Technische Universität Dortmund
Fakultät Statistik
Jun.-Prof. Dr. Andreas Groll
M.Sc. Hendrik van der Wurp

Wintersemester 2018/2019
24. Januar 2019

Übungen zur Vorlesung Wahrscheinlichkeitsrechnung und Mathematische Statistik (für Informatiker)

Blatt 7 - Lösungsvorschlag

Aufgabe 19: (Erneut, per Hand)

(a) Ein Viertel der 20 Angestellten eines kleinen IT-Unternehmens beschweren sich auf Nachfrage über zu hohe Arbeitsbelastung. Ein Unternehmensberater befragt drei zufällig ausgewählte Angestellte dieses Unternehmens. Wie groß ist die Wahrscheinlichkeit, dass
   (i) sich keiner beschwert?
   (ii) sich höchstens zwei beschweren?
   (iii) sich höchstens einer nicht beschwert?

(b) Jeweils ein Viertel der Angestellten von drei IT-Unternehmen beschweren sich auf Nachfrage über zu hohe Arbeitsbelastung. Ein Unternehmensberater befragt jeweils einen zufällig ausgewählten Angestellten pro Unternehmen. Wie groß ist die Wahrscheinlichkeit, dass
   (i) sich keiner beschwert?
   (ii) sich höchstens zwei beschweren?
   (iii) sich höchstens einer nicht beschwert?

Lösungsvorschlag:

(a) Situation: Urnenmodell mit insgesamt 20 Kugeln. Davon sind 5 Rot und 15 schwarz. Wir ziehen daraus drei Kugeln ohne zurücklegen. Die Summe der gezogenen **roten** Kugeln ist dann hypergeometrisch verteilt mit

$$X \sim \text{Hyp}(n, r, s) \;=\; \text{Hyp}(3, 5, 15)$$

(i) $$P(X = 0) \;=\; \frac{\binom{r}{x} \cdot \binom{s}{n-x}}{\binom{r+s}{n}} \;=\; \frac{\binom{5}{0} \cdot \binom{15}{3}}{\binom{20}{3}} \;=\; \frac{1 \cdot 455}{1140} \;=\; 0.399$$

Beachte Binomialkoeffizienten: $\binom{n}{k} \;=\; \frac{n!}{k! \cdot (n-k)!}$

(ii) $$P(X \leq 2) \;=\; P(X = 0) + P(X = 1) + P(X = 2) \;=\; 0.399 + \frac{\binom{5}{1} \cdot \binom{15}{3-1}}{\binom{20}{3}} + \frac{\binom{5}{2} \cdot \binom{15}{3-2}}{\binom{20}{3}}$$

$$= \; 0.399 + \frac{5 \cdot 105}{1140} + \frac{10 \cdot 15}{1140} \;=\; 0.399 + 0.461 + 0.132 \;=\; 0.992$$

(iii) $P(X \geq 2) \;=\; P(X=2) + P(X=3) \;=\; 0.132 + \dfrac{\binom{5}{3} \cdot \binom{15}{3-3}}{\binom{20}{3}} \;=\; 0.132 + \dfrac{10 \cdot 1}{1140}$

$= \; 0.132 + 0.009 \; = \; 0.141$

Abgesehen vom Rundungsfehler hätte man $P(X = 3)$ natürlich auch direkt über $1 - 0.992$ ausrechnen können, da $\{0, 1, 2, 3\}$ der gesamte Träger von X ist. Besonders bei größeren (n, r, s) ist die Hypergeometrische Verteilung durchaus rechenintensiv. Gängige Programmiersprachen haben die Verteilungen natürlich implementiert. In `R` lässt sich die Zähldichte über `dhyper(x, r, s, n)` für $x = 0, 1, 2, 3$ berechnen. Die Reihenfolge der Parameter variiert je nach Software und muss beachtet werden.

(b) Da hier keine Summe an Angestellten vorgegeben ist, gehen wir von einer großen Zahl aus, sodass der Unterschied zwischen '*mit Zurücklegen*' und '*ohne Zurücklegen*' verschwindend gering ist (für $n \to \infty$ vollständig verschwindet). Ein Urnenmodell mit Zurücklegen lässt sich über die Binomialverteilung modellieren. Es werden drei Kugeln mit Zurücklegen gezogen. Bei jedem Zug ist die Wahrscheinlichkeit für eine rote Kugel $\frac{1}{4}$ und für eine schwarze Kugel $\frac{3}{4}$. Die Summe gezogenen roten Kugeln ist dann binomial verteilt mit

$$X \sim \text{Bin}(n, p) \;=\; \text{Bin}\left(3,\; \frac{1}{4}\right)$$

(i) $P(X = 0) \;=\; \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x} \;=\; \binom{3}{0} \cdot 0.25^0 \cdot 0.75^3 \;=\; 1 \cdot 1 \cdot 0.422 \;=\; 0.422$

(ii) $P(X \leq 2) \;=\; P(X=0) + P(X=1) + P(X=2) \;=\; 1 - P(X=3)$

$= \; 1 - \left(\binom{3}{3} \cdot 0.25^3 \cdot 0.75^0\right) \;=\; 1 - \left(1 \cdot \frac{1}{64} \cdot 1\right) \;=\; 1 - \frac{1}{64} \;=\; \dfrac{63}{64}$

(iii) $P(X \geq 2) \;=\; P(X=2) + P(X=3) \;=\; P(X=2) + \dfrac{1}{64} \;=\; \binom{3}{2} \cdot 0.25^2 \cdot 0.75^1 + \dfrac{1}{64}$

$= \; 3 \cdot 0.0625 \cdot 0.75 + \dfrac{1}{64} \;=\; 0.156$

Für $P(X \leq x)$ muss man softwareseitig natürlich nicht jeden Wert über die Zähldichte berechnen (`dyhper(x, r, s, n)` oder `dbinom(x, n, p)`), sondern kann auch direkt die Verteilungsfunktion bestimmen (`phyper(q, r, s, n)` oder `pbinom(q, n, p)`).

Aufgabe 20: (per Hand)

Betrachten Sie die Zufallsvariable X mit dem Träger $T = \{0, 1, 2\}$. Es gelte

$$p_X(0) = p_X(1) = 1/4, \qquad p_X(2) = 1/2$$

(a) Berechnen Sie die Verteilung von $Y = |X - 1|$.

(b) Bestimmen Sie die bivariate Zähldichte und die bivariate Verteilungsfunktion von (X, Y).

(c) Bestimmen Sie die Erwartungswerte und die Varianzen von X und Y.

Lösungsvorschlag:

(a) Bestimme zuerst alle möglichen Realisierungen:

X	0	1	2
Y	1	0	1
p	0.25	0.25	0.5

$$p_Y(y) = \begin{cases} 0.25, & \text{für } y = 0 \\ 0.75, & \text{für } y = 1 \\ 0, & \text{sonst} \end{cases} \qquad \text{und} \qquad F_Y(y) = \begin{cases} 0, & \text{für } y < 0 \\ 0.25, & \text{für } 0 \leq y < 1 \\ 1, & \text{für } y \geq 1 \end{cases}$$

(b) Auch bivariat existieren nur diese drei Realisierungsmöglichkeiten. Alle anderen Kombinationen haben die Wahrscheinlichkeit 0. Daher:

$p_{X,Y}$		Y = 0	Y = 1
X	0	0	0.25
	1	0.25	0
	2	0	0.5

und

$F_{X,Y}$		Y = 0	Y = 1
X	0	0	0.25
	1	0.25	0.5
	2	0.25	1

(c)

$$\begin{aligned}
E(X) &= 0.25 \cdot 0 + 0.25 \cdot 1 + 0.5 \cdot 2 && = 1.25 \\
E(Y) &= 0.25 \cdot 0 + 0.75 \cdot 1 && = 0.75 \\
\mathrm{Var}(X) &= 0.25 \cdot (0 - 1.25)^2 + 0.25 \cdot (1 - 1.25)^2 + 0.5 \cdot (2 - 1.25)^2 && = 0.6875 \\
\mathrm{Var}(Y) &= 0.25 \cdot (0 - 0.75)^2 + 0.75 \cdot (1 - 0.75)^2 && = 0.1875
\end{aligned}$$

Aufgabe 21: (per Hand)

Ein neuer Algorithmus wird vom Erfinder damit beworben, kompromittierte Einträge und Teile von großen Datenbanken zu finden. Der Algorithmus arbeitet selbstständig und gibt für alle Sektionen einer Datenbank die Warnstufen:

- Rot - Kompromittierung vermutet
- Gelb - Kompromittierung möglich
- Grün - Kompromittierung ausgeschlossen

Das Unternehmen *Datenbanken, Datenbanken, Datenbanken, Datenbanken* benutzt den Algorithmen schon länger. Wenn der Algorithmus Rot und Gelb bewertet, muss ein Mitarbeiter die Sektion umfangreich prüfen. Nachdem der Algorithmus Rot meldete, ist erfahrungsgemäß in 85% der Fälle tatsächlich eine Kompromittierung vorhanden. Nach einer Warnung Gelb entpuppen sich hingegen 75% als falscher Alarm. Bei Grünen Meldungen wird keine manuelle Prüfung durchgeführt. Es wird jedoch davon ausgegangen, dass die Kompromittierungswahrscheinlichkeit nach einer Bewertung Grün bei 0.1% liegt.

Insgesamt hat das Unternehmen festgestellt, dass neue Datenbanksektionen mit Wahrscheinlichkeit 2% als Rot und 7% als Gelb kategorisiert werden.

(a) Notieren Sie alle angegebenen Wahrscheinlichkeiten mathematisch angemessen.

(b) Wie groß ist die Wahrscheinlichkeit, dass eine unbekannte und neue Datenbanksektion eine Kompromittierung aufweist?

(c) Wie groß ist die Wahrscheinlichkeit, dass bei einer vorliegenden Kompromittierung der Algorithmus die Warnstufe Rot ausgibt?

Lösungsvorschlag:

Sei zunächst $K :=$ kompromittiert, $R :=$ Rot, $Y :=$ Gelb, $G :=$ Grün. Aus dem Text erhalten wir folgende Wahrscheinlichkeiten. Die jeweiligen Gegenwahrscheinlichkeiten notieren wir parallel:

(a)

$$\begin{aligned} P(K|R) &= 0.85 & \qquad P(K^C|R) &= 0.15 \\ P(K|Y) &= 0.25 & P(K^C|Y) &= 0.75 \\ P(K|G) &= 0.001 & P(K^C|G) &= 0.999 \\ P(R) &= 0.02 & P(Y) &= 0.07 \\ P(G) &= 0.91 \end{aligned}$$

(b) Gesucht: $P(K)$. Über den Satz der totalen Wahrscheinlichkeit:

$$\begin{aligned} P(K) = & \; P(K|R) \cdot P(R) + P(K|Y) \cdot P(Y) + P(K|G) \cdot P(G) \\ = & \; 0.85 \cdot 0.02 + 0.25 \cdot 0.07 + 0.001 \cdot 0.91 \\ = & \; 0.03541 \end{aligned}$$

(c) Gesucht: $P(R|K)$. Über den Satz von Bayes:

$$P(R|K) = \frac{P(K|R) \cdot P(R)}{P(K)} = \frac{0.85 \cdot 0.02}{0.03541} = 0.480$$

Aufgabe 22: (per Hand)

Zum Thema *Wichtige Wahrscheinlichkeitsverteilungen*:
Überlegen Sie sich zu jeder der folgenden diskreten Wahrscheinlichkeitsverteilungen ein anschauliches Beispiel. Geben Sie dazu an, was die jeweiligen Parameter in Ihrem Beispiel bedeuten würden.

(a) $X \sim \mathrm{G}(x_1, x_2, x_3)$

(b) $X \sim \mathrm{Bin}(n, p)$

(c) $X \sim \mathrm{Hyp}(n, r, s)$

(d) $X \sim \mathrm{Geo}(p)$

(e) $X \sim \mathrm{Poi}(\lambda)$

Lösungsvorschlag:

(a) Diskrete Gleichverteilung:
X könnte das Ergebnis eines Würfelwurfs sein, wobei die Ereignisse $x_1 = \{1, 2\}$, $x_2 = \{3, 4\}$ und $x_3 = \{5, 6\}$ seien. Wichtig ist, dass die Ereignisse x_1, x_2, x_3 alle mit identischen Wahrscheinlichkeiten auftreten.

(b) Binomialverteilung:
X könnte die Summe von *erwischten* grünen Ampeln während einer Fahrt sein. Dabei wäre n die Gesamtzahl an passierten Ampeln und p die Wahrscheinlichkeit, bei einer Ampel grün zu erwischen, also nicht halten zu müssen. Wichtig dabei: Es wurde allen Ampeln dieselbe Wahrscheinlichkeit p zugeordnet.

(c) Hypergeometrische Verteilung:
Wenn wir im Hörsaal eine Stichprobe der Größe n ziehen (ohne Zurücklegen), könnte X die Zahl der gezogenen Informatiker sein. Dabei wäre r die Zahl der Informatiker im Hörsaal insgesamt und s die Zahl der übrigen Anwesenden.

(d) Geometrische Verteilung:
X könnte die Anzahl an fehlgeschlagenen Klausurversuchen sein, die ein Student benötigt, eine bestimmte Klausur zu bestehen. Dabei wäre p die Wahrscheinlichkeit für jeden einzelnen Versuch zu bestehen. (Eine maximale Anzahl an Versuchen wird dabei außer Acht gelassen).

(e) Poisson-Verteilung:
X könnte die Anzahl an Toren sein, die ein Fußballspieler pro Saison schießt. Dabei wäre λ der Erwartungswert der geschossenen Tore.

Der letzte Übungstermin am **31.01.2019** wie üblich im EF50-HS1 wird als Fragestunde stattfinden. Dabei können zuerst Fragen im Stil einer Globalübung, anschließend in kleineren Gruppen gestellt werden. Beide Lehrenden stehen dafür zur Verfügung.