

Mathematische Methoden der Computerlinguistik III:

Statistische Methoden

Sommersemester 2015

Matt Crocker, Enrico Lieblang

Übung 2, 18.5.2015, 31 Punkte

1. Aufgabe Kombinatorik (6 Punkte)

- (a) 5 Personen sollen auf die 5 Plätze eines Autos verteilt werden. Wie viele Möglichkeiten der Sitzverteilung gibt es? (2)
- (b) Wie viele Möglichkeiten gibt es, wenn nur 3 einen Führerschein besitzen? (2)
- (c) Wie viele Möglichkeiten gibt es, wenn nur die mit Führerschein mitfahren wollen? (2)

2. Aufgabe ( 7 Punkte)

Für die Prüfung im Fach Statistische Methoden stehen 5 Prüfer bereit. 2 davon sind scharf, 3 milde. 2 Prüfer halten jeweils die Prüfung ab. Die Wahrscheinlichkeit für die Teilnehmer, die Prüfung zu bestehen (Ereignis B), hängt von der Anzahl X der scharfen Prüfer ab, nämlich es gilt:

$$P(B | X = 0) = 0,9; \quad P(B | X = 1) = 0,5 \quad P(B | X = 2) = 0,3.$$

- (a) Mit welcher Wahrscheinlichkeit wird eine Studentin die Prüfung bestehen, wenn sie sich die Prüfer frei auswählen kann? (1)
- (b) Mit welcher Wahrscheinlichkeit besteht eine Studentin die Prüfung, wenn die Prüfer durch Los bestimmt werden? (2)
- (c) Eine Studentin hat die Prüfung bestanden. Mit welcher Wahrscheinlichkeit wurde sie nur von milden Prüfern geprüft? (2)
- (d) Eine Studentin hat die Prüfung bestanden. Mit welcher Wahrscheinlichkeit wurde sie nur von scharfen Prüfern geprüft? (2)

3. Aufgabe (4 Punkte)

Es sei A das Ereignis, dass ein Wort W die (sehr seltene)lexikalische Kategorie N besitzt (1 Prozent aller Fälle). B sei das Ereignis, dass ein Tagger einem Wort die Kategorie N zuweist. In 98 Prozent der Fälle weist der Tagger N richtig zu, in 3 Prozent der Fälle weist er einem Wort fälschlicherweise N zu. Wie sicher kann man sein, dass, wenn der Tagger N zuweist, auch tatsächlich die Kategorie N vorliegt?

4. Aufgabe (10 Punkte)

Eine Klausur bestehe aus 10 Fragen, wobei bei jeder Frage genau eine von 4 Antwortmöglichkeiten richtig ist.

- (a) Wie groß ist die Wahrscheinlichkeit, dass bei zufälligem Ankreuzen der Antworten alle Fragen bzw. keine Fragen korrekt gelöst sind? (2)
- (b) Berechnen Sie die Wahrscheinlichkeit, durch zufälliges Ankreuzen der Antworten mindestens 3 Fragen korrekt zu beantworten. (2)
- (c) Wie viele korrekte bzw. falsche Antworten kann man bei beliebigem Ankreuzen erwarten? (2)
- (d) Geben Sie die Verteilungsfunktion der zugrunde liegenden Zufallsvariablen an. (2)

5. Aufgabe (2+2+2=6 Punkte)

Das Auszählen eines Corpus habe für die lexikalischen Kategorien Artikel (Art), Nomen (N), Verb (V) und Präposition (P) folgende Häufigkeiten für die Kategorienfolgen ergeben:

	Art	N	V	P
$\emptyset$	213	87	0	0
Art	0	558	10	0
N	0	108	358	367
V	194	75	31	0
P	226	81	0	0

Ein Auszählen der Kategorien einzelner Wörter ergab folgendes:

Wort	N	V	Art	P
das	0	0	50	0
der	0	0	76	0
im	0	0	0	26
essen	5	30	0	0
zahlt	0	12	0	0
Herr	5	0	0	0
sonstige	899	258	432	281

Hierdurch ist nun ein einfaches Sprachmodell gegeben.

- (a) Bestimmen Sie alle Kategorienfolgen, die den Satz 'Das Essen zahlt der Herr' erzeugen können.
- (b) Berechnen Sie  $P('Das Essen zahlt der Herr' | Art N V Art N)$ .
- (c) Wie kann man die Wahrscheinlichkeit für 'Das Essen zahlt der Herr' berechnen? (nichts ausrechnen)