2. Aufgabenblatt zur Statistik-Vorlesung

Basisaufgaben, um das Grundverständnis zu testen

Aufgabe 2.1 (Notation von Zufallsvariablen, einfach)

Seien X und Y Zufallsvariablen, die angeben, wie viel Rechenzeit in Sekunden (Variable X) und wie viel Speicherplatz in MB (Variable Y) an einem Web-Service ankommenden Anfragen benötigen. Geben Sie für die folgenden Größen jeweils die formalen Bezeichner an:

a) Wahrscheinlichkeit, dass die Rechenzeit über 5s beträgt.
b) Wahrscheinlichkeit, dass höchstens 7MB Speicher benötigt wird.
c) Wahrscheinlichkeit, dass mehr als 5s Rechenzeit oder mehr als 7MB Speicher benötigt werden.
d) Wahrscheinlichkeit, dass die Rechenzeit weniger als k Sekunden beträgt $(k \in \mathbb{R}^+)$
e) Wahrscheinlichkeit, dass die Rechenzeit in Sekunden größer als der Speicherverbrauch in MB ist.
f) Wahrscheinlichkeit, dass bei einer Anfrage die mehr als 7MB Speicher benötigt, die Rechenzeit über 5s beträgt.

Aufgabe 2.2 (einfach, Grundverständnis der Begriffe)

Ergänzen Sie jeweils mit einem der folgenden Fachbegriffe zu einem vollständigen Satz. Geben Sie auch die Parameter des jeweiligen Begriffs mit an (siehe spitze Klammern):

- <x%>-*Quantil* <von welcher Variable>
- *Mittelwert* <von welcher Variable>
- *relative Häufigkeit* <von was> <an was>
- *Anteil* <von was><an was>

Beispiel:
1 Mio. erfasste deutsche Haushalte verdienten 2014 brutto insgesamt 3.5 Mrd. €.
Also ist „3 500 €" was?
→ **der *Mittelwert* des deutschen Haushaltseinkommens**.

a) 10% aller deutschen Haushalte haben ein Jahreseinkommen von höchstens 12000€.
(i) „12000€" ist …   (ii) 10% ist …
b) 10% deutschen Haushalte haben ein Jahreseinkommen von über 85000€.
(i) „85000€" ist …   (ii) 10% ist …
c) 80% der deutschen Haushalte haben ein Jahreseinkommen zwischen 12000€ und 85000€.
(i) „80%" ist …

Aufgabe 2.3 (einfach bis mittel, zum Üben der Basiskennzahlen)

Die Last L an einem Server wurde zu verschiedenen Zeiten gemessen. Dabei ergaben sich folgende Werte:

0; 5; 1; 0; 0; 7; 50; 0; 5; 1

a) Erstellen Sie eine Tabelle der (Häufigkeits-)Verteilung von L

b) Erstellen Sie eine Tabelle der kumulierten Verteilungsfunktion von L

c) Berechnen Sie den Mittelwert von L unter Verwendung der Formel $\overline{w} = \frac{1}{n}\sum_{i=1}^{n} w_i$ von Folie 2-21

(Beim Rechenweg genügt, die ersten beiden und den letzte Summanden anzugeben)

d) Berechnen Sie den Mittelwert von L unter Verwendung der Formel $\overline{w} = \sum_{i=1}^{m} x_i \cdot f_i$ (2-22)

e) Bestimmen Sie den Median von L

i. direkt aus der Werteliste

ii. aus der Häufigkeitsverteilung aus (a)

f) Bestimmen Sie das 10%-Quantil von L

g) Bestimmen Sie den Modus von L

Aufgabe 2.4 (Fortsetzung 4.1) (einfach bis mittel, zum Üben der Standardabweichung)

a) Berechnen Sie die(Stichproben-) Varianz von L unter Verwendung der Formel

$s^2 = \frac{1}{n}\sum_{i=1}^{n}(\overline{w} - w_i)^2$ von Folie 2-47

(Beim Rechenweg genügt, die ersten beiden und den letzte Summanden anzugeben)

b) Berechnen Sie die (Stichproben-)Varianz von L aus der in Aufgabe 2.3a erstellten Häufigkeitstabelle unter Verwendung der Formel

$s^2 = \sum_{j=1}^{m} f_j \cdot \left(x_j - \overline{w}\right)^2$ von Folie 2-48

c) Berechnen Sie die (Stichproben-)Standardabweichung von L .

Aufgabe 2.5 (einfach, Anwendung)

Bei einem Würfelspiel zahlen Sie 13 Cent Einsatz. Dafür dürfen Sie ein Mal würfeln und bekommen das Quadrat der Augenzahl in Cent ausbezahlt. (also 1 Cent im schlechtesten und 36 im besten Fall)

a) Gewinnen oder verlieren Sie im langfristigen Mittel? Wie viel pro Spiel?

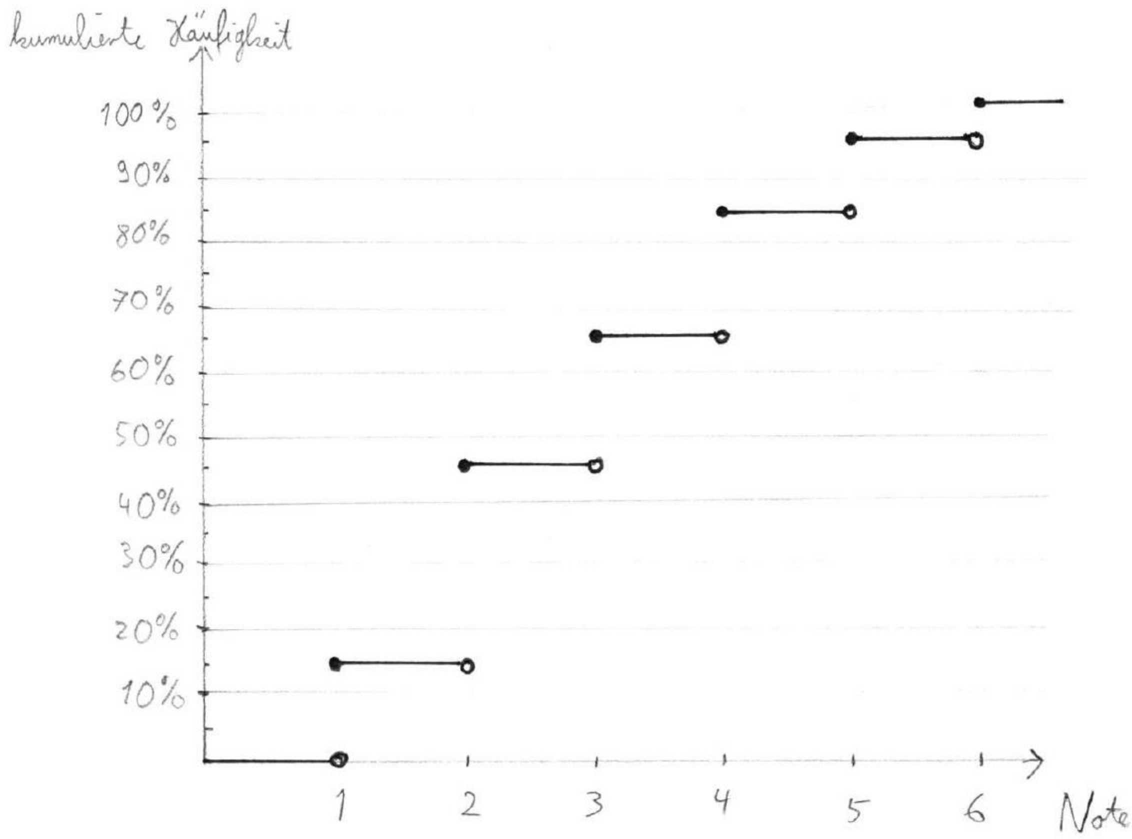
b) Können Sie (halbwegs zuverlässig) den Gewinn/Verlust pro Spiel abschätzen, den Sie bei 1000 Spielen machen werden?

c) Können Sie das auch für 5 Spiele zuverlässig abschätzen?

Weitere Übungsaufgaben für das Grundverständnis

Aufgabe 2.6 (einfach, zum Üben der Grundbegriffe)

In einer Schulprüfung ergibt sich die dargestellte kumulierte Häufigkeitsverteilung der Noten *N.*



a) Skizzieren Sie die (relative) Häufigkeitsverteilung der Noten in graphischer Form.

b) Bestimmen Sie die folgenden Kennzahlen. Geben Sie dabei jeweils den Rechenweg oder eine Begründung an, bzw. zeichnen Sie in das entsprechende Diagramm ein, was Sie wo abgelesen haben:

   i. **Median** der Noten
   ii. **75%-Quantil** der Noten
   iii. **25%-Quantil** der Noten
   iv. **Quartilsabstand** der Noten
   v. Die **Spannweite** der Noten
   vi. **Modus** der Noten
   vii. **Durchfallquote.** (Noten 1-4 sind bestanden, 5-6 nicht bestanden)
   viii. Spitzenquote.

Aufgaben, die über die Grundbegriffe hinausgehen

Aufgabe 2.7 (einfach, Anwendung Mittelwert und Standardabweichung)
Bei einer Softwareanschaffung in einer Firma stehen drei Preismodelle zur Auswahl:

(A, Flatrate): Pauschale (hohe) Gebühr, unabhängig von der Anzahl an Nutzern.
(B, Mischvariante): Mittelhohe Grundgebühr plus eine mittelhohe Gebühr pro Nutzer.
(C, linear): Nur eine (höhere) Gebühr pro Nutzer.

Die Gesamtkosten (in Geldeinheit GE) hängen davon ab, wie viele Nutzer die Software nutzen werden:

Anzahl Nutzer	hoch	mittel	niedrig
Preismodell A	6 GE	6 GE	6 GE
Preismodell B	7 GE	5 GE	3 GE
Preismodell C	11 GE	4.5 GE	1 GE

Sie schätzen die Wahrscheinlichkeit für eine hohe Benutzerzahl auf 20%, für eine mittlere auf 50%, und für eine niedrige auf 30%.

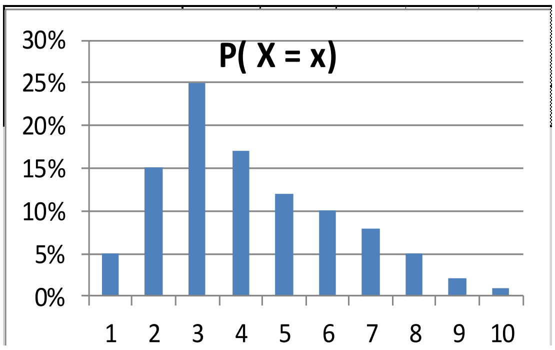
a) Bestimmen Sie für jede der Varianten (A), (B), und (C) jeweils den Erwartungswert der Kosten.
`Man nennt so etwas auch einen` *bedingten Erwartungswert*`: Erwartungswert der Kosten` K`, unter der Bedingung, dass Modell A gewählt wird:` $E(K \mid A)$

b) Bestimmen Sie für jede der Varianten (A), (B), und (C) jeweils die ***Standardabweichung*** der Kosten.

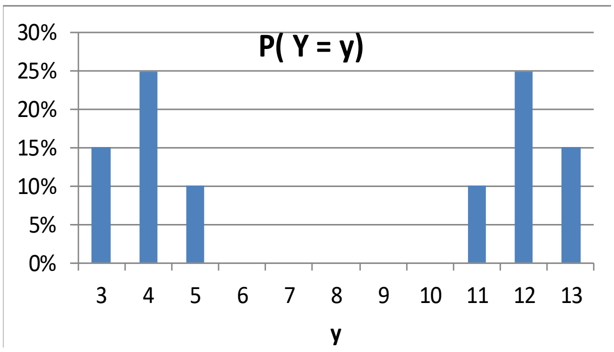
c) Bewerten Sie das Ergebnis. Welche Entscheidung würden Sie empfehlen und warum?

Aufgabe 2.8
(mittel, Verständnis Wahrscheinlichkeitsfunktion, Intuition zu Erwartungswert, Standardabweichung)
Schätzen Sie nach Augenmaß (d.h. ***ohne Rechnung***) **Erwartungswert** und **Standardabweichung** der Zufallsvariablen X bzw. Y. Als Genauigkeit genügt, wenn Sie jeweils angeben, welcher der folgenden Werte am ehesten in Frage kommt: -16; -8; -4; -2; 0; 1; 2; 4; 8; 16

a)



b)



Aufgabe 2.9 (mittel, Lesen der Notation von Formeln)

Die Kennzahl $K(X)$ für eine beliebige Zufallsvariable X mit Realisationen $x_1, \ldots, x_m$ sei definiert als

$$K(X) \coloneqq \sqrt{\left|\sum_{j=1}^{m} (x_j^2 - E(Z))^3 \cdot P(X = x_j)\right|}$$

wobei die senkrechten Striche für den Absolutbetrag stehen.

Für einen Datensatz, in dem zu einem Merkmal M die Werteliste $w_1, \ldots, w_n$ beobachtet wurde, sei die Kennzahl $K(M)$ definiert als

$$K(M) \coloneqq \sqrt{\left|\frac{1}{n} \cdot \sum_{j=1}^{n} (x_j^2 - \overline{w})^3\right|}$$

a) Die Zufallsvariable Z habe folgende Wahrscheinlichkeitsverteilung:

z:	-10	-1	0	6	25
P(Z = z):	5%	10%	10%	45%	30%

Berechnen Sie $K(Z)$.

b) Sie haben Ihren engeren Freunde gefragt, wie viele alte ungenutzte Handys bei ihnen jeweils zu Hause herumliegen, und die Antworten 3, 0, 2 bzw. 1 erhalten. Berechnen Sie für diesen Freundeskreis die Kennzahl K(*Anzahl herumliegender alter Handys*).

Und wer Quantil oder Standardabweichungsberechnung nochmal üben will:

c) Bestimmen Sie das 20%-Quantil von Z.

d) Bestimmen Sie $\sigma(Z)$

Aufgabe 2.10 (einfach bis mittel, Verständnis Kennzahlen)

Welche der folgenden Aussagen sind (immer) korrekt?

a) Der Mittelwert von n Zahlen liegt stets zwischen ihrem Minimum und Maximum.

b) Die Standardabweichung von n Zahlen liegt stets zwischen dem Abstand des dem Mittelwert nächsten und entferntesten Element vom Mittelwert.

c) Der Erwartungswert einer Zufallsvariablen ist nie negativ.

d) Wenn man in einer Werteliste nur ein einziges Element verändert und zwar vergrößert…
   i. … wird der Mittelwert größer
   ii. … wird die Standardabweichung größer

e) Die Standardabweichung einer Zufallsvariablen ist nie negativ.

f) Der Median einer Zufallsvariablen ist nie negativ.

g) Der Median einer Zufallsvariable ist nie kleiner als ihr 40%-Quantil.

h) Der Median einer Zufallsvariable ist nie kleiner als ihr Erwartungswert.

i) Angenommen, die kumulierte Verteilungsfunktion einer Zufallsvariablen X ist umkehrbar; die Umkehrfunktion sei mit F^{-1} bezeichnet. Den Wert $F^{-1}(0.2)$ nennt man dann auch ……………………………………………………… von X.

(Satz unter Verwendung eines passenden Fachbegriffs komplettieren)

Aufgabe 2.11 (mittel, Notation zum Erwartungwert, Varianz & Co)

a) Eine Prüfung besitze zwei Teile, die getrennt bepunktet werden. Ihre Ergebnisse in beiden Teilen, X bzw. Y sind aus Ihrer Sicht Zufallsvariablen, unter anderem weil sie davon abhängen, welche Aufgabentypen drankommen. Welcher der folgenden Ausdrücke steht für die Standardabweichung Ihrer Gesamtpunktzahl (=Summe aus beiden Teilen)?

i) $\sigma(X) + \sigma(Y)$ oder ii) $\sigma(X + Y)$

b) Berechnen Sie für die Zufallsvariable Y mit folgender Wahrscheinlichkeitsverteilung

y:	0	2
P(Y = y):	70%	30%

die Kennzahlen

i) $E(Y)$ sowie $\sigma(Y)$

ii) $E(Y^2)$ sowie $\sigma(Y^2)$

Aufgabe 2.12 (Anwendung zum Erwartungswert, wird in der VL behandelt)
Sie wetten 1€ (=Ihr *Einsatz*) auf das Eintreten eines Ereignisses F, das mit Wahrscheinlichkeit p eintreten wird. Wenn F eintritt, erhalten Sie das w-Fache den Einsatzes ausbezahlt (=***Bruttowettquote***), andernfalls ist Ihr Einsatz weg. Wie hoch muss w sein, damit die Wette fair ist? (Formel abhängig von p angeben)

Aufgabe 2.13 (wird in der VL behandelt, subjektive Wahrscheinlichkeiten schätzen)
Wir wollen ermitteln, mit welcher (subjektiven) Wahrscheinlichkeit Sie glauben, dass der Flughafen BER im Jahr 2017 in Betrieb geht (= **Ereignis** ***F***).

a) Wie hoch müsste die Bruttowettquote w_f („f" wie „*da**f**ür*") mindestens sein, damit Sie 1€ **darauf wetten** würden, dass der Flughafen BER im Jahr 2017 **in Betrieb** geht?
`(Falls Sie sagen, Sie würden niemals darauf wetten: Würden Sie wirklich nicht 1€ setzen, wenn Sie im Gewinnfall 100 Mio € bekommen würden? Na also. Und bei 10 Mio, 1Mio, …, wo ist Ihre Grenze?)`

b) Wie hoch müsste die Bruttowettquote w_g („g" wie „*da**g**egen*") mindestens sein, damit Sie 1€ **darauf wetten** würden, dass der Flughafen BER im Jahr 2017 ***NICHT*** **in Betrieb** geht?

c) Benutzen Sie das Ergebnis aus der vorigen Aufgabe, um zu ermitteln, **wie hoch Sie P(F)** …

i. … offenbar **einschätzen**, wenn Ihr Wert aus (a) gerade die Grenze zwischen Vorteilhaftigkeit und Nachteilhaftigkeit bildet.

ii. … offenbar **einschätzen**, wenn Ihr Wert aus (b) gerade die Grenze zwischen Vorteilhaftigkeit und Nachteilhaftigkeit bildet.

iii. Interpretieren Sie die Ergebnisse aus (i) und (ii). Sind Sie anscheinend risikoaffin oder risikoavers?

Schwere Aufgaben

Aufgabe 2.14 (schwer, Anwendung)
Bei einem Spiel würfeln Sie nacheinander mehrmals so lange, bis das erste Mal eine Sechs kommt, oder bis Sie freiwillig aufhören. Endet das Spiel mit einer Sechs, bekommen Sie nichts. Wenn Sie rechtzeitig freiwillig aufgehört haben, bekommen Sie <*Anzahl der getätigten Würfe zum Quadrat*> € ausbezahlt.
Nach welcher Zahl *n* von Würfen ohne Sechs sollten Sie freiwillig aufhören? (Berechnen Sie den Erwartungswert des Gewinns abhängig von der Anzahl Würfe, nach der sie spätestens abbrechen)

Aufgabe 2.15 (mittel bis schwer, Intuition und Verständnis Kennzahlen)
Kann eine Zufallsvariable Y eine Wahrscheinlichkeitsverteilung besitzen, die jeweils die folgenden Eigenschaften besitzt? Falls ja, bitte eine solche angeben; falls nein, kurze Begründung.

a) $P(Y < 0) = 0$, $E(Y) = 15$, und $\sigma(Y) = 15$

b) $P(Y < 0) = 0$, $E(Y) = 15$, und $\sigma(Y) \geq 20$

c) $P(Y < 0) = 0$, $E(Y) = 0$, und $\sigma(Y) \geq 5$