Übungsblatt 7 **Statistik**

Aufgabe 7.1 Kennzahlen von Stichproben

Gegeben sei die folgende Stichprobe Z von (bereits sortierten) Messwerten:

z	0.1	0.15	0.2	0.2	0.25	0.3	0.31	0.35
	0.7	1.0	1.0	1.1	1.5	2.5	2.5	2.7
	2.8	2.95	3.1	3.1	3.15	3.15	3.2	4.0
	4.2	5.2	7.5	10.5	10.7	20.0	25.5	31.5

(a) -

(b) Berechnen Sie Mittelwert, Median und Standardabweichung für diese Stichprobe.

(c) Wie ändern sich Mittelwert, Median und Standardabweichung, wenn statt z=0.25 ein Wert z=100.25 gemessen wird?

(d) Können Sie eine Kennzahl angeben, die wie die Standardabweichung die Breite einer Verteilung misst, aber durch den einzelnen Ausreißer 100.25 kaum gestört wird?

(e) Zeichnen Sie den Boxplot für Z (Geben Sie dabei die notwendigen Berechnungsschritte an)

(f) Zeichnen Sie den Boxplot für die gemäß (c) abgeänderte Stichprobe Z‘! Wie ändern sich Box und Whisker?

Aufgabe 7.2 Kombinatorik I

(a) Wie viele Wörter über dem Alphabet {v,w,x,y,z} mit 10 Buchstaben gibt es, die genau 8 z's enthalten? Wie viele Wörter enthalten mehr als 7 z's?

(b) 5 Würfel werden gleichzeitig geworfen. Wieviel Wurfergebnisse gibt es?

(c) Wie groß ist die Wahrscheinlichkeit, mit 3 Würfeln mindestens 2 Fünfer zu erzielen?

(d) Ein Würfel wird fünfmal geworfen. Wie viele Ergebnisse gibt es? Wie wahrscheinlich ist es, mit einer Eins zu beginnen und einer Fünf zu enden?

Aufgabe 7.3 Kombinatorik II

(a) Bei der geheimen Wahl zum Dekan treten drei Kandidaten D, E, F an, die von 9 Stimmen gewählt werden. Wie viele Wahlausgänge gibt es insgesamt? Wie viele mit der absoluten Mehrheit für Kandidat E? Wie viele mit einem "Patt"?

(b) Bei einem Hunderennen mit 9 Hunden sind auf dem Wettschein Platz 1., 2., 3. und 4. zu notieren. Wie groß ist die Wahrscheinlichkeit, durch zufälliges Raten die richtigen Namen (in richtiger oder falscher Reihenfolge) auf dem Wettschein zu haben UND auf Platz 2. auch genau den richtigen Hundenamen stehen zu haben?

[Zu (a): "Patt" = kein Kandidat kann sich durchsetzen]

Aufgabe 7.4

Erklären (beweisen) Sie, dass die Anzahl der Teilmengen einer n-elementigen Ereignismenge genau 2^n beträgt.

[Hinweis: Dies ist auch die Anzahl der möglichen Rucksäcke, die man packen kann, wenn n unterschiedliche Gegenstände verfügbar sind.]

Aufgabe 7.5 Geburtstag + Hashtabelle

a) Wie groß ist die Wahrscheinlichkeit, dass von 10 Personen mindestens 2 am gleichen Tag Geburtstag haben? Erst "aus dem Bauch heraus" schätzen, dann ausrechnen! Wie lautet die Formel allgemein für k Personen? Ab welcher Gruppengröße lohnt sich im Schnitt eine Wette: "Wetten, dass in dieser Gruppe mindestens 2 Personen am gleichen Tag Geburtstag haben?"

b) Eine **Hashtabelle** der Länge n speichert d Daten mit beliebigen Schlüsseln $k \in K$, indem die Schlüssel mit einer **Hashfunktion** $h(k) : K \rightarrow \{0,\ldots,n-1\}$ in einen Tabellenindex umgerechnet werden.[1] Eine **Kollision** liegt vor, wenn zwei verschiedene Schlüssel auf einen gleichen Tabellenindex abgebildet werden. Machen Sie sich die Analogie zum Geburtstagsproblem klar und schreiben Sie eine kleine Prozedur {in MAPLE oder Excel oder Java oder …}, mit der Sie Hashtabellen und Datenmengen variabler Größe simulieren können. Beantworten Sie dann folgende Fragen:

   i. Wie wahrscheinlich ist bei n=10000 und d=40 mindestens eine Kollision?

   ii. Bis zu welcher Zahl d ist bei n=10000 die Kollisionswahrscheinlichkeit unter 25%?

[Hinweis zu a): Betrachte komplementäres Ereignis "Es gibt **keine** 2 Personen, die am gleichen Tag Geburtstag haben".]

Aufgabe 7.6 Hellseher

Ein Hellseher wirbt mit folgender Anzeige:

Sage werdenden Eltern das Geschlecht ihres Kindes voraus. Bei Nichteintreffen Geld zurück!

Die Vorhersage kostet 100€, der Hellseher wirft eine Münze. Die Wahrscheinlichkeit einer Mädchengeburt sei 0.465. Berechnen Sie das mittlere Jahreseinkommen des Hellsehers, wenn 100 Anfragen je Monat eintreffen. Mit welcher Strategie kann er sein Jahreseinkommen (bei gleichem Tarif) verbessern?

[Hinweis: Stellen Sie eine Zufallsvariable "Gewinn" auf und berechnen Sie ihren Erwartungswert.]

Aufgabe 7.7 Zufallsvariable I

Eine diskrete Zufallsvariable X habe Ausprägungen x_i mit Wahrscheinlichkeiten p_i gemäß

x_i	1	5	10	20	100	
p_i	0.55	0.25	0.1	0.05	0.05	

a) Berechnen Sie Erwartungswert E(X) und Varianz Var(x).

b) Berechnen Sie Erwartungswert E(2X+5) und Varianz Var(2X+5).

[1] Ein Beispiel für eine solche Hashfunktion ist h(k) = k mod n (wenn k beliebig ganzzahlig ist).

Aufgabe 7.8 Zufallsvariable II: Bitfolgen

a) Es werden Bitfolgen der Länge 3 zufällig generiert, wobei **0** und **1** gleichwahrscheinlich sind. Ermitteln Sie die Wahrscheinlichkeiten $p_k = P(X=k)$ der Zufallsvariablen

X = *Anzahl der Einsen in der Bitfolge*

ferner den Erwartungswert, die Varianz und die Standardabweichung von X. Was ist $P(X\geq 1)$?

b) Wiederholen Sie alle Berechnungen aus a) für Bitfolgen der Länge 4! Erkennen Sie eine Systematik bei den Wahrscheinlichkeiten? Können Sie raten, wie $P(X=k)$ für die Länge 5 aussieht?

c) Wiederholen Sie alle Berechnungen aus a) für Bitfolgen der Länge 2, wobei nun aber bei der Generierung p(**0**)=0.2 und p(**1**)=0.8 gilt. Können Sie den Rechenweg durch einen Baum visualisieren?

Aufgabe 7.9 Verteilungen

Eine Serienproduktion von Transistoren hat einen gleichbleibenden Ausschuss-Anteil von $p=30\%$. Wie groß ist die Wahrscheinlichkeit, dass in einer Stichprobe von $n=10$ entnommenen Transistoren

(a) keine, (b) genau neun, (c) mehr als acht

fehlerhafte Bauteile sind? Überlegen Sie zunächst, welche Verteilung Sie anwenden!

Wie groß sind Erwartungswert und Varianz der Anzahl X der fehlerhaften Bauteile in der Stichprobe?

Aufgabe 7.10 Spam-Filter

In Ihrer Mailbox sind 25% der neu eintreffenden Nachrichten SPAM.
Sie stellen aufgrund der Inhalte Ihrer Mailbox folgende Tabelle für das Vorkommen bestimmter Wörter in Ihren Spam- oder Nicht-Spam-Mails (SPAM oder HAM) auf:

	„Gewinn“	„Million“
SPAM	19%	25%
HAM	1%	10%

Diese Tabelle ist wie folgt zu lesen: Die Wahrscheinlichkeit, in einer SPAM-Nachricht das Wort W_1=“Gewinn“ zu finden, beträgt $P(W_1|S)=19\%$.

(a) Eine neu eintreffende Nachricht enthält das Wort „Gewinn“. Mit welcher Wahrscheinlichkeit ist sie SPAM?

(b) Eine neu eintreffende Nachricht enthält das Wort „Million“. Mit welcher Wahrscheinlichkeit ist sie SPAM?

(c) Eine neu eintreffende Nachricht enthält beide Worte „Gewinn“ UND „Million“. Mit welcher Wahrscheinlichkeit ist sie SPAM?

[Hinweise: 1) Benutzen Sie die Bayes-Formel.

2) Machen Sie bei (c) die (nicht unbedingt realistische, aber das Leben stark vereinfachende) Annahme, dass das Auftreten der beiden Wörter W_1 und W_2 in beiden Mail-Kategorien KAT $\in$ {SPAM, HAM} statistisch unabhängig sei, also

$$P(W_1 \cap W_2|KAT) = P(W_1|KAT) \cdot P(W_2|KAT)$$

.]

Bereiten Sie die Aufgaben für den 26.04.2021 so vor, dass Sie in der Lage sind, Ihre Lösungen vorzutragen.

Aufgabe 7.11 Mammographie

Die Wahrscheinlichkeit für Frauen, an Brustkrebs zu erkranken, beträgt 1%. Ein Mammographie-Test mit Ergebnis „pos" oder „neg" hat folgende Zuverlässigkeitsquoten:

	krank	gesund
pos	95%	10%
neg	5%	90%

also fällt der Test für kranke Frauen zu 95% positiv und zu 5% negativ aus, d.h. P(pos|krank) = 95%.

Wenn für eine zufällig ausgewählte Frau der Test positiv ausfällt, mit welcher Wahrscheinlichkeit ist sie dann krank?

Berechnen Sie dies auf zwei Arten:

a) mit bedingten Wahrscheinlichkeiten und dem Satz von Bayes

b) mit natürlichen Häufigkeiten nach [Gigerenzer09] und einem entsprechenden Baum

Aufgabe 7.12 Normalverteilung 1

Es sei bekannt, dass in einem Rechnernetz pro 1000 übertragener Pakete die Anzahl X der fehlerhaft übertragenen Pakete binomialverteilt sind mit Erwartungswert $\mu=80$ und Standardabweichung $\sigma=15$. Die Binomialverteilung kann durch die Normalverteilung angenähert werden (Satz von Moivre-Laplace). Berechnen Sie die Wahrscheinlichkeit dafür,

(a) dass weniger als 50 aus 1000 Paketen

(b) dass mehr als 100 aus 1000 Paketen

fehlerhaft übertragen werden.

[Hinweis: Tabelle der Funktion $\Phi(x)$ aus Formelsammlung benutzen.]

Aufgabe 7.13 Normalverteilung 2

Eine Maschine produziere Werkstücke mit einem Sollmaß von 800 mm. Die Produktion sei **normalverteilt** mit μ= 800 mm und σ= 5 mm. Es sollen Werte zwischen (800-t) und (800+t) mm eingehalten werden. Wie groß muss $t>0$ mindestens sein, damit der mittlere Ausschussanteil höchstens 5% ist?

Aufgabe 7.14 Bedingte Wahrscheinlichkeiten

In einer Urne befinden sich 10 Kugeln, darunter 4 schwarze und 6 weiße. Es werden 2 Kugeln gezogen. Sei

A = "Die 1. gezogene Kugel ist schwarz"

B = "Die 2. gezogene Kugel ist schwarz"

a) Wir entnehmen die Kugeln <u>ohne</u> Zurücklegen.
   i. Wie lauten die Wahrscheinlichkeiten P(A) und P(B|A)?
   ii. Sind A und B **statistisch unabhängig**?
   iii. Berechnen Sie die Wahrscheinlichkeit, 2 schwarze Kugeln zu ziehen, $P(A\cap B)$.

b) Beantworten Sie die gleichen Fragen wie bei a), wenn man die Kugeln <u>mit</u> Zurücklegen entnimmt.

[Tipp: Zur Berechnung der Wahrscheinlichkeit P(B) (die Sie auch brauchen), ist **S10-17** (mit einer geeigneten Zuordnung für die A_i in **S10-17**) nützlich. Oder Sie stellen eine passende **graphische Überlegung** an.]