

Probe-Klausur Mathematische Methoden der Computerlinguistik, Teil Statistik
III: Statistische Methoden

Enrico Lieblang

Sommersemester 2015

1. Aufgabe Wahrscheinlichkeitstheorie (4 Punkte)

Ein Leser L einer Zeitschrift erkennt in 95 Prozent aller Fälle korrekt den Autor M eines Artikels, also in den Fällen, in denen tatsächlich M der Autor ist. Mit 8 Prozent Wahrscheinlichkeit weist L einen Artikel in der Zeitschrift fälschlicherweise dem Autor M zu. Autor M verfasst 40 Prozent der Artikel der Zeitschrift. Mit welcher Wahrscheinlichkeit stammt ein Artikel von Autor M, wenn L sagt, dass M ihn geschrieben habe?

2. Zufallsvariable (4 Punkte)

Die Gewinnchance für A gegen B in einem Tennisspiel sei 0,6.

- (a) Mit welcher Wahrscheinlichkeit gewinnt B mehr als die Hälfte der Spiele gegen A, wenn sie 5 mal gegeneinander spielen?
- (b) Mit welcher Erwartung kann B in die Spiele gehen?

3. Aufgabe (6 Punkte)

Es werde 2 mal mit einem fairen Würfel gewürfelt. X bezeichne das Ergebnis des ersten Wurfes, Y das Ergebnis des zweiten Wurfes. Eine Zufallsvariable Z sei wie folgt definiert:

$$Z := \begin{cases} X - Y & : & X \geq Y \\ 0 & : & X < Y \end{cases}$$

Geben Sie die Verteilung von Z, die Verteilungsfunktion und den Erwartungswert von Z an.

4. Aufgabe (4 Punkte)

Die Zufallsvariablen X und Y seien wie folgt gegeben:

X	0	0	1	1
Y	0	1	0	1
P(X=x, Y=y)	0,32	0,08	0,48	0,12

Man gebe die Randverteilungen

und die bedingte Verteilung $P(X=x|Y=1)$ an. Sind die Zufallsvariablen X und Y unabhängig verteilt?

5. Aufgabe (4 Punkte)

Es soll überprüft werden, ob das Wortpaar 'gutes Omen' eine Kollokation darstellt. Hierbei wurden folgende Häufigkeiten festgestellt und in einer Kontingenztafel aufgelistet:

	gutes	nicht gutes	Summe
Omen	5	2	7
nicht Omen	23	43600	43623
Summe	28	43602	57166

Man führe einen Chi-Quadrat-Test zum Niveau $\alpha = 0,05$ zur Überprüfung der Hypothese durch, dass 'gutes Omen' eine Kollokation darstellt.

6. Aufgabe Hidden-Markov-Modelle (7 Punkte)

Es werde der Satz 'das alte Schloss' betrachtet. Die möglichen Kategorien für die einzelnen Wörter seien

das	Art (Artikel)	RP (Relativpronomen)
alte	Adj (Adjektiv)	NN (Nomen)
Schloss	NN (Nomen)	Vfin (finites Verb)

Die Kategorien seien als Zustände eines Markov-Modells codiert, die Wörter seien als Ausgaben codiert. Folgende Übergangsmatrix sei gegeben:

	Art	Adj	RP	NN	Vfin
Art	0	0,4	0,1	0,5	0
Adj	0,1	0,3	0	0,6	0
RP	0,2	0	0	0,8	0
NN	0	0	0,2	0,4	0,4
Vfin	0,6	0	0,1	0,3	0

Die Anfangsverteilung sei wie folgt gegeben:

	Art	Adj	RP	NN	Vfin
P	0,7	0	0,1	0,2	0

Die Ausgabewahrscheinlichkeiten besitzen folgende Werte:

	Art	Adj	RP	NN	Vfin
das	0,8	0	0,2	0	0
alte	0	0,7	0	0,3	0
Schloss	0	0	0	0,6	0,3

- Berechnen Sie die Wahrscheinlichkeit dieses Satzes mit Hilfe des Forward-Algorithmus.
- Geben Sie die Kategorienfolge an, die bei der vorliegenden Beobachtung die größte Wahrscheinlichkeit besitzt.