheinlichkeit – zwischen beiden Grenzen übrig bleibt.

$$F(1{,}5) - F(-0{,}8) = F(1{,}5) - \big(1 - F(0{,}8)\big) = 0{,}9332 - (1 - 0{,}7881) = 0{,}721$$

Lösung zu Aufgabe 4: Bei dieser Rechnung benötigt man die Formel für den notwendigen Stichprobenumfang bei Schätzung des Anteilswertes p:

$$n = \frac{(z^2 * p(1-p))}{(\delta * p)^2}$$

Der nicht in der Aufgabenstellung zu findende Z-Wert kann der Tabelle der Standardnormalverteilung entnommen werden. Da nach einem zweiseitigen Konfidenzintervall gefragt wird, muss α noch halbiert werden.

In die obige Formel eingesetzt erhält man:

$$n = \frac{2{,}5758^2 * 0{,}5 * 0{,}5}{0{,}01^2} = 16586{,}9$$

Es müssen demnach mindestens 16.587 Personen befragt werden, um die geforderte Sicherheit beim Konfidenzintervall gewährleisten zu können.

Fehler entdeckt?

Die Musterlösungen und Hinweise in diesem Manuskript wurden von mir nach bestem Wissen und Gewissen zusammengestellt. Sollte sich dennoch irgendwo ein Fehler eingeschlichen haben, bin ich über Hinweise an

creinboth@hs-harz.de

dankbar. Bei zukünftigen Überarbeitungen dieses Skripts werde ich sämtliche Rückmeldungen und Verbesserungsvorschläge natürlich gerne berücksichtigen.