

7. GESETZ DER GROSSEN ZAHLEN, ZENTRALER GRENZWERTSATZ, POISSONAPPROXIMATION

(7.0) Überblick

In diesem Kapitel werden wir drei Theoreme behandeln, bei denen in der exakten, mathematischen Formulierung die Anzahl der involvierten Versuche oder Zufallsgrößen gegen unendlich geht, $n \rightarrow \infty$. Für die Anwendungen heisst das oft “ n gross”. “ n gross” heisst aber auch: es kostet Zeit und Geld.

Es gehört zur Allgemeinbildung von Akademikerinnen und Akademikern, zu wissen, was das “Gesetz der grossen Zahlen” und “Zentraler Grenzwertsatz” aussagen.

7.1. DAS GESETZ DER GROSSEN ZAHLEN

(7.1.0) Überblick

Sei X_1, X_2, \dots eine Folge von unabhängigen, identisch verteilten (englisch: independent and identically distributed (iid)) Zufallsgrößen mit Erwartungswert μ (und $E[|X_1|] < \infty$). Dann genügt diese Folge dem Gesetz der grossen Zahlen (englisch: Law of Large Numbers LLN), das heisst, es gilt für jedes $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| < \varepsilon \right] = 1.$$

Einsatz: Schätzen von Parametern; siehe auch Kapitel 8, 9 und 10, weil $\frac{1}{n} \sum_{i=1}^n X_i$ (Theorie) bzw. $\frac{1}{n} \sum_{i=1}^n x_i$ (Daten) ein zentral wichtiger Schätzer für den Erwartungswert μ ist.

(7.1.1) Konvergenz relativer Häufigkeiten gegen Wahrscheinlichkeiten

Eine einfache Form des Gesetzes der grossen Zahlen stellt eine Beziehung zwischen relativen Häufigkeiten und Wahrscheinlichkeiten her. Wir illustrieren dies zuerst an einem einfachen Beispiel. In (7.1.2) werden wir das verallgemeinern.

Bei einem ehrlichen Würfel ist die Wahrscheinlichkeit für eine Sechs gleich $p = 1/6$. Nun werfen wir diesen Würfel n -mal und S_n^* sei die Anzahl der Sechsen unter diesen n Würfeln. Die Zufallsgrösse S_n/n beschreibt dann die relative Häufigkeit der Sechs. Ist nämlich beispielsweise $n = 120$ und nimmt S_n den Wert 15 an, so ist $S_n/n = 15/120 = 0.125$ die angegebene relative Häufigkeit. Aufgrund unserer Vorstellung als idealisierte relative Häufigkeit erwarten wir nun, dass diese Zufallsgrösse S_n/n einen Wert nahe bei $p = 1/6$ annimmt, und zwar sollte diese Annäherung umso besser sein, je grösser n ist. Diese Vorstellung lässt sich nun mathematisch präzise formulieren.

Wir gehen nun allgemeiner davon aus, dass wir einen Versuch n -mal wiederholen, wobei jedes Mal ein Ereignis E mit der Wahrscheinlichkeit p eintritt und die Wiederholungen unabhängig voneinander sind. Die Zufallsgrösse S_n gibt an, wie oft das Ereignis E eintritt. Sie ist binomial verteilt mit den Parametern n und p . Ferner sei $\varepsilon > 0$ eine (beliebig kleine) positive Zahl. Dann gilt (der Beweis folgt unten)

$$(1) \quad P \left(\left| \frac{S_n}{n} - p \right| < \varepsilon \right) \geq 1 - \frac{p(1-p)}{\varepsilon^2 n}.$$

Da die rechte Seite der Ungleichung (bei festem ε und p) mit wachsendem n gegen 1 strebt, kann die Aussage (1) auch so formuliert werden: Für alle $\varepsilon > 0$ ist

$$(2) \quad \lim_{n \rightarrow \infty} P \left(\left| \frac{S_n}{n} - p \right| < \varepsilon \right) = 1.$$

Diese Aussage besagt in Worten: Unter den angegebenen Voraussetzungen strebt die Wahrscheinlichkeit dafür, dass die relative Häufigkeit S_n/n des Eintreffens von E und die Wahrscheinlichkeit p höchstens um ε voneinander abweichen, mit wachsendem n gegen 1.

Ein Zahlenbeispiel. Wir betrachten unseren Würfel mit $p = 1/6$ und wählen $\varepsilon = 0.05$. Die Formel (1) besagt dann

$$P \left(\left| \frac{S_n}{n} - \frac{1}{6} \right| < 0.05 \right) \geq 1 - \frac{\frac{1}{6} \cdot \frac{5}{6}}{0.05^2 \cdot n} = 1 - \frac{1}{n} \frac{2000}{36} = 1 - \frac{55.56}{n}.$$

Für $n > 556$ ist

$$P \left(\left| \frac{S_n}{n} - p \right| < 0.05 \right) \geq 1 - \frac{55.56}{556} > 0.9.$$

Wählt man also $n > 556$, so ist man also zu mindestens 90% sicher, dass sich die relative Häufigkeit um weniger als 0.05 von $1/6$ unterscheidet, d.h., im Bereich von 0.117 bis 0.217 liegt.

* S für Summe und da es im Folgenden auf die Zahl n ankommt, schreiben wir hier S_n statt wie üblich bloss X oder S .

Will man zu 99% sicher sein, so muss $n > 5556$ sein, und durch entsprechende Wahl von n lässt sich diese Wahrscheinlichkeit beliebig nahe an 100% bringen. Die Abweichung $\varepsilon = 0.05$ kann man, unter entsprechenden “Kosten” für n , ebenfalls beliebig verkleinern.

Nun zum Beweis der Formel (1). Die Zufallsgrösse S_n ist binomial verteilt. Ferner sind die Aussagen $|\frac{S_n}{n} - p| < \varepsilon$ und $|S_n - np| < n\varepsilon$ sicher gleichwertig. Weiter ist $np = \mu$, und dies ist der Erwartungswert von S_n . Es ist also

$$P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = P(|S_n - \mu| < n\varepsilon).$$

Für die Wahrscheinlichkeit rechts haben wir aber in Formel (1) von (5.3) eine Abschätzung gefunden. Wir müssen darin nur noch $n\varepsilon$ in der Form $r\sigma$ schreiben, d.h., wir müssen $r = \varepsilon n / \sigma$ setzen. Dann ist nach der erwähnten Formel

$$P(|S_n - \mu| < n\varepsilon) = P(|S_n - \mu| < r\sigma) \geq 1 - \frac{1}{r^2} = 1 - \frac{\sigma^2}{\varepsilon^2 n^2} = 1 - \frac{np(1-p)}{\varepsilon^2 n^2} = 1 - \frac{p(1-p)}{\varepsilon^2 n}.$$

Dabei wurde noch die Beziehung $\sigma^2 = npq$ für die Varianz der Binomialverteilung gebraucht.

Es sei betont, dass wir nicht bewiesen haben, dass die relative Häufigkeit h_n von E im Sinn eines Grenzwerts von Folgen gegen p strebt, d.h., dass $\lim_{n \rightarrow \infty} h_n = p$ gilt. Vielmehr wurde gezeigt, dass ein grosser Unterschied zwischen h_n und p mit wachsendem n immer unwahrscheinlicher wird. In der höheren Mathematik führt die erste Vorstellung übrigens zum “Starkes Gesetz der grossen Zahlen”; wir haben hier lediglich das “Schwache Gesetz der grossen Zahlen” untersucht.

(7.1.2) Konvergenz des arithmetischen Mittelwerts gegen den Erwartungswert

In (7.1.1) haben wir eine Beziehung zwischen relativen Häufigkeiten und Wahrscheinlichkeiten hergeleitet. Dies werden wir jetzt verallgemeinern. Dazu untersuchen wir, statt S_n/n , ab jetzt allgemeiner $\bar{X} := \frac{1}{n} \sum_{k=1}^n X_k$, weil dies das arithmetische Mittel ist.

Die Fragestellung lautet: Sei X_1, X_2, \dots eine Folge von unabhängigen, identisch verteilten (englisch: independent and identically distributed (iid)) Zufallsgrössen mit Erwartungswert μ (und $E[|X_1|] < \infty$). Was geschieht mit $\bar{X} := \frac{1}{n} \sum_{k=1}^n X_k$ falls $n \rightarrow \infty$ und welche Voraussetzungen werden wir sinnvollerweise machen? Im Beispiel des fairen Würfels, fragen wir uns vielleicht mit einer Stichprobe von n Würfeln, was mit dem arithmetischen Mittel \bar{x} geschieht, wenn n immer grösser wird. Inwiefern gilt vielleicht $\bar{x} \rightarrow \mu = 3.5$?

Wir repetieren kurz die Rechenregeln für Erwartungswert und Varianz von (5.11):

$$E(aX + bY) = aE(X) + bE(Y). \quad (\text{Erwartungswert ist linear})$$

Wegen der Linearität des Erwartungswerts gilt dann für eine Summe von X_i 's mit $E[X_i] = \mu$:

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n \mu = n\mu.$$

Für den Erwartungswert des arithmetischen Mittels erhalten wir damit

$$E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) = \mu.$$

Für die Varianz gilt bei Unabhängigkeit der Zufallsgrößen (siehe Kapitel 8 für eine exakte Definition der Unabhängigkeit von Zufallsgrößen):

$$V(aX + bY) = a^2V(X) + b^2V(Y).$$

Mit $V[X_i] = \sigma^2$ erhalten wir für die Varianz des arithmetischen Mittels damit:

$$V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{1}{n} \sigma^2.$$

Für die Standardabweichung des arithmetischen Mittels erhalten wir damit:

$$\text{sd}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{\sqrt{n}} \sigma.$$

Das heisst, \bar{X} (auch \bar{x}) wird um μ herum schwanken, mit immer kleinerer Schwankung, je grösser das n . Der Grund liegt darin, dass die Varianz (auch die Standardabweichung) mit n gegen unendlich gegen 0 geht und Varianz und Standardabweichung Masse für die Variabilität um den Erwartungswert sind. Wir wollen also eine Aussage der Art machen:

$$\frac{1}{n} \sum_{i=1}^n X_i$$

konvergiert gegen den Erwartungswert von X_1 (wir können hier, weil alle X_i die gleiche Verteilung haben, einfach X_1 schreiben). **Hier Bild mit R** Die exakte Formulierung lautet:

Theorem 7.1 [Satz von Kolmogoroff (LLN)] Sei X_1, X_2, \dots eine Folge von unabhängigen, identisch verteilten (sog. iid) Zufallsgrößen mit Erwartungswert μ (und

$E[|X_1|] < \infty$). Dann genügt diese Folge dem Gesetz der grossen Zahlen (englisch: Law of Large Numbers LLN), das heisst, es gilt für jedes $\varepsilon > 0$:

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{1}{n} \sum_{i=1}^n X_i - \mu \right| < \varepsilon \right] = 1.$$

In diesem Theorem ist es extrem wichtig, dass man *zuerst* ein $\varepsilon > 0$ wählt und *erst dann* mit n gegen unendlich geht. Den Ausdruck $P[|\frac{1}{n} \sum_{i=1}^n X_i - \mu| < \varepsilon]$ liest man zudem besser von innen nach aussen: $\frac{1}{n} \sum_{i=1}^n X_i$ ist unser interessierendes Objekt und wir nehmen dann erstmal die (absolute) Differenz zum Zielwert (Mittelwert μ): $|\frac{1}{n} \sum_{i=1}^n X_i - \mu|$. Dann untersucht man das Ereignis, dass diese (absolute) Differenz weniger als ε gross ist: $|\frac{1}{n} \sum_{i=1}^n X_i - \mu| < \varepsilon$, die Wahrscheinlichkeit dieses Ereignisses muss dann schliesslich gegen 1 gehen.

Wir betrachten die beiden zentralen Anwendungsgebiete des LLN:

1. Arithmetisches Mittel

Von der beschreibenden Statistik in Kapitel 2 kennen wir den Stichprobenmittelwert \bar{x} . Dürfen man nun den LLN anwenden? Ist die Stichprobe iid (unabhängig, identische Verteilung); wie kann man das wissen?

Wenn das Wort "Stichprobe", englisch "sample," fällt, dann ist das per definitionem immer unabhängig, ausser es wird explizit anders erwähnt. Ansonsten muss *a priori, vor dem Versuch*, die Substanzwissenschaft (hier die Naturwissenschaften, also Sie!) und nicht die Mathematik und Statistik, entscheiden, ob man Unabhängigkeit und gleiche Verteilung voraussetzen darf. *A posteriori, nach dem Versuch*, kann man als Statistiker/in dann auch untersuchen, ob es Hinweise gibt, dass die Daten doch nicht unabhängig oder identisch verteilt sind ("Model checking").

Wenn Sie dann iid als Modellannahme akzeptieren, benutzen Sie die Wortwahl: " \bar{x} konvergiert wegen Theorem 7.1 (wegen des LLN, Gesetzes der grossen Zahlen) gegen μ ; zB beim Würfel gegen 3.5."

2. Relative Häufigkeiten

Bei den relativen Häufigkeiten gibt es eine kleine Klippe, die wir umschiffen müssen: Theorem 7.1 und obiges Beispiel 1 behandeln ja die Konvergenz gegen einen *Erwartungswert* μ . Wie ist das zu verstehen, wenn man bei relativen Häufigkeiten wie dem Anteil 6er beim Würfel in (7.1.1) ebenfalls Formulierungen benutzt wie: *wegen des Gesetzes*

der grossen Zahlen konvergiert der Anteil 6er gegen $1/6$? $1/6$ und relative Häufigkeiten sind ja *Wahrscheinlichkeiten* und keine *Erwartungswerte*?

Um dieses Problem zu lösen müssen wir eine Übersetzungsleistung vollbringen, welche wir mit einem Beispiel begleiten. Wenn Sie 6000 Mal würfeln, dann erwarten Sie (wegen des LLN), dass der Anteil der 6er etwa $1/6 = 1000/6000$ ist. Dazu führen wir neue Zufallsgrössen Y_i ein und definieren $Y_i = 1$ wenn es eine 6 gibt bei Wurf i und $Y_i = 0$ wenn es keine 6 gibt bei Wurf i . Y_i ist dann natürlich $\text{Be}(1/6)$ -verteilt und hat damit Erwartungswert $1/6$: $E[Y_i] = p = 1/6$. Damit ist $\sum_{i=1}^n Y_i$ die *Anzahl* 6er; $\frac{1}{n} \sum_{i=1}^n Y_i$ ist der *Anteil* 6er. Wegen des LLN gilt nun

$$\frac{1}{n} \sum_{i=1}^n Y_i \rightarrow \mu = E[Y_1] = p = 1/6.$$

Damit ist die Verbindung zwischen (7.1.1) und (7.1.2) hergestellt: (7.1.2) beinhaltet den Fall (7.1.1).

Nebenbemerkung: Zum LLN können wir keine Übungsaufgaben stellen, weil es ein rein theoretisches Resultat ist.

Grenzen des Einsatzes des LLN: In den Naturwissenschaften, also im Labor, in der Technik, ist der Einsatz unbestritten, weil wir konstante Naturgesetze haben. Beispiel: fairer Würfel: $\bar{x} \rightarrow 3.5$ - iid ist erfüllt. Bei anderen Gebieten wie Noten von Schüler/innen, Zinsen und weiteres - also in den Sozialwissenschaften - ist "iid" oft verletzt (wegen Lernprozessen hat man im Zeitablauf nicht die gleiche Verteilung. Auch die Unabhängigkeit ist oft verletzt). Der LLN kann, oft in Abwandlungen, manchmal doch vorsichtig verwendet werden.

Bei der Diskussion um die Klimaerwärmung haben wir auch nicht iid; d.h. der LLN ist nicht anwendbar (gleiche Verteilung ist verletzt (iid) - das ist ja gerade der Punkt: die Temperatur steigt!). Hingegen darf man \bar{X} und \bar{x} immer berechnen (sobald Intervallskala, zum Beispiel Durchschnittstemperatur der 1990er Jahre), aber es ist eine falsche Erwartung, dass $\bar{x} \rightarrow \mu$ für irgendein "richtiges" μ .

7.2. ZENTRALER GRENZWERTSATZ

(7.2.0) Überblick

Sei $X_k, k \geq 1$, eine Folge von iid Zufallsgrößen mit $E[X_1] =: \mu$ und $0 < V[X_1] =: \sigma^2 < \infty$. Dann gilt für $a \in \mathbb{R}$ beliebig:

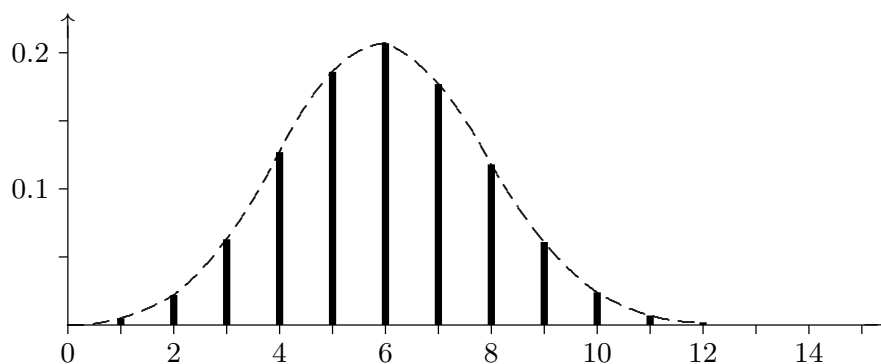
$$P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq a\right] \rightarrow P[\mathcal{N}(0, 1) \leq a]$$

wenn $n \rightarrow \infty$. Einsatz: Approximative Berechnungen, zum Beispiel für Tests; siehe auch Kapitel 8, 9, 10.

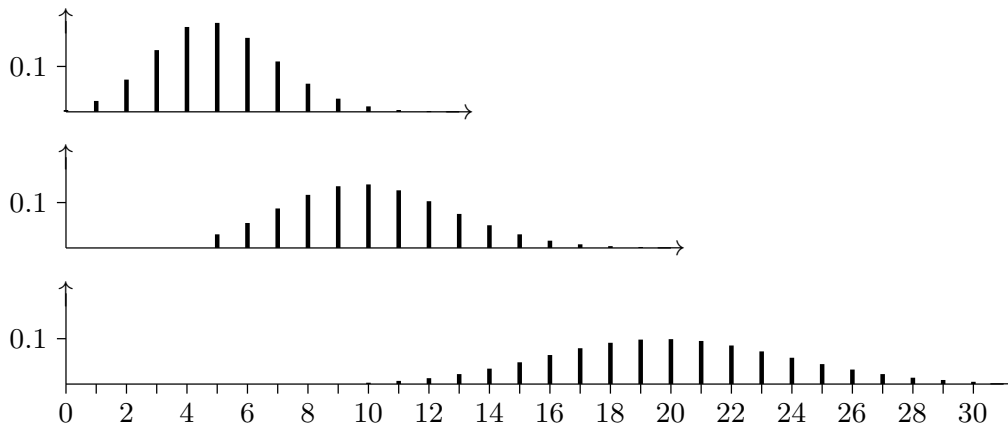
Wir beginnen in (7.2.1) mit dem Spezialfall der Binomialverteilung; in (7.2.2) folgt dann der allgemeine Fall.

(7.2.1) Approximation der Binomialverteilung durch die Normalverteilung

Zeichnet man das Stabdiagramm einer Binomialverteilung für grössere Werte von n , und verbindet man die Enden der “Stäbe”, so gewinnt man den Eindruck, dass ungefähr eine Glockenkurve entstanden sei. Dieser Eindruck wird umso besser, je grösser n ist.



Allerdings verschiebt sich diese Kurve mit wachsendem n (und bei festgehaltenem p) immer mehr nach rechts, was nicht überrascht, da der Erwartungswert einer binomial verteilten Zufallsgrösse $= np$ ist (vgl. (5.5)).



Man kann nun aber, ähnlich wie in (5.10.3) die Zufallsgrösse “standardisieren”. Sie hat dann stets den Erwartungswert 0 und die Standardabweichung 1, unabhängig von n und p . Nun kann man n gegen unendlich streben lassen, ohne dass sich die Lage der Kurve ändert.

Dies wollen wir etwas genauer durchführen. Wir wählen eine Wahrscheinlichkeit p ($0 < p < 1$) fest und bezeichnen mit S_n eine Zufallsgrösse, die einer Binomialverteilung mit den Parametern n und p folgt. Diese Zufallsgrösse hat den Erwartungswert $\mu = np$ und die Standardabweichung $\sigma = \sqrt{npq}$ (vgl. (5.5), aber Vorsicht: in (7.2.2) werden μ und σ die Kennziffern von einzelnen Zufallsgrössen sein, welche wir aufsummieren). Wir führen nun die neue Zufallsgrösse

$$Z_n = \frac{S_n - \mu}{\sigma} = \frac{S_n - np}{\sqrt{npq}}$$

ein. (In (4.1.3), Bemerkung 6), ist erwähnt worden, wie man neue Zufallsgrössen aus alten bildet.) Man kann nun (unter Verwendung von Bemerkung 5) aus (5.1) sowie Bemerkung 3) aus (5.2)) zeigen, dass für alle Werte von n die Beziehungen $E(Z_n) = 0$ und $V(Z_n) = 1$ gelten. Ist nun $F_n(x)$ die Verteilungsfunktion von Z_n , so lässt sich beweisen (was wir hier aber nicht tun wollen), dass diese Verteilungsfunktion mit wachsendem n gegen die Verteilungsfunktion $\Phi(x)$ der Standard-Normalverteilung strebt. In Formeln

$$\lim_{n \rightarrow \infty} F_n(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt .$$

Weiter gilt

$$(1) \quad \lim_{n \rightarrow \infty} P(r \leq Z_n \leq s) = \frac{1}{\sqrt{2\pi}} \int_r^s e^{-x^2/2} dx = \Phi(s) - \Phi(r) ,$$

wobei Φ immer noch die Verteilungsfunktion der Standard-Normalverteilung ist. Für grosse Werte von n lässt sich wegen dieser Grenzwertbeziehung die Binomialverteilung durch die Standard-Normalverteilung approximieren. Wegen

$$Z_n = \frac{S_n - np}{\sqrt{npq}}$$

ist

$$a \leq S_n \leq b$$

gleichbedeutend mit

$$(2) \quad \frac{a - np}{\sqrt{npq}} \leq Z_n \leq \frac{b - np}{\sqrt{npq}}.$$

Kombiniert man (1) und (2), so kommt man — fast — auf die Formel

$$(3) \quad P(a \leq S_n \leq b) \approx \Phi\left(\frac{b - np + 1/2}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - 1/2}{\sqrt{npq}}\right).$$

“Fast” deshalb, weil noch Korrekturterme $\pm 1/2$ eingeschoben worden sind, welche — wie die Theorie zeigt — der Tatsache Rechnung tragen, dass die Binomialverteilung diskret, die Normalverteilung aber stetig ist. Man nennt diese Korrektur auch “Stetigkeitskorrektur” (englisch “Continuity-correction”). Wir begnügen uns mit einem einleuchtenden Bild: **R-Bild mit unteren Werten Beispiel 7.2.1.A und n=120 und k=19, 20, 21; p=1/6 und dann damit die Korrekturterme erklären**

Die Begründung der Formel (3) scheint vielleicht etwas theoretisch (auf einen strengen Beweis gehen wir ohnehin nicht ein), aber sie ist in der Praxis gut brauchbar, wenn es darum geht, Binomialverteilungen mit grossen Werten von n auszurechnen. Die exakte Auswertung der Ausdrücke $\binom{n}{k}p^kq^{n-k}$ gibt ja vor allem wegen der Binomialkoeffizienten viel zu rechnen.

Als Faustregel kann man sich merken: Die Formel (3) liefert hinreichend genaue Werte, wenn $npq > 9$ ist.

Beispiel 7.2.1.A

Wir würfeln 120-mal mit einem unverfälschten Würfel. Man wird ungefähr 20 Sechsen erwarten, aber auch kleine Abweichungen zugestehen. Wie gross ist die Wahrscheinlichkeit dafür, dass 19-, 20- oder 21-mal eine Sechs fällt?

Natürlich haben wir eine Binomialverteilung mit $n = 120$, $p = \frac{1}{6}$ vor uns. **Das ist genau das obige Bild.** Die Berechnung von Binomialkoeffizienten wie $\binom{120}{19}$ usw. schreckt sicher ab. Wir arbeiten deshalb mit der Formel (3), was nach der Faustregel gestattet ist. Es ist $\mu = np = 120 \cdot \frac{1}{6} = 20$, $\sigma = \sqrt{npq} = \sqrt{120 \cdot \frac{1}{6} \cdot \frac{5}{6}} = 4.0825$. Die Formel (3) ergibt

$$\begin{aligned} P(19 \leq X \leq 21) &\approx \Phi\left(\frac{21 - 20 + 0.5}{4.0825}\right) - \Phi\left(\frac{19 - 20 - 0.5}{4.0825}\right) \\ &= \Phi(0.367) - \Phi(-0.367) \approx 0.6432 - 0.3568 = 0.2864. \end{aligned}$$

Grob gesagt beträgt diese Wahrscheinlichkeit also etwa 30%. Exakt gibt es mit \mathbb{R} berechnet den Wert 0.2863054. \square

(7.2.2) Zentraler Grenzwertsatz, englisch: Central Limit Theorem (CLT)

Wir orientieren uns am Spezialfall der Binomialverteilung: Formel (3) in (7.2.1) lautet

$$(3) \quad P(a \leq S_n \leq b) \approx \Phi\left(\frac{b - np + 1/2}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - 1/2}{\sqrt{npq}}\right).$$

Wir werden jetzt die Fragestellung verallgemeinern.

In einem ersten Schritt erinnern wir uns, dass eine binomialverteilte Zufallsgrösse S_n auch als *Summe* von unabhängigen Bernoulli-Zufallsgrössen X_i aufgefasst werden kann, also in obiger Notation $S_n := \sum_{i=1}^n X_i$. Der Erwartungswert einer Bernoulli-Zufallsgrösse ist p und die Varianz $pq = p(1-p)$ (der Erwartungswert von S_n ist damit np und die Varianz $npq = np(1-p)$). Vorsicht: Wir haben ab jetzt als Untersuchungsobjekt eine Summe und im Unterschied zu (7.2.1) werden sich in (7.2.2) sowohl der Erwartungswert

wie auch die Varianz (und die Standardabweichung) immer auf die einzelnen Summanden (und nicht auf die Summe) beziehen:

$$\mu := E(X_i) \quad \sigma^2 := V(X_i) \quad \sigma = \sqrt{V(X_i)}.$$

Weiter geht die Verallgemeinerung in die Richtung, dass wir neben der Binomialverteilung alle möglichen Verteilungen zulassen wollen (sofern die Varianz existiert). Wir haben also erstmal als Untersuchungsobjekt eine Summe von iid Zufallsgrößen X_1, \dots, X_n mit $E(X_1) = \mu$ und $V(X_1) = \sigma^2$:

$$\sum_{i=1}^n X_i$$

Auch hier wollen wir jetzt wie in (7.2.1) zentrieren ($\sum_{i=1}^n X_i - n\mu$), der Mittelwert ist damit 0. Dann normieren wir auch hier (zusammen nennt man die Schritte “standardisieren”); wir erhalten:

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}.$$

Wie in (7.2.1) gilt jetzt auch hier (ohne Beweis): dieser Ausdruck konvergiert für $n \rightarrow \infty$ gegen eine Standardnormalverteilung; es gilt genauer:

Theorem 7.2 [Zentraler Grenzwertsatz] Sei $X_k, k \geq 1$, eine Folge von iid Zufallsgrößen mit $E[X_1] =: \mu$ und $0 < V[X_1] =: \sigma^2 < \infty$. Dann gilt für $a \in \mathbb{R}$ beliebig:

$$P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq a\right] \rightarrow P[\mathcal{N}(0,1) \leq a]$$

wenn $n \rightarrow \infty$.

Beispiel 7.2.2.A

In einem ersten, diskreten Beispiel arbeiten wir zuerst ohne die Stetigkeitskorrektur, welche in (7.2.1) eingeführt wurde ($\pm 1/2$), dann noch mit Korrektur und vergleichen.

Wie gross ist die Wahrscheinlichkeit, dass bei 100 Würfeln mit einer fairen Münze mindestens 65 mal Kopf erscheint? Um den Überblick zu bewahren: der korrekte Wert mit R ist: $1 - \text{pbinom}(64, 100, 0.5) = 0.001758821$.

Wir haben wieder eine Binomialverteilung, welche eine Summe von Bernoulli-Verteilungen (Kopf oder Zahl) ist. Mit der Konstruktion, welche wir in (7.1.2) bei den relativen Häufigkeiten eingeführt haben, gelangen wir zu folgender Rechnung mit $E(X_1) = p = 0.5$; $\sigma^2 = V(X_1) = p(1-p) = 0.5 \cdot 0.5$, womit $\sigma = 0.5$:

$$P\left(\sum_{i=1}^{100} X_i \geq 65\right) = P\left(\frac{\sum_{i=1}^{100} X_i - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5}} \geq \frac{65 - 100 \cdot 0.5}{\sqrt{100 \cdot 0.5}}\right).$$

Um Theorem 7.2 anwenden zu können, mussten wir in einem ersten Schritt also standardisieren.

$$P\left(\frac{\sum_{i=1}^{100} X_i - 100 \cdot 0.5}{\sqrt{100} \cdot 0.5} \geq \frac{65 - 100 \cdot 0.5}{\sqrt{100} \cdot 0.5}\right) = P\left(\frac{\sum_{i=1}^{100} X_i - 100 \cdot 0.5}{\sqrt{100} \cdot 0.5} \geq 3\right).$$

Wegen Theorem 7.2 können wir jetzt approximativ

$$\frac{\sum_{i=1}^{100} X_i - 100 \cdot 0.5}{\sqrt{100} \cdot 0.5}$$

durch eine standardnormalverteilte Zufallsgrösse Z ersetzen und erhalten

$$P(Z \geq 3) = 0.001349898.$$

Der Wert 0.001349898 ist von R; mit der Tabelle aus (6.2.3) haben wir den unpräziseren Wert 0.0013.

Wir wollen schauen, ob wir mit der Stetigkeitskorrektur besser fahren und untersuchen dazu (Vorsicht: 0.5 muss hier *abgezogen* werden, siehe Bild dazu in (7.2.1)!):

$$P\left(\sum_{i=1}^{100} X_i \geq 65 - 0.5\right) = P\left(\frac{\sum_{i=1}^{100} X_i - 100 \cdot 0.5}{\sqrt{100} \cdot 0.5} \geq \frac{64.5 - 100 \cdot 0.5}{\sqrt{100} \cdot 0.5}\right)$$

Dank des CLT gilt damit approximativ: $P(Z \geq 2.9) \doteq 0.001865813$. Zusammengefasst haben wir:

- 0.001758821 (exakter Wert)
- 0.001349898 (CLT ohne Stetigkeitskorrektur)
- 0.001865813 (CLT mit Stetigkeitskorrektur).

Die Korrektur hat also viel geholfen!

☒

Beispiel 7.2.2.B

Im zweiten Beispiel nehmen wir *stetige* Zufallsgrössen. Die Stetigkeitskorrektur macht dann keinen Sinn (siehe unten). Sei $\lambda = 2$ und $(X_k)_{k=1}^{50}$ iid $\exp(\lambda)$. Berechnen Sie $P(\sum_{k=1}^{50} X_k > 22)$. Wir repetieren von Kapitel 6: $E(X_1) = 1/\lambda = 0.5$ und $V(X_1) = 1/\lambda^2 = 0.25$, womit $\sigma := \sqrt{V(X_1)} = 1/\lambda = 0.5$. Damit sind wir gerüstet, um die Aufgabe zu lösen:

$$P\left(\sum_{k=1}^{50} X_k > 22\right) = P\left(\frac{\sum_{k=1}^{50} X_k - 50 \cdot 0.5}{\sqrt{50} \cdot 0.5} > \frac{22 - 50 \cdot 0.5}{\sqrt{50} \cdot 0.5}\right) \doteq P(Z > -0.8485281)$$

Beim letzten Gleichheitszeichen (\doteq) haben wir den CLT benutzt. Z sei dabei wie immer standardnormalverteilt. Diese Wahrscheinlichkeit ist gleich 0.801928 - der exakte Wert ohne CLT von R ist 0.7988194.

Wenn wir ein stetiges Beispiel haben, spielt es offenbar keine Rolle, ob wir

$$P\left(\sum_{k=1}^{50} X_k > 22\right) \quad \text{oder} \quad P\left(\sum_{k=1}^{50} X_k \geq 22\right)$$

untersuchen, weil $P(\sum_{k=1}^{50} X_k = 22) = 0$. Deshalb macht hier auch die Stetigkeitskorrektur keinen Sinn.

□

Theorem 7.2 ist ein theoretisches Limesresultat. In der Praxis wird aber wohl selten direkt gefragt, was die Wahrscheinlichkeit

$$P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq a\right]$$

für einen Wert annimmt. Es wird wohl eher nach einem Ausdruck wie

$$P\left[\sum_{k=1}^n X_k \leq a\right]$$

gefragt. Dann muss man die Summe zuerst standardisieren und kann dann Theorem 7.2 anwenden. Wie in (7.2.1) formen wir deshalb den Ausdruck $P[\sum_{k=1}^n X_k \leq a]$ abstrakt soweit um, dass wir direkt die Wahrscheinlichkeiten in einer \mathcal{N} -Tabelle, wie in (6.2.3), ablesen können:

$$P\left[\sum_{k=1}^n X_k \leq a\right] = P\left[\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq \frac{a - n\mu}{\sqrt{n}\sigma}\right] \approx \Phi\left(\frac{a - n\mu}{\sqrt{n}\sigma}\right).$$

Für ein Intervall von $[a, b]$ haben wir entsprechend

$$P\left[a \leq \sum_{k=1}^n X_k \leq b\right] \approx \Phi\left(\frac{b - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a - n\mu}{\sqrt{n}\sigma}\right).$$

Mit Stetigkeitskorrektur (Continuity Correction), welche nur bei diskreten Zufallsgrößen Sinn macht, lauten die entsprechenden Formeln:

$$P\left[\sum_{k=1}^n X_k \leq a\right] \approx \Phi\left(\frac{a + 1/2 - n\mu}{\sqrt{n}\sigma}\right).$$

Für ein Intervall von $[a, b]$ haben wir entsprechend

$$P\left[a \leq \sum_{k=1}^n X_k \leq b\right] \approx \Phi\left(\frac{b + 1/2 - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a - 1/2 - n\mu}{\sqrt{n}\sigma}\right).$$

Seien Sie vorsichtig und machen Sie am Besten eine Skizze wie in (7.2.1) bei der Frage, ob und wo Sie $1/2$ abziehen oder dazuzählen, je nachdem, ob man $<, \leq, \geq, >$ untersucht (zum Beispiel im Zusammenhang mit Gegenwahrscheinlichkeiten). Weiter verliert die Stetigkeitskorrektur an Bedeutung, je grösser das n ist.

Zum Schluss, gut zu wissen: Wenn $X \sim \mathcal{N}(\mu, \sigma^2)$, dann gilt bekanntlich (siehe Kapitel 5) $P[X \in [\mu \pm 2\sigma]] \doteq 0.95$. Dank des CLT können wir auch für andere Verteilungen folgende Praktikerregeln schliessen:

* $Y \sim \text{Bin}(n, p) \Rightarrow P[Y \in [np \pm 2\sqrt{np(1-p)}]] \doteq 0.95$ wenn $np(1-p) \geq 9$.

* $Z \sim \text{Po}(\lambda) \Rightarrow P[Z \in [\lambda \pm 2\sqrt{\lambda}]] \doteq 0.95$ wenn λ gross.

(7.2.3) Weshalb die Normalverteilung “überall” auftaucht

Es zeigt sich, dass sehr viele durch Messungen in der Natur gefundene (so genannte empirische) Verteilungen, wie Gewichte, Körperlängen, aber auch Messfehler bei physikalischen Experimenten, sich zumindest angenähert durch eine Normalverteilung beschreiben lassen. Hierzu gibt es auch eine mathematische Begründung. Messwerte der genannten Art werden ja von einer Vielzahl von Faktoren beeinflusst (das Gewicht eines Menschen etwa durch Erbanlagen, Ernährungsgewohnheiten usw.), so dass der Messwert als Summe von vielen Zufallsgrössen aufgefasst werden kann (“ $\sum_k X_k$ ”). Der CLT besagt nun, dass unter passenden Voraussetzungen eine solche Summe annähernd (und im Grenzfall sogar genau) normal verteilt ist. Dabei ist Unabhängigkeit der Summanden wichtiger als die gleiche Verteilung. Dies erklärt einigermaßen die Bedeutung der Normalverteilung in den Anwendungen.

7.3. POISSONAPPROXIMATION

(7.3.0) Überblick

Die diskrete Zufallsgrösse X folgt der *Poisson-Verteilung* mit dem Parameter $\lambda > 0$, wenn sie die (abzählbar unendlich vielen) Werte $0, 1, 2, 3, \dots$ annimmt und wenn die zugehörigen Wahrscheinlichkeiten durch

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

gegeben sind.

Für den Erwartungswert und die Varianz der Poisson-Verteilung mit dem Parameter λ gilt:

$$E(X) = V(X) = \lambda .$$

Die Poisson-Verteilung kann in gewissen Fällen als bequem zu berechnende Annäherung an die Binomialverteilung verwendet werden: Für grosses n (n etwa > 10) und kleines p (p etwa < 0.05) ist in guter Näherung

$$\binom{n}{k} p^k q^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!} \quad \text{mit} \quad \lambda = np .$$

(7.3.1) Definition der Poisson-Verteilung

Die Poisson-Verteilung ist ein weiteres wichtiges Beispiel einer diskreten Verteilung. Im Gegensatz zur Binomialverteilung ist es hier aber so, dass die Zufallsgrösse X abzählbar unendlich viele Werte (die natürlichen Zahlen $0, 1, 2, \dots$) annimmt. Wir führen in diesem Abschnitt gleich die allgemeine Definition an; eine Motivation folgt in (7.3.2). Dabei benützen wir, dass eine diskrete Zufallsgrösse dadurch bestimmt ist, dass man ihre Werte mitsamt den zugehörigen Wahrscheinlichkeiten angibt. Dies werden wir nun tun.

Es sei noch vorausgeschickt, dass die Poisson-Verteilung von einer vorgegebenen Zahl $\lambda > 0$ abhängt; es gibt also dementsprechend nicht nur eine, sondern unendlich viele Möglichkeiten für die Poisson-Verteilung.

Die diskrete Zufallsgrösse X folgt der *Poisson-Verteilung** mit dem Parameter λ ($\lambda > 0$), wenn sie die (abzählbar unendlich vielen) Werte $0, 1, 2, \dots$ annimmt und wenn für die zugehörigen Wahrscheinlichkeiten gilt:

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Der Parameter λ kann eine beliebige reelle Zahl > 0 sein.

Damit wir sicher sind, dass die angegebenen Wahrscheinlichkeiten tatsächlich eine Verteilung beschreiben, müssen wir zeigen, dass ihre (hier unendliche) Summe $= 1$ ist (4.1.5). Dies ist auch der Fall:

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

denn die letzte Reihe ist ja die Exponentialreihe (19.8.a) aus dem ersten Band.

Die Wahrscheinlichkeiten $P(X = k)$ sind mit dem Taschenrechner leicht zu bestimmen; man kann sie aber auch in Abhängigkeit von λ und k tabellieren (vgl. Tabelle (6.1.4)).

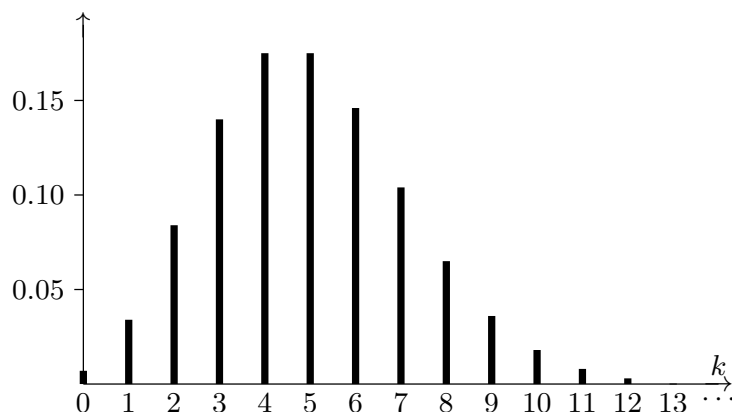
Als konkretes Beispiel geben wir die Werte

$$P(X = k) = e^{-5} \frac{5^k}{k!}$$

für die Poisson-Verteilung mit dem Parameter $\lambda = 5$ in gerundeter Form an:

k	0	1	2	3	4	5	6	7	8	...
$P(X = k)$	0.007	0.034	0.084	0.140	0.175	0.175	0.146	0.104	0.065	...

Das zugehörige Stabdiagramm sieht so aus:



* Benannt nach S.D. POISSON, 1781–1840.

(7.3.2) Herleitung der Poisson-Verteilung aus der Binomialverteilung

Die Poisson-Verteilung ist in (7.3.1) einfach als abstrakt gegebene Verteilung eingeführt worden. Wir wollen nun die fehlende Motivation nachholen. Dabei setzen wir die Poisson-Verteilung mit der Binomialverteilung in Beziehung.

Bekanntlich (siehe (5.5)) ist der Erwartungswert λ der Binomialverteilung mit den Parametern n und p gegeben durch $\lambda = np$ (zielgerichtet verwenden wir gleich λ statt μ). Dieser Wert λ kann natürlich auf verschiedene Arten zustande kommen. Auf $\lambda = 2$ beispielsweise kommt man etwa durch die Wahl $n = 4$, $p = \frac{1}{2}$, aber auch durch $n = 100$, $p = \frac{1}{50}$.

Berechnet man nun bei festgehaltenem λ die Wahrscheinlichkeiten $P(X = k)$ der Binomialverteilung für verschiedene n und p (aber immer mit $np = \lambda$), so stellt man fest, dass bei grossem n (und entsprechend kleinem p) die Werte $P(X = k)$ für ein festes k ungefähr gleich gross sind.

Das folgende Zahlenbeispiel mit $\lambda = 2$ soll diesen Sachverhalt illustrieren, wobei wir uns auf die Wahrscheinlichkeiten $P(X = 0)$ und $P(X = 2)$ beschränken.

	$n = 50,$ $p = \frac{1}{25}$	$n = 100,$ $p = \frac{1}{50}$	$n = 200,$ $p = \frac{1}{100}$	$n = 500,$ $p = \frac{1}{250}$	$n = 1000,$ $p = \frac{1}{500}$
$P(X = 0)$	0.1299	0.1326	0.1340	0.1348	0.1351
$P(X = 2)$	0.2762	0.2734	0.2720	0.2712	0.2709

Wir werden nun rechnerisch zeigen, dass die zur Binomialverteilung gehörende Wahrscheinlichkeit

$$\binom{n}{k} p^k q^{n-k}$$

bei festgehaltenem λ für jeden einzelnen Wert von k ($0 \leq k \leq n$) gegen einen bestimmten Grenzwert strebt. Es gilt nämlich für jedes k die folgende Beziehung:

$$(*) \quad \lim_{n \rightarrow \infty} \binom{n}{k} p^k q^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad (\lambda = np, \text{ fest}).$$

Dies soll jetzt bewiesen werden. Weil $np = \lambda$ festgehalten wird, ist $p = \frac{\lambda}{n}$, ferner ist $q = 1 - p = 1 - \frac{\lambda}{n}$. Nun formen wir um, wobei wir p und q durch die eben angegebenen Ausdrücke ersetzen:

$$\begin{aligned} \binom{n}{k} p^k q^{n-k} &= \frac{n(n-1)(n-2) \cdots (n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \frac{n(n-1)(n-2) \cdots (n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \cdot 1 \cdot \left(1 - \frac{1}{n}\right) \cdot \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right). \end{aligned}$$

Im letzten Ausdruck lassen wir n gegen ∞ streben (gleichzeitig geht dann p gegen 0). Der erste Faktor $\frac{\lambda^k}{k!}$ ändert sich nicht. Der zweite Faktor $(1 - \frac{\lambda}{n})^n$ strebt, wie interessierte Leute in (7.3.6) nachlesen können, gegen $e^{-\lambda}$. Alle übrigen Faktoren dagegen streben gegen 1, denn $\frac{1}{n} \rightarrow 0$, $\frac{2}{n} \rightarrow 0$ usw. Infolgedessen gilt

$$\lim_{n \rightarrow \infty} \binom{n}{k} p^k q^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}, \quad (\lambda = np \text{ fest}),$$

was wir behauptet haben.

Die Grenzwertbeziehung (*) lässt sich, wenigstens für genügend grosse n , auch als Näherung schreiben:

$$(**) \quad \binom{n}{k} p^k q^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}, \quad (\lambda = np).$$

Diese Approximation ist dann brauchbar, wenn n gross und p klein ist. Als Faustregel kann man sich etwa merken, dass $n > 10$ und $p < 0.05$ sein sollte.

In der folgenden Tabelle werden als Illustration die beiden Seiten von (**) verglichen. Sie bezieht sich auf die Werte $n = 100$, $p = 0.01$, $\lambda = 1$.

In der Zeile B steht jeweils der Wert der Binomialverteilung, also von $\binom{n}{k} p^k q^{n-k}$, in der Zeile P jener von $e^{-\lambda} \frac{\lambda^k}{k!}$, also der Poisson-Verteilung.

k	0	1	2	3	4	5
B	0.366	0.370	0.185	0.0610	0.0149	0.0029
P	0.368	0.368	0.184	0.0613	0.0153	0.0031

Zu beachten ist, dass die Werte der Binomialverteilungen nur für $0 \leq k \leq n$ definiert sind (für $k > n$ könnte man $P(X = k) = 0$ setzen); dagegen ist $e^{-\lambda} \frac{\lambda^k}{k!}$ für alle $k \geq 0$ definiert und > 0 . Allerdings wird dieser Ausdruck mit wachsendem k immer kleiner, so dass dieser Unterschied zur Binomialverteilung fast nichts ausmacht.

Schliesslich sei darauf hingewiesen, dass die rechte Seite in der Näherung (**) leichter zu berechnen ist als die linke Seite, wo die doch etwas unhandlichen Binomialkoeffizienten vorkommen.

(7.3.3) Erwartungswert und Varianz der Poisson-Verteilung

Die Poisson-Verteilung kann gemäss (7.3.2) aus Binomialverteilungen mit dem konstant gehaltenen Erwartungswert $\lambda = np$ hergeleitet werden. Es überrascht deshalb nicht, dass gilt:

Die Poisson-Verteilung mit dem Parameter λ hat den Erwartungswert λ :

$$E(X) = \lambda .$$

Auch für die Varianz können wir eine anschauliche Überlegung anstellen: Nach (5.5) hat die Binomialverteilung die Varianz $V(X) = npq = np(1-p)$. Schreiben wir wie in (7.3.2) $p = \lambda/n$ und $q = 1 - \lambda/n$, so ergibt sich $V(X) = \lambda(1 - \lambda/n)$, und mit $n \rightarrow \infty$ strebt dieser Ausdruck gegen λ . Somit hat die Poisson-Verteilung, als Grenzfall der Binomialverteilung, die Varianz

$$V(X) = \lambda .$$

Diese Herleitungen sind plausibel, aber mathematisch nicht ganz korrekt, denn $E(X)$ und $V(X)$ sind ja als Summen definiert, und zwar hier, wo X abzählbar unendlich viele Werte annimmt, als unendliche Summen, d.h., als Reihen. Korrekte Beweise stützen sich deshalb auf die Definitionen von (5.1) und (5.2):

Für den Erwartungswert sieht die Rechnung so aus:

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} k && \text{(Definition von } E(X)) \\ &= \sum_{k=1}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} k && \text{(Summand mit } k=0 \text{ hat den Wert 0)} \\ &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} && (k! = k(k-1)!) \\ &= \lambda e^{-\lambda} \sum_{h=0}^{\infty} \frac{\lambda^h}{h!} && \text{(Setze } h = k-1) \\ &= \lambda e^{-\lambda} e^{\lambda} && \text{(Exponentialreihe!)} \\ &= \lambda . \end{aligned}$$

Für die Herleitung der Formel für die Varianz sei auf den Anhang verwiesen **noch anpassen** (49.4).

(7.3.4) Beispiele zur Approximation der Binomialverteilung durch die Poisson-Verteilung

In (7.3.2) wurde gezeigt, dass die Poisson-Verteilung eine bequem zu berechnende Approximation der Binomialverteilung ist. Für die Praxis ist, wie bereits erwähnt, die Näherung brauchbar, wenn n gross (d.h. n etwa > 10) und p klein (d.h. p etwa < 0.05) ist.

Zur weiteren Erläuterung dienen die folgenden Beispiele:

Beispiel 7.3.4.A

In 5 kg Teig befinden sich 200 Rosinen. Der Bäcker stellt daraus Brötchen von 50 g her. Wie gross ist die Wahrscheinlichkeit dafür, dass ein Brötchen keine Rosinen enthält?

Wir stellen folgende Überlegung an: Wir betrachten ein bestimmtes Brötchen. Die Wahrscheinlichkeit, dass eine gewisse Rosine gerade in dieses Brötchen gerät, ist $p = 0.01$, denn es sind ja 100 Brötchen vorhanden.

Nun nehmen wir uns die Rosinen der Reihe nach vor und sehen nach, ob sie in unserem gewählten Brötchen sind (“Erfolg”) oder nicht (“Misserfolg”). Dies liefert uns $n = 200$ unabhängige Wiederholungen eines “Versuchs” mit zwei möglichen Ausgängen. Gefragt ist nach der Wahrscheinlichkeit von 0 Erfolgen. Die Binomialverteilung liefert

$$P(X = 0) = \binom{200}{0} 0.01^0 \cdot 0.99^{200} = 0.13397 \dots$$

Dabei bezeichnet die Zufallsgrösse X die Anzahl Rosinen pro Brötchen. Da n gross und p klein ist, dürfen wir auch die Poisson-Verteilung gebrauchen, mit $\lambda = np = 2$. Wir erhalten damit

$$P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-2} = 0.13533 \dots$$

Die Approximation ist genügend gut.

Wir haben also gefunden: Von den 100 Brötchen werden ca. 14 keine Rosine enthalten. \boxtimes

Hier machte übrigens die Berechnung der Binomialverteilung nicht viel mehr Mühe als jene der Poisson-Verteilung. Dies liegt daran, dass $k = 0$ war. Für grosse n und k ist aber der Binomialkoeffizient $\binom{n}{k}$ etwas mühsam zu berechnen.

Beispiel 7.3.4.B

Wir setzen Beispiel 7.3.4.A fort und fragen, wieviele Rosinen man braucht, damit mit 95%-iger Wahrscheinlichkeit in einem beliebigen Brötchen mindestens eine Rosine zu finden ist.

Wie vorhin ist $p = 0.01$ gegeben. Die Zahl n der Rosinen und damit λ ist unbekannt. Der Parameter λ ist so zu bestimmen, dass

$$P(X = 0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda} \leq 0.05 \text{ ist.}$$

Durch Logarithmieren erhält man

$$-\lambda \leq \ln 0.05 = -2.9957$$

$$\lambda \geq 2.9957$$

$$\text{und schliesslich } n = \frac{\lambda}{p} = 100\lambda \geq 299.57.$$

Es sind also mindestens 300 Rosinen zu verwenden. \boxtimes

Beispiel 7.3.4.C

Eine Fabrik produziert Schrauben, die mit einer Wahrscheinlichkeit von $p = 0.001$ defekt sind. Wie gross ist die Wahrscheinlichkeit, in einer Schachtel von 500 Schrauben zwei oder mehr defekte zu finden?

Im Grunde liegt eine Binomialverteilung mit $n = 500$ und $p = 0.001$ vor, die wir unbesorgt mit einer Poisson-Verteilung approximieren. Es ist dann $\lambda = 500 \cdot 0.001 = 0.5$. Gesucht ist $P(X \geq 2)$, wobei die Zufallsgrösse X die Zahl der defekten Schrauben in einer Schachtel bezeichnet. Es ist hier sinnvoll, die Gegenwahrscheinlichkeit zu berechnen. Man hat

$$P(X \geq 2) = 1 - P(X < 2)$$

und

$$\begin{aligned} P(X < 2) &= P(X = 0) + P(X = 1) \\ &= e^{-\lambda} \frac{\lambda^0}{0!} + e^{-\lambda} \frac{\lambda^1}{1!} \\ &= e^{-0.5} + 0.5e^{-0.5} = 0.9097. \end{aligned}$$

Für die gesuchte Wahrscheinlichkeit erhalten wir schliesslich

$$P(X \geq 2) = 1 - 0.9097 = 0.0903. \quad \square$$

Beachten Sie bei dieser Gelegenheit, dass stets

$$\frac{\lambda^0}{0!} = 1 \quad \text{und} \quad \frac{\lambda^1}{1!} = \lambda, \quad \text{also} \quad P(X = 0) = e^{-\lambda} \quad \text{und} \quad P(X = 1) = \lambda e^{-\lambda}$$

ist.

(7.3.5) Weitere Beispiele zur Poisson-Verteilung

In (7.3.4) haben wir die Poisson-Verteilung zur Approximation der Binomialverteilung benützt. Es gibt aber auch Situationen, wo man von vornherein eine Poisson-Verteilung annimmt.

Dies trifft besonders bei “seltenen Ereignissen” zu, d.h., bei Ereignissen, die mit einer geringen Wahrscheinlichkeit eintreffen. Beispiele von solchen seltenen Ereignissen sind etwa:

- Radioaktiver Zerfall,
- Verteilung der Rosinen in Rosinenbrötchen,
- Anzahl Druckfehler pro Seite eines Buchs (von den vielen Buchstaben auf einer Seite sind die allermeisten richtig),
- Anzahl Pannen pro Jahr eines bestimmten Automodells,
- Anzahl nicht keimender Kartoffeln in einer Sendung Saatkartoffeln,
- etc. etc.

Die Überlegungen sind in allen diesen Fällen analog; wir führen sie im nächsten Beispiel konkret durch.

Beispiel 7.3.5.A Radioaktiver Zerfall

Wir untersuchen eine radioaktive Substanz und interessieren uns dafür, wieviele Atome in einem gegebenen Zeitintervall einer bestimmten Länge (z.B. einer Sekunde) zerfallen.

Jedes Atom hat dieselbe (sehr kleine) Wahrscheinlichkeit p , in diesem Zeitintervall zu zerfallen. Die totale Anzahl n der vorhandenen Atome ist sehr gross. Wir fixieren nun ein solches Zeitintervall und gehen in diesem Intervall (in Gedanken) jedes einzelne der Atome durch. Zerfällt es in diesem Intervall, notieren wir einen Erfolg, zerfällt es nicht, liegt ein Misserfolg vor.

Mit X bezeichnen wir die Zufallsgrösse

“Anzahl Zerfälle pro Zeitintervall*”.

Die Grösse X nimmt die Werte $0, 1, 2, \dots$ an. Unter der Voraussetzung, dass die Atome unabhängig voneinander zerfallen, würde eigentlich eine Binomialverteilung vorliegen. Um hier die Wahrscheinlichkeiten $P(X = k)$ zu berechnen, müsste man aber die Anzahl n der Atome und die Wahrscheinlichkeit p kennen, was aber nicht der Fall ist.

Das Problem ist jedoch lösbar, wenn man zur Poisson-Verteilung übergeht. Dies ist im Sinne einer Annäherung an die Binomialverteilung erlaubt, da n sehr gross und p sehr klein ist. Vor allem ist aber wichtig, dass man zwar n und p nicht einzeln kennt, wohl aber ihr Produkt $\lambda = np$, und dies ist ja gerade der Parameter, der für die Poisson-Verteilung bekannt sein muss. Wie wir nämlich von (7.3.3) her wissen, ist λ der Erwartungswert der Poisson-Verteilung, und in unserem Beispiel entspricht der Erwartungswert gerade der durchschnittlichen Anzahl Zerfälle pro Zeitintervall und kann somit bestimmt (oder zum mindesten geschätzt) werden.

Aufgrund dieser Vorbetrachtungen nehmen wir an, dass beim radioaktiven Zerfall die Anzahl der Zerfälle pro Zeiteinheit einer Poisson-Verteilung folgt. (Dies wird übrigens experimentell gestützt. Wir kommen anlässlich der Besprechung des χ^2 -Tests **noch anpassen** [Beispiel 47.4.B] auf derartige Fragen zurück.)

Nach all dieser Theorie nun ein Zahlenbeispiel:

Bei einem Experiment mit radioaktivem Zerfall wurden im Durchschnitt (über eine lange Zeitspanne hinweg) 4.4 Zerfälle pro Sekunde gezählt. Welches ist die Wahrscheinlichkeit dafür, dass in einem Zeitintervall von einer Sekunde Länge

* Wir haben hier also eine “Verteilung pro Zeiteinheit” vor uns, im Gegensatz zur “Rosinenaufgabe” 7.3.4.A, wo eine “Verteilung pro Raumeinheit” vorliegt.

- a) genau vier Atome zerfallen,
 b) mindestens ein Atom zerfällt?

Aufgrund des oben Gesagten nehmen wir hier eine Poisson-Verteilung mit Parameter $\lambda = 4.4$ an (denn dies ist gerade der Erwartungswert). Dann findet man sofort:

a) $P(X = 4) = e^{-4.4} \frac{4.4^4}{4!} = 0.1917$.

- b) Hier berechnet man natürlich zuerst die Gegenwahrscheinlichkeit $P(X = 0)$.

Die Wahrscheinlichkeit dafür, dass im gewählten Zeitintervall kein Zerfall auftritt, ist

$$P(X = 0) = e^{-4.4} \frac{\lambda^0}{0!} = e^{-4.4} = 0.0123 .$$

Die Wahrscheinlichkeit für mindestens einen Zerfall ist daher gleich

$$1 - P(X = 0) = 1 - 0.0123 = 0.9877 .$$

Beispiel 7.3.5.B

Ein Buch weist im Mittel pro Seite 3 Druckfehler auf. Wie gross ist die Wahrscheinlichkeit dafür, dass auf einer zufällig aufgeschlagenen Seite höchstens ein Druckfehler ist?

Antwort: Nach den eingangs dieses Abschnitts gemachten Feststellungen dürfen wir annehmen, dass die Zufallsgrösse

“ $X =$ Anzahl Druckfehler pro Seite”

Poisson-verteilt mit $\lambda = 3$ ist. Wir erhalten deshalb für die gesuchte Wahrscheinlichkeit

$$P(X = 0) + P(X = 1) = e^{-3} \frac{3^0}{0!} + e^{-3} \frac{3^1}{1!} = e^{-3}(1 + 3) = 0.1991 . \quad \boxtimes$$

Wir überlegen uns noch etwas genauer, warum wir eine Poisson-Verteilung annehmen durften. Auf einer Buchseite hat es sehr viele Buchstaben (ihre Anzahl n ist gross), und jeder einzelne ist mit einer sehr geringen Wahrscheinlichkeit p falsch. Prüfen wir nun der Reihe nach jeden Buchstaben auf seine Richtigkeit hin, so führen wir ein Bernoulli-Experiment im Sinn von (4.2.2) durch, mit der Interpretation “Erfolg = Druckfehler”. Demnach liegt eigentlich eine Binomialverteilung mit den Parametern n, p vor, die wir aber getrost durch eine Poisson-Verteilung approximieren. Es bleibt uns auch nichts anderes übrig, denn n und p sind unbekannt (nicht aber $\lambda = np = 3$).

Bei der Annahme, dass eigentlich eine Binomialverteilung vorliege, haben wir stillschweigend vorausgesetzt, dass die Druckfehler voneinander unabhängig sind. Dies trifft vielleicht nicht immer zu, z.B. im Fall von vertauschten Buchstaben. Gewöhnlich wird dies aber als ein einziger Fehler gezählt, so dass dieses Problem etwas entschärft wird.

Beispiel 7.3.5.C

Eine etwas andere Anwendung ist die folgende: Eine Blutprobe wird im Mikroskop unter einem Raster aus 400 Quadraten betrachtet. Es wird festgestellt, dass 15 Felder

keine Erythrozyten (rote Blutkörperchen) enthalten. Wieviele Erythrozyten sind in der Blutprobe vorhanden?

Es scheint auf den ersten Blick, als ob diese Aufgabe infolge ungenügender Angaben unlösbar sei. Wir machen nun aber die aufgrund des bisher gesagten vernünftige Annahme, die Zufallsvariable

$$"X = \text{Anzahl Erythrozyten pro Quadrat}"$$

gehorche einer Poisson-Verteilung.

Eigentlich liegt eine Binomialverteilung mit unbekanntem n und bekanntem $p = \frac{1}{400}$ vor. (Der Sachverhalt ist genau derselbe wie bei der "Rosinen-Aufgabe" 7.3.4.A; die Blutkörperchen entsprechen den Rosinen, die einzelnen Felder den Brötchen.) Da p klein und n gross ist, verwenden wir jedoch ohne Skrupel eine Poisson-Verteilung mit vorläufig unbekanntem λ .

Wir wissen, dass 15 Felder frei von Erythrozyten sind, d.h., dass

$$P(X = 0) = \frac{15}{400}$$

ist. Somit ist auch

$$e^{-\lambda} \frac{\lambda^0}{0!} = e^{-\lambda} = \frac{15}{400} = 0.0375$$

und daraus lässt sich λ durch Logarithmieren bestimmen:

$$\lambda = -\ln 0.0375 = 3.283 .$$

Mit $\lambda = np$ und $p = 0.0025$ ergibt sich dann schliesslich für die Anzahl der Erythrozyten

$$n \approx 1313 . \quad \boxtimes$$

(7.3.6) Eine Grenzwertdarstellung der Exponentialfunktion

Bei der Herleitung der Poisson-Verteilung aus der Binomialverteilung wurde in (7.3.2) die folgende Formel (mit $r = -\lambda$) benützt, die wir nun beweisen wollen:

$$e^r = \lim_{n \rightarrow \infty} \left(1 + \frac{r}{n}\right)^n \quad \text{für alle } r \in \mathbb{R} .$$

Um den "Exponenten herunterzuholen", betrachten wir anstelle der Folge

$$a_n = \left(1 + \frac{r}{n}\right)^n$$

die Folge

$$\ln a_n = \ln \left(1 + \frac{r}{n}\right)^n = n \cdot \ln \left(1 + \frac{r}{n}\right) .$$

Wir setzen noch $h_n = \frac{r}{n}$ (und damit $n = \frac{r}{h_n}$) und erhalten so

$$\ln a_n = r \cdot \frac{\ln(1 + h_n)}{h_n} = r \cdot \frac{\ln(1 + h_n) - \ln(1 + 0)}{h_n - 0}.$$

Der kleine Trick mit der Addition von $\ln(1 + 0) = 0$ im Zähler, bzw. von 0 im Nenner hat zur Folge, dass rechts der Differenzenquotient (siehe 4.3.b im ersten Band) der Funktion $\ln(1 + x)$ an der Stelle 0 (mit $x = h_n$) steht. Wenn nun $n \rightarrow \infty$ strebt, dann strebt die Folge h_n gegen 0. Dann aber strebt der Differenzenquotient

$$\frac{\ln(1 + h_n) - \ln(1 + 0)}{h_n - 0}$$

gegen die Ableitung der Funktion $\ln(1 + x)$ an der Stelle $x = 0$, und nach den bekannten Regeln hat diese Ableitung den Wert 1. Damit folgt

$$\lim_{n \rightarrow \infty} \ln a_n = r \cdot 1 = r,$$

und da die Exponentialfunktion stetig ist, ist dann weiter

$$\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} e^{\ln a_n} = e^{\lim_{n \rightarrow \infty} \ln a_n} = e^r,$$

womit die gewünschte Formel bewiesen ist.

(7.∞) Aufgaben

Zu (7.1), LLN, gibt es keine Aufgaben, weil die Resultate rein theoretisch sind.

Aufgaben zum zentralen Grenzwertsatz (7.2) - mit Stetigkeitskorrektur zu lösen

- 7-1 Eine faire Münze wird 500 mal geworfen. Berechnen Sie die Wahrscheinlichkeit dafür, dass die Anzahl der Köpfe zwischen 245 und 255 (Grenzen eingeschlossen) liegt.
- 7-2 Eine Maschine produziert Teile, die mit einer Wahrscheinlichkeit von 10 % Ausschuss sind. Mit welcher Wahrscheinlichkeit sind unter den 1000 Teilen höchstens 100 Ausschussteile?
- 7-3 Sei $(X_i)_{i=1}^{100}$ eine iid Folge von $\text{Po}(\lambda)$ -Zufallsgrößen, deren Summe bezeichnen wir mit $S := \sum_{i=1}^{100} X_i$. Schätzen Sie mit Hilfe des CLT $P[S \leq 110]$ ab, wenn $\lambda = 1$ ist.

Aufgaben zum zentralen Grenzwertsatz (7.2) - ohne Stetigkeitskorrektur zu lösen (stetig)

- 7-6 Wie gross ist die Wahrscheinlichkeit, dass eine Summe von 900 iid χ_3^2 -Zufallsgrößen grösser als 2680 ist? Benutzen Sie den CLT zur approximativen Berechnung. In Kapitel 6 finden Sie die notwendigen Kennziffern (Erwartungswerte, Varianzen).
- 7-7 Sei $(X_i)_{i=1}^{121}$ eine iid Folge von t_7 -Zufallsgrößen. Schätzen Sie mit dem CLT $P[\sum_{i=1}^{121} X_i \leq 8]$ ab.
- 7-8 Sie geben in R den Befehl `(sum(rexp(900, rate=4))-a)/(30*b)` ein (**a** und **b** seien konkrete, reelle Zahlen). Welche Werte müssen Sie für **a** und **b** wählen, damit Sie den CLT anwenden können? Wie gross ist die Wahrscheinlichkeit, dass `sum(rexp(900, rate=4))` grösser als 232 ist? Benutzen Sie den CLT zur approximativen Berechnung.

Aufgaben zur Poissonapproximation (7.3)

- 7–11 Zeichnen Sie in derselben Skizze die Stabdiagramme der Binomialverteilung mit $n = 5, p = 0.4$ und der Poisson-Verteilung mit $\lambda = 2$.
- 7–12 Ein Medikament hat mit einer Wahrscheinlichkeit von 1.5% lästige Nebenwirkungen (Übelkeit). 200 Personen nehmen dieses Medikament ein. Die Zufallsgrösse X bezeichne die Anzahl der Personen, die von der Nebenwirkung betroffen werden.
- Bestimmen Sie den Erwartungswert von X . Was bedeutet er konkret?
 - Berechnen Sie exakt die Wahrscheinlichkeit dafür, dass es höchstens zwei Personen übel wird.
 - Lösen Sie b) näherungsweise mittels der Poisson-Verteilung.
 - Wieviele Personen darf die Gruppe höchstens umfassen, wenn mit einer Wahrscheinlichkeit von mindestens 20% keine Übelkeitsfälle auftreten sollen? Verwenden Sie sowohl die Binomial- als auch die Poisson-Verteilung.
- 7–13 3% aller Menschen können mit den Ohren wackeln. Wie gross ist die Wahrscheinlichkeit dafür, dass in einer Gruppe von 100 Leuten a) keine, b) zwei, c) vier Personen diese nützliche Fähigkeit haben? Lösen Sie diese Aufgabe sowohl mit der Binomial- als auch mit der Poisson-Verteilung. Geben Sie die Antwort in Prozenten, mit zwei Stellen nach dem Dezimalpunkt, an.
- 7–14 Beim Roulette hat die Null (wie jeder andere Ausgang) die Wahrscheinlichkeit $1/37$. Wie gross ist die Wahrscheinlichkeit dafür, dass in 50 Spielen die Null höchstens zweimal auftritt? Lösen Sie die Aufgabe einmal exakt (Binomialverteilung) und einmal näherungsweise (Poisson-Verteilung).
- 7–15 Der berühmte Mathematiker Prof. Dr. Wurzel-Zieher hält wieder einmal einen seiner stets ausverkauften Vorträge über “General Abstract Nonsense”. Die Plätze im Hörsaal, der 90 Personen fasst, müssen vorgängig im Mathematischen Institut reserviert werden. Die Sekretärin weiss aus Erfahrung, dass 3% aller Leute, die reserviert haben, doch nicht auftauchen und verteilt deshalb Karten an 93 Personen. Wie gross ist die Wahrscheinlichkeit dafür, dass alle Platz finden? a) Exakte Lösung mit Binomialverteilung. b) Näherungslösung mit Poissonverteilung. Tipp: Betrachten Sie die Anzahl der nicht erscheinenden Personen.
- 7–16 Im Verlauf einer Stunde erhält eine Telefonzentrale durchschnittlich 60 Anrufe. Mit welcher Wahrscheinlichkeit treffen in einem Zeitintervall von a) 30 Sekunden, b) 2 Minuten keine Anrufe ein?
- 7–17 Ein Frosch fängt im Durchschnitt 3 Fliegen pro Stunde. Mit welcher Wahrscheinlichkeit fängt er in der nächsten Stunde a) keine Fliege, b) mehr als drei Fliegen?
- 7–18 Eine radioaktive Substanz gibt im Verlauf von 10 Sekunden im Mittel 5 α -Teilchen ab. Wie gross ist die Wahrscheinlichkeit dafür, dass sie im nächsten 10-Sekunden-Intervall a) mindestens ein, b) genau 5 α -Teilchen emittiert?
- 7–19 Ein Buch von 500 Seiten Umfang enthält 1200 Druckfehler. Wir nehmen an, die Zufallsgrösse “Anzahl Druckfehler pro Seite” folge einer Poissonverteilung.
- Ich schlage eine Seite zufällig auf. Wie gross ist die Wahrscheinlichkeit dafür, dass auf dieser Seite a₁) kein Druckfehler ist, a₂) mindestens zwei Druckfehler sind?
 - Wie gross müsste die totale Zahl der Druckfehler sein, damit auf der aufgeschlagenen Seite mit 50% Wahrscheinlichkeit kein Druckfehler zu finden ist?
- 7–20 “Entsteinte” Kirschen sind dies nicht immer. Die Erfahrung zeigt, dass pro Stück einer Kirschenwähe im Mittel eine Kirsche mit Stein vorkommt. a) Mit welcher Wahrscheinlichkeit hat es auf einem Wähenstück höchstens 3 Kirschen mit Stein? b) Vier Leute haben unabhängig voneinander ein Wähenstück gekauft; auf jedem hatte es mindestens eine Kirsche mit Stein. Wie gross ist die Wahrscheinlichkeit für ein solches Ereignis?

- 7–21 Sepps Souvenirgeschäft verkauft pro Tag im Mittel zwei Sennenkäppli. Nachdem ein allseits beliebter Sportler am Samstag im Fernsehen ein Sennenkäppli trug, verkaufte Sepp am Montag sechs solcher Dinger. Könnte dies etwas mit dem Fernsehauftritt zu tun haben? Berechnen Sie die Wahrscheinlichkeit dafür, dass an einem normalen Tag sechs oder mehr Sennenkäppli verkauft werden und urteilen Sie.
- 7–22 Ein Städtchen feiert in einem halben Jahr sein tausendjähriges Bestehen. Der hochwohllöbliche Magistrat beschliesst, jedem Kind, das am Jubeltag zur Welt kommt, 1000 Dukaten zu schenken. Der Schatzmeister wird beauftragt, einen Betrag zu budgetieren, der mit einer Wahrscheinlichkeit von 90% ausreicht. Wie gross muss dieser Betrag sein, wenn in unserm Städtchen im Mittel zwei Kinder pro Tag geboren werden?
- 7–23 Ein Vorplatz ist mit 500 gleichgrossen Steinplatten gepflastert. Es regnet ganz kurz, so dass die einzelnen Tropfen noch zu erkennen sind. 10 Platten sind vollständig trocken geblieben. Wieviel Regentropfen sind schätzungsweise insgesamt auf die 500 Platten herniedergegangen?
- 7–24 Fortsetzung der “Rosinenaufgabe” 7.3.4.B: Wieviele Rosinen muss man in den Teig tun, damit in jedem Brötchen mit 98%-iger Wahrscheinlichkeit mindestens zwei Rosinen sind? (falls bekannt, verwenden Sie ein numerisches Verfahren zur Lösung von Gleichungen; sonst stehen lassen und mit der Lösung vergleichen.)