Weitere Aufgaben zur Prüfungsvorbereitung

12.1 Zufallsvariable

Geben Sie zwei reelle Zufallsvariablen X und Y, für welche gilt $\mathrm{E}[X] > \mathrm{E}[Y]$ und $\mathrm{E}[X \cdot Y] \neq \mathrm{E}[X] \cdot \mathrm{E}[Y]$.

12.2 Zufallswahl

Im folgenden Algorithmus mit Eingabe $k > 0$ wählt die Funktion *random*(a, b) einen zufälligen Integer in $[a, b]$. Modellieren Sie die Ausgabe des Algorithmus als eine Zufallsvariable Y, beschreiben Sie diese, und begründen Sie Ihre Antwort.

```
x ← 0
y ← 2^k
while x < y − 1 do
    r ← random(x, y − 1)
    m ← x + (y−x)/2
    if m ≤ r then
        x ← m
    else
        y ← m
return y
```

12.3 Bitte anschnallen

In ein Flugzeug mit r Sitzreihen steigen g Passagiere auf Geschäftsreise und f Passagiere auf dem Weg in die Ferien. Jeder Passagier wählt zufällig und uniform eine Reihe und nimmt dort Platz, wobei die Geschäftsreisenden sich auf die vorderen 2/3 des Flugzeugs beschränken und die Ferienpassagiere nur in den hinteren 2/3 aller Reihen Platz nehmen (r ist durch 3 teilbar).

a) Wie viele Passagiere sitzen im Mitteldrittel durchschnittlich in einer Reihe?
b) Angenommen g und h sind viel kleiner als $\sqrt{r}$. Geben Sie eine Abschätzung für die Wahrscheinlichkeit, dass im Mitteldrittel nirgends zwei oder mehr Passagiere in derselben Reihe sitzen.
c) Unter derselben Annahme wie vorher, geben Sie eine Abschätzung für die Wahrscheinlichkeit, dass alle Passagiere im Flugzeug allein in einer Reihe sitzen.

Hinweis: Führen Sie Indikator-ZV ein; daraus konstruieren Sie eine ZV für die gesuchte Anzahl doppelt besetzter Reihen; auf diese wenden Sie dann zur Abschätzung die Markov-Ungleichung an.

$\Longrightarrow$

12.4 Warteschlange

In einem Supermarkt mit c Kassen wählen k Kunden unabhängig und uniform verteilt je eine Kasse und stellen sich an.

a) Wie viele Kunden stehen durchschnittlich an jeder Kasse an?

b) Geben Sie eine Abschätzung unter Verwendung der Chebyshev-Ungleichung für die Wahrscheinlichkeit, dass an der ersten Kasse x oder mehr Kunden anstehen.

c) Basierend auf der letzten Teilaufgabe, geben Sie eine Abschätzung für die Wahrscheinlichkeit dafür, dass in irgendeiner Warteschlange x oder mehr Kunden stehen.

12.5 Quiz

In einem Quiz gibt es rote, grüne, blaue und weisse Fragen. Die Gewinnchancen unterscheiden sich je nach Farbe einer Frage und sind:

Farbe	rot	grün	blau	weiss
Chance	0.4	0.3	0.2	0.1

Das Quiz wird n Mal gespielt und Fragen aller Farben kommen gleich häufig vor. Für jeden Gewinn gibt es einen Preis. Die Quizmaster errechnen die erwartete Anzahl Gewinne, erhöhen diese als Puffer um 20% und kaufen so viele Preise ein. Berechnen Sie eine Schranke abhängig von n für die Wahrscheinlichkeit, dass die bereitgestellten Preise trotzdem nicht ausreichen.

12.6 Vergleichen durch Hashing

Anna und Barbara möchten zwei grosse Datensätze $D_A, D_B \in \mathcal{D}$ miteinander vergleichen, wobei Anna D_A kennt und Barbara D_B. Anna und Barbara können miteinander kommunizieren, aber die Datensätze sind zu gross, um sie auszutauschen.

Die beiden nehmen dazu eine ideale Hashfunktion $H : \mathcal{D} \to [1, n]$. Für die Anwendung von H wird der Datensatz zuerst in eine kanonische Ordnung gebracht. In der Analyse wird die Hashfunktion so modelliert, dass H für jeden einzelnen Datensatz eine uniform verteilte Zahl in $[1, n]$ ausgibt. In diesem Modell wirkt die Hashfunktion wie ein Orakel, welches aber wichtige Eigenschaften konkreter Hashfunktionen hat. Wenn ein Wert $d \in \mathcal{D}$ zum ersten Mal angefragt wird (von einem beliebigen Teilnehmer), so wählt H eine zufällige Zahl h in $[1, n]$; dies wiederspiegelt, dass die Hashfunktion einen scheinbar zufälligen Wert ausgibt. Jeder weitere Aufruf von $H(d)$, egal ob von demselben Teilnehmer oder einem anderen, ergibt wiederum das gleiche h, entsprechend einer realen, deterministischen Hashfunktion.

a) Beschreiben Sie einen Algorithmus für den Vergleich. Das Ergebnis soll möglichst sicher bestimmt werden und höchstens mit Wahrscheinlichkeit ϵ falsch sein.

b) In welcher Art kann sich dieser Algorithmus irren?

c) Wie viele Aufrufe von H sind nötig mit $n = 10$, damit Anna und Barbara sich höchstens mit Wahrscheinlichkeit 2^{-10} irren?

12.7 Entropie

Sei $Y = f(X)$ für eine Zufallsvariable X und eine deterministische Funktion f. Was können Sie allgemein aussagen über Grössen oder Beziehungen zwischen $H(X)$, $H(Y)$ und $H(X|Y)$, $H(Y|X)$?

$\Longrightarrow$

12.8 Entropie zweier Zufallsvariablen

Berechnen Sie folgende Grössen für die Verteilung P_{XY} von $X, Y \in \{0,1\}^2$ unten:

a) $H(X)$, $H(Y)$, $H(XY)$;

b) $H(X|Y)$, $H(Y|X)$;

c) $H(X) - H(X|Y)$, $H(Y) - H(Y|X)$.

X \ Y	0	1
0	$\frac{1}{8}$	$\frac{1}{4}$
1	$\frac{7}{16}$	$\frac{3}{16}$

(Ohne Taschenrechner benutzen Sie hierzu Ausdrücke wie $\log(5)$ oder $h(\frac{1}{4})$; Logarithmen sind zur Basis 2.)

12.9 Bernoulli-Verteilung erzeugen

Gegeben $p \in [0,1]$ soll eine Bernoulli-Zufallsvariable X erzeugt werden mit $P_X(1) = p$. Zur Verfügung steht eine Quelle von uniform verteilten und unabhängigen Zufallsbits $Z_1, Z_2, \ldots$.

Hinweis: Stellen Sie p im Binärsystem dar. Für rationale p siehe auch Übung 6, aber hier geht es um $p \in \mathbb{R}$.

12.10 Code

Betrachten Sie die folgenden Codes und geben Sie jeweils an, ob ein Code (1) präfixfrei und/oder (2) eindeutig decodierbar ist:

$$\begin{array}{ll} (a) & [0, 01, 001] \\ (b) & [00, 01, 100, 101, 11] \\ (c) & [0, 00, 000, 0000] \end{array}$$

12.11 Shannon- und Huffman-Code

Gegeben eine Zufallsvariable X mit der Verteilung

x	a	b	c	d
$P_X(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{12}$

a) Finden Sie einen Huffman-Code für X.

b) Zeigen Sie, dass zwei optimale Codes gibt, nämlich einen Code C_1 mit den Codewort-Längen $[1,2,3,3]$ und einen Code C_2 mit den Längen $[2,2,2,2]$.

c) Ist C_1 oder C_2 ein Shannon-Code? Warum?