Technische Universität München **Sommersemester 2009**
Theoretische Informatik **Übungsblatt 5**
Prof. J. Esparza / S. Kiefer / M. Luttenberger

Übungen zu Diskrete Wahrscheinlichkeitstheorie

Dieses Blatt ist am 8. Juni bis 10:15 abzugeben und wird am 8./9. Juni besprochen.

Beachten Sie: Alle Ergebnisse sind zu begründen!

Aufgabe 5.1 **1P+1P+1.5P+1.5P**

Betrachten Sie wieder das Szenario mit den Filmdateien und der Serverfarm von Aufgabe 4.3 vom letzten Aufgabenblatt. Sie haben dort festgestellt, dass $X_1 \sim \text{Bin}(200, 0.01)$ gilt und (nach einigem Aufwand) $\Pr[X_1 > 11] \leq 0.0001$.

(a) Berechnen Sie $\Pr[X_1 > 6]$ approximativ mit der Poisson-Verteilung.
Hinweis: Benutzen Sie die Poisson-Verteilung nur in dieser Teilaufgabe.

(b) Geben Sie $\mathbb{E}[X_1]$ an und bestimmen Sie nun mit der Markov-Ungleichung ein möglichst kleines k, sodass $\Pr[X_1 > k] \leq 0.0001$ gilt.

(c) Berechnen Sie $\text{Var}[X_1]$ und bestimmen Sie nun mit der Chebyshev-Ungleichung ein möglichst kleines k, sodass $\Pr[X_1 > k] \leq 0.0001$ gilt.

(d) Bestimmen Sie nun mit der Chernoff-Schranke ein möglichst kleines k, sodass $\Pr[X_1 > k] \leq 0.0001$ gilt.

Aufgabe 5.2 **1P+1P+2P+1P+1P**

Sie benutzen eine E-Mail-Adresse von GMX. Eine Mail ist mit W'keit $p = 0.4$ Spam. Die Zufallsvariablen M bzw. S geben an, wie viele Mails insgesamt bzw. wie viele Spam-Mails Sie täglich bekommen.

(a) Geben Sie die W'keit an, dass Sie k Spam-Mails bekommen, unter der Bedingung, dass Sie insgesamt 8 Mails bekommen. In anderen Worten: Berechnen Sie $\Pr[S = k \mid M = 8]$.

(b) Sie bekommen im Mittel $\lambda = 10$ Mails, also $\mathbb{E}[M] = \lambda$. Laut Folien gilt somit $\sum_{i=1}^{\infty} \Pr[M \geq i] = \lambda$. Zeigen Sie mithilfe von Indikatorvariablen: $\mathbb{E}[S] = \lambda p = 4$.
Hinweis: Sie werden in A5.6 dieses Resultat mit Hilfe von erzeugenden Funktionen zeigen. Betrachten Sie für diese Aufgabe die Indikatorvariablen I_{A_i} für die Ereignisse

$$A_i = \text{„}M \geq i \text{ und die } i\text{-te Mail ist Spam“}$$

und deren Summe über alle $i \in \{0, 1, \ldots\}$.

(c) Genauer gesagt ist Ihr tägliches Mail-Aufkommen M Poisson-verteilt mit Parameter λ. Zeigen Sie mit dem Satz von der totalen W'keit, dass S Poisson-verteilt ist mit Parameter λp.
Hinweis: Benutzen Sie die Exponentialreihe $\displaystyle\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$.

(d) Alle Spam-Mails landen im Ordner Spamverdacht. Wenn Sie an einem Tag mindestens eine Spam-Mail bekommen, stellt Ihnen GMX eine Zusammenfassung Ihrer Spam-Mails zu. Berechnen Sie die W'keit, dass GMX Ihnen morgen eine solche Zusammenfassung schickt.
Hinweis: Benutzen Sie (c), auch wenn Sie (c) nicht gelöst haben.

(e) Angenommen, Sie bekommen eine solche Zusammenfassung. Wie viele Spam-Mails erwarten Sie darin?
Hinweis: Die Zahl „4“ ist nicht die richtige Antwort.

Aufgabe 5.3 **2P+2P+1P**

(a) Die Firma "Poissonous Cookies" stellt unter anderem Kekse mit Rosinen her. Hierfür werden $\lambda \cdot N$ Rosinen in den Teig für N Kekse gegeben, woraufhin die ganze Masse so gut durchmischt wird, dass wir annehmen können, dass jede einzelne Rosine mit derselben W'keit in einem der Kekse landet.

Wie groß muss die durchschnittliche Zahl λ von Rosinen pro Kekse sein, wenn höchstens jeder hunderste Keks keine Rosinen enthalten darf.

Hinweis: N ist unbekannt, aber groß.

(b) Zwei Bekannte lesen Ihre Bachelorarbeit unabhängig von einander Korrektur. Bekannter Nr. 1 gibt Ihenn eine Liste mit k_1 vielen Fehlern, Bekannter Nr. 2 eine Liste mit k_2 vielen Fehlern. Die Listen stimmen in k_{12} vielen Einträgen überein.

Es sei n die Anzahl der tatsächlichen Fehler in Ihrer Bachelorarbeit. Wir nehmen an, dass beide Bekannte bei jedem der n Fehler mit W'keit p_1 bzw. p_2 auf diesen aufmerksam werden.

i) Es sei X_i die Zufallsvariable, die die Anzahl der von Bekannten Nr. i gefundenen Fehler angibt.

Wie ist dann X_i verteilt? Für welche Werte von n gilt insbesondere

$$\Pr\left[\left|\frac{1}{n}X_i - p_i\right| \geq 0.1\right] \leq 0.01?$$

Hinweis: Überlegen Sie sich, dass stets $\mathrm{Var}[X_i] \leq 1/4$ gilt.

ii) Nehmen Sie an, dass n groß genug ist, damit die Ungleichung aus (i) erfüllt ist.

Begründen Sie damit, dass $k_i \approx n \cdot p_i$ eine vernünftige Annahme ist.

Leiten Sie hiermit die Schätzung $n \approx \frac{k_1 k_2}{k_{12}}$ her.

Aufgabe 5.4 **1P+1P+1P+2P+1P**

Ein Kind surft zufällig im Internet. Es startet bei Seite A.

Auf Seite A klickt es mit W'keit $\frac{2}{3}$ Seite B an, mit W'keit $\frac{1}{3}$ Seite C.

Auf Seite B klickt es mit W'keit $\frac{1}{4}$ Seite C an, mit W'keit $\frac{3}{4}$ Seite A.

Sei X die Zahl der Klicks, bis das Kind Seite C *erstmals* erreicht.

(a) Zeichnen Sie ein Markov-Diagramm, das das Verhalten des Kindes darstellt.

(b) Bestimmen Sie zunächst $\Pr[X = 0]$, $\Pr[X = 1]$ und $\Pr[X = 2]$.

(c) Geben Sie nun für $k \geq 1$ $\Pr[X = k + 2]$ in Abhängigkeit von $\Pr[X = k]$ an.

(d) Zeigen Sie: Für die erzeugende Funktion G_X von X gilt:

$$G_X(z) = \frac{2z + z^2}{6 - 3z^2}$$

Hinweis: Benutzen Sie die Definition von $G_X(z)$ sowie (b) und (c), um eine Gleichung für $G_X(z)$ aufzustellen.

(e) Benuten Sie (d), um $\mathbb{E}[X]$ zu berechnen.

Aufgabe 5.5 **1P+1P+1P+1P**

In der Vorlesung wurde die erzeugende Funktion $G_X(z)$ einer Zufallsvariable X mit Werten in $\mathbb{N}_0$ definiert als

$$G_X(z) := \mathbb{E}[z^X] = \sum_{k=0}^{\infty} z^k \cdot \Pr[X = k] \text{ für } z \in [0,1].$$

Im Weiteren seien X und Y unabhängige Zufallsvariablen. Sowohl X als auch Y nehmen nur Werte in $\mathbb{N}_0$ an.

(a) Zeigen Sie für beliebige Konstanten $a, b \in \mathbb{R}$, dass auch a^X und b^Y unabhängig sind.

(b) Begründen Sie mit Hilfe von (a) und den Ergebnissen aus der Vorlesung, dass

$$G_{X+Y}(z) = G_X(z) \cdot G_Y(z)$$

für alle $z \in [0,1]$.

(c) Es seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit Werten in $\mathbb{N}_0$ und $a_1, \dots, a_n \in \mathbb{N}$ Konstanten.

Verallgemeinern Sie das Resultat aus (b) auf

$$G_{\sum_{i=1}^n a_i X_i}(z) = G_{X_1}(z^{a_1}) \cdots G_{X_n}(z^{a_n}).$$

(d) Bestimmen Sie die erzeugende Funktion einer negativ binomialverteilten Zufallsvariable.

Hinweis: Die Beweise zu (a) und (b) ergeben sich direkt aus Sätzen aus der Vorlesung.

Aufgabe 5.6 **1P+1P+1P+1P**

Es seien $N, X_1, X_2, X_3, \dots$ unabhängige Zufallsvariablen, die Werte in $\mathbb{N}_0$ annehmen. Weiterhin seien die X_i identisch verteilt. Wir definieren hiermit die diskrete Zufallsvarialbe $S := \sum_{i=1}^N X_i$.

Man kann zeigen (siehe A5.7 (b)), dass dann für die erzeugende Funktion $G_S(z)$ von S gilt:

$$G_S(z) = G_N(G_{X_1}(s)). \tag{*}$$

(a) Wir nehmen an, dass $N \sim \text{Poi}(\lambda)$ und $X_i \sim$ Bin(1;p) gilt.

  i) Bestimmen Sie explizit $G_S(z)$.

  ii) Wie kann man direkt aus (i) das Ergebnis aus A5.2 (c) ableiten?

(b) Für jedes feste $n \in \mathbb{N}$ gilt nach A5.5:

$$G_{\sum_{i=1}^n X_i}(z) = G_{X_1}(z)^n.$$

Wie folgt dieses Resultat aus (*)?

(c) Zeigen Sie mittels (*), dass $\mathbb{E}[S] = \mathbb{E}[N] \cdot \mathbb{E}[X_1]$ gilt.

Hinweis: Die Folien zu den erzeugenden Funktionen sind sehr hilfreich.

Aufgabe 5.7 **(freiwillige Zusatzaufgabe, nicht abzugeben/wird nicht korrigiert)** **-**

Wir betrachten die Verbreitung eines Computerwurms. Zu Beginn existiert genau eine Kopie des Wurms. Wir gehen weiterhin davon aus, dass jeder Rechner mit einem Virenscanner ausgerüstet ist, der den Wurm schließlich unschädlich macht. Bis zu dem Erkennen des Wurms vebreitet dieser eine zufällige Anzahl $X \in \mathbb{N}_0$ an Kopien über das Internet, die sich absolut identisch verteilen.

Um das Modell zu vereinfachen, nehmen wir im Weiteren an, dass jede Kopie des Wurms genau eine Zeiteinheit existiert, bevor sie durch den Virenscanner entfernt wird. Weiterhin nehmen wir an, dass genau zum Zeitpunkt seines Entfernens der Wurm X Kopien von sich über das Internet unabhängig von andern Kopien verteilt.

Wir können also von Generationen von Kopien sprechen. Für $t \in \mathbb{N}_0$ sei N_t die Anzahl der Kopien des Wurms in der t-ten Generation. Die 0-te Generation besteht aus genau $N_0 = 1$ Kopien. Die 1-te Generation besteht mit W'keit $\Pr[X = k]$ aus genau $N_1 = k$ Kopien. Für die $t + 1$-te Generation gilt dann

$$N_{t+1} = \sum_{i=1}^{N_t} X_i^{(t)},$$

wobei die $X_i^{(t)}$ unabhängige Zufallsvariablen sind, die identisch zu X verteilt sind. Wir nehmen auch an, dass die $X_i^{(t)}$ von N_t unabhängig sind.

Beispiel:

Zum Zeitpunkt $t = 0$ existiert genau $N_0 = 1$ Kopie des Wurms. Diese wird zum Zeitpunkt $t = 1$ gelöscht, kann aber noch kurz davor $X_1^{(0)} = 3$ Kopien von sich verschicken. Zum Zeitpunkt $t = 2$ wird jede dieser $N_1 = 3$ Kopien entfernt, jede von ihnen kann jedoch $X_1^{(1)} = 2$ bzw. $X_2^{(1)} = 5$ bzw. $X_3^{(1)} = 1$ Kopie von sich verschicken, so dass insgesamt zum Zeitpunkt $t = 2$ genau $N_2 = 8$ Kopien des Wurms existieren, usw.

Man kann die Verteilung von X auffassen als eine Beschreibung der Güte des Virenscanners: Ein guter Virenscanner lässt mit hoher W'keit nur eine geringe Anzahl an Kopien zu. Wir untersuchen im Folgenden den Zusammenhang zwischen N_t und X.

(a) Begründen Sie, dass

$$\mathbb{E}[N_{t+1} | N_t = k] = k \cdot \mathbb{E}[X]$$

gilt.

Führen Sie damit $\mathbb{E}[N_t]$ auf $\mathbb{E}[X]$ und t zurück.

Hinweis: Leiten Sie zunächst eine Beziehung zwischen $\mathbb{E}[N_{t+1}]$ und $\mathbb{E}[N_t]$ her. Leiten Sie dann mit Hilfe vollständiger Induktion einen geschlossenen Ausdruck für $\mathbb{E}[N_t]$ her, der nur von $\mathbb{E}[X]$ und t abhängt.

(b) Zeigen Sie, dass

$$G_{N_{t+1}}(z) = G_{N_t}(G_X(z))$$

gilt.

Hinweis: Beginnen Sie mit der Definition von $G_{N_{t+1}}(z)$; wenden Sie dann den Satz der totalen W'keit an, um nach den Werten von N_t zu bedingen; verwenden Sie anschließend, dass die Reihen für $z \in [0, 1]$ absolut konvergent sind und daher umgeordnet werden dürfen. Beachten Sie auch A5.3.

(c) Folgern Sie aus (b), dass $G_{N_t}(z)$ die t-fache Komposition von $G_X(z)$ mit sich selbst ist, also

$$G_{N_t}(z) = G_X^t(z) := \underbrace{G_X \circ \ldots \circ G_X}_{t-\text{mal}}(z).$$

Überprüfen Sie hiermit Ihr Ergebnis aus (a).

(d) Die W'keit, dass der Wurm zum Zeitpunkt t erfolgreich durch den Virenscanner bekämpft wurde, ist gerade $\Pr[N_t = 0]$. Die W'keit, dass der Wurm *schließlich* vollkommen gestoppt wird, ist daher $z^* := \Pr\left[\bigcup_{t \geq 0} N_t = 0\right]$. Man kann zeigen, dass z^* die kleinste nicht negative Lösung der Gleichung $z = G_X(z)$ ist.

Wir betrachten einen Virenscanner, der genau zwei oder genau null Kopien zulässt. Es sei daher $p := \Pr[X = 2]$ und damit $1 - p = \Pr[X = 0]$ mit $0 < p < 1$.

Zeigen Sie, dass für diese Verteilung von X die Gleichung $z = G_X(z)$ nur die Lösungen 1 und $\frac{1-p}{p}$ hat. Für welche Werte von p wird der Wurm somit schließlich mit W'keit 1 gestoppt?

(e) Zeigen Sie, dass z^* die kleinste nicht negative Lösung von $z = G_X(z)$ ist.