Serie 4

Aufgabe 4.1

Für die Körpergrösse von 18-20jährigen Männern ergibt sich ein Mittelwert von 1.80 m bei einer Standardabweichung von 7.4 cm. Die Körpergrösse kann als normalverteilt angesehen werden.

a) Mit welcher Wahrscheinlichkeit ist ein zufällig ausgewählter Mann dieser Altersgruppe grösser als 1.85 m bzw. zwischen 1.70 m und 1.80 m gross?

b) In welchem symmetrischen Bereich um den Mittelwert liegen die Grössen von 50 % der Körpergrössen?

c) Wie gross muss ein Mann sein, damit er zu den 5 % grössten Männern gehört?

Aufgabe 4.2

Die Lebensdauer (in Jahren) eines Radios ist exponentialverteilt mit dem Parameter $\lambda = \frac{1}{8}$.

a) Wie gross ist der Erwartungswert für die Lebensdauer?

b) Wie gross ist die Wahrscheinlichkeit, dass es nach 10 Jahren noch funktioniert?

Aufgabe 4.3

Aufgrund langjähriger Untersuchungen ist bekannt, dass der Bleigehalt X in einer Bodenprobe annähernd normalverteilt ist. Ausserdem weiss man, dass der Erwartungswert 32 ppb beträgt und dass die Standardabweichung 6 ppb beträgt.

a) Machen Sie eine Skizze der Dichte von X, und zeichnen Sie die Wahrscheinlichkeit, dass eine Bodenprobe zwischen 26 und 38 ppb Blei enthält, in die Skizze ein.

b) Wie gross ist die Wahrscheinlichkeit, dass eine Bodenprobe höchstens 40 ppb Schwermetall enthält? *Hinweis*: Benützen Sie die **Python**-Funktion

```
norm.cdf()
```

Überprüfen Sie das Resultat, indem Sie die standardisierte Zufallsvariable Z einführen und mit der Standardnormalverteilung die Wahrscheinlichkeit berechnen.

c) Wie gross ist die Wahrscheinlichkeit, dass eine Bodenprobe höchstens 27 ppb Schwermetall enthält?

d) Welcher Bleigehalt wird mit einer Wahrscheinlichkeit von 97.5 % unterschritten? Das heisst, bestimmen Sie dasjenige c, so dass die Wahrscheinlichkeit, dass der Bleigehalt kleiner oder gleich c ist, genau 97.5 % beträgt. *Hinweis*: Benützen Sie die **Python** -Funktion

```
norm.ppf()
```

e) Welcher Bleigehalt wird mit einer Wahrscheinlichkeit von 10 % unterschritten?

f) Wie gross ist die Wahrscheinlichkeit, die in Aufgabe a) eingezeichnet wurde?

Aufgabe 4.4

Wir gehen davon aus, dass bei der Detektion eines digitalen Signals das Hintergrundrauschen einer Normalverteilung folgt, und zwar mit einem Mittelwert von 0 Volt und einer Standardabweichung von 0.45 Volt. Das System geht davon aus, dass eine digitale 1 übertragen worden ist, wenn die Spannung 0.9 Volt überstiegen worden ist.

a) Wie gross ist die Wahrscheinlichkeit, dass eine 1 detektiert worden ist, wenn in Wahrheit keine gesendet wurde? Wie interpretieren Sie diese Wahrscheinlichkeit?

b) Bestimmen Sie die symmetrischen Grenzen um 0 Volt, die 99 % des Hintergrundrauschens einschliessen.

c) Nehmen Sie an, eine digitale 1 wird als eine Verschiebung des Mittelwertes um 1.8 Volt im Vergleich zum Rauschsignal dargestellt. Die Standardabweichung des Signals ist ebenfalls 0.45 Volt. Wie gross ist die Wahrscheinlichkeit, dass eine digitale 1 nicht detektiert wird?

Aufgabe 4.5

Die monatlichen Aufwendungen X [CHF] für den Wasserverbrauch eines 2-Personenhaushalts seien durch eine Zufallsvariable mit der folgenden Wahrscheinlichkeitsdichtefunktion gegeben:

$$f(x) = \begin{cases} cx(15 - \frac{x}{4}) & \text{falls} \quad 0 \leq x \leq 60 \\ 0 & \text{sonst} \end{cases}$$

a) Welchen Wert muss c annehmen?

b) Geben Sie die Verteilungsfunktion F der Zufallsvariablen X an.

c) Welcher Wert der Aufwendungen wird nur mit 10 % Wahrscheinlichkeit überschritten?

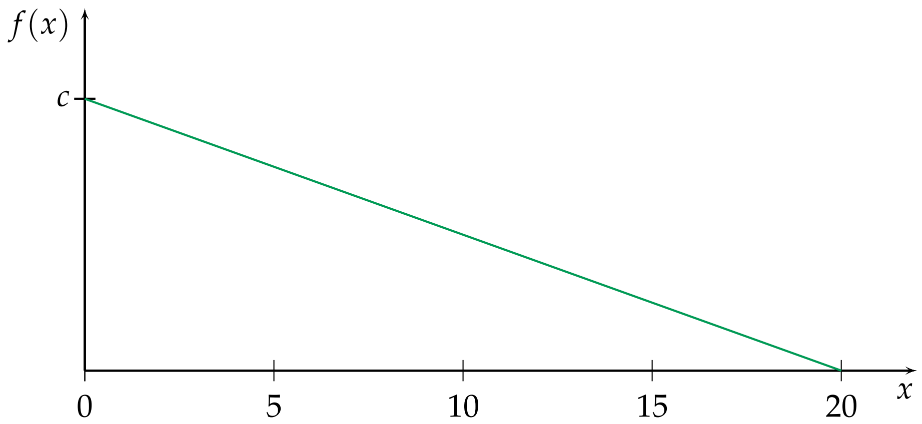
Hinweis: Zum lösen dieser Aufgabe brauchen Sie einen Solver, wie

https://www.wolframalpha.com/calculators/equation-solver-calculator

d) Wie hoch sind die erwarteten monatlichen Aufwendungen für den Wasserverbrauch eines 2-Personenhaushalts?

Aufgabe 4.6

In der Stadt Luzern gibt es bekanntlich viele Baustellen. Die Dauer X der Arbeiten bei einer Baustelle liege zwischen 0 und 20 Wochen. Die Dichte $f(x)$ habe die folgende Form.



a) Begründen Sie, warum $c = 0.1$ ist und schreiben Sie die Dichte $f(x)$ explizit auf.

b) Berechnen Sie die Wahrscheinlichkeit, dass die Bauzeit X weniger als

   a) 5

   b) 10

   Wochen beträgt.

c) Skizzieren Sie die kumulative Verteilungsfunktion.

d) Berechnen Sie den Erwartungswert, den Median und die Standardabweichung der Dauer X.

e) $K = 40'000 \cdot \sqrt{X}$ entspreche dem Betrag in Franken, den die Arbeiten bei einer Baustelle kosten. Wie gross ist die Wahrscheinlichkeit, dass die Arbeiten bei einer Baustelle höchstens 120'000.- Fr. kosten?

Die vorgeschlagene Verteilung ist nur ein Modell. Man könnte die Dauer der Bauarbeiten zum Beispiel auch als exponentialverteilt annehmen.

f) Für welchen Parameter λ hat die Exponentialverteilung denselben Erwartungswert wie die bisherige angenommene Verteilung?

g) Berechnen Sie mit der gefundenen Exponentialverteilung nochmals Teilaufgabe e).

Kurzlösungen vereinzelter Aufgaben

A 4.1:

a) 25 %, 41.2 %

b) $[175, 185]$

c) mindestens 192.2 cm

A 4.4:

a) 0.02275

b) $[-1.16, 1.16]$

c) 0.02275

A 4.6:

a) $c = \frac{1}{10}$

b) $F(x) = \frac{x}{10} - \frac{x^2}{400}$ und $P[X < 5] = 0.4375$ und $P[X < 10] = 0.75$.

d) $\mathrm{E}(X) = \frac{20}{3}$ und $\mathrm{Var}(X) = \frac{200}{9}$ und Median $\tilde{m} = 5.858$

e) 0.6975

f) $\lambda = \frac{3}{20}$

g) $P[K \leq 120'000] = 0.741$

Musterlösungen zu Serie 4

Lösung 4.1

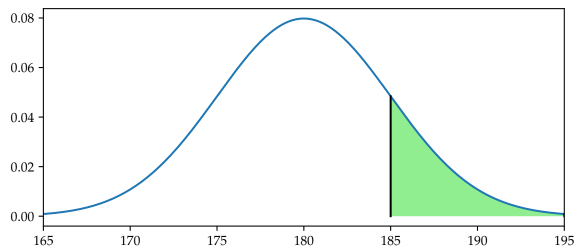
Die Zufallsvariable X misst die Körperlänge einer zufällig ausgewählten Person. Die Verteilung von X sieht wie folgt aus:

$$X \sim \mathcal{N}(1.8, 0.074^2)$$

Machen Sie für die folgenden Teilaufgaben jeweils eine Skizze und zeichnen Sie die gesuchte Grösse ein.

a) Gesucht ist die Wahrscheinlichkeit

$$P(X \geq 1.85) = 0.2496$$



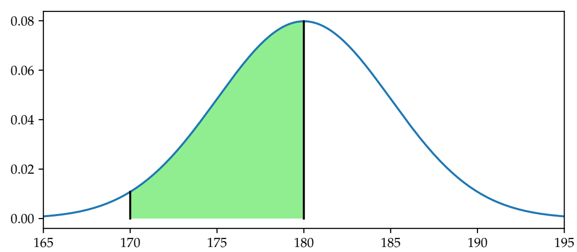
Also etwa 25 % der 18-20 Jahre alten Männer sind grösser als 1.85 m.

```
from scipy.stats import uniform, expon, norm
1 - norm.cdf(x=1.85, loc=1.80, scale=0.074)

## 0.24962329143760953
```

Gesucht ist die Wahrscheinlichkeit

$$P(1.70 \leq X \leq 1.80) = 0.4117$$

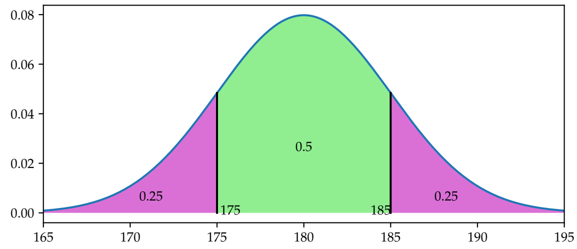


Also etwa 41 % der 18-20 Jahre alten Männer sind zwischen 1.70 m und 1.80 m.

```
from scipy.stats import uniform, expon, norm
norm.cdf(x=1.8, loc=1.80, scale=0.074) - norm.cdf(x=1.7, loc=1.80, scale=0.074)

## 0.4117085451462671
```

b) Gesucht sind die Quantile $q_{0.25}$ und $q_{0.75}$ (das sind gerade das untere und obere Quartil):

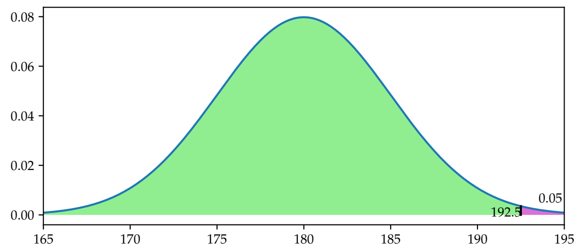



```python
from scipy.stats import uniform, expon, norm
norm.ppf(q=[0.25, 0.75], loc=1.80, scale=0.074)

## array([1.75008776, 1.84991224])
```

Das heisst, 50 % der Männer sind zwischen 1.75 m und 1.85 m gross.

c) Gesucht ist das Quantil $q_{0.95}$




```python
from scipy.stats import uniform, expon, norm
norm.ppf(q=0.95, loc=1.80, scale=0.074)

## 1.921719168394409
```

Das heisst, 5 % der Männer sind grösser als 1.92 m.

Lösung 4.2

Sei X die Zufallsvariable für die Lebensdauer. Also gilt

$$X \sim \text{Exp}(\lambda) = \text{Exp}\left(\frac{1}{8}\right)$$

a) Es gilt

$$\text{E}(X) = \frac{1}{\lambda} = \frac{1}{\frac{1}{8}} = 8$$

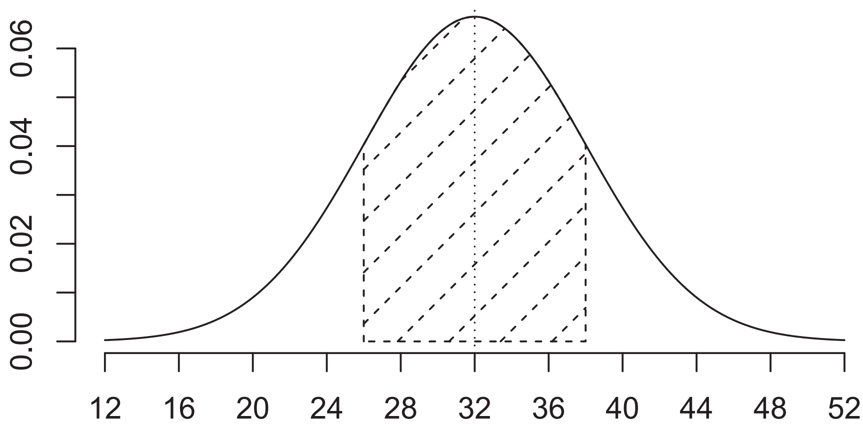
b) Die Wahrscheinlichkeit, dass das Radio nach 10 Jahren noch funktioniert, ist 28.6 %

```
from scipy.stats import uniform, expon, norm
1- expon.cdf(x=10, loc=0, scale=8)

## 0.28650479686019015
```

Lösung 4.3

a) Skizze:



b) X bezeichne den Bleigehalt. Es gilt:

$$X \sim \mathcal{N}(\mu, \sigma^2) \text{ mit } \mu = 32 \text{ und } \sigma^2 = 6^2$$

Mit **Python** kann die Wahrscheinlichkeit $P(X \leq 40)$ direkt (ohne Transformation) berechnet werden: (zu **R**)

```
from scipy.stats import norm

norm.cdf(x=40, loc=32, scale=6)

## 0.9087887802741321

print(norm.cdf(x=40, loc=32, scale=6))

## 0.9087887802741321
```

Standardisieren wir die Zufallsvariable X, so lautet die standardisierte Zufallsvariable $Z = (X - \mu)/\sigma$. Es gilt: $Z \sim \mathcal{N}(0,1)$.

$$P[X \leq 40] = P\left[Z \leq \frac{40-32}{6}\right] = P[Z \leq 1.33] = \Phi(1.33)$$

Mit **Python** finden wir: (zu **R**)

```
from scipy.stats import norm

norm.cdf(x=(40-32)/6)
```

```
## 0.9087887802741321
print(norm.cdf(x=(40-32)/6))
## 0.9087887802741321
```

c) Wir können $P[X \leq 27]$ direkt mit **Python** berechnen (zu **R**)

```
from scipy.stats import norm

norm.cdf(x=27, loc=32, scale=6)
## 0.20232838096364308
print(norm.cdf(x=27, loc=32, scale=6))
## 0.20232838096364308
```

oder durch Standardisieren der Zufallsvariablen X, also $P[X \leq 27] = P[Z \leq -0.83] = \Phi(-0.83)$ (zu **R**)

```
from scipy.stats import norm

norm.cdf(x=-0.83)
## 0.2032693918280684
print(norm.cdf(x=-0.83))
## 0.2032693918280684
```

d) Mit **Python** lässt sich die 97.5 %-Quantile berechnen mit (zu **R**)

```
from scipy.stats import norm

norm.ppf(q=0.975, loc=32, scale=6)
## 43.759783907240326
print(norm.ppf(q=0.975, loc=32, scale=6))
## 43.759783907240326
```

Alternativ lässt sich die 97.5 %-Quantile, die wir hier mit c bezeichnen, folgendermassen berechnen: es gilt $P[X \leq c] = 0.975 = P[Z \leq \frac{c-32}{6}] = \Phi(\frac{c-32}{6})$. Mit Hilfe von **Python** findet man (zu **R**)

```
from scipy.stats import norm

norm.ppf(q=0.975)
## 1.959963984540054
print(norm.ppf(q=0.975))
## 1.959963984540054
```

also $\Phi(1.96) = 0.975$. Also muss gelten:

$$\frac{c-32}{6} = 1.96 \text{ und deshalb } c = 32 + 1.96 \cdot 6 = 43.76$$

e) Mit **Python** ergibt sich: (zu **R**)

```python
from scipy.stats import norm

norm.ppf(q=0.1, loc=32, scale=6)

## 24.310690606732397

print(norm.ppf(q=0.1, loc=32, scale=6))

## 24.310690606732397
```

f) Direkt mit **Python** berechnet finden wir (zu **R**)

```python
from scipy.stats import norm

norm.cdf(x=32+6, loc=32, scale=6) - norm.cdf(x=32-6, loc=32, scale=6)

## 0.6826894921370859

print(norm.cdf(x=32+6, loc=32, scale=6) - norm.cdf(x=32-6, loc=32, scale=6))

## 0.6826894921370859
```

Andererseits gilt $\Phi(1) - \Phi(-1) = 2 \cdot \Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 0.6826$, wobei wir $\Phi(1)$ folgendermassen berechnen (zu **R**)

```python
from scipy.stats import norm

norm.cdf(x=1)

## 0.8413447460685429

print(norm.cdf(x=1))

## 0.8413447460685429
```

Lösung 4.4

a) Wir bezeichnen mit der Zufallsvariablen N die Spannung des Hintergrundrauschens. Die gesuchte Wahrscheinlichkeit lautet dann $P(N > 0.9)$ für die Verteilung $\mathcal{N}(0, 0.45^2)$: (zu **R**)

```python
from scipy.stats import norm

1-norm.cdf(x=0.9, loc=0, scale=0.45)

## 0.022750131948179211

print(1-norm.cdf(x=0.9, loc=0, scale=0.45))

## 0.022750131948179211
```

Diese Wahrscheinlichkeit kann als die Wahrscheinlichkeit einer falschen Detektion aufgefasst werden.

Oder mit Standardisieren: (zu **R**)

$$P(N > 0.9) = P\left(\frac{N-0}{0.45} > \frac{0.9-0}{0.45}\right)$$
$$= P(Z > 2) = 1 - P(Z \leq 2) = 1 - 0.97725 = 0.02275$$

```
from scipy.stats import norm

1-norm.cdf(2)

## 0.022750131948179209

print(1-norm.cdf(2))

## 0.022750131948179209
```

b) Wir müssen also ein c finden, so dass $P(-c < N < c) = 0.99$. Dies heisst nichts anderes als

$$P(-c < N) = 0.005 \qquad \text{und} \qquad P(N < c) = 0.995$$

Damit erhalten wir für $-c$ (zu **R**)

```
from scipy.stats import norm

norm.ppf(q=0.005, loc=0, scale=0.45)

## -1.1591231865970053

print(norm.ppf(q=0.005, loc=0, scale=.45))

## -1.1591231865970053
```

Daraus erhalten wir $c = 1.16$ und das Intervall, in dem 99 % des Rauschsignals enthalten ist, lautet:

$$[-1.16, 1.16]$$

Oder mit Standardisieren

$$\begin{aligned} P(-c < N < c) &= P\left(\frac{-c-0}{0.45} < \frac{N-0}{0.45} < \frac{c-0}{0.45}\right) \\ &= P\left(\frac{-c}{0.45} < Z < \frac{c}{0.45}\right) \\ &= 0.99 \end{aligned}$$

Aus der Symmetrie der Standardnormalverteilung folgt $P(Z < \frac{c}{0.45}) = 0.005$. Mit (zu **R**)

```
from scipy.stats import norm
norm.ppf(0.005)

## -2.5758293035489

print(norm.ppf(0.005))

## -2.5758293035489
```

Oder mit Standardisierung finden wir, dass $P(Z < 2.58) = 0.005$ und folglich ist $\frac{c}{0.45} = 2.58$.

c) Wir bezeichnen mit S die Spannung, wenn eine digitale 1 übertragen wird. Es gilt $S \sim \mathcal{N}(1.8, 0.45^2)$ und wir erhalten für $P(S < 0,9)$ (zu **R**)

```
from scipy.stats import norm

norm.cdf(x=0.9, loc=1.8, scale=0.45)

## 0.022750131948179195

print(norm.cdf(x=0.9, loc=1.8, scale=0.45))

## 0.022750131948179195
```

Diese Wahrscheinlichkeit kann interpretiert werden als die Wahrscheinlichkeit, ein Signal zu verpassen.

Oder mit Standardisieren (zu **R**)

$$P(S < 0.9) = P\left(\frac{S - 1.8}{0.45} < \frac{0.9 - 1.8}{0.45}\right) = P(Z < -2) = 0.02275$$

```
from scipy.stats import norm
norm.cdf(-2)

## 0.022750131948179195

print(norm.cdf(-2))

## 0.022750131948179195
```

Lösung 4.5

a) Die Integration der Wahrscheinlichkeitsdichtefunktion über ihren gesamten Wertebereich muss 1 ergeben:

$$1 = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = c\int_{0}^{60} x(15 - \frac{x}{4})\,\mathrm{d}x = c\left[\frac{15}{2}x^2 - \frac{1}{12}x^3\right]_0^{60}$$

also ist

$$1 = c\left((27000 - 18000) - 0\right)$$

und daher

$$c = \frac{1}{9000}$$

b) In Aufgabe a) haben wir die Funktion bereits integriert: Für $0 \leq x \leq 60$ haben wir

$$F(x) = \frac{1}{9000}\left(\frac{15}{2}x^2 - \frac{1}{12}x^3\right)$$

Über alle Wertebereiche lautet die kumulative Verteilungsfunktion:

$$F(x) = \begin{cases} 0 & \text{falls } x < 0 \\ \frac{1}{9000}(\frac{15}{2}x^2 - \frac{1}{12}x^3) & \text{falls } 0 \leq x \leq 60 \\ 1 & \text{falls } x > 60 \end{cases}$$

c) Wir suchen ein a, so dass gilt $P(X > a) = 0.1$, d. h. $P(X \leq a) = 0.9$. Es muss also gelten $F(a) = 0.9$. Daraus folgt mit der Formel für F aus Aufgabe b), dass

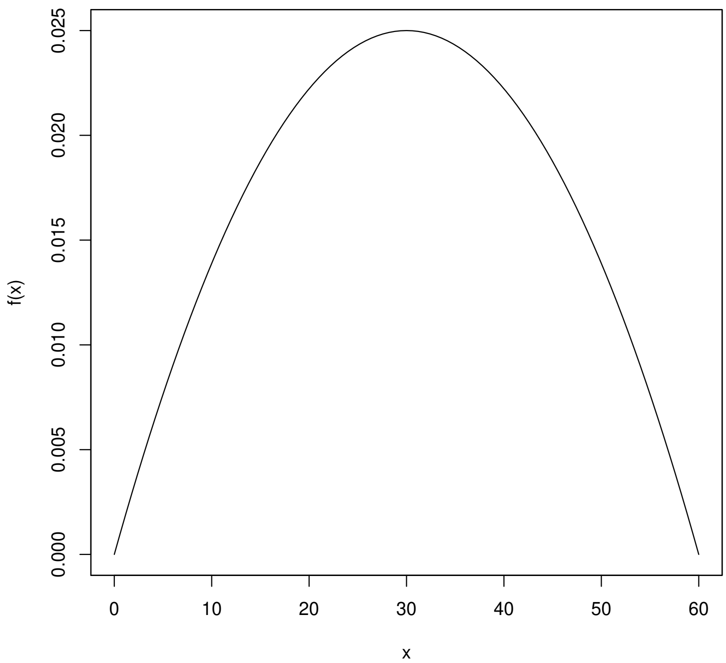
$$\frac{1}{12}a^3 - \frac{15}{2}a^2 + 8100 = 0$$

Diese kubische Gleichung hat 3 Lösungen, es liegt jedoch nur eine davon im Bereich $[0, 60]$, nämlich $a = 48.25$. a ist gerade das sogenannte 90 %-Quantil von F.

d) Der Erwartungswert kann wie folgt berechnet werden:

$$\begin{aligned} E(X) = &\int_{-\infty}^{\infty} x f(x) \, \mathrm{d}x \\ =&\frac{1}{9000} \int_0^{60} x^2 \left(15 - \frac{x}{4}\right) \, \mathrm{d}x \\ =&\frac{1}{9000} \left[5x^3 - \frac{1}{16}x^4\right]_0^{60} \\ =&\frac{1}{9000}(1080000 - 810000) \\ =&30 \end{aligned}$$

Alternativ: Da die Dichte um den Wert 30 symmetrisch ist, folgt sofort, dass der Erwartungswert 30 ist (da der Erwartungswert physikalisch gesehen der Schwerpunkt ist).



Lösung 4.6

a) Damit $f(x)$ eine Dichte ist, muss die Fläche des Dreiecks gleich 1 sein. Es muss also gelten

$$\frac{c \cdot 20}{2} = 1$$

Daraus folgt $c = \frac{1}{10}$. Die Dichte lässt sich somit durch die Funktion

$$f(x) = \begin{cases} 0 & x < 0 \text{ und } x > 20 \\ \frac{1}{10}(1 - \frac{x}{20}) & 0 \leq x \leq 20 \end{cases}$$

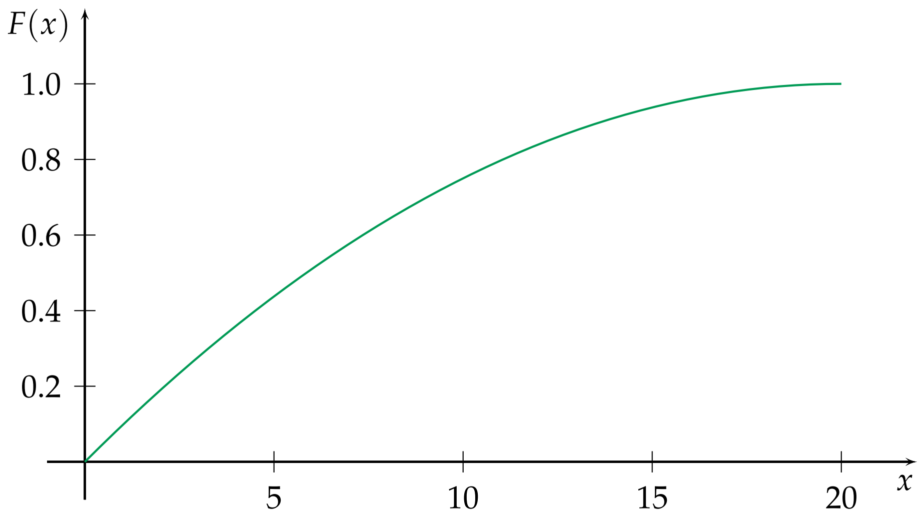
beschreiben.

b) Die kumulative Verteilungsfunktion von X lässt sich durch Integration der Dichtefunktion berechnen: Für $0 \leq x \leq 20$ gilt:

$$F(x) = P[X \leq x] = \int_0^x f(t)\,\mathrm{d}t = \int_0^x \left(\frac{1}{10} - \frac{t}{200}\right)\,\mathrm{d}t = \frac{x}{10} - \frac{x^2}{400}$$

Für $x \leq 0$ ist $F(x) = 0$ und für $x \geq 20$ gilt $F(x) = 1$. Insbesondere gilt: $P[X < 5] = F(5) = 0.4375$ und $P[X < 10] = F(10) = 0.75$.

c) Die kumulative Verteilungsfunktion wurde bereits in b) berechnet. Skizze von $F(x)$:



d)

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_{0}^{20} x\left[\frac{1}{10}\left(1 - \frac{x}{20}\right)\right]\,\mathrm{d}x = \frac{1}{10}\left(\frac{x^2}{2} - \frac{x^3}{60}\right)\Bigg|_0^{20} = \frac{20}{3}$$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$$

$$\mathrm{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \int_{0}^{20} x^2 \frac{1}{10}\left(1 - \frac{1}{20}\right)\,\mathrm{d}x = \frac{1}{10}\left(\frac{x^3}{3} - \frac{x^4}{80}\right)\Bigg|_0^{20} = \frac{200}{3}$$

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \frac{200}{3} - \left(\frac{20}{3}\right)^2 = \frac{200}{9}$$

Also ist die Standardabweichung $\sigma_X = \sqrt{\mathrm{Var}(X)} = \sqrt{2} \cdot \frac{10}{3} \approx 4.71$. Für den Median $\tilde{m}$ muss gelten: $F(\tilde{m}) \stackrel{!}{=} 0.5$. Der Median liegt sicher im Intervall $[0, 20]$ und somit haben wir

$$\frac{\tilde{m}}{10} - \frac{\tilde{m}^2}{400} \stackrel{!}{=} 0.5 \quad \Rightarrow \quad \tilde{m} = 20 - 10\sqrt{2} \approx 5.858$$

e)

$$\begin{aligned} P[K \le 120'000] &= P[40'000 \cdot \sqrt{X} \le 120'000] \\ &= P[\sqrt{X} \le 3] \end{aligned}$$

$$
\begin{aligned}
&= P[X \leq 9] \\
&= F(9) \\
&= \frac{9}{10} - \frac{9^2}{400} \\
&= 0.6975
\end{aligned}
$$

f) Die Exponentialverteilung hat die Dichte $g(x)$:

$$
g(x) = \begin{cases} 0 & x \leq 0 \\ \lambda \exp(-\lambda x) & x > 0 \end{cases}
$$

Wenn X exponentialverteilt ist, dann ist der Erwartungswert

$$
E(X) = \frac{1}{\lambda}
$$

Für $\lambda = \frac{3}{20}$ erhalten wir somit denselben Erwartungswert wie in der bisherigen Verteilung.

g) Die kumulative Verteilungsfunktion $G(x)$ ist für $x > 0$

$$
G(x) = P[X \leq x] = 1 - \exp(-\lambda x)
$$

Daher

$$
\begin{aligned}
P[K \leq 120'000] &= P[40'000 \cdot \sqrt{X} \leq 120'000] \\
&= P[\sqrt{X} \leq 3] \\
&= P[X \leq 9] \\
&= G(9) \\
&= 1 - \exp\left(-\frac{3}{20} \cdot 9\right) \\
&= 1 - 0.259 \\
&= 0.741
\end{aligned}
$$

Wenn die Dauer der Baustellen als exponentialverteilt angenommen wird, ist die Wahrscheinlichkeit also grösser, dass die Kosten einer Baustelle unter 120'000 Fr. liegen, verglichen mit der ursprünglich angenommenen Verteilung, obwohl der Erwartungswert für die Dauer der Baustellen identisch ist für beide Verteilungen.