Centrum für Informations- und Sprachverarbeitung (CIS)
Übungen zum Masterkurs P2 ,,Vertiefung Grundlagen der Computerlinguistik“
WS 2016/17 Klaus U. Schulz
Blatt 2

Aufgabe 1.1 In einem Experiment werden ein roter und ein grüner Würfel geworfen. Beide Würfel sind fair. Wie sieht der dazu passende Wahrscheinlichkeitsraum aus? Welche Wahrscheinlichkeiten haben die folgenden Ereignisse:

1. A: Die Summe der Augenzahlen ist gerade,
2. B: Die Summe der Augenzahlen ist durch 3 teilbar,
3. C: Die Augensumme ist eine Quadratzahl,

Bilden Sie Vereinigungen und Durchschnitte über je zwei, über alle drei Ereignisse und berechnen Sie deren Wahrscheinlichkeit.

Aufgabe 1.2 Ein Professor gibt einen Morgenkurs und einen Mittagskurs. Sei A das Ereigniss “der Morgenkurs war schlecht”, sei B das Ereignis “der Mittagskurs war schlecht”. Es sei $P(A) = 0,3$, $P(B) = 0,2$, $P(A \cap B) = 0,1$. Im folgenden bezeichne A' das Komplement zu A, analog für B. Berechnen Sie $P(B \mid A)$, $P(B \mid A')$ und $P(B' \mid A)$.

Aufgabe 1.3 Wenn ein Multiple-Choice-Test 15 Fragen umfasst und es für jede Frage 5 mögliche Antworten gibt, wie groß ist die W, dass bei einer rein zufälligen Auswahl der Antworten höchstens 8 (genau 8, mindestens 8) Fragen richtig beantwortet werden?

Aufgabe 1.4 Es sei X eine Bernoulli-Variable, d.h. eine ZV mit den möglichen Werten 0 und 1. Wenn p die Wahrscheinlichkeit für 0 ist, wie ist der Erwartungswert von X? Wie die Varianz, die Standardabweichung?

Aufgabe 1.5 Die ZV X habe die drei möglichen Werte 3 (Wahrscheinlichkeit $0,3$), 4 (Wahrscheinlichkeit $0,4$) und 5 (Wahrscheinlichkeit $0,3$). Berechnen Sie Erwartungswert und Varianz.

Aufgabe 1.6 In einer Bibliothek darf ein Buch maximal 2 Stunden zum Lesen ausgeliehen werden. Es sei X die Ausleihzeit des Buches. Die Dichtefunktion von X sei $f(x) = x/2$ für $0 \leq x \leq 2$ und $f(x) = 0$ sonst. Wie groß ist $P(X \leq 1)$, $P(0,5 \leq X \leq 1,5)$? Wie sieht die kumulative Verteilungsfunktion von X aus? Was ist der Erwartungswert, die Varianz von X?

Aufgabe 1.7 Die Zufallsvariable X habe die möglichen Werte 100, 250, die Zufallsvariable Y die möglichen Werte 0, 100, 200. Die gemeinsame Verteilung sei gegeben durch $p(100, 0) = 0,2$, $p(100, 100) = 0,1$, $p(100, 200) = 0,2$, $p(250, 0) = 0,05$, $p(250, 100) = 0,15$ und $p(250, 200) = 0,3$. Sind X und Y unabhängig? Wie hoch ist die Kovarianz von X und Y, wie der Korrelationskoeffizient?

Aufgabe 1.8 Nach Angaben im Skript von Michaela Geierhos hat das Buch “Tom Sawyer” 71.370 Tokens und 8018 Types. Die höchste Rangzahl ist damit 8018. Es sei angenommen, dass dies die Rangzahl der Types mit nur einem Vorkommen ist. Geben Sie eine Abschätzung, wieviele Types im Buch genau 2 mal vorkommen. Anleitung: Berechnen Sie zunächst die sich hier ergebenden Parameter k und c (vgl. Geierhos Skript).

Aufgabe 1.9 In einem Korpus mit 10 Millionen Einträgen komme das Wort U 20.000 mal vor, das Wort V 2.000 mal und das Paar UV 7 mal. Wie kann man die Werte der Binomialverteilung (die Sie nicht ermitteln müssen) verwenden, um zu prüfen, ob das Wortpaar UV mit Signifikanz 0,1 eine Kollokation darstellt? Stellen Sie den Rechenweg dar.