TU Berlin - Institut für Mathematik
Sommersemester 2024
Dozent: Dr. Nikolas Tapia
Assistentin: M.Sc. Claudia Drygala

Stochastik für Informatik(er) – Lösungsvorschlag Übung 4

Abgabe bis Freitag, den 24.05.2024 um 23:59

Hinweise zur Bearbeitung des Übungsblattes:

- Das Übungsblatt enthält Haus- und Tutoriumsaufgaben.
- Die Tutoriumsaufgaben werden in den Tutorien der KW 20 besprochen (13.05.-17.05.).
- Die Abgabe der Hausaufgaben erfolgt über ISIS in festen Gruppen von 2-3 Personen. Die Gruppen bilden sich aus Studierenden, die das gleiche Tutorium besuchen bzw. mindestens dieselbe/denselben Tutor*in haben. Laden Sie Ihre handschriftlichen Lösungen (z.B. Scan Ihrer Lösungen oder Erstellung Ihrer Lösungen über Tablet) als eine PDF-Datei bei dem enstprechenden Übungsblatt hoch. LaTeX-Abgaben sind auch willkommen (in diesem Fall die kompilierte PDF)! Achten Sie darauf, dass die Abgaben gut lesbar und verständlich verfasst sind. Bitte Namen, Matrikelnummer und Übungsgruppe auf der Abgabe mit angeben!

Tutoriumsaufgaben

Tutoriumsaufgabe 4.1

Ein Bewohner einer sehr kleinen Insel ist Mitglied eines sozialen Netzwerkes, und die Anzahl seiner Freunde in diesem Netzwerk ist zipfverteilt mit Parameter 3.

(i) Erinnern Sie sich an die Definition einer Zufallsvariablen mit Zipf-Verteilung mit Parameter $a > 1$.

(ii) Was ist wahrscheinlicher: Er hat nur einen Freund im Netzwerk oder er hat mindestens 3 Freunde im Netzwerk?
Hinweis: Sie können mit $\sum_{k=1}^{\infty} k^{-3} = 1.2021$ rechnen (dies ist eigentlich nur eine Rundung des korrekten Wertes).

Lösung für Tutoriumsaufgabe 4.1

(i) Eine Zufallsvariable X heißt Zipfverteilt mit Parameter $a > 1$, falls gilt: $X(\Omega) = \mathbb{N}^*$ und

$$\mathbb{P}(X = k) = \frac{k^{-a}}{Z(a)},$$

wobei

$$Z(a) := \sum_{k=1}^{\infty} k^{-a}$$

(ii) Sei X die Anzahl der Freunde des Bewohners und $Z(3) := \sum_{k=1}^{\infty} k^{-3} = 1,2021$, dann gelten

$$\mathbb{P}(X=1) = \frac{1^{-3}}{Z(3)} = \frac{1}{Z(3)} \approx 0,8,$$
$$\mathbb{P}(X=2) = \frac{2^{-3}}{Z(3))} = \frac{1}{8Z(3)},$$

und deshalb

$$\mathbb{P}(X \geq 3) = 1 - \mathbb{P}(X=1) - \mathbb{P}(X=2) = \frac{Z(3) - 1 - \frac{1}{8}}{Z(3)} \approx 0,0641\ .$$

Es ist deutlich wahrscheinlicher, dass der Einwohner nur einen Freund hat, als dass er mindestens 3 Freunde hat.

Tutoriumsaufgabe 4.2

Die Anzahl der Studierenden, die eine Sprechstunde für Stochastik für Informatiker besuchen, ist poissonverteilt mit Parameter $\lambda > 0$. Diese Studierenden studieren jeweils mit Wahrscheinlichkeit p unabhängig voneinander Informatik, wobei $p \in (0,1)$.

(i) Bestimmen Sie die Wahrscheinlichkeit, dass genau k Informatikstudierende zur Sprechstunde kommen, falls insgesamt n Studierende zur Sprechstunde kommen, für $k, n \in \mathbb{N}$.

(ii) Wie ist die Anzahl der Informatikstudierenden bei der Sprechstunde verteilt? Geben Sie die Verteilung dieser Zufallsvariable explizit an.

Lösung für Tutoriumsaufgabe 4.2

Sei Y die Anzahl der Studierenden und X die Anzahl der Informatikstudierenden bei der Sprechstunde.

(i) Gesucht ist $\mathbb{P}(X = k|Y = n)$. Falls $n < k$, $\mathbb{P}(X = k|Y = n) = 0$. Sonst, gegeben das Ereignis $\{Y = n\}$ ist X binomialverteilt mit Parameter n und p. Damit gilt $\mathbb{P}(X = k|Y = n) = \binom{n}{k} p^k (1-p)^{n-k}$.

(ii) Gesucht ist $\mathbb{P}(X = k)$. Laut der Formel von der Gesamtwahrscheinlichkeit,

$$\mathbb{P}(X = k) = \sum_{n=0}^{\infty} \mathbb{P}(X = k|Y = n)\mathbb{P}(Y = n) = \sum_{n=k}^{\infty} \mathbb{P}(X = k|Y = n)\mathbb{P}(Y = n).$$

Damit gilt

$$
\begin{aligned}
&\mathbb{P}(X=k) = \\
&= \sum_{n=k}^{\infty} \mathbb{P}(Y=n)\mathbb{P}(X=k|Y=n) = \sum_{n=k}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{n=k}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{n=k}^{\infty} \frac{\lambda^k p^k}{k!} e^{-\lambda p} \frac{\lambda^{n-k}(1-p)^{n-k}}{(n-k)!} e^{-\lambda(1-p)} \\
&= \frac{\lambda^k p^k}{k!} e^{-\lambda p} \sum_{n=k}^{\infty} \frac{\lambda^{n-k}(1-p)^{n-k}}{(n-k)!} e^{-\lambda(1-p)} = \frac{\lambda^k p^k}{k!} e^{-\lambda p} \sum_{n=0}^{\infty} \frac{\lambda^n (1-p)^n}{n!} e^{-\lambda(1-p)} \\
&= \frac{\lambda^k p^k}{k!} e^{-\lambda p}.
\end{aligned}
$$

Deshalb ist X poissonverteilt mit Parameter λp.

Tutoriumsaufgabe 4.3

Die Wahrscheinlichkeit, dass jemand aus der Bevölkerung in einem Jahr bestimmten Unfallarten ausgesetzt ist liegt bei $p = \frac{1}{1000}$. Es ist auch sinnvoll, eine Unabhängigkeit zwischen den einzelnen Unfällen anzunehmen. Eine Versicherungsgesellschaft hat 10.000 Personen aus der Bevölkerung versichert und möchte nun die Wahrscheinlichkeit finden, dass höchstens 2 Personen einen Unfall erleiden werden.

(i) Sei X die Zufallsvariable der Anzahl der Personen, die den Unfall in einem Jahr erleiden. Was ist die Verteilung von X?

(ii) Drücken Sie die von der Versicherungsgesellschaft gesuchte Wahrscheinlichkeit mit Hilfe der Zufallsvariablen X aus und berechnen Sie sie.

(iii) Der Probabilist, der diese Wahrscheinlichkeit berechnen muss, nimmt an, dass X Poisson-verteilt mit Parameter 10 ist. Berechnen Sie das Ergebnis des Wahrscheinlichkeitsrechners und vergleichen Sie es mit der Antwort in (ii). Was fällt Ihnen auf? Begründen Sie die erzielten Ergebnisse.

Lösung für Tutoriumsaufgabe 4.3

(i) Das Experiment besteht aus aufeinander folgenden, unabhängigen Bernoulli-Experimenten mit Erfolgsparameter $p = 1/1000$. Dann X ist eine binomiale Zufallsvariable mit den Parametern $n = 10.000$ und $p = 1/1000$.

(ii) Die gesuchte Wahrscheinlichkeit wird durch das Ereignis bestimmt

$$\mathbb{P}(X \leq 2) = \mathbb{P}(X=0) + \mathbb{P}(X=1) + \mathbb{P}(X=2)$$

$$= (1-p)^{10000} + 10000p(1-p)^{9999} + \binom{10000}{2} p^2 (1-p)^{9998} \approx 0.002760$$

.

(iii) Falls X eine Poisson-Zufallsvariable mit Parameter $\lambda = 10$ ist, dann:

$$\mathbb{P}(X \leq 2) = \mathbb{P}(X=0) + \mathbb{P}(X=1) + \mathbb{P}(X=2)$$

$$= \exp(-10)\left(1 + \frac{10}{1} + \frac{10^2}{2!}\right) \approx 0,002769\,.$$

Die Werte unterscheiden sich erst in der 6. Nachkommastelle, d.h. sie sind gleich bei einer Toleranz von 10^{-6}! Eine derartige Übereinstimmung ist zu erwarten: Da $\lambda = 10 = 10000 \cdot \frac{1}{1000} = np$ ist, gilt die Poisson-Approximation für das Binom (für großes n und kleines p).

Tutoriumsaufgabe 4.4

20% der Bewerber für einen Job verfügen über fortgeschrittene Computerkenntnisse. Bewerber führen nacheinander Vorstellungsgespräche und sind zufällig unabhängig voneinander aus einer Bewerberliste gewählt. X bezeichne die Anzahl der Vorstellungsgespräche, die geführt werden müssen, um den ersten Bewerber mit fortgeschrittenen Computerkenntnissen zu finden.

(i) Wie groß ist die Wahrscheinlichkeit, dass der erste Bewerber mit fortgeschrittenen Computerkenntnissen im dritten Vorstellungsgespräch gefunden wird?

(ii) Wie ist X verteilt?

(iii) $p \in (0,1)$ beschreibe die Wahrscheinlichkeit, dass ein Bewerber für den Job über fortgeschrittene Computerkenntnisse verfügt. Welches $p \in (0,1)$ maximiert die Wahrscheinlichkeit, dass die Firma den ersten Bewerber mit fortgeschrittenen Computerkenntnissen im dritten Vorstellungsgespräch findet?

Lösung für Tutoriumsaufgabe 4.4

Gehen wir von einer unendlich langen Liste aus, so ist X die Anzahl der Versuche bis zum ersten Erfolg inklusive diesem, und damit geometrisch verteilt auf $\mathbb{N}^* = \{1, 2, \ldots\}$. Ist die Liste nur endlich (aber ausreichend) lang, so ist X im Prinzip genauso verteilt, nur besteht die Möglichkeit, dass X nicht definiert ist (wenn die Liste keinen geeigneten Bewerber enthält). Das stört hier wenig. Grundsätzlich sei die Entscheidung, ob Bewerber i fortgeschrittene Kenntnisse hat, ein Bernoulli-Experiment, und verschiedene Bewerber seien unabhängig.

(i) Für einen Erfolg im dritten Versuch braucht es genau zwei Fehlversuche in Folge und danach einen Erfolg, also

$$\mathbb{P}(X = 3) = 0.8^2 \cdot 0.2 = 0.128\ .$$

(ii) Mit verallgemeinerter Begründung $\mathbb{P}(X = k) = 0.8^{k-1} \cdot 0.2$.

(iii) Setze $f(p) := \mathbb{P}(X = 3) = (1-p)^2 \cdot p$, dann ist f zu maximieren unter der Bedingung, dass $p \in [0,1]$. Es gelten dann

$$\begin{aligned} f'(p) &= (1-p)^2 - 2(1-p)p = (1-p)(1-3p) \\ f''(p) &= -3(1-p) - (1-3p) = 6p - 4. \end{aligned}$$

Nullsetzen der Ableitung liefert $f'(p) = 0$ genau dann, wenn $p \in \{\frac{1}{3}, 1\}$ und es gilt $f''(\frac{1}{3}) = -2$, $f''(1) = 2$. Daher liegt das Maximum der Funktion bei $p = \frac{1}{3}$, welches daher der eindeutige Parameter ist, der die gesuchte Wahrscheinlichkeit maximiert.

Hausaufgaben

Hausaufgabe 4.1 (2=1+1 Punkte)

Wir nehmen an, dass in einem kleinen Vorort lediglich 6 weibliche Erwachsene wohnen, während in der Hauptstadt abzählbar unendlich viele erwachsene Frauen leben. In einem Frauennetzwerk tauscht sich eine Bewohnerin aus der Hauptstadt regelmäßig mit anderen Frauen aus. Dabei ist die Bewohnerin mit X Freundinnen aus dem realen Leben in diesem Netzwerk vernetzt, wobei X zipfverteilt mit Parameter 4 ist. Außerdem wissen wir, dass die Bewohnerin mit keiner Freundin aus der Haupstadt in dem Frauennetzwerk vernetzt ist, falls nicht auch alle Bewohnerinnen des Dorfes ihre Freundinnen sind und in dem Frauennetzwerk aktiv sind.

(i) Beschreiben Sie mit der Zufallsvariablen X die Wahrscheinlichkeit, dass die Bewohnerin mindestens eine Freundin aus der Hauptstadt in dem Frauennetzwerk hat.

(ii) Berechnen Sie die Wahrscheinlichkeit aus (i).

Hinweis: Sie können die Gleichung $\sum_{k=1}^{\infty} k^{-4} = \frac{\pi^4}{90}$ verwenden.

Hausaufgabe 4.2 (6=1+3+2 Punkte)

Eine Urne enthält 8 weiße Kugeln mit den Nummern $1, 2, \ldots, 8$ und 2 schwarze Kugeln mit den Nummern 9 und 10. Zwei Kugeln werden zufällig ohne Zurücklegen ausgewählt.

(i) Beschreiben Sie einen Wahrscheinlichkeitsraum, um diese Ziehung zu modellieren.

(ii) Sei die Anzahl der weißen Kugeln durch die Zufallsvariable X beschrieben, Bestimmen Sie die Verteilung von X.

(iii) Seien X_1, X_2 die gezogenen Kugeln und sei Z das Minimum über die Nummern der ausgewählten Kugeln, also $Z = \min(X_1, X_2)$. Berechnen Sie die Wahrscheinlichkeit, dass $\mathbb{P}(Z \geq 7)$.

Hausaufgabe 4.3 (6=2+2+2 Punkte)

In einer wissenschaftlichen Zeitung sind auf 500 Seiten 100 Druckfehler zufällig und unabhängig voneinander verteilt.

(i) Wir interessieren uns für die Anzahl X der Druckfehler auf Seite 347. Welche Verteilung hat die Zufallsvariable X? Bestimmen Sie die Wahrscheinlichkeit, dass drei oder mehr Druckfehler auf Seite 347 auftreten.

(ii) Berechnen Sie die Wahrscheinlichkeit aus (i) mit einer geeigneten Approximation. Warum ist diese Approximation zulässig?

(iii) Sei Z die Seitenzahl der Seite, auf der der erste Fehler auftritt. Wie groß ist die Wahrscheinlichkeit, dass kein Druckfehler auf den Seiten 1–3 auftritt?

Hausaufgabe 4.4 (6=4+2 Punkte)

In dieser Übung wollen wir "gedächtnislose" diskrete Zufallsvariablen untersuchen. Eine Zufallsvariable $X\colon \Omega \to \mathbb{N}^+ = \{1\,,\cdots,\}$ heißt gedächtnislos, wenn für jedes $n, m \geq 0$ gilt

$$\mathbb{P}(X > n + m | X > n) = \mathbb{P}(X > m)\,.$$

(i) Angenommen, X ist eine geometrisch-verteilte Zufallsvariable mit dem Parameter $p \in (0, 1]$. Das heißt, $X\colon \Omega \to \mathbb{N}^+$ erfüllt für jede ganze Zahl $k \geq 1$

$$p_X(k) = p(1-p)^{k-1}\,.$$

Beweisen Sie, dass X eine gedächtnislose Zufallsvariable ist.

(ii) Sei X eine Zufallsvariable mit Werten in $\mathbb{N}^+$, so dass die Eigenschaft der Gedächtnislosigkeit gilt. Dann lässt sich zeigen Sie, dass X geometrisch verteilt ist mit Parameter $p := P(X = 1)$. Für $n \in \mathbb{N}$ sei X_n nun geometrisch verteilt mit Parameter $p_n = a/n$. Beweisen Sie für jedes $b \in \mathbb{N}$:

$$\lim_{n \to +\infty} P(X_n \geq bn) = e^{-ab}\,.$$