Peter Hartmann

Mathematik für Informatiker

7. Auflage

Lösungsskizzen zu Teil 3: Wahrscheinlichkeitsrechnung und Statistik

19 Wahrscheinlichkeitsräume

Verständnisfragen

1. Welches sind die charakteristischen Eigenschaften eines Laplace-Raums?

Der Raum ist endlich, die Wahrscheinlichkeit für alle Elementarereignisse ist gleich.

2. Unter welchen Bedingungen gilt $p(A \cap B) = p(A)p(B)$. Unter welchen Bedingungen gilt $p(A \cup B) = p(A) + p(B)$?

$p(A \cap B) = p(A)p(B)$ gilt, wenn A und B unabhängig sind, $p(A \cup B) = p(A) + p(B)$ gilt, wenn A und B disjunkt sind.

3. Wenn A, B und A, C jeweils voneinander unabhängige Ereignisse sind, sind dann auch B und C voneinander unabhängig?

Das muss nicht sein. Es könnte ja z.B. $B = C$ sein.

4. Durch ein Labyrinth mit zwei Ausgängen können Ratten geschickt werden. Hinter einem der Ausgänge ist Futter zu finden. Ist das Ereignis „Ratte findet das Futter" unter folgenden Bedingungen ein Bernoulliexperiment:

   a) Eine (kluge) Ratte wird 50 Mal durch das Labyrinth geschickt.

   b) 50 Ratten werden nacheinander durch das Labyrinth geschickt

a) ist kein Bernoulliexperiment, denn die Ratte lernt dazu und die Wahrscheinlichkeit ändert sich. b) ist ein Bernoulliexperiment, da die einzelnen Stufen unabhängig voneinander sind.

5. In einer sehr großen Urne mit sehr vielen Kugeln sind die Experimente „Ziehen mit Zurücklegen" und „Ziehen ohne Zurücklegen" sehr ähnlich. Warum?

Weil sich die Ausgangssituation nach einem Zug nur geringfügig ändert.

6. Eine Wahlumfrage kann auch als Urnenexperiment angesehen werden. Handelt es sich hier um ein Experiment mit oder ohne Zurücklegen?

Eigentlich „ohne Zurücklegen", denn jeder Wähler wird nur einmal befragt. Rechnerisch ist es aber einfacher mit dem Experiment „mit Zurücklegen" zu arbeiten.

Übungsaufgaben

1. Lassen Sie Ihren Rechner das Buffon'sche Nadelexperiment durchführen und bestimmen Sie die Zahl π auf 4 Dezimalstellen genau. Wie viele Versuche benötigen Sie? Führen Sie das Experiment mehrmals durch und vergleichen Sie jeweils die Anzahl der durchgeführten Versuche.

   Das folgende Programm führt das Buffon'sche Nadelexperiment durch und gibt den Näherungswert für π aus. Später lernen Sie, die Genauigkeit dieses Wertes einzuschätzen.

```
def buffon(versuche):
    treffer = 0
    for i in range(0,versuche):
        winkel = RR.random_element(0,1000)
        abstand = RR.random_element(0,1)
        if abstand < abs(sin(winkel)):
            treffer += 1
    return 2*versuche/treffer.n()
```

   Bei der Zufallszahl für den Winkel habe ich irgendeine Zufallszahl zwischen 0 und 1000 gewählt. Es kommt ja nur auf den Sinus an.

2. In einer Schachtel mit 100 Losen befinden sich 40 Nieten. Sie kaufen 6 Lose. Wie groß ist die Wahrscheinlichkeit, dass Sie mindestens 3 Gewinne erhalten?

   Es handelt sich um ein Experiment "Ziehen ohne Zurücklegen". Die 40 Nieten sind die schwarzen Kugeln von insgesamt 100 Kugeln, 6 werden gezogen:

$$p(\text{Anzahl der Nieten } = k) = \binom{40}{k}\binom{60}{6-k} \Big/ \binom{100}{6}.$$

$$p(\text{Anzahl Nieten} = 0) = 0.0420,\ p(\text{Anzahl Nieten} = 1) = 0.1832,$$

$$p(\text{Anzahl Nieten} = 2) = 0.3191,\ p(\text{Anzahl Nieten} = 3) = 0.2836.$$

$$p(\text{Anzahl Nieten} <= 3) = 0.83 \text{ (die Summe)}.$$

3. Nehmen Sie die Hypothese als wahr an, dass eine Toastscheibe beim Herunterfallen genauso oft auf die Butterseite wie auf die ungebutterte Seite fällt. Wie groß ist dann die Wahrscheinlichkeit, dass bei 100 Versuchen der Toast mehr als 52 Mal auf die Butterseite fällt?

   Dies ist ein Bernoulliexperiment vom Umfang 100, das heißt:

$$p(A = k) = b_{100,1/2}(k) = \binom{100}{k} \cdot \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^{100-k}$$. Damit ergibt sich:

   $p(X = 51) = 0.0780$, $p(X = 52) = 0.0735$.

   Weiter ist $p(X < 50) + p(X > 50) + p(X = 50) = 1$. Aus Symmetriegründen gilt $p(X > 50) = p(X < 50)$, das heißt $p(X > 50) = (1 - p(X = 50))/2 = 0.4602$. Dann ist $p(X > 52) = p(X > 50) - p(X = 51) - p(X = 52) = 0.3087$.

4. Eine Krankheit tritt bei 0.5 % der Bevölkerung auf. Ein Test findet 99% der Kranken (Test positiv) spricht aber auch bei 2% der gesunden Bevölkerung an. Wie groß ist die Wahrscheinlichkeit, dass eine getestete Person krank ist, wenn der Test positiv ausgefallen ist?

   Wenden Sie den Satz von Bayes an: Die Ereignisse sind: B_1 = Patient gesund, B_2 = Patient krank, A = Test positiv. Es ist $p(B_1) = 0.995$, $p(B_2) = 0.005$, $p(A|B_1) = 0.02$, $p(A|B_2) = 0.99$. Eingesetzt in die Formel von Bayes erhalten Sie $p(B_2|A) = 0.199$.

5. In einer Klausur wird ein Multiple-Choice-Test durchgeführt. Auf eine Frage sind n Antworten möglich, genau eine ist richtig. Gut vorbereitete Studenten kreuzen die richtige Antwort an, schlecht vorbereitete kreuzen zufällig an. $(p\cdot 100)$% der Studenten sind gut vorbereitet. Mit welcher (bedingten) Wahrscheinlichkeit stammt ein richtiges Ergebnis von einem gut vorbereiteten Studenten?

   Nochmal der Satz von Bayes, jetzt mit den Ereignissen B_1 = gut vorbereitet, B_2 = schlecht vorbereitet, A = richtiges Ergebnis und den zugehörigen Wahrscheinlichkeiten $p(B_1) = p$, $p(B_2) = 1-p$, $p(A|B_1) = 1$, $p(A|B_2) = 1/n$. Eingesetzt in die Formel von Bayes ergibt sich:

   $$p(B_1|A) = \frac{p}{p + \frac{1}{n}(1-p)}.$$

6. Bei 4000 Ziehungen im Zahlenlotto 6 aus 49 wurde die Zahlenreihe 15, 25, 27, 30, 42, 48 zweimal gezogen: am 20.12.1986 und am 21.6.1995. Dies erregte unter den Lottospielern ziemliches Aufsehen. Rechnen Sie nach, ob dieses Ereignis wirklich unwahrscheinlich war.

   Dies ist das Geburtstagsproblem (siehe Abschnitt 19.4) mit $\binom{49}{6} = 13983816$ verschiedenen möglichen Ergebnissen (den Tagen) und 4000 Versuchen. Die Formel ist einfach, man tut sich nur mit dem Taschenrechner etwas schwer. Das folgende kleine Programm berechnet die Wahrscheinlichkeit, dass bei 4000 Versuchen zweimal das gleiche Resultat gezogen wird zu 0.436.

```
erg = binomial(49,6)
anz = 4000

w = 1
for i in range(0,anz):
    w = w*(erg-i)/erg
show(1-w.n())
```

7. Berechnen Sie die Wahrscheinlichkeit dafür im Lotto einen Dreier, Vierer oder Fünfer zu erhalten.

   Nochmal ein Urnenexperiment "Ziehen ohne Zurücklegen". Die Urne enthält 49 Kugeln, 6 davon sind weiß – das sind die 6 gezogenen Kugeln. Die Wahrscheinlichkeit, dass mein Tipp k weiße Kugeln enthält (also k richtige) ist

   $$\binom{43}{6-k}\binom{6}{k}\Big/\binom{49}{6}.$$

   Daraus ergibt sich:

   $$p(3 \text{ Richtige}) = 1.77\cdot 10^{-2},\ p(4 \text{ Richtige}) = 9.69\cdot 10^{-4},\ p(5 \text{ Richtige}) = 1.84\cdot 10^{-5}.$$

8. Das folgende Spiel wurde in ähnlicher Form in einer amerikanischen Fernsehquizsendung durchgeführt: Der Kandidat steht vor drei Türen. Hinter einer davon befindet sich ein Auto, hinter den anderen beiden ein Schaf. Der Kandidat darf eine der Türen wählen, aber noch nicht öffnen. Der Quizmaster öffnet eine der beiden anderen Türen, und zwar eine hinter der ein Schaf steht. Nun darf sich der Kandidat zwischen den beiden geschlossenen Türen noch einmal wählen. Er erhält, was sich hinter dieser Tür befindet. Berechnen Sie die Wahrscheinlichkeiten für einen (Auto-)Gewinn bei den folgenden Vorgehensweisen für die zweite Entscheidung:

   a) Der Kandidat wirft eine Münze.
   b) Der Kandidat bleibt immer bei seiner ursprünglichen Entscheidung.
   c) Der Kandidat ändert immer seine ursprüngliche Wahl.

Strategie a): $p = 1/2$.

Strategie b): Mit $p = 1/3$ erwischt der Kandidat das Auto bei der ersten Wahl. Da er bei seiner Entscheidung bleibt, bleibt auch p erhalten, also $p = 1/3$.

Strategie c): Mit $p = 2/3$ steckt das Auto hinter einer der beiden am Anfang nicht gewählten Türen, also auch mit $p = 2/3$ hinter der vom Quizmaster nicht geöffneten Tür, also $p = 2/3$.

Wer's nicht glaubt: Schreiben Sie ein Programm

20 Zufallsvariable

Verständnisfragen

1. Erklären Sie was eine (diskrete) Zufallsvariable, eine Verteilung und eine Verteilungsfunktion ist.

Diskrete Zufallsvariable: Funktion vom Wahrscheinlichkeitsraum nach $\mathbb{R}$ mit abzählbarem Wertebereich. Verteilung: Eine Abbildung von der (diskreten) Wertemenge nach [0,1], die jedem Wert seine Wahrscheinlichkeit zuordnet. Verteilungsfunktion: Eine Funktion von $\mathbb{R}$ nach [0,1], die jedem x die Wahrscheinlichkeit zuordnet, dass die Zufallsvariable einen Wert $\leq x$ annimmt.

2. Warum dürfen in einer diskreten Zufallsvariablen höchstens abzählbar viele Funktionswerte auftreten?

Weil man die Funktionswerte aufaddieren können muss.

3. Beim Würfeln mit 2 Würfeln soll die Zufallsvariable X_i das Ergebnis des i-ten Würfels sein. Sind die Zufallsvariablen X_1 und X_2 voneinander unabhängig? Sind die Zufallsvariablen X_1 und $(X_1 + X_2)$ voneinander unabhängig?

X_1 und X_2 sind voneinander unabhängig, X_1 und $(X_1 + X_2)$ nicht.

4. Hat jede diskrete Zufallsvariable einen Erwartungswert?

Nein, bei abzählbar unendlichem Wertebereich kann es sein, dass die unendliche Summe, mit der der Erwartungswert gebildet wird, nicht existiert. Etwas präziser: die Reihe ist nicht konvergent.

5. Welche Art der Abweichung vom Erwartungswert beschreibt die Varianz einer Zufallsvariablen?

Die mittlere quadratische Abweichung.

6. Seien X und Y unabhängige Zufallsvariable. Gilt dann $\mathrm{Var}(X \cdot Y) = \mathrm{Var}(X) \cdot \mathrm{Var}(Y)$?

Im Allgemeinen nicht.

Übungsaufgaben

1. Berechnen Sie den Erwartungswert, die Varianz und die Standardabweichung der Augensumme beim Würfeln mit 5 Würfeln.

   Erwartungswert beim Würfeln mit 1 Würfel ist 3.5. Ist X_i die Zufallsvariable "Würfeln mit Würfel i", so ist die Augensumme $S = X_1 + X_2 + ... + X_5$ und $E(S) = 5 \cdot 3.5 = 17.5$. Da die X_i alle unabhängig sind, kann auch die Varianz von S als Summe der Varianzen berechnet werden, das heißt $\mathrm{Var}(S) = 5 \cdot \mathrm{Var}(X)$, und $\mathrm{Var}(X) = 1/6 \cdot (1^2 + 2^2 + ... + 6^2) - 3.5^2 = 2.9167$ nach Satz 20.15. Daraus: $\mathrm{Var}(S) = 14.5833$. Die Standardabweichung ist die Wurzel daraus.

2. Die Zufallsvariable P beschreibt das Produkt der Augen bei zweimaligem Würfeln. Zeichnen Sie in einem Histogramm die Verteilung von P. Berechnen Sie den Erwartungswert und die Varianz von P.

   Zum Zeichnen bleibt Ihnen nichts anderes übrig, als alle möglichen Produkte auszurechnen und jeweils zu zählen wie oft es vorkommt. Eine ziemlich langweilige Arbeit. Den Erwartungswert erhalten Sie durch Multiplikation der Erwartungswerte, also $E(P) = 3.5^2 = 12.25$, da die beiden Würfe unabhängig sind. Die Varianz ist nicht multiplikativ, Sie müssen Sie entsprechend dem Satz 20.15 ausrechnen! Zur Kontrolle: $\mathrm{Var}(P) = 79.965$.

3. Die Zufallsvariable X beschreibe die Anzahl der Buben im Skat nach dem Austeilen (32 Karten mit vier Buben, 3 Spieler erhalten je 10 Karten, 2 Karten sind im Skat). Bestimmen Sie die Verteilung und den Erwartungswert von X,

   a) ohne Information über die Kartenverteilung bei den 3 Spielern,
   b) mit dem Wissen, dass der erste Spieler keinen Buben erhalten hat.

   Hier die Berechnung der Wahrscheinlichkeiten im Fall a). B_1 sei das Ereignis, dass die erste Karte des Skats ein Bube ist, B_2 die, dass die zweite ein Bube ist. Dann ist

   $$p(B_1 \cap B_2) \underset{\substack{\uparrow \\ \text{Def. 19.6}}}{=} p(B_1) \cdot p(B_2 \mid B_1) = \frac{4}{32} \cdot \frac{3}{31} \approx 0.012.$$

   Genauso: $$p(\overline{B}_1 \cap \overline{B}_2) \underset{\substack{\uparrow \\ \text{Def. 19.6}}}{=} p(\overline{B}_1) \cdot p(\overline{B}_2 \mid \overline{B}_1) = \frac{28}{32} \cdot \frac{27}{31} \approx 0.782.$$

   Die Wahrscheinlichkeit von einem Buben im Skat ist die Summe $p(\overline{B}_1 \cap B_2) + p(B_1 \cap \overline{B}_2)$.
   Es ist $p(\overline{B}_1 \cap B_2) = p(\overline{B}_1) \cdot p(B_2 \mid \overline{B}_1) = \frac{28}{32} \cdot \frac{4}{31} = p(B_1 \cap \overline{B}_2)$, daher $p(1\text{ Bube}) \approx 0.226$.

   Damit ist der Erwartungswert $E(X) = 0 \cdot 0.782 + 1 \cdot 0.226 + 2 \cdot 0.012 = 0.25$.

   Für b) muss jeweils nur 32 durch 22 ersetzt werden.

   Aufgaben dieser Art lassen sich vorteilhaft mit Hilfe von Baumdiagrammen lösen, auf die ich im Buch leider nicht eingehen konnte.

4. Beim Roulettespiel setzt ein Spieler 10 € auf Rot. Gewinnt er (18 von 37 Feldern sind rot), so erhält er 20 € Gewinn (Reingewinn = 10 €).Verliert er, so verdoppelt er seinen Einsatz solange bis er gewinnt oder bis das Einsatzlimit der Bank von 1000 € erreicht ist. Beschreiben Sie die Zufallsvariable "Reingewinn" bei dieser Strategie und errechnen Sie Erwartungswert und Varianz dieses Spiels.

   Sofern vor dem Limit rot kommt erhält der Spieler: 10 € Reingewinn. Wenn 7 Mal hintereinander nicht rot kommt macht er 1270 € Verlust (10+20+40+80+160+320+640). Die Wahrscheinlichkeit für 7 Mal nicht rot ist $(19/37)^7 = 0.009416$, die Wahrscheinlichkeit für wenigstens einmal rot ist $1 - 0.009416 = 0.9906$. Damit erhalten wir:

   $E(X) = 10\cdot 0.9906 - 1270\cdot 0.009416 = -2.05$,

   $\mathrm{Var}(X) = 10^2\cdot 0.9906 + 1270^2\cdot 0.009416 - 2.05^2 = 15290$.

5. Verwenden Sie ein Tabellenkalkulationsprogramm oder ein Mathematiktool um für verschiedene Werte von n und p Histogramme für die Funktion $b_{n,p}(k)$ zu zeichnen.

   Hier ist keine Mathematik gefragt, sondern Ihre Fähigkeit mit Excel oder einem Zeichenprogramm umzugehen. Darum habe ich hier auch keinen Lösungsvorschlag.

6. Die folgende Anzeige erschien vor einigen Jahren in der Süddeutschen Zeitung:

   Geschlechtsvorhersage mit Erfolgsgarantie!
   Aus ca. 10 Zeilen Handschrift der werdenden Mutter sagen wir Ihnen schon im 2. Monat mit 100-%iger Sicherheit das Geschlecht voraus. Honorar nur DM150.– Geld garantiert zurück, wenn unsere Vorhersage nicht zutreffen sollte.

   Inst. f. wissenschaftl. Vorhersagen
   Postfach XXXXX

   Nehmen wir an, dass die wissenschaftliche Vorhersage aus einem Münzwurf besteht. Die Wahrscheinlichkeit einer Mädchengeburt ist 0.465. Berechnen Sie das ungefähre Jahreseinkommen des Hellsehers, wenn ca. 100 Anfragen pro Monat eintreffen. Wie kann er sein Einkommen (bei gleichem Tarif) verbessern? (Stellen Sie hierzu eine Zufallsvariable "Gewinn" auf und berechnen Sie den Erwartungswert.)

   Die Zufallsvariable X sei der Gewinn: $X = 100$ falls Kopf und Mädchen oder Zahl und Junge, $X = 0$ sonst. Daher ist $p(X = 100) =(0,5\cdot 0.465 + 0.5\cdot 0.535) = 0.5$ und $E(X) = 0.5\cdot 100 = 50$. Damit beträgt das Jahreseinkommen ca. $1200\cdot 50 = 60000$ €.

   Das Einkommen verbessert sich, wenn er immer Junge vorhersagt: $p(X{=}100) = 0.535$, $E(X) = 53.5$.

7. Berechnen Sie die Entropie und die mittlere Wortlänge des Huffman Codes aus der Aufgabe 6 in Kapitel 11.

Der Code für das Alphabet $\{a, b, c, d, e, f, g, h\}$ bei der Häufigkeitsverteilung $\{4, 6, 7, 8, 10, 15, 20, 30\}$ ($a = 4\%$, $b = 6\%$, ... , $h = 30\%$) lautete:

a = 0000, b = 0001, c = 1000, d = 1001, e = 001, f = 101, g = 01, h = 11.

Die Formel für die Entropie lautet: $H(Q) := \sum_{i=1}^{n} p_i \cdot \log_2(1/p_i)$, damit ist hier

$$H(Q) = 0.04 \cdot \log_2\left(1/0.04\right) + \ldots + 0.3 \cdot \log_2\left(1/0.3\right) = 2.7176$$

Die mittlere Codewortlänge beträgt $L(Q) = \sum_{i=1}^{n} p_i \cdot l(x_i)$, das ergibt hier

$$L(Q) = 0.04 \cdot 4 + \ldots + 0.3 \cdot 2 = 2.75$$

21 Wichtige Verteilungen

Verständnisfragen

1. Unter welchen Umständen kann die Binomialverteilung durch die Poisson-Verteilung angenähert werden?

Großes n und kleines p.

2. Unter welchen Umständen kann die Binomialverteilung durch die Normalverteilung angenähert werden?

Großes n und nicht zu kleines p, $(1-p)$. Genauer: $np > 5$, $n(1-p) > 5$.

3. Warum tritt die Normalverteilung in vielen Anwendungsfällen auf?

Verteilungen die sich aus vielen einzelnen Einflüssen additiv zusammensetzen sind oft normalverteilt (der zentrale Grenzwertsatz).

4. Was sind die Gemeinsamkeiten von geometrischer Verteilung und Exponentialverteilung?

Beide sind gedächtnislos.

5. Die Wartezeit auf ein zufällig vorbeikommendes Taxi ist exponentialverteilt. Wie sieht es mit der Wartezeit auf einen Linienbus aus?

Die ist nicht exponentialverteilt, da der Bus einem Fahrplan folgt.

6. Wie hängen Binomialverteilung und hypergeometrische Verteilung zusammen?

Je größer die Grundgesamtheit wird, umso ähnlicher werden die beiden Verteilungen.

7. Die Normalverteilung ist die Grenzverteilung der Binomialverteilung, wenn n immer größer wird. Wie verhält sich diese Konvergenz in Abhängigkeit vom Parameter p der Binomialverteilung?

Je näher p an 0.5 liegt, umso besser ist die Konvergenz.

Übungsaufgaben

1. Bei einem leichten Nieselregen fallen auf einen Gitterrost auf dem Boden 100000 Wassertropfen. Der Gitterrost ist 1 m^2 groß, die Löcher im Gitter 1 cm^2.

   Gehen Sie bei den folgenden Aufgaben davon aus, dass die Anzahlen der Tropfen, die auf verschiedene Flächenstücke des Gitters fallen, voneinander unabhängig sind.

   Beschreiben Sie für jede Teilaufgabe ausführlich, welches "Experiment" durchgeführt wird und für welche Ereignisse Wahrscheinlichkeiten berechnet werden. Beschreiben Sie die auftretenden Zufallsvariablen und begründen Sie, welche Verteilung Sie für welche Teilaufgabe verwenden.

   a) Bestimmen Sie die Wahrscheinlichkeit, dass in ein bestimmtes Gitterloch genau 10 Regentropfen fallen und die Wahrscheinlichkeit dafür, dass in jedes Loch mindestens 1 Tropfen fällt.

   b) Das Gitter wird in 100 Teile unterteilt. Berechnen Sie die Wahrscheinlichkeit, dass in ein solches Flächenstück zwischen 900 und 1100 Tropfen fallen, sowie die Wahrscheinlichkeit, dass es ein Stück gibt, auf das mehr als 1050 bzw. mehr als 1100 Tropfen fallen.

   c) Auf 10 nebeneinanderliegende Löcher fallen zufällig genau 100 Tropfen. Berechnen Sie die Wahrscheinlichkeit, dass in eines dieser 10 Löcher genau 10 Tropfen fallen.

Dies ist ein Bernoulliexperiment mit 100 000 Versuchen (die 100 000 Tropfen). Das untersuchte Ereignis A und $p(A)$ ist in verschiedenen Teilaufgaben verschieden, jedenfalls ist die Zufallsvariable $X = k \Leftrightarrow$ "Ereignis A tritt k-mal ein" binomialverteilt.

Zu a): Es sei das Ereignis A = "Tropfen fällt in ausgewähltes Loch". $p(A) = 1/10\,000$ (da es 10 000 Löcher gibt). Es ist $np = 10$ und $n > 1500 \cdot p$, daher kann durch Poissonverteilung angenähert werden:

$$p(X = 10) = (10^{10}/10!)e^{-10} = 0.1251.$$

Mit der Poissonverteilung erhalten wir die Näherung $p(X = 0) = e^{-10}$ also ist die Wahrscheinlichkeit $p(X \geq 1) = 1 - e^{-10}$. Die Wahrscheinlichkeit dass in das 1. Loch und in das 2. Loch und ... in das 10 000. Loch mehr als 0 Tropfen fallen ist wegen der Unabhängigkeit $(1 - e^{-10})^{10000} = 0.63507$.

Hier kann man auch mit der Binomialverteilung direkt rechnen: $p(X = 0) = 1 \cdot p^0 \cdot p^{100000} =: q$. Dann ist $(1 - q)^{10000} = 0.63522$.

Zu b): Wieder ist Ereignis A = "Tropfen fällt in ausgewähltes Flächenstück", aber jetzt $p(A) = 1/100$. $np = 1000$, $n(1 - p) = 99000$. Daher kann hier durch die Normalverteilung angenähert werden:

$$p(900 \leq X \leq 1100) = \Phi((1100 - 1000 + 0{,}5)/\sqrt{990}) - \Phi((900 - 1000 - 0{,}5)/\sqrt{990}) =$$
$$\Phi(3.19) - \Phi(-3.19) = 2\Phi(3.19) - 1 = 0.9986$$

Zur zweiten Hälfte von b): Aufgepasst, berechnen Sie nicht $p(X > 1050)$ sondern $p(X \leq 1050) = \Phi((1050 - 1000 + 0,5)/\sqrt{990}) = 0.946$. Dann ist die Wahrscheinlichkeit, dass in alle 100 Flächenstücke höchstens 1050 Tropfen fallen = 0.946^{100} und die Wahrscheinlichkeit, dass in mindestens eines mehr fallen $1 - 0.946^{100} = 0.996$.

Ebenso ist p(mehr als 1100 Tropfen in ein Flächenstück) = $1 - \Phi(3.19)^{100} = 0.067$.

Zu c): Sei zunächst eines der 10 Löcher fest. Jetzt ist im Bernoulliexperiment $n = 100$, A = "Tropfen trifft das ausgewählte Loch", $p = 0.1$. Hier darf keine Näherung genommen werden sondern nur die Binomialverteilung selbst:

$$p(X{=}10) = \binom{100}{10} 0.1^{10} \cdot 0.9^{90} = 0.1319\,.$$

2. Ein Bäcker bäckt 100 Olivenbrote. In den Teig wirft er 400 ganze Oliven. Bestimmen Sie die Wahrscheinlichkeit dafür, dass zwei bis sechs Oliven in einem Brot sind und dafür, dass Sie ein Brot ohne Oliven erwischen.

   Wieder liegt Binomialverteilung vor: $n = 400$, $p = 1/100$. $np \leq 10$ und $n \geq 1500 \cdot p$, es kann also mit der Poissonverteilung gerechnet werden. $\lambda = np = 4$. Damit $p(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Für $k = 0$ erhalten wir: $p(X = 0) = 0.018$, Die Summe $p(X = 2) + p(X = 3) + \ldots + p(X = 6)$ ergibt etwa 0.80.

3. Eine Fluggesellschaft weiß, dass im Schnitt 10% der gebuchten Flugplätze storniert werden und überbucht daher um 5%. Für ein Flugzeug mit 100 Plätzen verkauft sie also 105 Tickets. Wie groß ist die Wahrscheinlichkeit, dass zum Abflug mehr Passagiere kommen als Plätze vorhanden sind?

   Ein Bernoulliexperiment mit 105 Versuchen, A ist das Ereignis "Passagier kommt zum Flug". $p(A) = 0.9$. $p(X = k)$ ist binomialverteilt. Annäherung durch die Normalverteilung ist nach der Rechenregel 21.17 möglich, ist aber hier nicht genau:

   Rechnung mit Näherung: ($n = 105$, $p = 0.9$) Rechenregel 21.17:

$$p(X \leq 100) = \Phi\left(\frac{100 - 105 \cdot 0,9 + 0,5}{\sqrt{105 \cdot 0,9 \cdot 0,1}}\right) = \Phi(1.95) = 0.9744\text{, also } p(X > 100) = 0.0256\,.$$

   Exakte Rechnung:

$$p(X > 100) = \sum_{k=101}^{105} b_{105;0,9}(k) = 0.0167\,.$$

4. An einer Autobahnmautstelle treffen zwischen 7.00 und 17.00 Uhr 1000 Autos zufällig verteilt ein. Wenn mehr als 5 in einer Minute eintreffen, ist die Mautstelle überlastet. In wie viel Prozent der Intervalle ist dies der Fall?

   Hier muss der Poisson-Prozess verwendet werden: λ = 100/h = 1.667/min. p(k Autos treffen in einer Minute ein) = $e^{-\lambda t}\dfrac{(\lambda t)^k}{k!}$ mit λt = 1.667/min·1 min = 1.667. $\Rightarrow p(0) = 0.189$, $p(1) = 0.315$, $p(2) = 0.262$, $p(3) = 0.146$, $p(4) = 0.061$, $p(5) = 0.020$. p(mehr als 5) = 1 – Summe $\approx$ 0.007.

5. Die telefonische Bestellzentrale eine Versandhändlers erhält vormittags zwischen 8 und 13 Uhr durchschnittlich 500 Bestellungen, nachmittags zwischen 14 und 18 Uhr 800 Bestellungen, jeweils etwa gleichverteilt über die 5 bzw. 4 Stunden. Wartezeiten treten auf, wenn mehr als 3 Anfragen in einer Minute eintreffen. Berechnen Sie für den Vormittag und den Nachmittag die Wahrscheinlichkeit, dass in einem Intervall von einer Minute dieser Fall eintritt.

   Noch einmal der Poisson-Prozess: $p(X=k)=\dfrac{(\lambda t)^k}{k!}e^{-\lambda t}$ Vormittag: λ = 100/h = 1.667/min, Nachmittag: λ = 200/h = 3.33/min, t = 1 min. Vormittag ist die Summe $p(X = 0)$ +... + $p(X = 3)$ = 0.912, Nachmittag ist diese Summe 0.572. Damit liegt am Vormittag Überlast in 8.8 % der Intervalle vor, am Nachmittag in 42.8 % der Intervalle.

6. Eine Maschine produziert Chips mit einem (zufällig verteilten) Ausschussanteil von 9%.

   a) Nach welcher Verteilung bestimmt sich die Anzahl der ungenießbaren Chips in einer Anzahl n von produzierten Chips. Begründen Sie dies.

   b) Berechnen Sie mit einer geeigneten Näherung die Wahrscheinlichkeit für 110 oder mehr defekte Chips in einer Produktion von 1000 Chips. Warum dürfen Sie die Näherung verwenden?

   Die Produktion stellt ein Bernoulliexperiment dar, nach Satz 18.14 ist daher die Anzahl der schlechten Chips $b_{n,0.09}$-verteilt. Für n = 1000 ist $n \cdot p$ = 90 und $n \cdot (1-p)$ = 910, nach der Rechenregel 20.16 kann durch die Normalverteilung angenähert werden und es ist

$$p(X \geq 110) = 1 - p(X < 110) = 1 - \Phi\left(\frac{110-90-0.5}{\sqrt{81{,}9}}\right) = 1 - \Phi(2.15) = 1 - 0.984 = 0.016 .$$

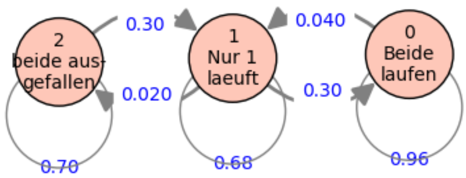
7. Ein Frachtschiff hat zwei Dieselgeneratoren, die in der Regel beide laufen. An einem Reisetag besteht die Wahrscheinlichkeit von 2%, dass eines der beiden Aggregate ausfällt. Das macht nichts, das Schiff kann auch mit einem Aggregat weiterfahren. Ist ein Aggregat ausgefallen, fällt aber mit 2% Wahrscheinlichkeit am nächsten Tag auch noch das zweite Aggregat aus. Ein ausgefallenes Aggregat können die Maschinisten bis zum nächsten Tag mit einer Wahrscheinlichkeit von 30% wieder instand setzen, sind beide ausgefallen reduziert sich die Reparaturwahrscheinlichkeit auf 15% pro Aggregat. Wir wollen vernachlässigen, dass an einem Tag auch beide Aggregate ausfallen könnten.

   a) Definieren Sie für diesen Prozess Zustände und Zustandsübergänge. Zeichnen Sie einen Zustandsgraphen.

   b) Wenn das Schiff losfährt, laufen beide Aggregate. Die Fahrt von China nach Europa dauert 48 Tage. Wie groß ist die Wahrscheinlichkeit, dass am letzten Tag ein Aggregat bzw. beide Aggregate ausgefallen sind? Verwenden Sie sage oder ein anderes Tool zur Berechnung.

   Anmerkung: Wenn beide Aggregate ausgefallen sind, gibt es noch ein Notstromaggregat, mit dem das Schiff zumindest manövrierfähig bleiben sollte und nicht vor Hamburg in der Elbe strandet

Ein mit sage gezeichnetes Zustandsübergangsdiagramm und die zugehörige Matrix könnte so ausschauen:



$$\begin{pmatrix} 0.960 & 0.0400 & 0.000 \\ 0.300 & 0.680 & 0.0200 \\ 0.000 & 0.300 & 0.700 \end{pmatrix}$$

Ist A die Zustandsübergangsmatrix, so ergibt:

```
a = vector([1,0,0]) # beide Aggregate laufen
b = a*(A^48) # Zustand nach 48 Tagen

# Ausgabe
show(b)

(0.875486443045213,0.116731473913417,0.00778208304136855)
```

also ist mit 11.7% ein Aggregat ausgefallen, mit 0.778% sind beide ausgefallen.

22 Statistische Verfahren

Verständnisfragen

1. Was versteht man unter einer Parameterschätzung?

Ist p ein unbekannter Parameter eines Zufallsexperiments, so sucht man eine Schätzfunktion für diesen Parameter, der selbst wieder eine Zufallsvariable ist. Die Auswertung der Schätzfunktion liefert dann einen Schätzwert für den Parameter. Die Schätzfunktion ist besonders nützlich, wenn sie erwartungstreu und konsistent ist.

2. Wenn eine Stichprobe durch ein Zufallsexperiment gewonnen wurde, gibt es einen Zusammenhang zwischen dem Erwartungswert und der Varianz der Zufallsvariablen und dem Mittelwert und der Varianz der Stichprobe. Wie lautet dieser Zusammenhang?

Mittelwert beziehungsweise Varianz der Stichprobe sind Schätzwerte für den Erwartungswert beziehungsweise die Varianz der Zufallsvariablen.

3. Sie wollen ein Konfidenzintervall berechnen. Wenn das Konfidenzniveau angehoben wird, wird dann das Konfidenzintervall kleiner oder größer? Wenn der Umfang der Stichprobe vergrößert wird, wird das Konfidenzintervall kleiner oder größer?

Höheres Konfidenzniveau heißt größeres Konfidenzintervall. Vergrößerung der Stichprobe hat eine Verkleinerung des Konfidenzintervalls zur Folge.

4. Was ist der Fehler erster und zweiter Art in einem Hypothesentest?

Fehler erster Art: die Hypothese wird zu Unrecht abgelehnt. Fehler zweiter Art: die Hypothese wird zu Unrecht nicht abgelehnt.

5. Wie sollte man eine Hypothese formulieren, wenn man eine Vermutung bestätigen will?

Immer versuchen die Gegenhypothese zu widerlegen.

6. Wenn in einem Hypothesentest die Irrtumswahrscheinlichkeit verkleinert wird, wird dann der Ablehnungsbereich kleiner oder größer?

Wenn die Irrtumswahrscheinlichkeit verkleinert wird, wird auch der Ablehnungsbereich kleiner.

7. Kann man einen Hypothesentest mehrmals durchführen, wenn einem das Ergebnis eines durchgeführten Tests nicht gefällt?

Das kann man natürlich schon – aber das hat nichts mehr mit dem Hypothesentest mit den vorher gewählten Parametern zu tun. Sehen Sie die letzte Seite des Kapitels an.

Übungsaufgaben

1. 10-maliges Würfeln mit 2 Würfeln ergibt die Paare (2,4), (5,6), (2,2), (6,2), (2,6), (5,1), (5,6), (6,4), (4,4), (3,1).
   Berechnen Sie Mittelwert und Varianz für die Augensummen dieser Stichprobe. Geben Sie einen Schätzwert für die Wahrscheinlichkeit eines Pasch (zwei gleiche Zahlen) und für den Erwartungswert der Augensumme ab.
   Warum können Sie mit unserer Standardformel kein Konfidenzintervall für p(Pasch) berechnen?
   Zeigen Sie, dass bei idealen Würfeln die Wahrscheinlichkeit dafür, dass bei 10 Würfen 0, 1, 2 oder 3 Pasch vorkommen, bei über 90 % liegt.

   Der Mittelwert ist 7.6, die Varianz = $\frac{1}{9}(\sum x_i^2 - 10(MW)^2)$ = 6.71. Schätzwert für Pasch ist 0.2 (= 2/10 = relative Häufigkeit des Paschs). Schätzwert für den Erwartungswert ist der Mittelwert. Die Standardformel ist nicht anwendbar, da k und $n - k < 30$ sind.

   Die Anzahl der Paschs ist binomialverteilt. Daher:

   p(Pasch tritt k-mal ein bei 10 Versuchen) = $\binom{10}{k}\left(\frac{1}{6}\right)^k\left(\frac{5}{6}\right)^{n-k}$ $\Rightarrow p(0 \text{ Pasch}) = 0.161$, $p(1 \text{ Pasch}) = 0.323$, $p(2 \text{ Pasch}) = 0.291$, $p(3 \text{ Pasch}) = 0.155$. Die Summe ist größer als 0.92.

2. Versuchen Sie mit Hilfe eines Mathematik-Tools die beiden Zahlenbeispiele in der Hauptkomponentenanalyse (Kapitel 22.2) nach zu vollziehen.

   Das folgende sage Programm rechnet das erste Beispiel durch. Das gleiche Programm können Sie auch mit anderen Merkmalsdaten verwenden.

```
def mw(X):          #Mittelwert
    return sum(X)/len(X);

def bmi(l,g):       #BMI, auf zwei Stellen gerundet
    return RR(round (g/(l/100)^2*100)/100)

def var (X):        #Varianz
    p = 0;
    for i in range(0,len(X)):
        p = p +(X[i]-mw(X))^2
    return p / (len(X)-1)


n = 13   # Anzahl Probanden
mm = 3   # Anzahl Merkmale

L = [161,180,174,156,165,191,184,162,183,171,159,166,193]
G = [51,89,75,63,54,75,76,52,79,75,66,61,84]
B = [bmi(L[i],G[i]) for i in range(0,n)]

A = matrix(RDF,[L,G,B])

show("Merkmalmatrix")
show(A.n(digits=4))

m = []
v = []
An = copy(A)
```

```
for i in range(0,mm):
    m.append(mw(An[i]))
    v.append(sqrt(var(An[i])))

for i in range(0,mm):
    for j in range(0,len(An[i])):
        An[i,j] = (An[i,j]-m[i])/v[i]

show("Normalisierte Matrix")
show(An.n(digits = 3))

C = 1/(n-1)*An*An.transpose()
show ("Kovarianzmatrix")
show(C.n(digits = 3))

erg = C.eigenvectors_right()

import numpy
L = [erg[i][0] for i in range(0,mm)]
K = numpy.argsort(L)  # die Indizes der Eigenwerte, der Größe der Eigenwerte nach geordnet

Sp = []
EV = []
for i in range(0,mm):
    EV.append(erg[K[i]][0][0])
    Sp.append(erg[K[i]][1][0])
Sp.reverse()
EV.reverse()

show( "sortierte Eigenwerte")
show(vector(EV).n(digits = 4))
show( "Transformationsmatrix - Die Spalten sind die sortierten Eigenvektoren")
T = Matrix(Sp).transpose()
show(T.n(digits = 3))

show( "Die neue Merkmalmatrix")
NA = T.transpose()*An
show(NA.n(digits = 3))

show("Zur Kontrolle: die neue Covarianzmatrix: ausserhalb der Diagonalen steht (fast) 0")
show("In der Diagonalen stehen die Eigenwerte")
N = 1/(n-1)*NA.transpose()
show(NA*N)
```

3. Bestimmen Sie den Ausschussanteil eines Massenartikels: Eine Stichprobe von 1000 Elementen liefert 94 Ausschussteile. Bestimmen Sie zum Konfidenzniveau 95 % ein Konfidenzintervall für die unbekannte Wahrscheinlichkeit p des Ausschusses.

   $\gamma = 0.95 \Rightarrow (1+\gamma)/2 = 0.975 \Rightarrow c = 1.96$. $n = 1000$, $k = 94$. Einsetzen in die Standardformel für Konfidenzintervalle (21.11) ergibt $p_u = 0.0759$, $p_o = 0.1120$.

4. Die Wahrscheinlichkeit einer Mädchengeburt:
   2017 wurden in der Deutschland 784 901 Kinder geboren, davon 382 374 Mädchen. Bestimmen Sie zum Konfidenzniveau 99 % ein Konfidenzintervall für die Wahrscheinlichkeit einer Mädchengeburt.

   $\gamma = 0.99 \Rightarrow (1+\gamma)/2 = 0.995 \Rightarrow c = 2.57$. $n = 784901$, $k = 382374$, $n - k = 396296$. Einsetzen in die Rechenregel 22.11 ergibt $p_u = 0.4857$, $p_o = 0.4886$.

5. In einer Umfrage soll der Prozentsatz der Bevölkerung ermittelt werden, der nicht an Statistiken glaubt. Für die Vorgaben

   a) Konfidenzniveau 98%, Konfidenzintervallbreite 4%,
   b) Konfidenzniveau 96%, Konfidenzintervallbreite 4% ,

   soll ausgerechnet werden welcher Stichprobenumfang notwendig ist.

   Die Befragung wird mit 3000 Personen durchgeführt. Davon glauben 1386 nicht an Statistiken. Bestimmen Sie zu den Konfidenzniveaus 96% und 98% die zugehörigen Konfidenzintervalle.

   Für a),b) ist Rechenregel 21.12 anzuwenden. Es ist $c_a = \Phi^{-1}(1.98/2) \approx 2.33$, $c_b = \Phi^{-1}(1.96/2) \approx 2.06$ und $B = 0.04$. Daraus ergibt sich für a) $n > 3394$, für b) $n > 2652$.

   Die Konfidenzintervalle erhält man mit Rechenregel 21.11:

$$p_{u/o} = \frac{1386}{3000} \pm \frac{c_{a/b}}{3000} \cdot \sqrt{\frac{1386 \cdot 1614}{3000}}.$$

   Für $c_a = 2.33$ (98%) ist das Konfidenzintervall [0.4408, 0.4832], für $c_b = 2.06$ (96%) lautet es [0.4432, 0.4807]. Das heißt: Mit 98% Sicherheit glauben zwischen 44.08 und 48.32 % der Bevölkerung nicht an Statistiken, mit 96% Sicherheit glauben zwischen 44.32 und 48.07 % der Bevölkerung nicht an Statistiken.

6. Beim Roulettespiel ist die Wahrscheinlichkeit für eine rote Zahl *p(*rot*)* = 18/37. Sie vermuten, dass dies in einer speziellen Trommel nicht stimmt. In Listen sind die letzten 5000 Wurfergebnisse ausgehängt. Dabei ergaben sich 2355 rote Zahlen. Stellen Sie eine geeignete Hypothese auf, um Ihre Vermutung zu überprüfen. Verwenden Sie die Irrtumswahrscheinlichkeiten 2% beziehungsweise. 5% und formulieren Sie präzise das errechnete Ergebnis.

   Die Hypothese lautet: "*p(*rot*)* = 18/37". Die Rechenregel 22.13 ist anzuwenden. Dabei ist $c_1 = \Phi^{-1}(1-0.02/2) = 2.33$ und $c_2 = \Phi^{-1}(1-0,05/2) = 1.96$. Weiter ist

$$n\delta = 5000 \cdot c_i \cdot \sqrt{\frac{18}{37} \cdot \frac{19}{37} \Big/ 5000} = \begin{cases} 82.35 & \alpha = 0.02 \\ 69.27 & \alpha = 0.05 \end{cases},$$

   und damit sind die Nichtablehnungsbereiche]2350, 2515[($\alpha = 0.02$) bzw.]2363, 2502[($\alpha = 0.05$). Mit Irrtumswahrscheinlichkeit 5% wird die Hypothese abgelehnt, mit 2% Irrtumswahrscheinlichkeit steht das Resultat nicht im Widerspruch zur Hypothese.

7. Bei 4964 Ausspielungen der Lottozahlen seit 1955 ergaben sich für die Zahlen 1 bis 49 der Reihe nach die folgenden Ziehungshäufigkeiten:

606	617	628	615	602	633	607	583	630	591	608	592	551
600	576	612	608	606	602	577	585	616	600	602	642	649
635	569	581	597	640	616	628	603	609	608	600	646	610
605	629	615	651	604	564	585	607	606	638			

Widerspricht dieses Ergebnis bei einer Irrtumswahrscheinlichkeit von 1 % der Hypothese der Gleichverteilung der 49 Zahlen? Schreiben Sie ein Testprogramm.

Das folgende kleine Programm liefert als Testergebnis für chi den Wert 38.81:

```
daten = [606,617,628,615,602,633,607,583,630,591,608,592,551,600,576,612,608,606,602,
577,585,616,600,602,642,649,635,569,581,597,640,616,628,603,609,608,600,646,610,605,
629,615,651,604,564,585,607,606,638]

p = 6/49      # Wahrscheinlichkeit für eine Zahl in einer Ziehung
n = 4964      # Anzahl der Ziehungen

chi = 0
for i in range(0,49):
    chi += (daten[i] - n*p)^2
chi /= n*p
show(chi.n())
```

Die Pearson'sche Testfunktion ist χ^2_{48}-verteilt, also annähernd N(48, 96) verteilt, daher ist für die Irrtumswahrscheinlichkeit α: $p(\chi^2 < \delta) = \Phi\left(\frac{\delta - 48}{\sqrt{96}}\right) = 1 - \alpha$. Für α = 1% (0.1%) ist $\Phi^{-1}(1-\alpha) = 2.33\,(3.09)$ und $\frac{\delta - 48}{\sqrt{96}} = 2.33\,(3.09)$. Daraus ergibt sich $\delta = 70.83$ bzw. $\delta = 78.27$. Das Testergebnis liegt also mit Irrtumswahrscheinlichkeit 1% und sogar 0.1% nicht im Ablehnungsbereich, die Hypothese der Gleichverteilung kann nicht abgelehnt werden.

http://www.springer.com/978-3-658-26523-6

Mathematik für Informatiker
Ein praxisbezogenes Lehrbuch
Hartmann, P.
2019, XII, 643 S. 186 Abb., Softcover
ISBN: 978-3-658-26523-6