Aufgabe 1

Eine Social-Media-Plattform modelliert die tägliche Screentime ihrer Nutzer, welche in Gruppe 1 und Gruppe 2 unterteilt wurden. Seien

X : „tägliche Screentime eines zufällig ausgewählten Nutzers aus Gruppe 1 (in Std.)“,

Y : „tägliche Screentime eines zufällig ausgewählten Nutzers aus Gruppe 2 (in Std.)“.

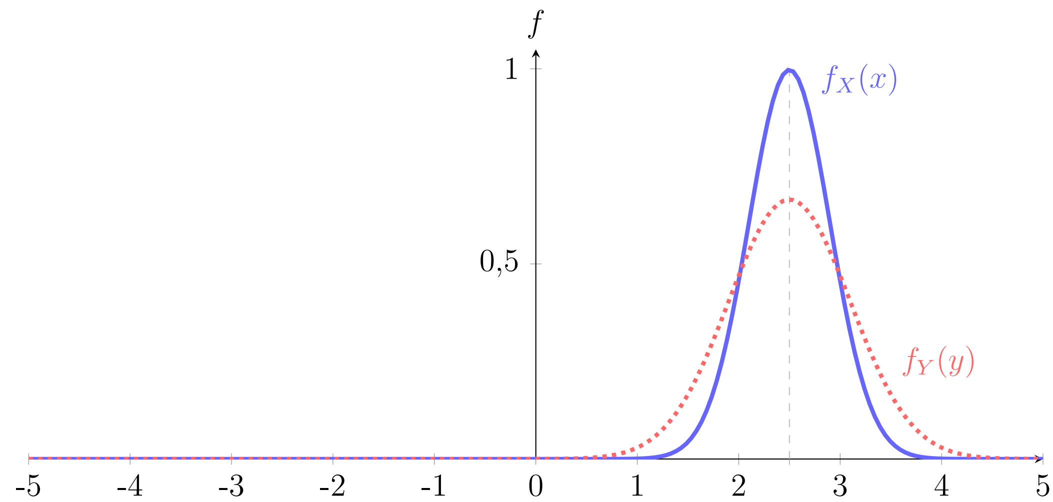
Es wird angenommen, dass die Screentime in beiden Gruppen normalverteilt ist. Für beide Gruppen liegt der Erwartungswert der täglichen Nutzungsdauer bei 2,5 Stunden ($\mu_X = \mu_Y$). Die beiden Gruppen unterscheiden sich jedoch in ihrer Streuung σ. Es gelten $\sigma_X = 0{,}4$ und $\sigma_Y = 0{,}6$.

Teil A

a) Geben Sie die Verteilung von X und Y jeweils an (Verteilungstyp und -parameter).

$X \sim \mathcal{N}(\mu; \sigma^2)$ mit $\mu = 2{,}5$ und $\sigma^2 = 0{,}4^2 = 0{,}16$
$Y \sim \mathcal{N}(\mu; \sigma^2)$ mit $\mu = 2{,}5$ und $\sigma^2 = 0{,}6^2 = 0{,}36$

b) Ordnen Sie die nachstehenden Dichtefunktionen den Zufallsvariablen X und Y zu.



Betrachten Sie nun die folgenden Transformationen der Zufallsvariable X:

i) $\tilde{X} := X - \mathbb{E}(X)$ ii) $Z := \dfrac{X - \mathbb{E}(X)}{\sqrt{Var(X)}}$

c) Wie sind die Zufallsvariable $\tilde{X}$ und Z jeweils verteilt (Verteilungstyp- und parameter)? Berechnen Sie hierfür den jeweiligen Erwartungswert und die Varianz. Geben Sie die Dichtefunktion von $\tilde{X}$ und Z explizit an und skizzieren Sie diese in dem Koordinatensystem in b). Wie nennt man diese Transformationen?

i) Diese Transformation heißt *Zentrierung* der Zufallsvariable X.

$$\mathbb{E}(\tilde{X}) = \mathbb{E}(X - \underbrace{\mathbb{E}(X)}_{Konstante}) = \mathbb{E}(X) - \mathbb{E}(X) = 0$$

$$Var(\tilde{X}) = Var(X - \underbrace{\mathbb{E}(X)}_{Konstante}) = Var(X) = \sigma_X^2 = 0{,}16$$

Es folgt $\tilde{X} \sim \mathcal{N}(0; 0{,}16)$. Die Dichtefunktion von $\tilde{X}$ ist

$$f_{\tilde{X}}(\tilde{x}) = \frac{1}{\sqrt{2\pi} \cdot 0{,}4} \cdot e^{\frac{-\tilde{x}^2}{2 \cdot 0{,}16}} \quad \text{für alle } \tilde{x} \in \mathbb{R}.$$

ii) Diese Transformation heißt *Standardisierung* der Zufallsvariable X.
(*Zusatz:* Sie impliziert die Zentrierung und die Normierung von X.)

$$\mathbb{E}(Z) = \mathbb{E}\Big(\tfrac{X - \mathbb{E}(X)}{\sqrt{Var(X)}}\Big) = \mathbb{E}\Big(\tfrac{\tilde{X}}{\sqrt{Var(X)}}\Big) = \tfrac{1}{\sqrt{Var(X)}}\mathbb{E}(\tilde{X}) = \tfrac{1}{\sqrt{Var(X)}} \cdot 0 = 0$$

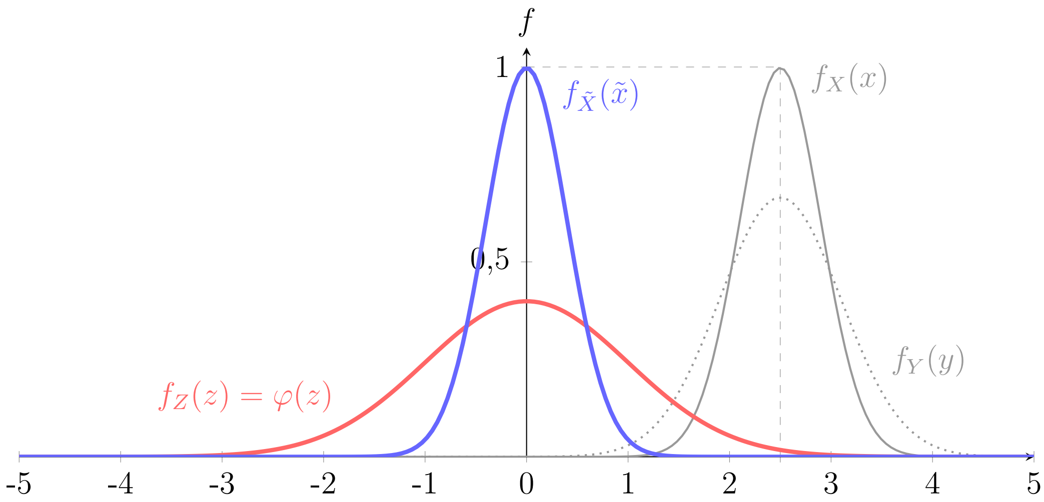
$$Var(Z) = Var\Big(\tfrac{X - \mathbb{E}(X)}{\sqrt{Var(X)}}\Big) = Var\Big(\tfrac{\tilde{X}}{\sqrt{Var(X)}}\Big) = \tfrac{1}{Var(X)} \cdot Var(\tilde{X}) = \tfrac{Var(X)}{Var(X)} = 1$$

Es folgt $Z \sim \mathcal{N}(0; 1)$. Die Zufallsvariable ist somit *standardnormalverteilt.* Die Dichtefunktion von Z ist

$$f_Z(z) = \varphi(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-z^2}{2}} \quad \text{für alle } z \in \mathbb{R}.$$

(*Nebenrechnung:* Für die Skizze von f_Z berechnen wir $f_Z(0) = \frac{1}{\sqrt{2\pi}} \approx 0{,}3989$.)

Skizze zu c)

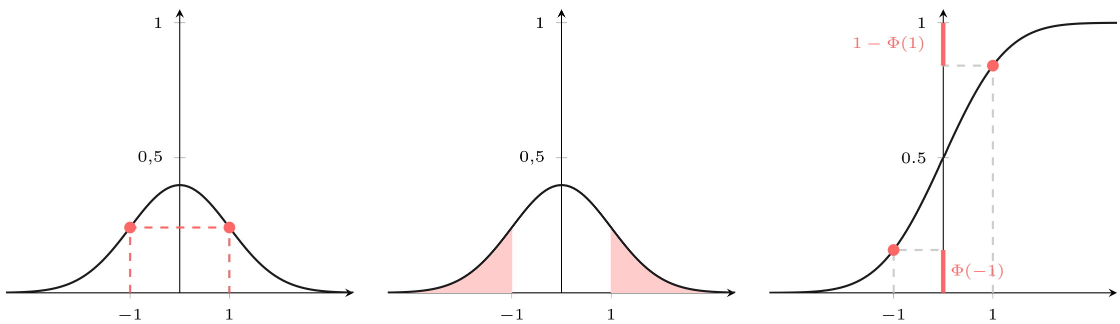


d) Für eine standardnormalverteilte Zufallsvariable Z kann formal gezeigt werden, dass die nachfolgenden Symmetrieeigenschaften gelten:

$$\text{i) } \varphi(z) = \varphi(-z) \quad \text{ii) } \mathbb{P}(Z \leq -z) = \mathbb{P}(Z \geq z) \quad \text{iii) } \Phi(-z) = 1 - \Phi(z)$$

für alle $z \in \mathbb{R}$, wobei φ die Dichtefunktion und Φ die Verteilungsfunktion von Z darstellen. Visualisieren Sie die nachfolgenden Eigenschaften jeweils für den Fall $\boldsymbol{z = 1}$ in den bereitgestellten Skizzen.

i) $\varphi(1) = \varphi(-1)$ ii) $\mathbb{P}(Z \leq -1) = \mathbb{P}(Z \geq 1)$ iii) $\Phi(-1) = 1 - \Phi(1)$



e) Übertragen Sie die in d) gegebenen Symmetrieeigenschaften der Standardnormalverteilung auf eine normalverteilte Zufallsvariablen X. Es ist keine formale Herleitung erforderlich.

Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit Verteilungsfunktion F und zugehöriger Dichtefunktion f.
Es gilt für alle $x \geq 0$:

i) $f(\mu + x) = f(\mu - x)$

ii) $\mathbb{P}(X \leq \mu - x) = \mathbb{P}(X \geq \mu + x)$

iii) $F(\mu - x) = 1 - F(\mu + x)$

f) Berechnen Sie für einen Nutzer aus Gruppe 1 die Wahrscheinlichkeit, dass die tägliche Screentime

i) weniger als 4 Stunden beträgt.

ii) mehr als 2 Stunden beträgt.

iii) zwischen 2 und 3 Stunden beträgt.

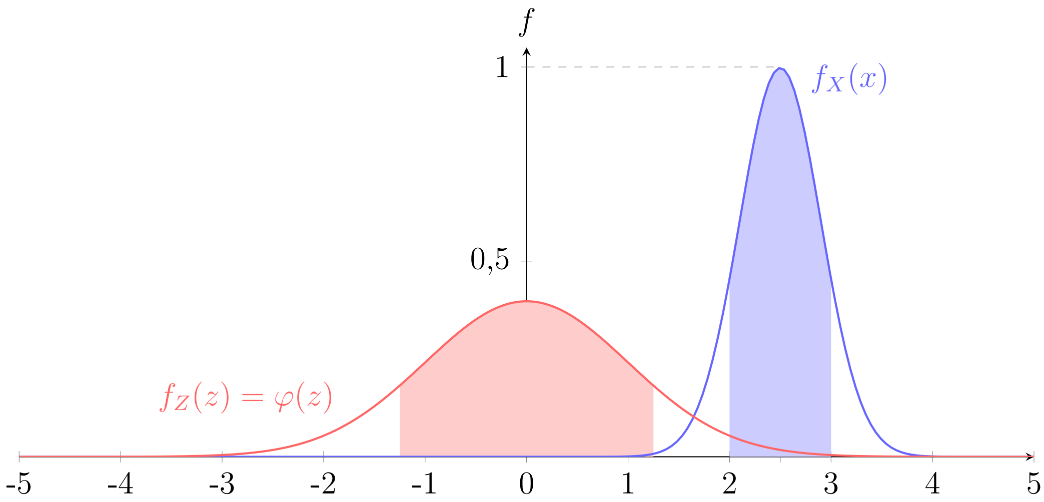
Veranschaulichen Sie die in iii) berechnete Wahrscheinlichkeit, indem Sie die zugehörige Wahrscheinlichkeitsmasse als Fläche unter der Dichtefunktion von X und Z in b) markieren.

i) $$\mathbb{P}(X < 4) = \mathbb{P}\left(\underbrace{\frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}}}_{=:Z \ mit \ Z \sim \mathcal{N}(0,1)} < \frac{4 - \mathbb{E}(X)}{\sqrt{Var(X)}}\right) = \mathbb{P}\left(\frac{X - 2{,}5}{\sqrt{0{,}16}} < \frac{4 - 2{,}5}{\sqrt{0{,}16}}\right)$$
$$= \mathbb{P}\left(Z < \frac{4 - 2{,}5}{0{,}4}\right) = \mathbb{P}(Z < 3{,}75)$$
$$= \mathbb{P}(Z \leq 3{,}75) = \Phi(3{,}75) = 0{,}999912$$

ii) $$\mathbb{P}(X > 2) = \mathbb{P}(Z > \tfrac{2 - 2{,}5}{0{,}4}) = \mathbb{P}(Z \geq -1{,}25)$$
$$= \mathbb{P}(Z \leq 1{,}25) = \Phi(1{,}25) = 0{,}894350$$

iii) $$\mathbb{P}(2 < X < 3) = \mathbb{P}\left(\frac{2 - 2{,}5}{0{,}4} < Z < \frac{3 - 2{,}5}{0{,}4}\right) = \mathbb{P}\left(-1{,}25 \leq Z \leq 1{,}25\right)$$
$$= \Phi(1{,}25) - \Phi(-1{,}25) = \Phi(1{,}25) - (1 - \Phi(1{,}25))$$
$$= 2 \cdot \Phi(1{,}25) - 1 = 2 \cdot 0{,}894350 - 1 = 0{,}7887$$

Skizze zu f)



Teil B

Ein Nutzer wird als „Digital Addict" eingestuft, sofern seine tägliche Screentime das 95te Perzentil der Verteilung von X überschreitet. Die Grundgesamtheit setzt sich dabei zu zwei Dritteln aus Nutzern der Gruppe 1 und zu einem Drittel aus Nutzern der Gruppe 2 zusammen.

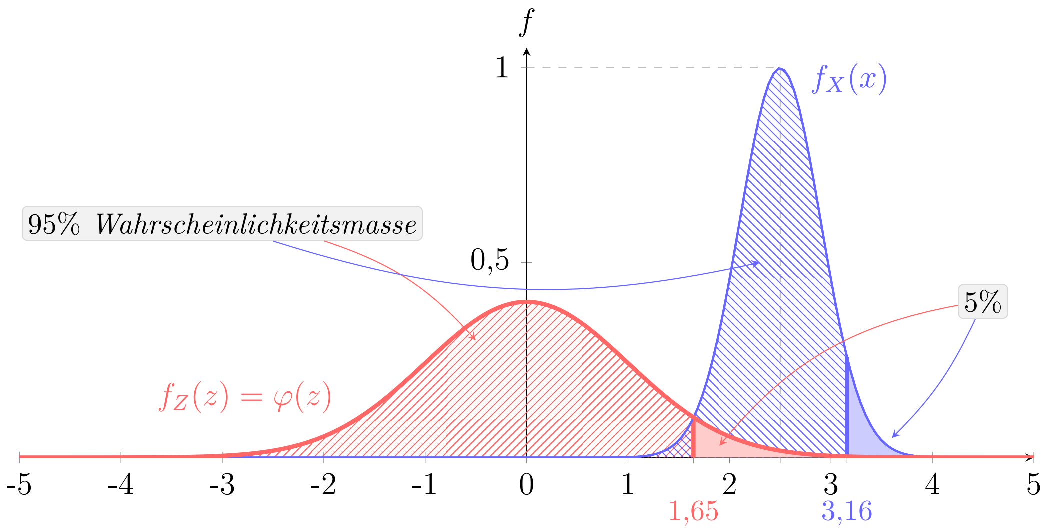
g) Berechnen Sie die tägliche Screentime, ab der ein Nutzer als „Digital Addict" gilt.

Vorüberlegung: Gesucht ist $x \in \mathbb{R}$, für welches $\mathbb{P}(X \leq x) = F_X(x) \overset{!}{=} 0{,}95$ gilt. Da F_X in der Formelsammlung nicht tabellarisiert ist, definieren wir $Z := \frac{X-2{,}5}{0{,}4}$. Für die Realisierungen von Z gilt $z = \frac{x-2{,}5}{0{,}4} \Leftrightarrow x = 2{,}5 + z \cdot 0{,}4 \quad (\star)$. Offensichtlich ist Z standardnormalverteilt, d.h. $Z \sim \mathcal{N}(0,1)$ (vgl. Teilaufgabe c) ii)). Es folgt

$$\mathbb{P}(X \leq x) = F_X(x) \overset{!}{=} 0{,}95 \quad \Leftrightarrow \quad \mathbb{P}(Z \leq z) = \Phi(z) \overset{!}{=} 0{,}95 \quad \Leftrightarrow \quad z = 1{,}65.$$

Eingesetzt in $(\star)$: $x = 2{,}5 + 1{,}65 \cdot 0{,}4 = 3{,}16$ (also ab ca. 3h und 9 Minuten)

Zusatz: Visualisierung zu g)



h) Mit welcher Wahrscheinlichkeit gilt ein zufällig ausgewählter Nutzer aus Gruppe 2 als „Digital Addict“?

$$\mathbb{P}(Y \geq 3{,}16) = \mathbb{P}(Z \geq \tfrac{3{,}16-2{,}5}{0{,}6}) = \mathbb{P}(Z \geq 1{,}10) = 1 - \mathbb{P}(Z < 1{,}10)$$
$$= 1 - \mathbb{P}(Z \leq 1{,}10) = 1 - \Phi(1{,}10) = 1 - 0{,}864334 = 0{,}135666$$

i) Mit welcher Wahrscheinlichkeit gilt ein zufällig ausgewählter Nutzer als „Digital Addict“?

Sei D:= „Nutzer ist Digital Addict“. Mit dem *Satz der Totalen Wahrscheinlichkeit* gilt nun

$$\mathbb{P}(D) = \mathbb{P}(X \geq 3{,}16) \cdot \frac{2}{3} + \mathbb{P}(Y \geq 3{,}16) \cdot \frac{1}{3} \approx 0{,}05 \cdot \frac{2}{3} + 0{,}1357 \cdot \frac{1}{3} = 0{,}0786.$$

Teil C

Es wird angenommen, dass jede studentische Wohngemeinschaft (WG) aus genau drei Personen besteht. Dabei gehören Person 1 und Person 2 der Gruppe 1 an, während Person 3 der Gruppe 2 angehört. Die Mediennutzungsdauer der einzelnen Personen erfolgt unabhängig voneinander. Sei

S : „durchschnittliche tägliche Screentime pro Kopf in einer zufällig ausgewählten WG“.

j) Wie ist die Zufallsvariable S verteilt (Verteilungstyp und -parameter)?

Vorüberlegung: Sei X_i :„tägliche Screentime des i-ten zufällig ausgewählten Nutzers aus Gruppe 1 (in Std.)“ mit $X_i \sim \mathcal{N}(2{,}5;\ 0{,}16)$ für $i = 1, 2$. Die Zufallsvariable S kann als Linearkombination von X_1, X_2 und Y beschrieben werden: $S := \frac{1}{3}(X_1 + X_2 + Y)$. Da jede Linearkombination von unabhängigen normalverteilten Zufallsvariablen erneut normalverteilt ist, folgt direkt

$$S \sim \mathcal{N}(\mu_S, \sigma_S^2)$$

mit

$$\begin{aligned}\mu_S &= \mathbb{E}(S) = \mathbb{E}(\frac{1}{3}(X_1 + X_2 + Y)) = \frac{1}{3}(\mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(Y)) \\ &= \frac{1}{3}(2{,}5 + 2{,}5 + 2{,}5) = 2{,}5 \quad \text{und} \\ \sigma_S^2 &= Var(S) = Var(\frac{1}{3}(X_1 + X_2 + Y)) \overset{(\star)}{=} \frac{1}{9}(Var(X_1) + Var(X_2) + Var(Y)) \\ &= \frac{1}{9}(0{,}16 + 0{,}16 + 0{,}36)) = 0{,}0756.\end{aligned}$$

Zusatz: In ($\star$) wird die vollständige stochastische Unabhängigkeit von Y, X_1, X_2 ausgenutzt.

k) Mit welcher Wahrscheinlichkeit ist die Screentime von Person 3 größer als die Screentime von Person 1 und 2 zusammen? Definieren Sie hierfür eine neue Zufallsvariable.

Sei $D := Y - X_1 - X_2$. Es folgt

$$D \sim \mathcal{N}(\mu_D, \sigma_D^2)$$

mit

$\mu_D = \mathbb{E}(D) = \mathbb{E}(Y - X_1 - X_2) = 2{,}5 - 2{,}5 - 2{,}5 = -2{,}5$ und

$\sigma_D^2 = Var(D) = Var(Y - X_1 - X_2) = Var(Y - 1 \cdot X_1 - 1 \cdot X_2)$

$\overset{(\star)}{=} Var(Y) + (-1)^2 \cdot Var(X_1) + (-1)^2 \cdot Var(X_2) = 0{,}36 + 0{,}16 + 0{,}16 = 0{,}68.$

Gesucht ist die Wahrscheinlichkeit des Ereignisses $\{D > 0\}$:

$$\begin{aligned}\mathbb{P}(D > 0) &= \mathbb{P}(Z > \tfrac{0-(-2{,}5)}{\sqrt{0{,}68}}) \approx \mathbb{P}(Z > 3{,}03) = 1 - \mathbb{P}(Z \leq 3{,}03)\\ &= 1 - \Phi(3{,}03) = 1 - 0{,}998777 = 0{,}001223.\end{aligned}$$

Zusatz: In $(\star)$ wird die vollständige stochastische Unabhängigkeit von Y, X_1, X_2 ausgenutzt.

Aufgabe 2

Ein Online-Händler weiß, dass 20% der 100 täglich versendeten Pakete als Retouren zurückgesendet werden. Sei

X : „Anzahl der an einem zufällig ausgewählten Tag retournierten Pakete".

a) Wie ist die Zufallsvariable X

   i) exakt,

   ii) approximativ

   verteilt (Verteilungstyp und -parameter)? Prüfen Sie die Approximationskriterien.

i) Exakt gilt $X \sim \mathrm{B}(n; \pi)$ mit $n = 100$ und $\pi = 0{,}2$.
ii) Approximativ gilt $X \overset{\text{appr.}}{\sim} \mathcal{N}(\mu; \sigma^2)$ mit $\mu = n \cdot \pi = 100 \cdot 0{,}2 = 20$ und $\sigma^2 = n \cdot \pi \cdot (1 - \pi) = 100 \cdot 0{,}2 \cdot 0{,}8 = 16$. Die Normal-Approximation ist zulässig, da $n \cdot \pi = 20 \geq 5$ und $n \cdot (1 - \pi) = 80 \geq 5$ erfüllt ist.

b) Wie groß ist die Wahrscheinlichkeit, dass genau 20 Pakete retourniert werden nach der exakten **und** approximativen Verteilung? Um wie viel Prozent überschätzt die Approximation die wahre Wahrscheinlichkeit?

Exakt: $\mathbb{P}(X = 20) = \binom{100}{20} \cdot 0{,}2^{20} \cdot 0{,}8^{80} \approx 0{,}0993$

Approximativ (über die Stetigkeitskorrektur):

$$\begin{aligned}
\mathbb{P}(X = 20) &= \mathbb{P}(19{,}5 \leq X \leq 20{,}5) = \mathbb{P}\Big(\tfrac{19{,}5-20}{\sqrt{16}} \leq Z \leq \tfrac{20{,}5-20}{\sqrt{16}}\Big) \\
&= \mathbb{P}(-\tfrac{1}{8} < Z \leq \tfrac{1}{8}) = \Phi(\tfrac{1}{8}) - \Phi(\tfrac{1}{8}) = \Phi(\tfrac{1}{8}) - (1 - \Phi(\tfrac{1}{8})) \\
&\approx 2 \cdot \Phi(0{,}13) - 1 = 2 \cdot 0{,}551717 - 1 = 0{,}103434 \approx 0{,}1034
\end{aligned}$$

Die Approximation überschätzt die wahre Wahrscheinlichkeit um

$$\frac{0{,}1034 - 0{,}0993}{0{,}0993} \cdot 100 \approx 4,13\%.$$

c) Mit welcher approximativen Wahrscheinlichkeit werden mehr als 30% der Pakete retourniert? Berechnen Sie die Wahrscheinlichkeit einmal mit **und** einmal ohne Stetigkeitskorrektur.

Mit Stetigkeitskorrektur:

$$\begin{aligned}\mathbb{P}(X > 30) &= \mathbb{P}(X > 29{,}5) = \mathbb{P}\Big(Z > \tfrac{29{,}5-20}{\sqrt{16}}\Big) \approx \mathbb{P}(Z > 2{,}38)\\ &= 1 - \mathbb{P}(Z \leq 2,38) = 1 - \Phi(2{,}38) = 1 - 0{,}991344 = 0{,}008656\end{aligned}$$

Ohne Stetigkeitskorrektur:

$$\begin{aligned}\mathbb{P}(X > 30) &= \mathbb{P}\Big(Z > \tfrac{30-20}{\sqrt{16}}\Big) = \mathbb{P}(Z > 2{,}5)\\ &= 1 - \mathbb{P}(Z \leq 2{,}5) = 1 - \Phi(2{,}5) = 1 - 0{,}993790 = 0{,}00621\end{aligned}$$

Aufgabe 3

In einem Berliner Bezirk werden pro Stunde durchschnittlich 30 Fahrten eines Car-Sharing Unternehmens beendet. Sei

Y : „Anzahl der in einer zufällig ausgewählten Stunde beendeten Fahrten“.

a) Wie ist die Zufallsvariable X

i) exakt,

ii) approximativ

verteilt (Verteilungstyp und -parameter)? Prüfen Sie die Approximationskriterien.

i) Exakt gilt $X \sim \mathrm{P}(\lambda)$ mit $\lambda = 30$.
ii) Approximativ gilt $X \overset{\text{appr.}}{\sim} \mathcal{N}(\mu; \sigma^2)$ mit $\mu = \sigma^2 = \lambda = 30$.
Die Normal-Approximation ist zulässig, da $\lambda = 30 \geq 5$ erfüllt ist.

b) Wie groß ist die Wahrscheinlichkeit, dass genau 30 Fahrten beendet werden nach der exakten **und** approximativen Verteilung?

Exakt: $\mathbb{P}(X = 30) = \dfrac{30^{30}}{30!} \cdot e^{-30} \approx 0{,}0726$

Approximativ (über die Stetigkeitskorrektur):

$$\begin{aligned}\mathbb{P}(X = 30) &= \mathbb{P}(29{,}5 \leq X \leq 30{,}5) = \mathbb{P}\Big(\tfrac{29{,}5-30}{\sqrt{30}} \leq Z \leq \tfrac{30{,}5-30}{\sqrt{30}}\Big)\\ &= \mathbb{P}(-0{,}09 \leq Z \leq 0{,}09) = \mathbb{P}(-0{,}09 < Z \leq 0{,}09)\\ &= \Phi(0{,}09) - \Phi(-0{,}09) = \Phi(0{,}09) - (1 - \Phi(0{,}09))\\ &= 2 \cdot \Phi(0{,}09) - 1 = 2 \cdot 0{,}535856 - 1 = 0{,}071712\end{aligned}$$

c) Mit welcher approximativen Wahrscheinlichkeit werden zwischen 25 und 35 Fahrten beendet? Berechnen Sie die Wahrscheinlichkeit einmal mit **und** einmal ohne Stetigkeitskorrektur.

Mit Stetigkeitskorrektur:

$$\begin{aligned}\mathbb{P}(25 \leq X \leq 35) &= \mathbb{P}(24{,}5 \leq X \leq 35{,}5) = \mathbb{P}(\tfrac{24{,}5-30}{\sqrt{30}} \leq X \leq \tfrac{35{,}5-30}{\sqrt{30}})\\ &\approx \mathbb{P}(-1 \leq X \leq 1) = \mathbb{P}(-1 < X \leq 1) = \Phi(1) - \Phi(-1)\\ &= \Phi(1) - (1 - \Phi(1)) = 2 \cdot \Phi(1) - 1 = 2 \cdot 0{,}841345 - 1 = 0{,}68269\end{aligned}$$

Ohne Stetigkeitskorrektur:

$$\begin{aligned}\mathbb{P}(25 \leq X \leq 35) &= \mathbb{P}\Big(\tfrac{25-30}{\sqrt{30}} \leq X \leq \tfrac{35-30}{\sqrt{30}}\Big) \approx \mathbb{P}(-0{,}91 \leq X \leq 0{,}91)\\ &= \mathbb{P}(-0{,}91 < X \leq 0{,}91) = \Phi(0{,}91) - \Phi(-0{,}91)\\ &= \Phi(0{,}91) - (1 - \Phi(0{,}91)) = 2 \cdot \Phi(0{,}91) - 1\\ &= 2 \cdot 0{,}818589 - 1 = 0{,}637178\end{aligned}$$