Kapitel 15

Wahrscheinlichkeitstheorie und Statistik

Verständnisfragen

Sachfragen

1. Erläutern Sie den Begriff der absoluten und relativen Häufigkeit einer Stichprobe!
2. Erläutern Sie den Begriff der Klassenhäufigkeit in der beschreibenden Statistik!
3. Erläutern Sie den Begriff der Summenhäufigkeit!
4. Erläutern Sie den Begriff der empirischen Dichte!
5. Erläutern Sie den Begriff des Modalwerts!
6. Erläutern Sie den Begriff des arithmetischen Mittels!
7. Erläutern Sie den Begriff des geometrischen Mittels!
8. Erläutern Sie den Begriff der Spannbreite, der Varianz und der Standardabweichung!
9. Was versteht man unter einer Kontingenztabelle?
10. Was ist die Stichproben-Kovarianz?
11. Wie ist der Bravais-Pearson-Korrelationskoeffizient definiert?
12. Was ist eine Ereignis-Algebra?
13. Erläutern Sie den Ansatz der Laplace-Wahrscheinlichkeit!
14. Was ist eine geometrische Wahrscheinlichkeit?
15. Was ist ein Wahrscheinlichkeitmaß?
16. Was versteht man unter einem Wahrscheinlichkeitsraum?
17. Was ist eine bedingte Wahrscheinlichkeit?
18. Was sagt der Satz von der totalen Wahrscheinlichkeit aus?

19. Wie lautet der Satz von Bayes?
20. Wann sind zwei Ereignisse unabhängig?
21. Was ist eine Zufallsvariable?
22. Was ist eine Verteilungsfunktion einer Zufallsvariablen?
23. Welche Eigenschaften hat eine Verteilungsfunktion?
24. Was ist eine diskrete Zufallsvariable?
25. Was ist eine stetige Zufallsvariable?
26. Erläutern Sie den Begriff der Wahrscheinlichkeitsdichte einer stetigen Zufallsvariablen!
27. Gibt es für eine stetige Zufallsvariable Ereignisse, die fast unmöglich sind?
28. Wie ist der Erwartungswert einer Zufallsvariablen definiert?
29. Wie ist die Varianz und die Standardabweichung einer Zufallsvariablen definiert?
30. Welche Rechenregeln für Erwartungswert und Varianz einer Zufallsvariablen kennen Sie?
31. Wann sind zwei Zufallsvariablen unabhängig?
32. Erläutern Sie den Korrelationskoeffizienten zweier Zufallsvariabler!
33. Wann sind zwei Zufallsvariable unkorreliert?
34. Erläutern Sie den Zusammenhang zwischen Unabhängigkeit und Unkorreliertheit!
35. Wie lautet die Ungleichung von Tschebyscheff?
36. Was ist ein Quantil einer Zufallsvariablen?
37. Nennen Sie wichtige diskrete und stetige Verteilungen!
38. Was ist eine Bernoulli-Variable?
39. Erläutern Sie Verteilungs- und Dichtefunktion der Normalverteilung!
40. Erläutern Sie den Zusammenhang zwischen $N(0,1)$ und einer beliebigen $N(\mu,\sigma)$-Verteilung!
41. Was ist ein $k\sigma$-Bereich?
42. Wie lautet das schwache Gesetz der großen Zahlen?
43. Was ist eine erwartungstreue Punktschätzung?
44. Was ist eine asymptotisch erwartungstreue Punktschätzung?
45. Nennen sie Beispiele für erwartungstreue und asymptotisch erwartungstreue Punktschätzer!
46. Erläutern Sie das Maximum-Likelihood-Prinzip!

Methodenfragen

1. Die empirische Dichte und die absoluten und relativen Häufigkeiten einer Stichprobe berechnen können.
2. Für eine Stichprobe Klassen bilden können.
3. Summenhäufigkeiten berechnen können.
4. Modalwert und Mittelwerte einer Stichprobe berechnen können.
5. Spannbreite, Varianz und Standardabweichung einer Stichprobe berechnen können.
6. Ein Kontingenztabelle für eine zweidimensionale Stichprobe aufstellen können.
7. Stichproben-Kovarianz und Bravais-Pearson-Korrelationskoeffizient berechnen können.
8. Ein Streuungsdiagramm erstellen und interpretieren können.
9. Für ein Zufallsexperiment eine geeignete Ereignis-Algebra konstruieren können.
10. Nachweisen können, dass eine gegebene Menge eine Ereignis-Algebra darstellt.
11. Laplace-Wahrscheinlichkeiten bestimmen können.
12. Die Rechenregeln für ein Wahrscheinlichkeitsmaß anwenden können.
13. Bedingte Wahrscheinlichkeiten berechnen und anwenden können.
14. Überprüfen können, ob zwei gegebene Ereignisse unabhängig sind.
15. Verteilungs- und Dichtefunktionen von Zufallsvariablen anwenden und bestimmen können.
16. Wahrscheinlichkeiten für Zufallsvariable berechnen können.
17. Erwartungswert, Varianz und Standardabweichung einer Zufallsvariablen berechnen können.
18. Rechenregeln für Erwartungswert und Varianz einer Zufallsvariablen anwenden können.
19. Zufallsvariable auf Unabhängigkeit überprüfen können.
20. Den Korrelationskoeffizienten bestimmen und interpretieren können.
21. Die Ungleichung von Tschebyscheff anwenden können.
22. Quantile bestimmen können.
23. Mit diskreten und stetigen Verteilungen arbeiten können.
24. Die typischen Anwendungsbereiche diskreter und stetiger Verteilungen kennen.
25. Mit der Normalverteilung arbeiten können.
26. $k\sigma$-Bereiche der Normalverteilung bestimmen können.
27. Punktschätzer anwenden und einschätzen können.
28. Das Maximum-Likelihood-Prinzip anwenden können.

Übungsaufgaben

1. Die folgenden Daten stellen eine Stichprobe einer Umfrage über das Alter von Eigenheimen in Jahren dar:

$$82, 70, 69, 19, 71, 70, 13, 70, 70, 23, 62, 53, 32, 65, 66, 55, 79, 15, 21, 69$$

$$59, 18, 21, 66, 72, 19, 65, 57, 24, 71, 51, 50, 82, 13, 13, 82, 82, 54, 64, 36.$$

   Stellen Sie die relativen und absoluten Häufigkeiten grafisch dar und bestimmen Sie die Lageparameter der Stichprobe!

   Lösung:
   Modalwerte für die Stichprobe sind die Ausprägungen 70 und 82. Das arithmetische Mittel der Stichprobe ist gegeben durch $\overline{x} = 51,825$, der Median durch $_med = 60,5$. Das geometrische Mittel der Stichprobe ist gegeben durch $x_g = 44,52$.

   Die absoluten Häufigkeiten finden sie in Abbildung 15.1, die relativen Häufigkeiten in Abbildung 15.2 und die empirische Verteilungsfunktion in Abbildung 15.3!

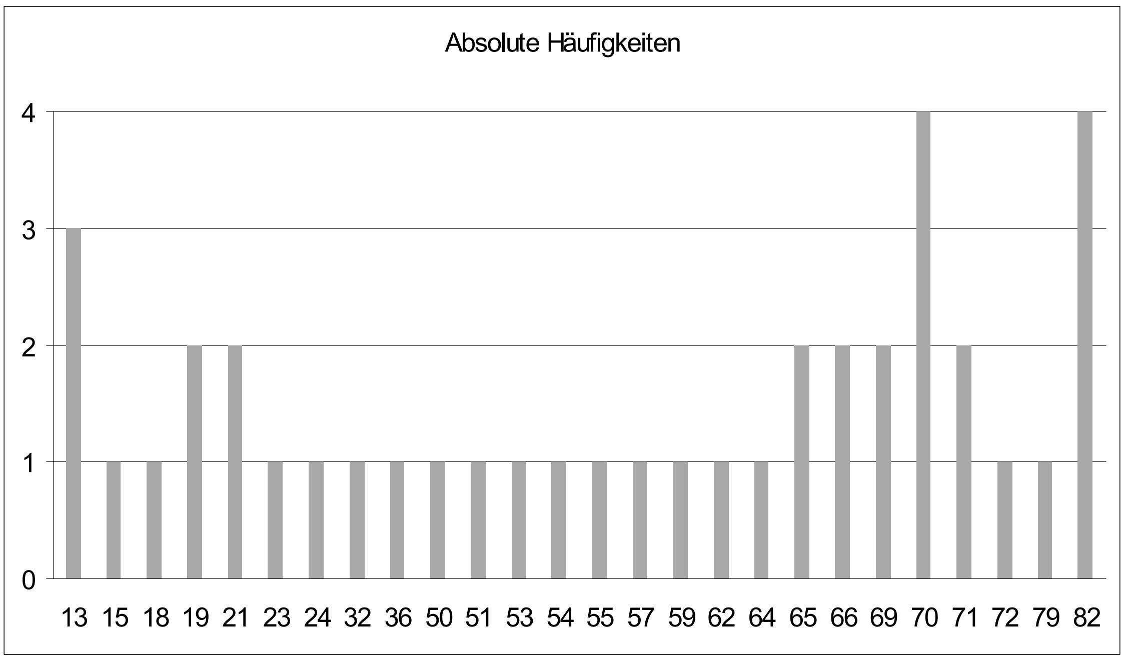

Abbildung 15.1: Die absoluten Häufigkeiten für die Daten in Aufgabe 1

2. Bestimmen Sie die relative Häufigkeit und die relative Summenhäufigkeit für eine Klassenbreite von $w = 50$ für die Lebensdauer-Stichprobe auf Seite 401, und stellen Sie diese grafisch dar!

   Lösung:
   Die relative Klassenhäufigkeiten bei Klassenbreite 50 finden Sie in Abbildung 15.4; die relative Summenhäufigkeit in Abbildung 15.5.

3. Bestimmen Sie die Lage- und Streuungsparameter für die Daten der der Lebensdauer-Stichprobe auf Seite 401 und der Stichprobe aus Aufgabe 1!

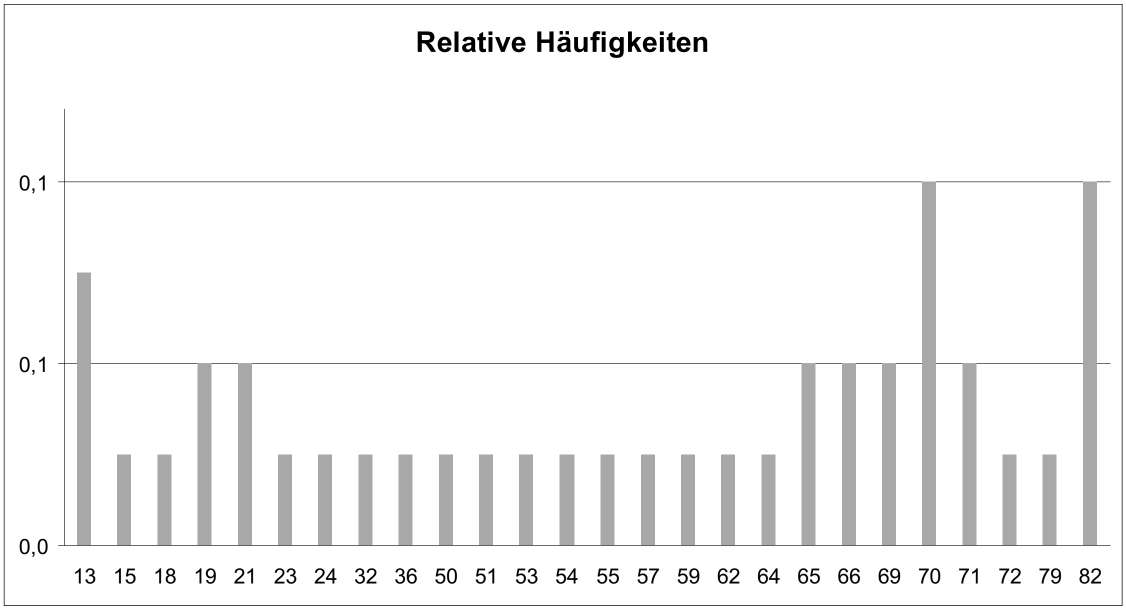

Abbildung 15.2: Die relativen Häufigkeiten für die Daten in Aufgabe 1

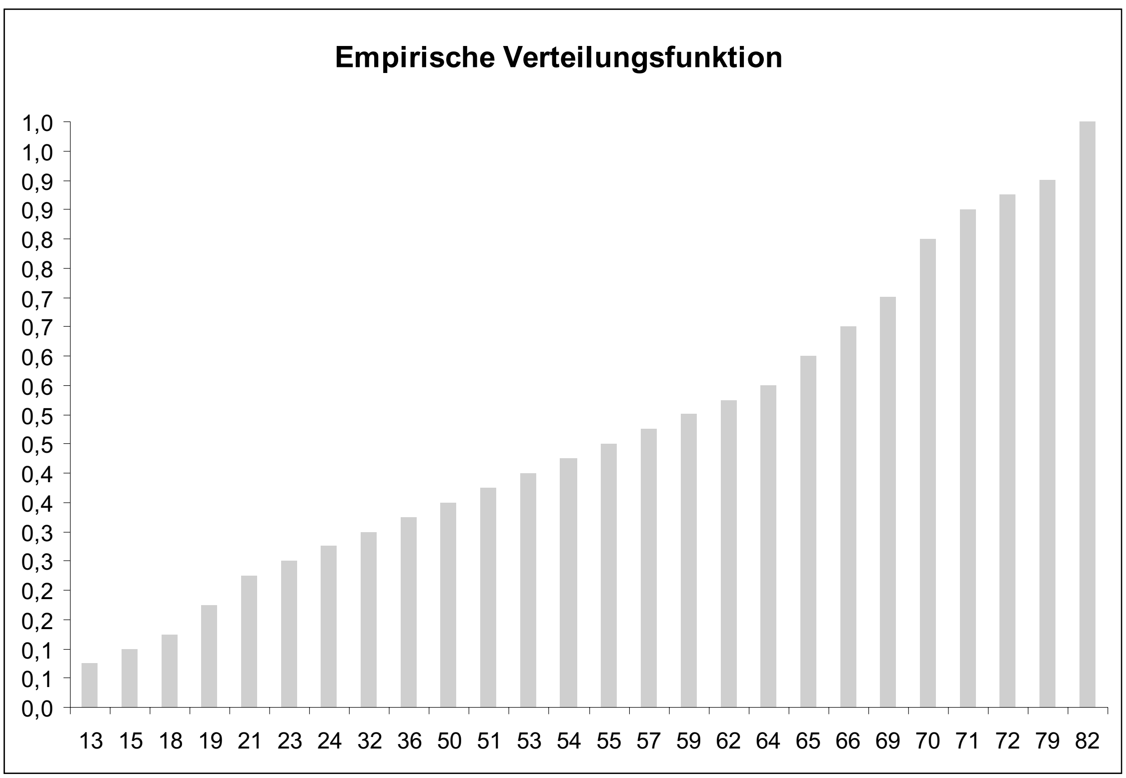

Abbildung 15.3: Die empirische Verteilungsfunktion für die Daten in Aufgabe 1

Lösung:
Die Ergebnisse für die Stichprobe der Lebensdauern auf Seite 401:
$\overline{x} \approx 313,166$, $x_g \approx 266,85$, $x_m = 325$. Die Spannbreite ist 470, die Varianz $s^2 \approx 2\,1283,59$ und

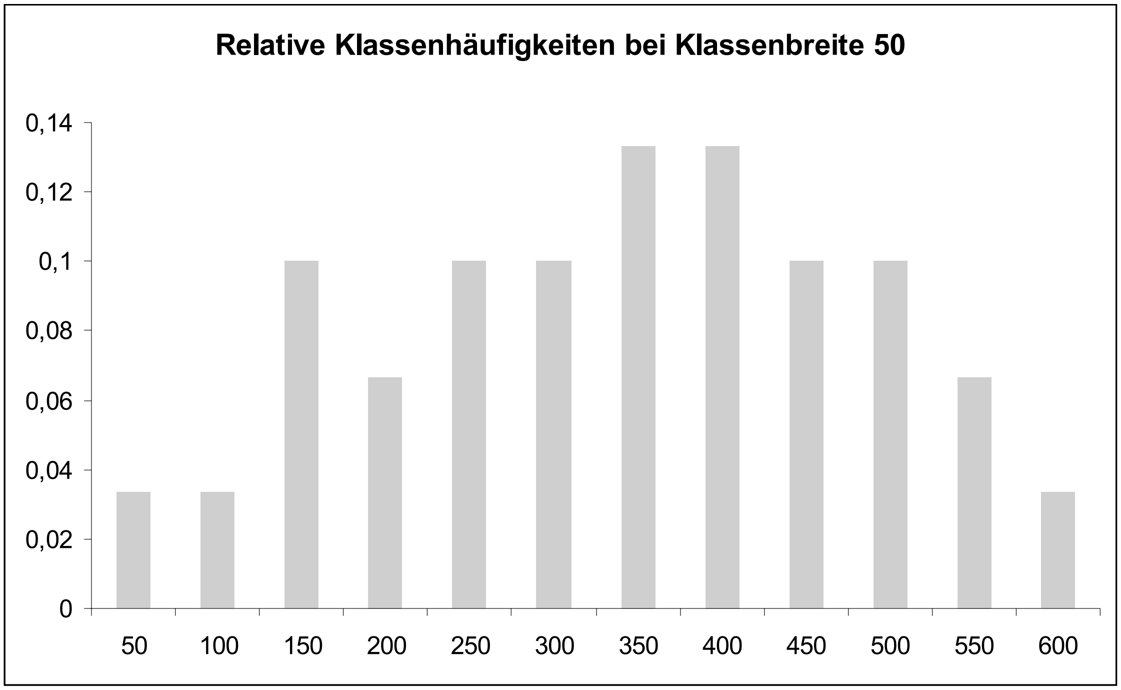

Abbildung 15.4: Die relativen Häufigkeiten für die Daten in Aufgabe 2

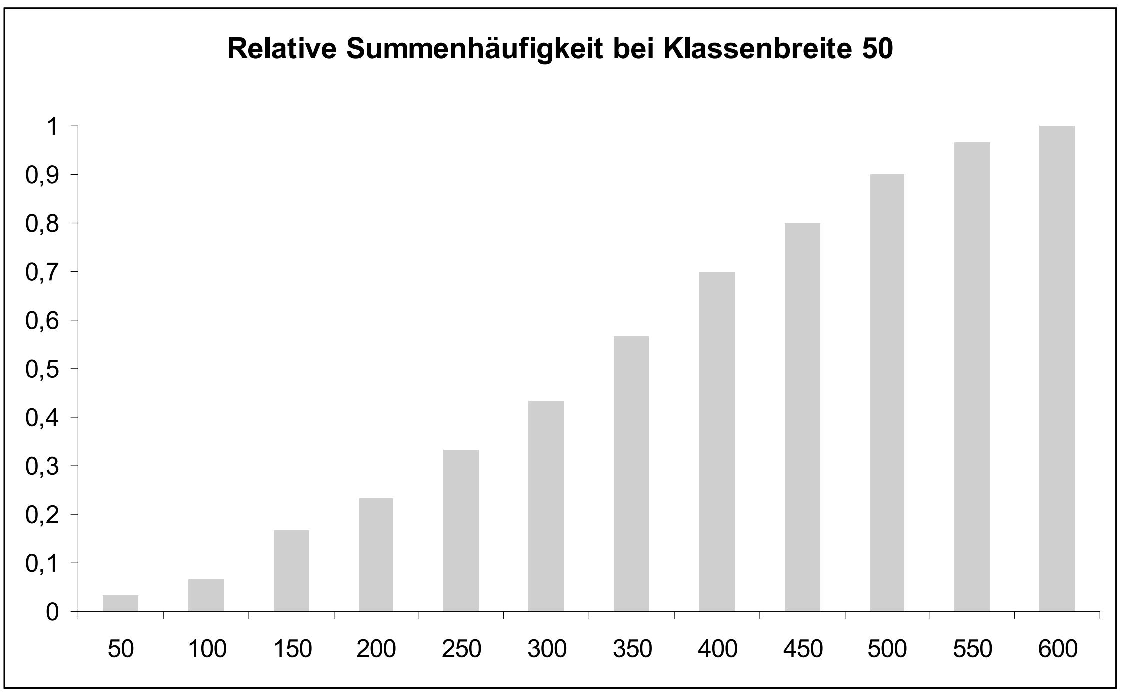

Abbildung 15.5: Die relativen Summenhäufigkeiten für die Daten in Aufgabe 2

die Standardabweichung $s \approx 145,89$.

Für die Stichprobe in Aufgabe 1 sind die Lageparameter: Modalwerte für die Stichprobe sind die Ausprägungen 70 und 82. Das arithmetische Mittel der Stichprobe ist gegeben durch $\overline{x} =$

51, 825, der Median durch $x_{med} = 60,5$. Das geometrische Mittel der Stichprobe ist gegeben durch $x_g = 44,52$.

Die Spannbreite der Stichprobe ist 6, die Varianz $s^2 \approx 560,507$ und die Standardabweichung ist $s \approx 23,675$.

4. Weisen Sie nach, dass der Median einer Stichprobe mit n Werten die Gleichungen

$$\sum_{i=1}^{n} \frac{x_i - x_{med}}{|x_i - x_{med}|} = 0, x_{med} = \frac{\sum_{i=1}^{n} \frac{x_i}{|x_i - x_{med}|}}{\sum_{i=1}^{n} \frac{1}{|x_i - x_{med}|}}$$

erfüllt!

Lösung:
Nehmen wir einmal an, dass es eine gerade Anzahl von Stichproben gibt. Dann war $x_{med} = \frac{1}{2}(x_k + x_{k+1})$ mit $k = \frac{n}{2}$. In der ersten Summe ist für alle Indices mit $i \leq k$ der Summand gleich -1, für die restlichen Summanden ist der Wert konstant 1, und die Summe ist ingesamt Null.

Ist die Anzahl der Stichproben n ungerade, dann war der Median gegeben durch die Zahl $x_{med} = x_k$ mit $k = \frac{n+1}{2}$. Für diesen Index ist der Summand 0; die Summanden mit $i < k$ sind wieder alle -1, die Summanden mit $i > k$ sind 1; und es gibt genauso so viele Summanden mit -1 wie mit 1.

Die zweite Gleichung für den Median folgt aus

$$x_{med} - \frac{\sum_{i=1}^{n} \frac{x_i}{|x_i - x_{med}|}}{\sum_{i=1}^{n} \frac{1}{|x_i - x_{med}|}} = \frac{\sum_{i=1}^{n} \frac{x_{med} - x_i}{|x_i - x_{med}|}}{\sum_{i=1}^{n} \frac{1}{|x_i - x_{med}|}} = 0.$$

Auf der Basis beider Gleichungen kann eine Berechungsroutine für den Median ohne Sortieren der Stichprobe aufgestellt werden. Beispielsweise kann das Bisektionsverfahren für die Bestimmung der Nullstelle eingesetzt werden.

5. Weisen Sie nach, dass das arithmetische Mittel für eine Stichprobe mit n Werten die Funktion $\sum_{i=1}^{n}(x_i - x)^2$ minimiert!

Lösung:
Für die Funktion $f(x) = \sum_{i=1}^{n}(x_i - x)^2$ ist die erste und die zweite Ableitung gegeben durch

$$f'(x) = -2\sum_{i=1}^{n} x_i + 2nx, f''(x) = 2n.$$

Für ein lokales Extremum der Funktion f muss $f'(x) = 0$ erfüllt sein. Die Nullstelle der ersten Ableitung ist gegeben durch

$$2nx = 2\sum_{i=1}^{n} x_i \Leftrightarrow x = \frac{1}{n}\sum_{i=1}^{n} = \overline{x}.$$

Die zweite Ableitung $f''(x) = 2n > 0$ ist immer positiv, also ist $\overline{x}$ ein lokales Minimum!

6. Bestimmen Sie die Streuungsparameter für die Daten der Lebensdauer-Stichprobe auf Seite 401 und der Stichprobe aus Aufgabe 1!

Lösung:
Die Lösung finden Sie in der Lösung von Aufgabe 3!

7. Wie groß ist die Wahrscheinlichkeit, dass in einer Gruppe von 60 Personen zwei Personen am gleichen Tag Geburtstag haben?

   Lösung:
   Auf Seite 408 finden sie im Beispiel die Berechnung für den Fall von $n = 30$ Personen; dort war die Wahrscheinlichkeit, dass zwei Personen am gleichen Tag Geburtstag haben (bei 365 Tagen im Jahr) $p = 0,706 = 1 - 0,294$.

   Jetzt ist die Wahrscheinlichkeit, dass alle Personen an einem verschiedenen Tag Geburtstag haben gegeben durch
   $$P(A^c) = \frac{\prod_{i=306}^{365} i}{365^{30}} \approx 0,005\,877.$$
   Daraus ergibt sich die gesuchte Wahrscheinlichkeit als
   $$P(A) = 1 - P(A^c) \approx 0,994\,123.$$

8. Weisen Sie nach, dass für zwei unabhängige Ereignisse A und B auch A^c und B^c sowie A und B^c unabhängig sind!

   Lösung:
   Es gilt
   $$A = (A \cap B) \cup (A \cap B^c),$$
   und damit folgt aus den Wahrscheinlichkeitsaxiomen und der vorausgesetzten Unabhängigkeit von A und B
   $$P(A) = P(A \cap B) + P(A \cap B^c) = P(A)P(B) + P(A \cap B^c)$$
   und damit
   $$P(A \cap B^c) = P(A)(1 - P(B)) = P(A)P(B^c).$$
   Dann sind A und B^c unabhängig.

   Wenn Sie in dieser Herleitung A durch A^c ersetzen erhalten Sie die Unabhängigkeit von A^c und B und auch von A^c und B^c.

9. Weisen Sie nach, dass für eine Zufallsvariable $P(X \in [\mu - 3\sigma; \mu + 3\sigma]) \geq \frac{8}{9}$ und $P(X \in [\mu - 4\sigma; \mu + 4\sigma]) \geq \frac{15}{16}$ gilt!

   Lösung:
   Allgemein gilt, das $P(x \in [\mu - k\sigma; \mu + k\sigma]) \geq 1 - \frac{1}{k^2}$. Also ist $P = \frac{8}{9}$ für $k = 3$ und $P = \frac{15}{16}$ für $k = 4$.

10. Für welchen Wert von a ist die Funktion $f(x) = a(x-1)$ für $x \in [1;3]$ und $f(x) = 0$ sonst Dichtefunktion einer Zufallsvariablen X? Bestimmen Sie die zugehörige Verteilungsfunktion, $P(X > 2)$, $E(X)$, $Var(X)$, das 75%- und das 90%-Quantil!

    Lösung:
    Für eine Dichtefunktion muss
    $$\int_{-\infty}^{\infty} f_X(t)dt = 1$$
    erfüllt sein. Für die gegebene Funktion f folgt daraus die Bedingung
    $$a\int_1^3 (x-1)dx = 1$$
    und $a = \frac{1}{2}$.

Dann ist die Verteilungsfunktion gegeben als

$$F_X(t) = \begin{cases} 0 & -\infty < t \leq 1, \\ \frac{1}{4}(t-1)^2 & 1 < t \leq 3, \\ 1 & t > 3. \end{cases}$$

Dann ist

$$P(X > 2) = 1 - P(X \leq 2) = 1 - F_X(2) = \frac{3}{4}.$$

Der Erwartungswert ist gegeben als

$$E(X) = \frac{1}{2}\int_1^3 x(x-1)dx = \frac{7}{3},$$

die Varianz als

$$Var(X) = \frac{1}{2}\int_1^3 (x-\frac{7}{3})^2(x-1)dx = \frac{2}{9}.$$

Das 75%-Quantil $t_{0,75}$ ist gegeben als Lösung der Gleichung

$$F_X(t_{0,75}) = 0,75.$$

Dadurch ergibt sich eine quadratische Gleichung für $t_{0,75}$ und die Lösung $t_{0,75} = 1 + \sqrt{3}$.
$t_{0,9}$ ist gegeben als Nullstelle der entsprechenden quadratischen Gleichung mit rechter Seite $0,9$ als $t_{0,9} \approx 2,64$.

11. Weisen Sie nach, dass die Varianz einer geometrischen Verteilung durch $Var(X) = \frac{1-p}{p^2}$ gegeben ist.

Lösung:
Wie im Fall des Erwartungswerts der geometrischen Verteilung auf Seite 426 kann hier die zweite Ableitung der geometrischen Reihe als Potenzreihe verwendet werden. Es gilt also

$$\sum_{i=1}^{\infty} i \cdot q^{i-1} = \frac{1}{(1-q)^2}, \sum_{i=2}^{\infty} i \cdot (i-1)q^{i-2} = \frac{2}{(1-q)^3}.$$

Die Varianz ist gegeben durch

$$Var(X) = E(X^2) - (E(X))^2,$$

also müssen wir zuerst $E(X^2)$ bestimmen. Es gilt

$$\begin{aligned} E(X^2) &= \sum_{i=1}^{\infty} i^2 p(1-p)^{i-1} \\ &= p\sum_{i=1}^{\infty} i^2(1-p)^{i-1} \\ &= p\sum_{i=1}^{\infty} (i + i(i-1))(1-p)^{i-1} \\ &= p\left[\sum_{i=1}^{\infty} i(1-p)^{i-1} + (1-p)\sum_{i=2}^{\infty} i(i-1)(1-p)i-2\right] \\ &= \frac{1}{p} + p(1-p)\frac{2}{p^3} \\ &= \frac{2-p}{p^2}. \end{aligned}$$

Dann folgt

$$Var(X) = E(X^2) - (E(X))^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

12. Bei der Produktion eines Speicherbausteins sind im Durchschnitt 5% Ausschuss. Die Losgröße beträgt 200 Stück. Wie groß ist die Wahrscheinlichkeit, dass in einer Stichprobe von 20 Stück genau 5 Bausteine Ausschuss sind, dass alle verwendbar oder höchstens 2 Bausteine defekt sind? Bestimmen Sie den Erwartungswert und die Varianz dieser Verteilung!

    Lösung:
    Es handelt sich um einer relativ kleine Stichprobe, das gezogene Stück wird nicht in die Grundgesamtheit zurückgelegt; dann verwenden wir die hypergeometrische Verteilung.

    Wir verwenden eine $Hyp(200; 10, 20)$-Verteilung. Dann ist die gesuchte Wahrscheinlichkeit, dass in der Stichprobe genau 5 Stück Ausschuss sind

    $$P_5 = \frac{\binom{10}{5}\binom{190}{15}}{\binom{200}{20}} \approx 0,001\,028.$$

    Die Wahrscheinlichkeit, dass unter den 20 gezogenen Stücken genau 5 Ausschuss sind beträgt rund $0,1\%$.

    Dass alle verwendbar, also kein Ausschuss in der Stichprobe ist, hat die Wahrscheinlichkeit

    $$P_0 = \frac{\binom{10}{0}\binom{190}{20}}{\binom{200}{20}} \approx 0,339\,78.$$

    Die Wahrscheinlichkeit, dass höchstens 2 Stück Ausschuss sind ist gegeben durch

    $$P(X \leq 2) = P(X=0) + P(X=1) \approx 0,934\,71.$$

    Der Erwartungswert ist $E(X) = n\frac{M}{N} = 1$, die Varianz ist $Var(X) = n\frac{M}{n}(1-\frac{M}{N})\frac{N-n}{N-1} \approx 0,859$.

13. Weisen Sie nach, dass für eine $P(\lambda)$-verteilte Variable $Var(X) = \lambda$ gilt!

    Lösung:
    Für den Erwartungswert gilt $E(X) = \lambda$. Dann gilt

    $$\begin{aligned}
    Var(X) &= \sum_{i=0}^{\infty}(i-\lambda)^2\frac{\lambda^i}{i!}e^{-\lambda} \\
    &= \sum_{i=0}^{\infty} i^2\frac{\lambda^i}{i!}e^{-\lambda} - 2\lambda\sum_{i=0}^{\infty} i\frac{\lambda^i}{i!}e^{-\lambda} + \lambda^2 e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} \\
    &= \sum_{i=0}^{\infty} i(i-1)\frac{\lambda^i}{i!}e^{-\lambda} + \sum_{i=0}^{\infty} i\frac{\lambda^i}{i!}e^{-\lambda} - 2\lambda\sum_{i=0}^{\infty} i\frac{\lambda^i}{i!}e^{-\lambda} + \lambda^2 e^{-\lambda}\sum_{i=0}^{\infty}\frac{\lambda^i}{i!} \\
    &= \sum_{i=2}^{\infty} i(i-1)\frac{\lambda^i}{i!}e^{-\lambda} + \lambda - 2\lambda^2 + \lambda^2 e^{\lambda}e^{-\lambda} \\
    &= \lambda^2 + \lambda - 2\lambda^2 + \lambda^2 = \lambda.
    \end{aligned}$$

14. Bei der Produktion von Stoff sind im Durchschnitt auf 100m 10 Fehler enthalten. Der Stoff wird zur Weiterverarbeitung in Abschnitte von jeweils 3m zerschnitten. Wie groß ist die Wahrscheinlichkeit, dass in einem Abschnitt kein Fehler enthalten ist?

Lösung:
Im Durchschnitt ergeben sich

$$\frac{m}{N} = 0,1 \text{ Fehler pro Meter.}$$

Für einen Abschnitt von 3 Meter folgt dann, dass es

$$n\frac{m}{N} = \lambda = 0,3$$

Fehler pro Abschnitt gibt.

Die Wahrscheinlichkeit, dass ein Abschnitt von 3 Metern keinen Fehler aufweist ist dann gegeben durch eine $P(0,3)$-Verteilung mit

$$P(X = 0) = \frac{0,3^0}{0!}e^{-0,3} \approx 0,741.$$

15. Stellen Sie die Poisson-Verteilung für $\lambda = 9$, $n = 1\,000$ und $\lambda = 1$, $n = 200$ als Stabdiagramm dar!

Lösung:
In Abbildung 15.6 finden Sie das Stabdiagramm für $\lambda = 9$, Abbildung 15.8 den Fall $\lambda = 1$. Man erkennt gut, dass im Fall $\lambda = 1$ die Verteilung unsymmetrisch ist. Mit wachsendem λ wird die Verteilung immer „symmetrischer".

Wie bereits im Buch vermerkt ist die Poisson-Verteilung eine gute Approximation einer Binomialverteilung. In der Aufgabenstellung ist neben λ auch ein Wert für n gegeben; daraus ergibt sich für die Binomialverteilung $p = \frac{\lambda}{n} = \frac{9}{1\,000}$ bzw. $p = \frac{\lambda}{n} = \frac{1}{1\,000}$. In den Abbildungen 15.7 und 15.9 finden Sie die Stabdiagramme für diese Binomialverteilungen.

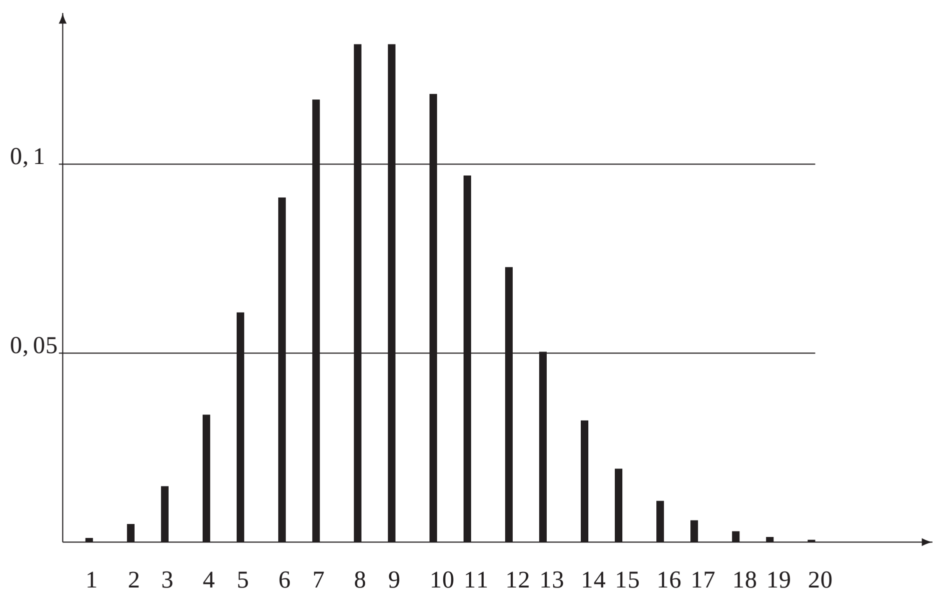

Abbildung 15.6: Das Stabdiagramm der Poisson-Verteilung mit $\lambda = 9$ für Aufgabe 15

16. Weisen Sie nach, dass für eine $E(\lambda)$-verteilte Zufallsvariable $E(X) = \frac{1}{\lambda}$ und $Var(X) = \frac{1}{\lambda^2}$ gilt!

Lösung:
Der Erwartungswert einer $E(\lambda)$-verteilten Zufallsvariablen X ist gegeben durch das Integral

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx.$$

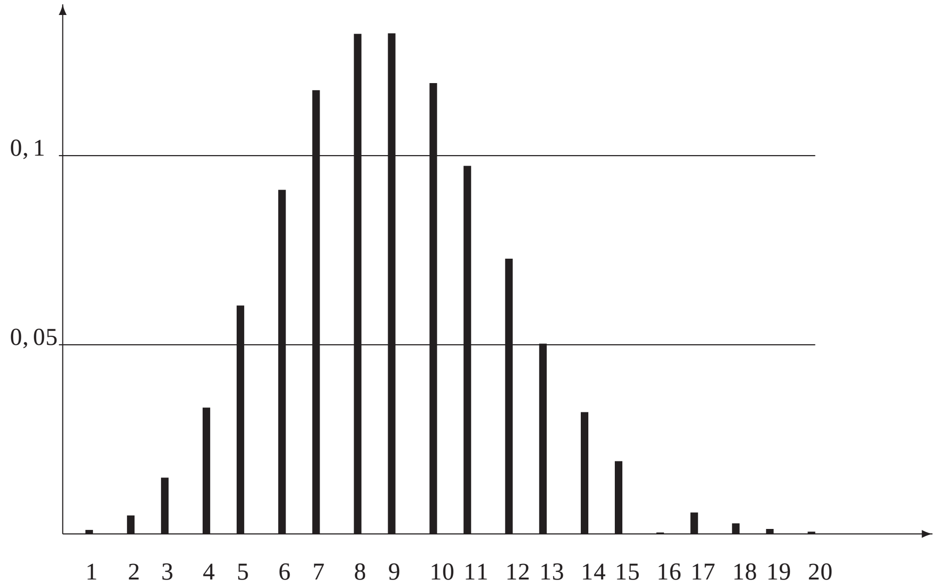

Abbildung 15.7: Das Stabdiagramm der Binomial-Verteilung mit $\lambda = 9$, $n = 1\,000$ und $p = \frac{\lambda}{n}$ für Aufgabe 15

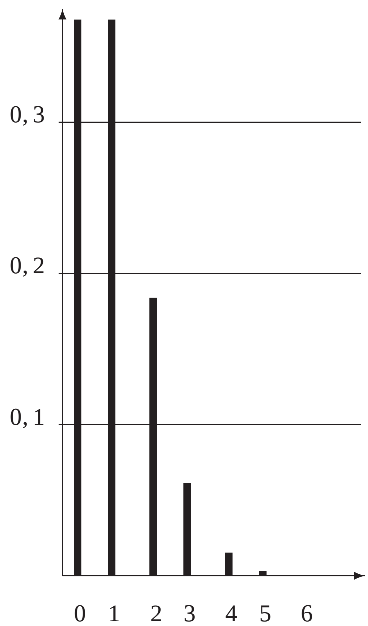

Abbildung 15.8: Das Stabdiagramm der Poisson-Verteilung mit $\lambda = 1$ für Aufgabe 15

Dieses Integral kann mit Hilfe von partieller Integration bestimmt werden. Dazu wählen wir

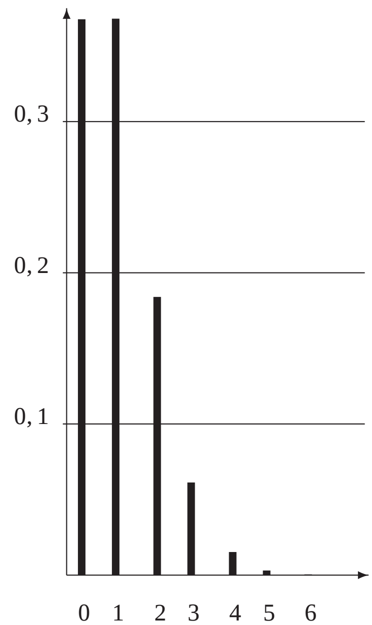

Abbildung 15.9: Das Stabdiagramm der Binomialverteilung mit $\lambda = 1$, $n = 1\,000$ und $p = \frac{1}{1\,000}$ für Aufgabe 15

$f(x) = x$, also ist $f'(x) = 1$ und $g'(x) = \lambda e^{-\lambda x}$, dann ist $g(x) = -e^{-\lambda x}$.

$$\begin{aligned} E(X) &= \int_0^\infty x\lambda e^{-\lambda x} dx \\ &= \left[-xe^{-\lambda x}\right]_0^\infty + \int_0^\infty e^{-\lambda x} dx \\ &= \left[-\frac{1}{\lambda} e^{-\lambda x}\right]_0^\infty \\ &= \frac{1}{\lambda}. \end{aligned}$$

Für die Bestimmung der Varianz berechnen wir $E(X^2)$, wieder mit partieller Integration:

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 \lambda e^{-\lambda x} dx \\ &= \left[x^2 e^{-\lambda x}\right]_0^\infty + 2\int_0^\infty xe^{-\lambda x} dx \\ &= \frac{2}{\lambda} \cdot \frac{1}{\lambda} = \frac{2}{\lambda^2}. \end{aligned}$$

Dann ist die Varianz gegeben durch

$$Var(X) = E(X^2) - E(X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

17. Erfahrungswerte zeigen, dass die Reparaturzeit eines PCs in einem Unternehmen exponentiell verteilt ist. Die mittlere Reparaturzeit beträgt 4 Stunden. Wie groß ist die Wahrscheinlichkeit, dass für einen PC zwischen 5 und 8 Stunden, höchstens 10 Stunden und mindestens 3 Stunden benötigt werden?

    Lösung:
    Wir müssen in einem ersten Schritt den Parameter λ der angenommenen Exponentialverteilung bestimmen. In der Aufgabenstellung finden Sie den Erwartungswert, nämlich 4 Stunden

für die Reparaturzeit. Der Erwartungswert der Exponentialverteilung ist $E(X) = \frac{1}{\lambda} = 4$, dann ist also $\lambda = 0,25$.

Es ist dann

$$P(5 < X < 8) = P(X < 8) - P(X < 5) = (1 - e^{-0,25\cdot 8}) - (1 - e^{-0,25\cdot 5}) \approx 0,151.$$

Dass eine Reparatur höchstens 10 Stunden dauert ist $0,9197$. Die Wahrscheinlichkeit, dass die Reparatur mindestens 3 Stunden dauert ist gegeben durch

$$P(X \geq 3) = 1 - P(X < 3) \approx 0,472.$$

18. Weisen Sie nach, dass $P(|X-\mu| \leq k\sigma) = 2\Phi(k) - 1$ gilt.

Lösung:
Zur Erinnerung: ist X eine $N(\mu, \sigma^2)$-verteilte Zufallsvariable, dann ist

$$Y = \frac{X-\mu}{\sigma}$$

$N(0,1)$-verteilt!

$$\begin{aligned} P(|X-\mu| \leq k\sigma) &= P(|\frac{X-\mu}{\sigma}| \leq k) \\ &= P(-k < \frac{X-\mu}{\sigma} < k) \\ &= P(Y < k) - P(Y < -k) \\ &= \Phi(k) - \Phi(-k) \\ &= \Phi(k) - (1 - \Phi(k)) \\ &= 2\Phi(k) - 1. \end{aligned}$$

19. Implementieren Sie die Approximation der Normalverteilung $\Phi(0;1)$ nach der Formel auf Seite 431 in der Programmiersprache Ihrer Wahl und vergleichen Sie die Ergebnisse mit Tafelwerken!

Lösung:
Die Approximation von Ponton (Sie finden Sie unter der URL *http://www.uiowa.edu/ grpporc/-crisp/crisp.3.6.htm*) dient zur Näherung der sogenannten „Error Function“

$$erfc(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{x^2}{2}}\, dx.$$

Die standardisierte Normalverteilung ist dann gegeben durch

$$\Phi(t) = \begin{cases} \frac{1}{2} + erfc(t), & t \geq 0, \\ \frac{1}{2} - erfc(t), & t < 0. \end{cases}$$

Dann ergibt sich der folgende `Java`-Code:

```
private static double ponton(double x) {
   double k, value = 0.0;

   k = 0.5 + (1.0/Math.sqrt(Math.PI)-0.5)*
         Math.exp(-x*x/Math.sqrt(2.0*Math.PI));
   value = k * Math.sqrt(1.0-Math.exp(-x*x/2.0));
```

```
    if (x>= 0.0)
        return 0.5 + value;
    else
        return 0.5 - value;
}
```

Die Tabellen, beispielsweise in Greiner / Tinhofer, Stochastik für Studienanfänger der Informatik aus dem Hanser-Verlag sind typischerweise so aufgebaut, dass in einer quadratischen Matrix Werte für $\Phi(t)$ stehen. In den Zeilen finden Sie Werte von t, in den Spalten die Überschriften $0.00, 0.01, \ldots, 0.09$. In der Zeile mit $t = 0.5$ und Spaltenüberschrift 0.04 können Sie dann den Wert für $t = 0.54$ ablesen: $\Phi(0.54) \approx 0.70540$. Der `Java`-Code liefert die Näherung 0.70521, das entspricht einem Fehler von $\delta = 1.866 \cdot 10^{-4}$. In Tabelle 15.1 finden Sie die berechneten Werte, die Tabellenwerte aus Greiner / Tinhofer und die Fehler für die komplette Zeile $t = 0.5$. Je näher die Argumente bei 0 liegen desto besser ist die Approximation von Ponton. Beispielsweise ist für $t = 0.05$ in Greiner / Tinhofer ein Wert von 0.51994 angegeben, Ponton berechnet $0,5199386205616786$, ein Fehler von $1,379 \cdot 10^{-6}$.

x	$\Phi(x)$	Greiner / Tinhofer	Fehler
$0,5$	$0,6913083625030305$	$0,69146$	$1,5163749696944784 \cdot 10^{-4}$
$0,51$	$0,6948119739173705$	$0,69497$	$1,5802608262949214 \cdot 10^{-4}$
$0,52$	$0,6982975085457394$	$0,69847$	$1,7249145426068146 \cdot 10^{-4}$
$0,53$	$0,7017647134242829$	$0,70194$	$1,7528657571708184 \cdot 10^{-4}$
$0,54$	$0,7052133408637061$	$0,7054$	$1,8665913629389852 \cdot 10^{-4}$
$0,55$	$0,7086431484939236$	$0,70884$	$1,9685150607640090 \cdot 10^{-4}$
$0,56$	$0,712053899306031$	$0,71226$	$2,0610069396898023 \cdot 10^{-4}$
$0,57$	$0,7154453616915879$	$0,71566$	$2,1463830841206288 \cdot 10^{-4}$
$0,58$	$0,7188173094792161$	$0,71904$	$2,2269052078394136 \cdot 10^{-4}$
$0,59$	$0,7221695219685125$	$0,7224$	$2,3047803148756874 \cdot 10^{-4}$

Tabelle 15.1: Die Approximation nach Ponton, Werte aus Greiner / Tinhofer und die Fehler für eine Zeile der tabellierten Standardnormalverteilung in Aufgabe 19

Ponton gibt eine noch einfachere Approximation an; er ersetzt die Näherung aus der Aufgabenstellung durch

$$K(t) = 0,5 + 0,064 \cdot e^{-0,4 \cdot t^2}, \Phi(t; 0, 1) \approx 0,5 \pm K(t)\sqrt{1 - e^{-\frac{t^2}{2}}}.$$

Für $t = 0.59$ ergibt sich damit ein Näherungswert von $0,72209537820118128$, eine Abweichung vom Tabellenwert von $3,046217988188493 \cdot 10^{-4}$.

20. Wie groß ist die Wahrscheinlichkeit, dass eine $N(50; 10)$-verteilte Zufallsvariable im Intervall $(45; 55)$, unterhalb von 45, oberhalb von 60 und um mehr als 5 vom Erwartungswert abweicht?

Lösung:
Es gilt

$$\begin{aligned} P(45 \le X \le 55) &= \Phi(55; 10, 0, 1) - \Phi(45; 10, 0, 1) \\ &= \Phi(0,5) - \Phi(-0,5) = 2\Phi(0,5) - 1 \\ &\approx 0,382\,92. \end{aligned}$$

Oberhalb von 60 liegt die Zufallsvariable mit Wahrscheinlichkeit

$$P(X > 60) = 1 - \Phi(1) \approx 0,158\,66.$$

Die Wahrscheinlichkeit, dass sie um mehr als 5 cm vom Erwartungswert abweicht ist

$$P(|X - \mu| > 5) = 1 - P(|X - \mu|) \le 5) = 1 - \left(2\Phi(\frac{5}{10}) - 1\right) \approx 0,617.$$

21. Weisen Sie nach, dass bei Binomial- und Poisson-Verteilung der Maximum-Likelihood-Schätzer mit dem arithmetischen Mittel übereinstimmt!

Lösung:
Die Maximum-Likelihood-Funktion für die Poisson-Verteilung ist

$$\begin{aligned} L(\mathbf{x};\lambda) &= \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \\ &= e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} \\ &= e^{-n\lambda} \frac{\lambda^{n\cdot\overline{x}}}{\prod_{i=1}^{n} x_i!}. \end{aligned}$$

Logarithmieren ergibt dann

$$LL(\lambda) = \ln\,(e^{-n\lambda} \frac{\lambda^{n\cdot\overline{x}}}{\prod_{i=1}^{n} x_i!}) = -n\lambda + n\overline{x}\ln\,(\lambda) - \ln\,(\prod_{i=1}^{n} x_i!).$$

Dann ist

$$LL'(\lambda) = -n + \frac{n\overline{x}}{\lambda},$$

Nullsetzen dieser Ableitung ergibt den Maximum-Likelihood-Schätzer $\lambda = \overline{x}$.

Die Maximum-Likelihood-Funktion für die Binomialverteilung (mit $p \neq 1, p \neq 0$) ist

$$L(\mathbf{x};p) = \prod_{i=1}^{n} \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}.$$

Logarithmieren dieser Funktion ergibt die Funktion

$$\begin{aligned} LL(p) &= \sum_{i=1}^{n} \ln\,(\binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}) \\ &= \sum_{i=1}^{n} \ln\,(\binom{n}{x_i}) + \ln\,(p)n\overline{x} + \ln\,(1-p)n(n-\overline{x}) \end{aligned}$$

Die erste Ableitung von LL ist gegeben durch

$$\begin{aligned} LL'(p) &= \frac{n\overline{x}}{p} - \frac{n(n-\overline{x})}{1-p} \\ &= \frac{n\overline{x} - pn^2}{p(1-p)} \end{aligned}$$

Die Nullstelle dieser Ableitung ist gegeben durch

$$p = \frac{\overline{x}}{n}.$$

Zur Erinnerung, $n \cdot p$ ist der Erwartungswert einer binomialverteilten Zufallsvariablen! Also ist das arithmetische Mittel $\overline{x}$ der Maximum-Likelihood-Schätzer für den Erwartungswert der Zufallsvariablen.