Vorlesung Statistik für Informatiker
Prof. Dr. Reimar Hofmann
Hochschule Karlsruhe – Technik und Wirtschaft

Aufgaben zu Kapitel 1 der Statistik-Vorlesung

`Hinweis: Nutzen Sie zur Lösung die` **`Vorlesungsfolien`** `und machen Sie sich die Verbindung zwischen Theorie und Anwendung klar.`

Zum Prüfen des Grundverständnisses

(Die Aufgaben sollten sehr schnell gehen, wenn Sie die Konzepte verstanden haben)

Aufgabe 1.A.1 (einfach, Grundbegriffe Kapitel 1.1)
Nebenstehende Tabelle zeigt die Ergebnisse einer Prüfung.
Bestimmen Sie für diese Prüfung …

Geschlecht	Bestanden?
M	B
W	NB
W	B
M	NB
M	B

a) … die absolute Häufigkeit von Frauen.
b) … die relative Häufigkeit von Frauen.
c) … h_B .
d) … f_B .
e) Geben Sie formale Bezeichner für die Größen aus (a) und (b) an.

Aufgabe 1.A.2 (einfach, Grundbegriffe Kapitel 1.1)
Worin besteht der Unterschied zwischen der relativen Häufigkeit und der Wahrscheinlichkeit eines Ereignisses und in welchem Zusammenhang stehen beide?

Aufgabe 1.A.3 (einfach, Grundbegriffe Kapitel 1.1)
Ergänzen Sie unter Verwendung des jeweiligen Fachbegriffs zu einem vollständigen (korrekten, verständlichen) Satz.

a) Von 200 Läufen gestern ist das Programm bei 50 abgestürzt. Also war …………………………………………………………………………………………… 25%.

b) Anscheinend ist also ………………………………………………………………………………, dass das Programm bei seinem nächsten Lauf abstürzen wird, ungefähr 25%.

Aufgabe 1.A.4 (einfach, Grundbegriffe Kapitel 1.1)
Welches sind zulässige Ergebnismengen für ein Fußballspielzwischen den Mannschaften A und B? (jew. kurze Begründung)

a) Ω = {0:0, 0:1, 1:0, 0:2, 1:1, 2:0, 0:3, 1:2, 2:1, 3:0, …}

b) Ω = {A gewinnt, unentschieden, B gewinnt}

c) Ω = {Tor in der letzten Minute, kein Tor in der letzten Minute}

d) Ω = {A gewinnt, unentschieden, B gewinnt , Tor in der letzten Minute, kein Tor in der letzten Minute}

e) Ω = {A gewinnt, B gewinnt, torlose Partie}

Aufgabe 1.A.5 (mittel, Kapitel 1.1)
Max spielt ein Glückspiel, bei dem gewürfelt wird. Bei den ersten 18 Versuchen ist keine einzige Sechs gekommen. Nun überlegt er sich Folgendes:

1. Nach dem Gesetz der großen Zahlen wird bei genügend vielen Versuchen die relative Häufigkeit von Sechs gegen 1/6 konvergieren.
2. Da bei den ersten 18 Versuchen keine Sechs dabei war, wird die Sechs also in Zukunft mit hoher Wahrscheinlichkeit häufiger als bei 1/6 der Würfe fallen.
3. Also sollte man für die folgenden Würfe mit mehr Sechsen rechnen, als wenn bei den ersten 18 Würfen 3 Sechsen gekommen wären.

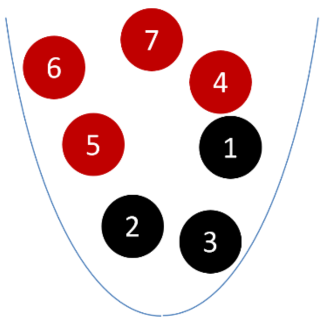
Wo liegt der Fehler in der Argumentation?

Aufgabe 1.A.6 (einfach, Kapitel 1.2)

a) Sie wollen ermitteln, mit welcher Wahrscheinlichkeit ein bestimmter Reißnagel mit der Spitze nach oben liegenbleibt, wenn er von Ihrem Tisch fällt. Wie können Sie dazu vorgehen?
b) Sie wollen Murphy's Law testen und ermitteln, mit welcher Wahrscheinlichkeit Ihr Pausen-Butterbrot mit der Butterseite nach unten landet, wenn es vom Tisch fällt. Wodurch wird dies schwieriger zu ermitteln sein, als bei (a)?

Aufgabe 1.A.7a (einfach, Kapitel 1.2)

Die Kugeln in einer Urne sind von 1 bis 7 durchnummeriert. Kugeln 1-3 sind schwarz, 4-7 sind rot. Wie wahrscheinlich ist es, dass eine zufällig gezogene Kugel (jede Kugel mit gleicher Wahrsch.)



a) rot ist?
b) geradzahlig ist?
c) rot *und* geradzahlig ist?
d) rot *oder* geradzahligist?

Begründen Sie dabei kurz Ihren Ansatz!

Aufgabe 1.A.7b (einfach, Kapitel 1.2)
Eine Urne enthält rote und schwarze Kugeln. Über die genaue Zusammensetzung sei nichts bekannt. Es sei aber bekannt, dass $P(R) = 60\%$, $P(G) = 40\%$ und $P(R \cap G) = 20\%$. Dabei steht ***R*** für das Ereignis „die gezogene Kugel ist **rot**", ***G*** analog für „die gezogene Kugel ist **geradzahlig**".
Wie wahrscheinlich ist es, dass eine gezogene Kugel

a) nicht rot ist?
b) rot *oder* geradzahlig ist?

Begründen Sie kurz Ihren Ansatz!

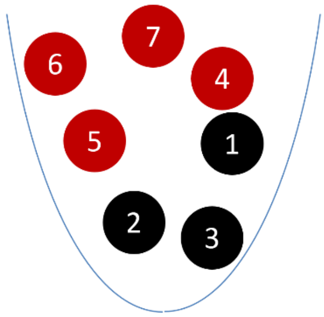
Aufgabe 1.A.8 (einfach, Kapitel 1.2 u. 1.3)
Sie werfen zwei Mal nacheinander eine Münze:

a) Wie wahrscheinlich ist es, dass beim **ersten** Wurf das Ergebnis „Kopf“ ist?
b) Wie wahrscheinlich ist es, dass beim **zweiten** Wurf das Ergebnis „Kopf“ ist?
c) Wie wahrscheinlich ist es, dass beim **ersten oder zweiten** Wurf das Ergebnis „Kopf“ ist?

Begründen Sie kurz Ihren Ansatz!

Aufgabe 1.A.9 (mittel, Kapitel 1.2)
Eine Website hat 90 000 kostenlos registrierte und 10 000 zahlende Nutzer.
Wie wahrscheinlich ist es, dass der nächste Nutzer, der sich am Server anmeldet, ein kostenlos registrierter Nutzer ist? (und es deshalb Sinn macht, neben der Login-Maske Werbung für ein Upgrade einzublenden?)

Aufgabe 1.A.10 (einfach, Kapitel 1.2 u. 1.3)



Die Kugeln in einer Urne sind von 1 bis 7 durchnummeriert.
Kugeln 1-3 sind schwarz, 4-7 sind rot.
Es werden ohne Zurückzulegen nacheinander **zwei Kugeln** gezogen. **Formalisieren** Sie die gesuchten Größen und **bestimmen** Sie ihren Wert. Verwenden Sie dabei die Ereignisse

R1 := „erste Kugel ist rot“,
R2 := „zweite Kugel ist rot“:

a) Wie wahrscheinlich ist es, *zwei schwarze* Kugeln zu ziehen?

b) Wie wahrscheinlich ist es, *zuerst eine rote, dann eine schwarze* zu ziehen?

c) Wie wahrscheinlich ist es, *eine rote und eine schwarze* zu ziehen? (Reihenfolge egal)

d) Wie wahrscheinlich ist es, dass die erste *oder* zweite Kugel rot ist?

Aufgabe 1.A.11 (einfach, Kapitel 1.3, Sprachverständnis bedingte W.)
Geben Sie bei den folgenden Teilaufgaben auch immer die korrekte **formale Bezeichnung** der gesuchten Größe an (z.B. P(rot | geradzahlig) oder P(rot ∪ geradzahlig)),
und bestimmen Sie ihren **Wert**:

Wir ziehen mit Zurücklegen aus der Urne aus der vorigen Aufgabe. Wie wahrscheinlich ist es, dass…

a) … eine rote geradzahlige Kugel gezogen wird?
b) … eine gezogene geradzahlige Kugel rot ist?
c) … eine gezogene rote Kugel geradzahlig ist?
d) … wenn eine geradzahlige Kugel gezogen wurde, sie rot ist?
e) … eine gezogene Kugel geradzahlig ist, wenn sie rot ist?
f) … eine gezogene Kugel, die rot ist, geradzahlig ist?
g) … eine gezogene Kugel rot und geradzahlig ist?
h) Welcher Anteil der gezogenen roten Kugeln, die rot sind, ist im langfristigen Mittel geradzahlig?
i) Welcher Anteil aller gezogenen Kugeln sind im langfristigen Mittel geradzahlige Kugeln, die rot sind?

Anwenden / Üben

Aufgabe 1.B.2 (mittel, Kapitel 1.2 u. 1,3)

Ein fairer Würfel wird nacheinander zwei Mal geworfen. Wie wahrscheinlich ist es, dass

a) erst eine Eins und dann eine Zwei kommt?

b) eine Eins und eine Zwei kommt, egal in welcher Reihenfolge? ("*Mäxchen*")

c) *beide Male* Sechs kommt? ("*Sechserpasch*")

d) mindestens eine Sechs kommt?

e) Wie viele mögliche Ergebnisse gibt es, und handelt es sich beim ZWEIMALIGEN Würfeln um ein Laplace-Experiment, …

i) … wenn man die Reihenfolge der Ergebnisse unterscheidet?

ii) … wenn man die Reihenfolge der Ergebnisse *nicht* unterscheidet?

(Tipp: Ergebnisse aus (b) und (c) vergleichen)

Aufgabe 1.B.3 (leicht bis mittel, Kapitel 1.2 u. 1,3)

Wie wahrscheinlich ist es, mit einem (fairen) Würfel …

… *nur Sechser* zu würfeln bei

a) zwei Würfen b) drei Würfen c) sechs Würfen

… *mindestens einmal Sechs* zu würfeln bei

d) zwei Würfen e) drei Würfen f) sechs Würfen ?

(Tipp: Betrachten Sie jeweils das Gegenereignis)

g) Warum wäre bei (d) falsch wie folgt zu rechnen:

Ereignisbezeichner: S_1: Erster Wurf Sechs, S_2: Zweiter Wurf Sechs,

$P(S_1 \cup S_2) = P(S_1) + P(S_2) = 1/6 + 1/6 = 1/3$

Aufgabe 1.B.5 (mittel, Kapitel 1.2)
Für die Projekte eines *Softwarebüros* betrachten wir die Ereignisse

R: Das Projekt wird rechtzeitig fertig
V: Das Projekt wird moderat verspätet fertig (Dauer *bis 20%* länger als geplant)
S: Das Projekt wird stark verspätet fertig (Dauer *über 20%* länger als geplant)
K: Die geplanten Kosten werden überschritten

Dabei sei $P(R) = \frac{1}{6}$, $P(V) = \frac{1}{3}$, $P(K) = 70\%$.
Bestimmen Sie, falls ohne weitere Informationen möglich:

a) $P(\overline{R})$
b) $P(R \cup V)$
c) $P(S)$
d) $P(S \cup K)$

Aufgabe 1.B.6 (mittel, Kapitel 1.2 und 1.3)
Zusätzlich zu den Angaben aus der vorigen Aufgabe sei jetzt auch noch bekannt:

$P(K \mid V) = 60\%$ und $P(K \mid S) = 90\%$.

Berechnen Sie, und beschreiben Sie bei (b) bis (e) die gesuchte Größe in Worten:

a) Stellen Sie die Angaben in einem Wahrscheinlichkeitsbaum dar
b) $P(V \cap K)$ c) $P(V \cup K)$ d) $P(V \mid K)$
e) $P(S \mid \overline{K})$ (Tipp: Berechnen Sie zuerst $P(\overline{K} \cap S)$)
f) $P(K \mid R)$ (Tipp: Berechnen Sie zuerst $P(K \cap R)$ mit Hilfe des Satzes von der totalen Wahrscheinlichkeit.)
g) Ist V von K stochastisch unabhängig? (Begründung!)

Aufgabe 1.B.7 (mittel, Kapitel 1.2 und 1.3)
Auf einem bestimmten Mail-Server sind 96% der ankommenden Mails Spam. Es ist aber ein Spam-Filter installiert, der alle Mails löscht, die er als Spam erkennt. Leider ist der Filter nicht perfekt:

Bei jeder ankommenden Spam-Mail beträgt die Wahrscheinlichkeit 95 %, dass sie auch als Spam erkannt wird (d.h. mit 5% Wahrscheinlichkeit wird eine Spam-Mail versehentlich durchgelassen). Die Wahrscheinlichkeit für den umgekehrten Fehler, nämlich dass eine echte Mail fälschlich als Spam betrachtet und gelöscht wird, betrage 10%.
Wie wahrscheinlich ist es, dass eine bei einem Benutzer ankommende Mail Spam ist? (d.h. welcher Anteil aller bei Benutzern ankommenden Mails sind Spam?)

a) Geben Sie die formalen Bezeichner der angegebenen Größen und der gesuchten Größe an.
b) Bestimmen Sie (mit rechenweg) die gesuchte Größe.

Mehr zum Üben

(Weitere Aufgaben zum Üben zu denselben Themen))

Aufgabe 1.B.1 (mittel, Kapitel 1.2 u. 1.3, in Vorlesung behandelt)
Hans hat einen ungünstigen Geburtstagstermin. Über die Jahre beträgt die Wahrscheinlichkeit, dass es an seinem Geburtstag

- regnet (Ereignis R): P(R)=60%
- kalt ist (Ereignis K): P(K)=70%
- regnet und kalt ist : $P(R \cap K) = 50\%$

Wie wahrscheinlich ist es, dass es an Hans' Geburtstag

a) regnet oder kalt ist?
b) nicht regnet oder nicht kalt ist?
c) regnet und nicht kalt ist?
d) Bestimmen Sie $P(\bar{R} \cap \bar{K})$.
e) Ist die Angabe nicht unlogisch: Wenn es langfristig an 60% der Termine regnet (zentraler Grenzwertsatz), und an 70% der Termine kalt ist, müsste es dann nicht an $0.6 \cdot 0.7 = 0.42 = 42\%$ der Termine regnen und kalt sein? Angegeben sind aber 50%.

Aufgabe 1.B.4 (einfach, Kapitel 1.2 u. 1,3)
In einer Schublade sind 4 rote und 6 blaue Socken. Wenn Sie in der Dunkelheit (also zufällig) zwei Socken aus der Schublade ziehen, wie groß ist die Wahrscheinlichkeit,

a) zwei rote b) zwei blaue c) zwei verschiedene d) zwei gleichfarbige Socken zu erwischen?
e) wie wahrscheinlich ist der als *zweites* gezogene Socken rot?
f) wie wahrscheinlich ist der als *zweites* gezogene Socken rot, wenn der erste blau war?

Aufgabe 1.C.1 (leicht bis mittel, Kapitel 1.3)
Die Aufnahmeprüfung einer Hochschule bestehe aus den Teilen A und B. Folgende Kreuztabelle zeige die gemeinsame Wahrscheinlichkeitsverteilung beider Prüfungsergebnisse.

	A bestanden	A nicht bestanden
B bestanden	62 %	13%
B nicht bestanden	12 %	13 %

a) Ist das Bestehen der Teilprüfungen A und B voneinander stochastisch unabhängig?

Formalisieren Sie jeweils die gesuchte Größe und bestimmen Sie:

b) Wie wahrscheinlich ist es, dass ein Bewerber *nicht beide* Prüfungen besteht?
c) Wie wahrscheinlich ist es, dass ein Bewerber *beide* Prüfungen *nicht* besteht?

Wie wahrscheinlich ist es, dass ein Bewerber, der Prüfung B nicht bestanden hat, auch Prüfung A nicht besteht?

Aufgabe 1.C.2 (mittel, Kapitel 1.2 und 1.3)
Von den Kunden eines Unternehmens sind 60% zufrieden, 40% unzufrieden.
Das Unternehmen verschickt an alle Kunden eine Zufriedenheitsumfrage in Form eines Fragebogens. Da unzufriedene Kunden üblicherweise stärker zur Teilnahme motiviert sind als zufriedene, beteiligen sich 2% der unzufriedenen aber nur 1% der zufriedenen Kunden an der Umfrage.

a) Definieren Sie geeignete Ereignisbezeichner und formalisieren Sie damit die angegebenen Informationen. Geben Sie auch an, auf welche Grundgesamtheit Sie sich beziehen. (d.h. betrachten Sie alle Kunden oder nur alle Teilnehmer?)
b) Welches Umfrageergebnis wird sich ergeben?

Zum Nachdenken / Vertiefen

Aufgabe 1.D.1 (mittel bis schwer, Grundlagenverständnis Kapitel 1.2 und 1.3)
Von einem Zufallsexperiment seien zu Ereignissen F und G nur P(G | F) sowie $P(\bar{F})$ bekannt.
Welche der folgenden Wahrscheinlichkeiten lassen sich mit diesen Informationen bestimmen, welche nicht? (Falls ja, Berechnungsformel angeben)
a) $P(F)$ b) $P(F \cap G)$ c) $P(G\,|\bar{F})$ d) $P(\bar{G}|F)$ e) $P(F\,|G)$

Aufgabe 1.D.2 (schwer, Verständnis 1.3)
Zwei faire Würfel werden im Würfelbecher geworfen und Sie wetten auf die Augensumme 2.

a) Angenommen, Sie haben vom Ergebnis schon mitbekommen, dass mindestens eine Eins dabei ist: Geben Sie die Wahrscheinlichkeit von Augensumme 2 unter der Bedingung „mindestens eine Eins" an.

b) Nehmen Sie an, die Würfel sind rot und blau, und Sie haben gesehen, dass der rote Würfel Eins ist; vom blauen haben Sie nichts gesehen. Wie Wahrscheinlich ist dann Augensumme 2?

c) Sie haben nichts gesehen, aber ein Schiedsrichter, der beide Würfel gesehen hat, sagt Ihnen, ob mindestens eine Eins dabei ist oder nicht. Wie wahrscheinlich ist Augensumme 2, wenn er das bejaht?

d) Welcher der beiden Situationen entspricht die Bedingung „mindestens eine Eins" aus (a) also, der aus (b) oder der aus (c)?

Aufgabe 1.D.3 (schwer, Vorbereitung des Bayes-Klassifikators)
Sie haben einen ***gezinkten*** Würfel, bei dem die Sechs mit Wahrscheinlichkeit 1/2 geworfen wird, sowie einen gleich aussehenden ***normalen*** Würfel. Sie nehmen wahllos einen davon, wissen nicht, um welchen der beiden es sich handelt, und würfeln zwei Mal nacheinander.

a) Wie wahrscheinlich ist es, dass Sie zwei Mal Sechs werfen? Begründen Sie Ihre Antwort mit Hilfe der Rechengesetze für Wahrscheinlichkeiten aus der Vorlesung.
TIPP: Benutzen Sie folgende Zwischenergebnisse:
   i. Wie wahrscheinlich ist es, dass Sie zwei Mal Sechs werfen, falls Sie den *normalen* Würfel erwischt haben?
   ii. Wie wahrscheinlich ist es, dass Sie zwei Mal Sechs werfen, falls Sie den *gezinkten* Würfel erwischt haben?

b) Angenommen, beim ersten Wurf kam eine Sechs. Wie wahrscheinlich ist es dann, dass auch beim zweiten Wurf eine Sechs kommt?

Aufgabe 1.D.4 (Bayes-Klassifikator, wird in der Vorlesung besprochen)

An einem Mail-Server seinen 96% der ankommenden Mails Spam. Es gibt drei alternative Spam-Filter:

- **Filter 1: Wortschatzuntersuchung**
- **Filter 2: Headeruntersuchung** (Absender, Empfängerliste)
- **Filter 3: Collaboratives Filtern** (Markieren andere Empfänger die Mail als Spam?)

Jedes dieser Verfahren für sich schafft es, 95% der Spam-Mails zu erkennen, markiert aber auch 10% der echten Mails als Spam. Das ist für praktische Zwecke unbrauchbar.

Deshalb werden die drei Filter per **Mehrheitsentscheidung** kombiniert. **Annahme (*): Fehlurteile** der drei Verfahren treten stochastisch **unabhängig** voneinander auf. Wie wahrscheinlich ist es dann, …

a) … dass eine Mail gelöscht wird, wenn es sich um Spam handelt?

b) … dass eine Mail gelöscht wird, wenn es sich nicht um Spam handelt?

c) Formalisieren Sie Annahme (*) so, wie Sie sie in (a) und (b) verwendet haben.

d) Berechnen Sie die Wahrscheinlichkeit, dass F1 die nächste ankommende Mail für Spam hält, unter der Bedingung, dass F2 sie für Spam hält. Sind die Urteile von Filter 1 und Filter 2 also voneinander stochastisch unabhängig? (ohne dass bekannt wäre, ob die geprüfte Mail Spam ist oder nicht)
Tipp:
i) Berechnen Sie zunächst die Wahrscheinlichkeit, dass F1 die nächste ankommende Mail für Spam hält, ohne die Bedingung zu F2. (Tipp: Satz v.d.totalen W.)
ii) Verallgemeinern Sie dann Ihr Vorgehen, um auch F2 einzubeziehen.