Wahrscheinlichkeitstheorie und Statistik

Lösungen Serie 8

Version 1 (30. April 2022)

Bitte stellt Fragen in den Übungen und/oder im Forum des Moodle-Kurs (und/oder (anonym) in diesem file https://docs.google.com/document/d/1CfTkwrNOhTKB8y8cVQW_deUahCejTbQWlm_BzrHIFYA/edit?usp=sharing)

Bitte stelle sicher, das du die Webseite https://kahoot.it/ in der Übung am **09. Mai** öffnen kannst (zB auf deinem smartphone).

Wir empfehlen die Aufgaben selbständig zu lösen und dann im Fach der entsprechenden Übungsgruppe im Raum HG G 53 abzugeben oder selbst mit dieser Lösung zu vergleichen am besten rechtzeitig vor der Übung am **09. Mai**.

Vergiss nicht am 16. Mai pünktlich um spätestens 13:15 in deine Übungsgruppe zur Lernkontrolle zu kommen im gleichen Raum wo auch die Lernkontrolle 1 war (wo sonst immer die Übungsgruppen statt finden, beziehungsweise in LFW C4 für Gruppe 6).

Aufgabe 8.1

(a) Vereinfachen Sie den Ausdruck $\sum_{j=0}^{k}\binom{m}{k-j}\binom{n}{j}$, wobei $m, n, k \geq 1$.

HINWEIS: Benützen Sie die Identität $(1+x)^{m+n} = (1+x)^m(1+x)^n$, $x \in \mathbb{R}$.

(b) Seien X und Y zwei binomial verteilte Zufallsvariablen mit Parametern (m,p) bzw. (n,p). Angenommen X und Y sind unabhängig, was ist die Verteilung von $X+Y$? Leiten sie dieses Resultat mithilfe von (a) her ohne Remark 2.13 aus dem Skript zu verwenden.

(c) Wir betrachten eine Urne mit 5 schwarzen und 5 weissen Kugeln. Wir ziehen $n \geq 1$ Kugeln mit Zurücklegen. Sei X' die Anzahl weisser Kugeln, welche gezogen wurden, und sei Y' die Anzahl schwarzer Kugeln, welche gezogen wurden.

Was sind die Verteilungen von X', Y' und $X'+Y'$? Widerspricht dies dem Resultat aus (b)?

Lösung 8.1

(a) Für $n \in \mathbb{N}$ gilt $(x+y)^n = \sum_{j=0}^{n}\binom{n}{j}x^j y^{n-j}$. Für $y=1$ erhalten wir

$$(1+x)^n = \sum_{j=0}^{n}\binom{n}{j}x^j.$$

Also gilt

$$\begin{aligned}
\sum_{k=0}^{m+n}\binom{m+n}{k}x^k &= (1+x)^{m+n} = (1+x)^m(1+x)^n \\
&= \left(\sum_{k=0}^{m}\binom{m}{k}x^k\right)\left(\sum_{j=0}^{n}\binom{n}{j}x^j\right) = \sum_{k=0}^{m}\sum_{j=0}^{n}\binom{m}{k}\binom{n}{j}x^{k+j} \\
&= \sum_{k=0}^{m+n}\left(\sum_{j=0}^{n}\binom{m}{k-j}\binom{n}{j}\right)x^k = \sum_{k=0}^{m+n}\left(\sum_{j=0}^{k}\binom{m}{k-j}\binom{n}{j}\right)x^k,
\end{aligned}$$

wobei wir für $j > n$, $\binom{n}{j} = 0$ setzen. Einen Koeffizientenvergleich liefert die Gleichung $\sum_{j=0}^{k}\binom{m}{k-j}\binom{n}{j} = \binom{m+n}{k}$.

BEMERKUNG: Die Gleichung wird Identität von Vandermonde genannt. Man kann die Gleichung auch kombinatorisch zeigen: wir betrachten $m+n$ Objekte und unterteilen sie in zwei Gruppen der Grössen m und n. Für $k \geq 1$ gibt es $\binom{m+n}{k}$ Möglichkeiten, k Objekte aus den $m+n$ Objekten zu wählen. Von den k Objekten sind j davon aus der Gruppe mit n Objekten und $k-j$ davon aus der Gruppe mit m Objekten. Die Anzahl Möglichkeiten k Objekte auf diese Weise aus den beiden Gruppen zu wählen, beträgt $\binom{m}{k-j}\binom{n}{j}$. Summieren wir über alle möglichen Werte von j, erhalten wir die gewünschte Identität.

(b) Weil die Mengen $\{Y=j\}$ für $j \in \mathbb{N}$ disjunkt sind, gilt für $k \leq m+n$,

$$\begin{aligned}
\mathbb{P}[X+Y=k] &= \sum_{j=0}^{\infty} \mathbb{P}[X+Y=k, Y=j] \\
&\overset{\substack{\{Y+X=k,Y=j\}=\varnothing \\ \text{für } j>k}}{=} \sum_{j=0}^{k} \mathbb{P}[X=k-j, Y=j] \\
&\overset{\text{Unabh.}}{=} \sum_{j=0}^{k} \binom{m}{k-j} p^{k-j}(1-p)^{m-k+j} \binom{n}{j} p^{j}(1-p)^{n-j} \\
&= \sum_{j=0}^{k} \binom{m}{k-j}\binom{n}{j} p^{k}(1-p)^{m+n-k} \\
&\overset{\text{(a)}}{=} \binom{m+n}{k} p^{k}(1-p)^{m+n-k}.
\end{aligned}$$

Also is $X+Y$ wieder binomial verteilt mit Parameter $(m+n,p)$.

Bemerkung 1: Alternativ kann das Resultat auch folgendermassen basierend auf Proposition 2.12 aus dem Skript bewiesen werden: Seien $X_1+\ldots+X_m, Y_1, \ldots Y_n$ i.i.d. Bernoulli verteilte Zufallsvariablen mit Parameter p.[1] Definiere $\tilde{X} := X_1+\ldots+X_m$ und $\tilde{Y} = Y_1+\ldots+Y_n$. Dann ist die Summe $\tilde{X}+\tilde{Y} = X_1+\ldots+X_m+Y_1+\ldots+Y_n$ binomial verteilt mit Parametern $(m+n,p)$ laut Proposition 2.12. Ausserdem ist wegen Proposition 2.12, $\tilde{X} = X_1+\ldots+X_m$ eine binomial verteilte Zufallsvariable mit Parametern (m,p) und $\tilde{Y} = Y_1+\ldots+Y_n$ binomial verteilt mit Parametern (n,p). Die beiden Zufallsvariablen $\tilde{X}$ und $\tilde{Y}$ sind auch unabhängig, weil $X_1+\ldots+X_m, Y_1, \ldots Y_n$ unbahängig sind. Jetzt muss man nur noch beweisen, dass die Verteilung von $X+Y$ gleich der Verteilung von $\tilde{X}+\tilde{Y}$ ist:

$$\begin{aligned}
\mathbb{P}[X+Y=k] &= \sum_{i+j=k} \mathbb{P}[X=i, Y=j] = \sum_{i+j=k} \mathbb{P}[X=i]\,\mathbb{P}[Y=j] = \sum_{i+j=k} \tilde{\mathbb{P}}[\tilde{X}=i]\tilde{\mathbb{P}}[\tilde{Y}=j] \\
\tilde{\mathbb{P}}[\tilde{X}+\tilde{Y}=k] &= \sum_{i+j=k} \tilde{\mathbb{P}}[\tilde{X}=i, \tilde{Y}=j] = \sum_{i+j=k} \tilde{\mathbb{P}}[\tilde{X}=i]\tilde{\mathbb{P}}[\tilde{Y}=j].
\end{aligned}$$

Bemerkung 2: Intuitiv kann die Essenz von Bemerkung 1 auch folgendermassen zusemmegefasst werden, auch wenn es dann kein vollständiger Beweis mehr ist und mathematisch teilweise nicht ganz korrekt formuliert ist (siehe Fußnoten 2–5): Wir schreiben X als Summe $X = X_1+\ldots+X_m$, wobei die Summanden X_i unabhängig Bernoulli verteilt sind mit Parameter p.[2] Ebenso schreiben wir Y als Summe $Y = Y_1+\ldots+Y_n$, wobei die Summanden Y_i ebenfalls unabhängig Bernoulli verteilt sind mit Parameter p.[4] Weil X und Y unabhängig sind, sind auch die Summanden X_i und Y_j alle unabhängig.[5] Daher ist $X+Y = X_1+\ldots+X_m+Y_1+\ldots+Y_n$ wieder binomial verteilt mit Parametern $(m+n,p)$.

[1] Wir können die Existenz von $X_1+\ldots+X_m, Y_1, \ldots Y_n$ leicht auf einem Wahrscheinlichkeitsraum $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ zeigen (z.B. ein Laplace-Modell auf $\tilde{\Omega} = \{0,1\}^{m+n}$) auch wenn X und Y möglicherweise auf einem anderen (beliebig unpraktischen) Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ definiert sind.

[2] Die Existenz von i.i.d. Bernoulli-Zufallsvariablen, die $X = X_1+\ldots+X_m$ ist für eine allgemeine binomialverteilte Zufallsvariable $X : \Omega \to \mathbb{R}$ nicht immer gegeben[3], aber man könnte beweisen, dass man jedes Problem einen anderen Wahrscheinlichkeitsraum übertragen kann, wo die Existenz gegeben ist. Dieser Beweis wäre ziemlich aufwendig obwohl die Aussage rein intuitiv so trivial klingt.

[3] Die Zufallsvariablen X und Y könnten auf einem Ω definiert sein mit $|\Omega| = mn$ und trotzdem alle in der Angabe geforderten Eigenschaften erfüllen, aber für die Existenz von i.i.d. Bernoulli-Zufallsvariablen $X_1+\ldots+X_m$ mit $p \in (0,1)$ benötigt man $|\Omega| \geq 2^m$.

[4] Siehe Fußnote 2

[5] Dieser "Weil"-Satz ist mathematisch falsch (die Implikation gilt im allgemeinen nur in die andere Richtung). Was hier eigentlich gemeint ist: Weil X und Y unabhängig sind, können die Bernoulli Zufallsvariablen $X_1+\ldots+X_m, Y_1, \ldots Y_n$ so gewählt werden, dass sie i.i.d. sind, wobei selbst diese Aussage nicht exakt richtig ist sondern

(c) Man sieht leicht, dass X' und Y' beide binomial verteilt sind mit Parametern $(n, 1/2)$. Natürlich gilt auch $X' + Y' = n$. Also ist $X' + Y'$ konstant und hat die Verteilung

$$\mathbb{P}\left[X' + Y' = k\right] = \begin{cases} 1 & \text{falls } k = n, \\ 0 & \text{sonst.} \end{cases}$$

Dies widerspricht dem Resultat aus (b) **nicht**, da X' und Y' nicht unabhängig sind.

Aufgabe 8.2 Ein Fabrikant verwendet Komponenten A, B, C um Chips herzustellen. Ein Chip wird aus einer Komponente A und einer Komponente B oder aus einer Komponente A und einer Komponente C hergestellt. Beide Möglichkeiten sind gleich wahrscheinlich. Die Komponenten A, B, C haben respektiv X, Y, Z Fehlerstellen. Wir nehmen an, dass X, Y, Z unabhängig und Poisson verteilt sind mit respektiven Parametern λ, μ und 2μ. Sei N die Anzahl Fehlerstellen in einem Chip.

(a) Berechnen Sie explizit die Verteilungen von $X + Y$ und $X + Z$.

(b) Berechnen Sie die Wahrscheinlichkeit, dass ein Chip die Komponente B enthält, falls $N = n$, für $n \geq 0$.

(c) Berechnen Sie den Erwartungswert von N.

Lösung 8.2

(a) Für $k \geq 0$ gilt

$$\begin{aligned}
\mathbb{P}[X + Y = k] &= \sum_{j=0}^{k} \mathbb{P}[X = k - j,\, Y = j] \overset{\text{Unabh.}}{=} \sum_{j=0}^{k} \mathbb{P}[X = k - j]\, \mathbb{P}[Y = j] \\
&= \sum_{j=0}^{k} e^{-\lambda} \frac{\lambda^{k-j}}{(k-j)!}\, e^{-\mu} \frac{\mu^j}{j!} = \frac{e^{-(\lambda+\mu)}}{k!} \sum_{j=0}^{k} \binom{k}{j} \lambda^{k-j} \mu^j \\
&= e^{-(\lambda+\mu)} \frac{(\lambda+\mu)^k}{k!},
\end{aligned}$$

d.h. $X + Y$ ist Poisson verteilt mit Parameter $\lambda + \mu$. Analog findet man, dass $X + Z$ Poisson verteilt ist mit Parameter $\lambda + 2\mu$.

Bemerkung: Allgemein gilt für zwei unabhängige Poisson verteilte Zufallsvariablen X und Y mit respektiven Parametern λ und μ, dass $X + Y$ wieder Poisson verteilt ist mit Parameter $\lambda + \mu$.

(b) Definiere das Ereignis $D = \{\text{Chip enthält Komponente } B\}$. Dann gilt mit dem Satz von Bayes

$$\begin{aligned}
\mathbb{P}[D | N = n] &= \frac{\mathbb{P}[N = n | D]\; \mathbb{P}[D]}{\mathbb{P}[N = n | D]\; \mathbb{P}[D] + \mathbb{P}[N = n | D^c]\; \mathbb{P}[D^c]} \\
&= \frac{\mathbb{P}[X + Y = n] \cdot \frac{1}{2}}{\mathbb{P}[X + Y = n] \cdot \frac{1}{2} + \mathbb{P}[X + Z = n] \cdot \frac{1}{2}} \\
&= \frac{1}{1 + e^{-\mu} \left(\frac{\lambda + 2\mu}{\lambda + \mu}\right)^n}.
\end{aligned}$$

(c) Es gilt $N = (X + Y)\mathbb{1}_D + (X + Z)\mathbb{1}_{D^c}$ und wegen der Unabhängigkeit von $X + Y$ und D (sowie von $X + Z$ und D) erhalten wir

$$\begin{aligned}
\mathbb{E}[N] &= \mathbb{E}[X + Y]\, \mathbb{P}[D] + \mathbb{E}[X + Z]\, \mathbb{P}[D^c] \\
&= \mathbb{E}[X + Y] \cdot \tfrac{1}{2} + \mathbb{E}[X + Z] \cdot \tfrac{1}{2} = \tfrac{1}{2}(\lambda + \mu + \lambda + 2\mu) = \lambda + \tfrac{3}{2}\mu.
\end{aligned}$$

gegebenenfalls einen Wechsel des Wahrscheinlichkeitsraumes benötigt wie in Fußnote 2 bereits erwähnt wurde. Der Beweis wäre ziemlich aufwendig (ähnlich wie Fußnote 2) obwohl die Aussage rein intuitiv so trivial klingt. Dieser Beweis ist um so vieles Aufwendiger als der aus Bemerkung 2, weil hier versucht wird direkt die möglicherweise sehr komplizierte Zufallsvaribale X als Summe von i.i.d. Bernoulli-ZV zu beschreiben. In Bemerkung 1 wird der Beweis für die ZV $\tilde{X}$ durchgeführt, welche eine einfachere Struktur hat, da sie per Definition als Summe von i.i.d. Bernoulli-ZV beschrieben werden kann. Und nur ganz am Ende des Beweise verwenden wir, dass die Verteilung von $X + Y$ nicht von der genauen Definition von $X : \Omega \to \mathbb{R}$ und $Y : \Omega \to \mathbb{R}$ abhängt sondern nur von der Verteilung gemeinsamen Verteilung von X und Y. Somit kann der Beweis auch für $\tilde{X}$ und $\tilde{Y}$ durchgeführt werden solange $\tilde{\mathbb{P}}[\tilde{X} = i, \tilde{Y} = j] = \mathbb{P}[X = i, Y = j]$ für alle i, j gilt.

Aufgabe 8.3 Wir betrachten einen Kreis mit zufälligem Radius R. Der Radius R sei exponentialverteilt mit Erwartungswert $1/\lambda$. Bestimmen Sie

(a) die Verteilungs- und Dichtefunktion des Flächeninhalts A des zufälligen Kreises;

(b) den Erwartungswert von A.

Lösung 8.3

(a) Sei X exponentialverteil mit Parameter μ, d.h. die Dichte von X ist $f_X(x) = \mu e^{-\mu x}$ für $x \geq 0$ und 0 sonst. Es folgt, dass

$$\mathbb{E}[X] = \int_0^\infty x\mu e^{-\mu x} dx = 1/\mu.$$

Also ist R exponentialverteilt mit Parameter $\mu = \lambda$.
Der Flächeninhalt des Kreises mit Radius R ist gegeben durch die Zufallsvariable $A = \pi R^2$. Die Verteilungsfunktion von A ist

$$F_A(x) = \mathbb{P}[A \leq x] = \mathbb{P}\left[R \leq \sqrt{x/\pi}\right] = \int_0^{\sqrt{x/\pi}} \lambda e^{-\lambda t} dt = 1 - e^{-\lambda\sqrt{x/\pi}}, \quad \text{falls } x \geq 0,$$

und 0 sonst. Die Dichtefunktion ist dann gegeben durch $f_A(x) = \frac{d}{dx} F_A(x) = \frac{\lambda}{2\sqrt{\pi x}} e^{-\lambda\sqrt{x/\pi}}$ falls $x \geq 0$ und 0 sonst.

(b)

$$\mathbb{E}[A] = \mathbb{E}\left[\pi R^2\right] = \int_0^\infty \pi t^2 f_R(t) dt = \pi\lambda \int_0^\infty t^2 e^{-\lambda t} dt \overset{(*)}{=} \frac{2\pi}{\lambda^2}.$$

(*) partielle Integration (2 Mal).

Bemerkung: Es ist auch möglich, den Erwartungswert mit Hilfe der Dichte f_A aus Aufgabe (a) zu bestimmen.

Aufgabe 8.4 Eine Zufallsvariable X habe die Dichtefunktion

$$f(x) = \begin{cases} \dfrac{c}{(1+x)^5}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

(a) Finden Sie c und die Verteilungsfunktion von X.

(b) Finden Sie $\mathbb{E}[X]$ und $\mathbb{E}\left[X^2\right]$.
Hinweis: Berechnen Sie zuerst $\mathbb{E}[1+X]$ und $\mathbb{E}\left[(1+X)^2\right]$.

(c) Was sind die Verteilungsfunktion und die Dichte von $Y = e^X$?

Lösung 8.4

(a) Wir setzen $\int_{-\infty}^\infty f(x)dx = 1$, also

$$\int_0^\infty \frac{c}{(1+x)^5} dx = c\left[-\frac{1}{4}(1+x)^{-4}\right]_0^\infty = \frac{c}{4} = 1,$$

und bestimmen so $c = 4$.
Die Verteilungsfunktion ist gegeben durch

$$\begin{aligned} F_X(x) = \mathbb{P}[X \leq x] &= \int_{-\infty}^x f(y)dy \\ &= \int_0^x \frac{4}{(1+y)^5} dy = \left[-(1+y)^{-4}\right]_0^x = 1 - \frac{1}{(1+x)^4}, \quad \text{für } x \geq 0, \end{aligned}$$

und 0 sonst.

(b) Wir berechnen zunächst

$$\mathbb{E}[1+X] = \int_{-\infty}^{\infty}(1+x)f(x)dx = \int_0^{\infty}\frac{4}{(1+x)^4}dx = 4\Big[-\frac{1}{3}(1+x)^{-3}\Big]_0^{\infty} = \frac{4}{3},$$
$$\mathbb{E}\left[(1+X)^2\right] = \int_0^{\infty}\frac{4}{(1+x)^3}dx = 4\Big[-\frac{1}{2}(1+x)^{-2}\Big]_0^{\infty} = 2.$$

Damit erhalten wir

$$\mathbb{E}[X] = \mathbb{E}[1+X]-1 = \frac{1}{3}, \text{ und}$$
$$\mathbb{E}\left[X^2\right] = \mathbb{E}\left[(1+X)^2\right] - 2\mathbb{E}[X] - 1 = 2 - \frac{2}{3} - 1 = \frac{1}{3}.$$

(c) Für $y < 1$ gilt für die Verteilungsfunktion F_Y von Y,

$$F_Y(y) = \mathbb{P}[Y \leq y] = \mathbb{P}\left[e^X \leq y\right] \leq \mathbb{P}\left[e^X < 1\right] = \mathbb{P}[X < 0] = 0.$$

Für $y \geq 1$ erhält man

$$F_Y(y) = \mathbb{P}\left[e^X \leq y\right] = \mathbb{P}[X \leq \log y] = F_X(\log y) = 1 - \frac{1}{(1+\log y)^4}.$$

Durch Differenzieren der Verteilungsfunktion erhalten wir die Dichte f_Y von Y:

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \begin{cases} 0 & \text{für } y < 1, \\ \frac{4}{y(1+\log y)^5} & \text{für } 1 \leq y. \end{cases}$$

Wenn du Feedback zum Übungszettel hast, schreibe bitte eine Mail an Jakob Heiss.