2.3 Stetige Merkmale / Zufallsvariablen

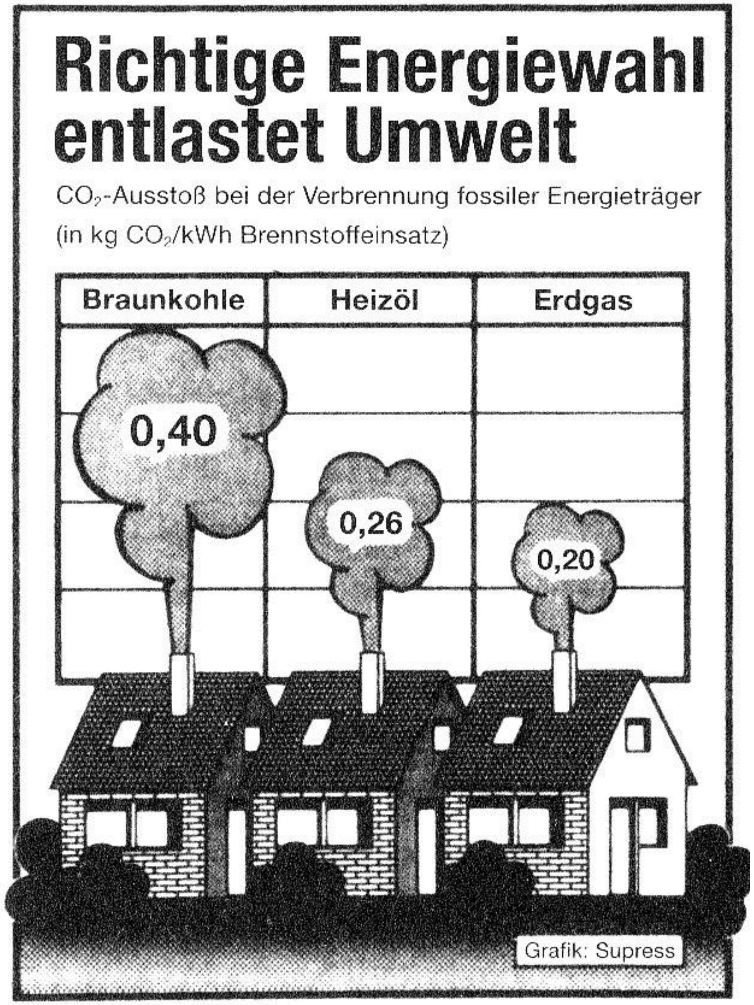
- **2.1 Grundbegriffe**
  - Typen von Merkmalen bzw. Zufallsvariablen
  - Häufigkeits- bzw. Wahrscheinlichkeits**verteilung**
  - **Kumulierte** Häufigkeits- bzw. Wahrscheinlichkeitsverteilung

- 2.2 Kennzahlen diskreter Merkmale / Zufallsvariablen
  - Arithmetischer Mittelwert / Erwartungswert
  - Andere Mittelwerte: geometrischer / harmonischer Mittelwert
  - Median, Quantil, Modus
  - Varianz / Standardabweichung

- 2.3 **Stetige** Merkmale / Zufallsvariablen
  - Wahrscheinlichkeitsdichten / Dichtefuntion
  - Übertragung der diskreten Kennzahldefinitionen

- 2.4 Wichtige **Standardverteilungen**:
  - Gleichverteilung
  - Binomialverteilung, Poissonverteilung
  - Exponentialverteilung
  - Normalverteilung

Visualisierung von stetigen, metrischen Merkmalen



Quelle: Walter Krämer: So lügt man mit Statistik

Manipulativ:

Die ***Höhe*** entspricht den Werten, aber das Auge bewertet die ***Fläche***.

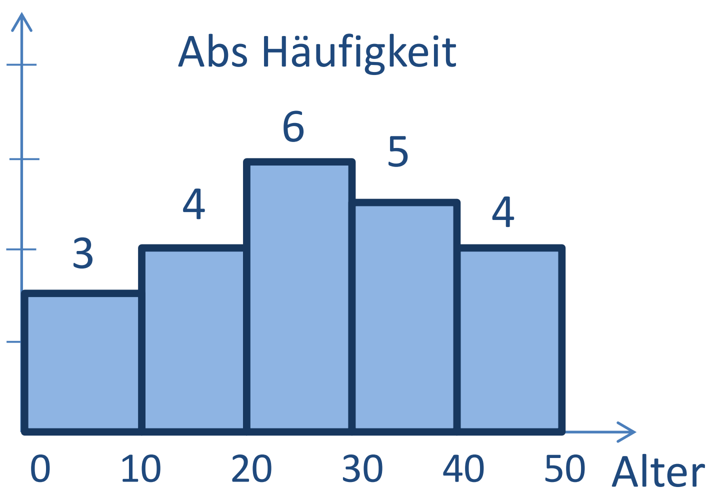
Merke:

Bei flächigen Visualisierung muss immer die Fläche – nicht der Durchmesser – proportional zur visualisierten Zahl sein.

Visualisierung von Häufigkeits- u. Wahrscheinlichkeits-verteilungen (1): **Histogramme**

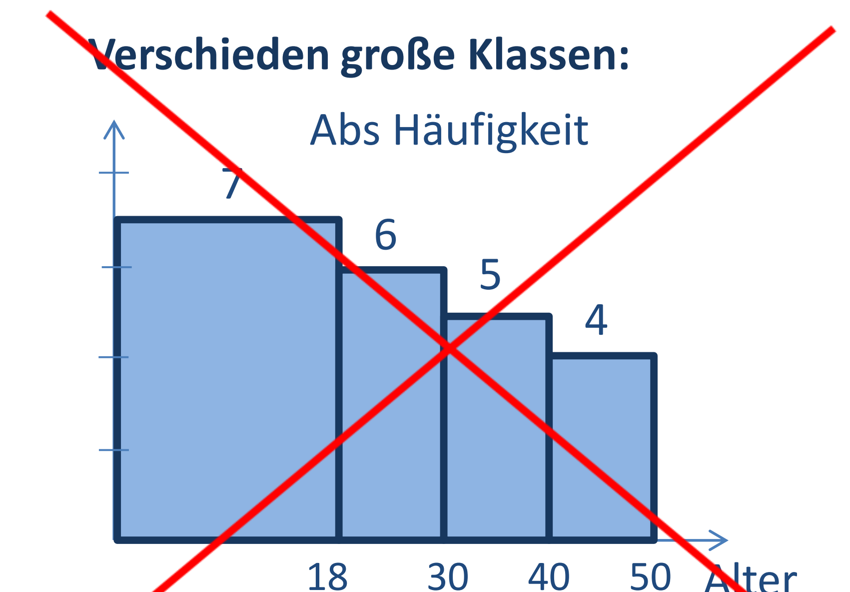
Zur Darstellung von *metrischen* Merkmalen bzw. stetigen Zufallsvariablen bietet sich an, die Wertebereichs-Klassen, die jeweils von einem Balken repräsentiert werden, durch die Balkenbreite zu visualisieren :

Normale Klasseneinteilung:



Grenzfälle wurden dem rechten Intervall zugerechnet

Verschieden große Klassen:



Grenzfälle wurden dem rechten Intervall zugerechnet

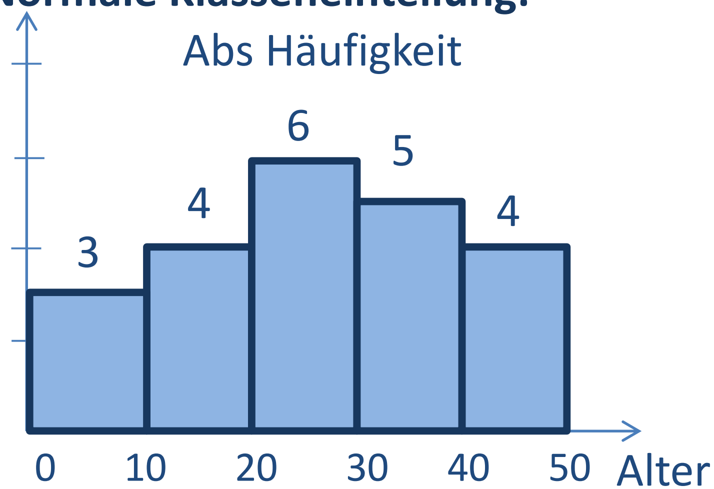
Visualisierung von Häufigkeits- u. Wahrscheinlichkeits-verteilungen (2): **Histogramme**

Zur Darstellung von *quantitativen* Merkmalen eignen sich Histogramme:

Definition: Beim **Histogramm** gibt die **Balkenbreite** die **Klassenbreite** wieder, und die **Balkenfläche** - nicht die Höhe - gibt die relative oder absolute **Häufigkeit** an.
Die Balkenhöhe berechnet sich deshalb als $\frac{\text{Häufigkeit}}{\text{Klassenbreite}}$ =: Häufigkeitsdichte

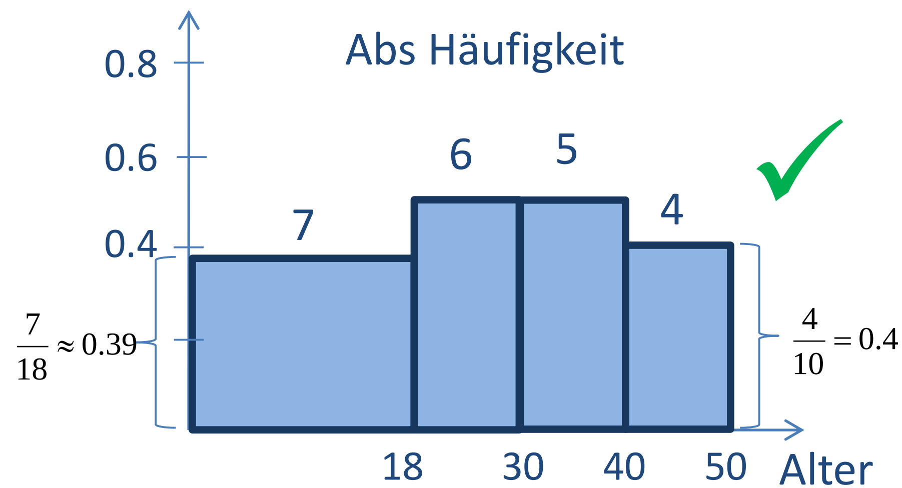
Vorteil: Unterschiedlich breite Klassen verzerren das Bild nicht

Normale Klasseneinteilung:



Grenzfälle wurden dem rechten Intervall zugerechnet

Verschieden große Klassen:

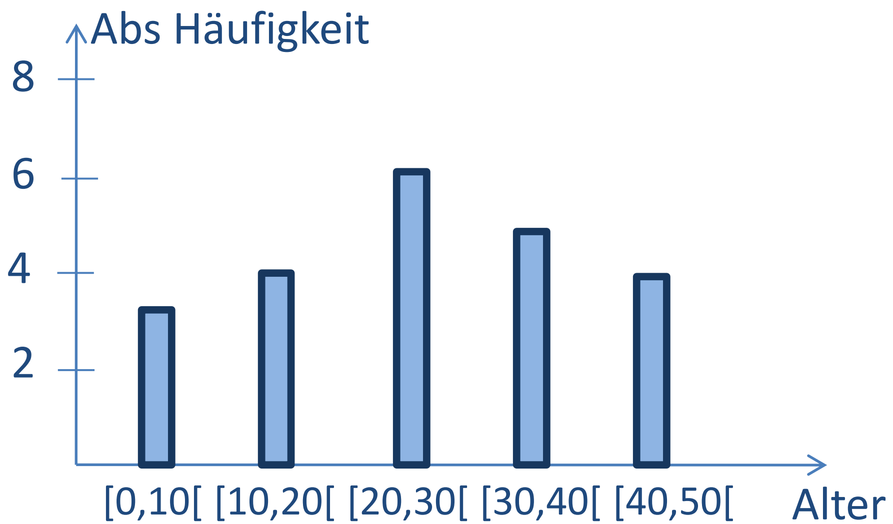


Grenzfälle wurden dem rechten Intervall zugerechnet

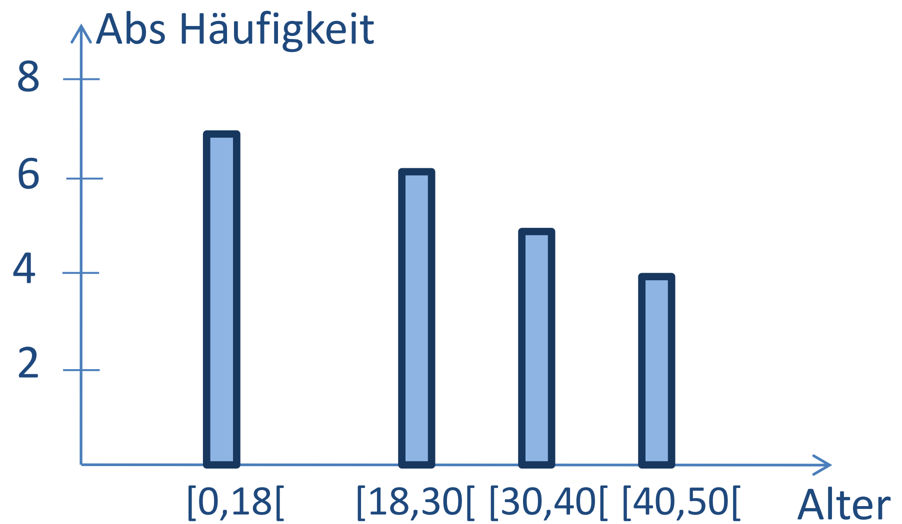
Visualisierung von Häufigkeits- u. Wahrscheinlichkeits-verteilungen (3): **Stabdiagramme**

Stabdiagrammen sind nur bei gleichmäßiger Klasseneinteilung optimal:

Normale Klasseneinteilung:



Verschieden große Klassen:



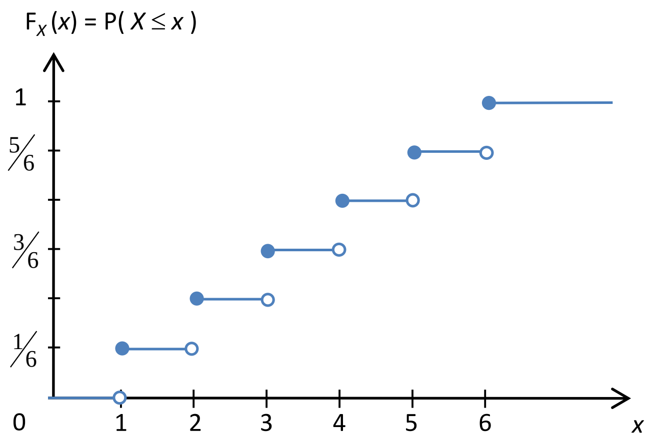
Unterschiedliche Klassenbreite führt bei Stabdiagrammen zu Fehleindrücken. Histogramme sind dann besser.

Umgang mit stetigen und quasistetigen Zufallsvariablen

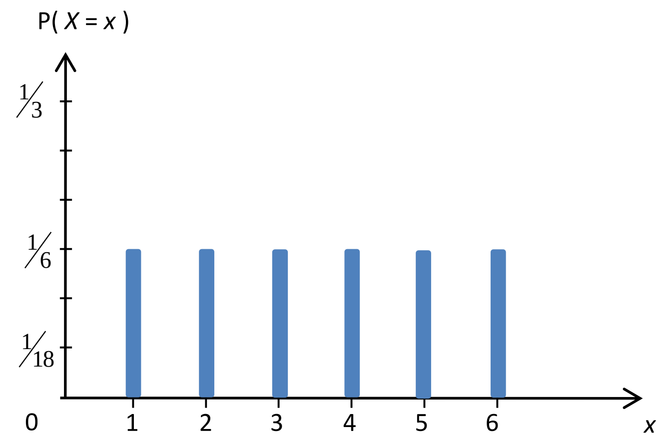
Wiederholung: Kumulierte Verteilungsfunktion

Beispiel: Ein Würfel wird geworfen: X := Augenzahl

Kumulierte Verteilungsfunktion von X



W-Funktion von X (Stabdiagramm)



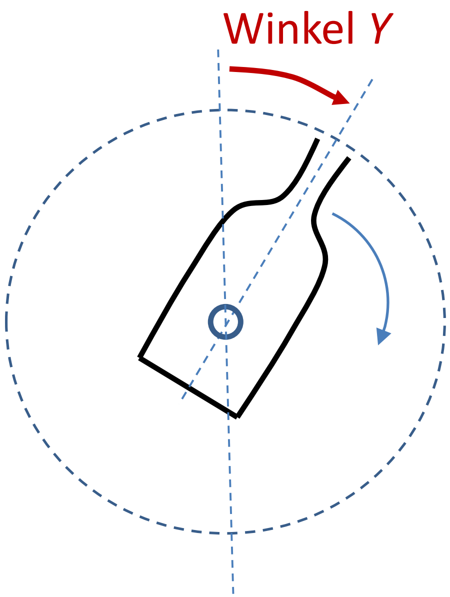
Satz 27.5

Für die kumulierte Verteilungsfunktion $F_X(x) := P(X \leq x)$ einer Zufallsvariablen X gilt:

- $F(x)$ wächst monoton von 0 bis 1.
- An Sprungstellen ist sie rechtsseitig stetig, d.h. der obere Wert ist der Funktionswert
- Für jedes a, b gilt: $\mathbf{P(a < X \leq b) = F_X(b) - F_X(a)}$
- **Die Höhe der Sprungstelle an einer Stelle x_0 ist die Wahrscheinlichkeit, mit der das Ereignis $X = x_0$ eintritt.**

Beispiel: Flaschendrehen

Beispiel: Eine liegende Flasche wird gedreht. Wenn sie liegen bleibt, wird der Winkel Y zwischen Endlage und Ausgangslage als Ergebnis notiert. $0°$ wird dabei als $360°$ notiert.



Variante 1: Y wird auf Vielfache von $120°$ gerundet,
also $Y \in \{ 120°; 240°; 360° \}$

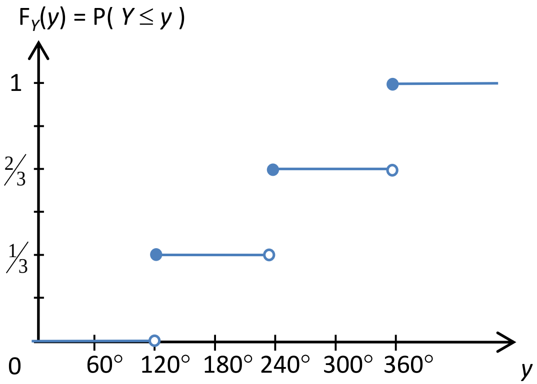
Variante 2: Y wird auf Vielfache von $60°$ gerundet,
also $Y \in \{ 60°; 120°; 180°; 240°; 300°; 360° \}$

Variante 3: Y wird nicht gerundet,
also $Y \in]\, 0; 360 \,]$

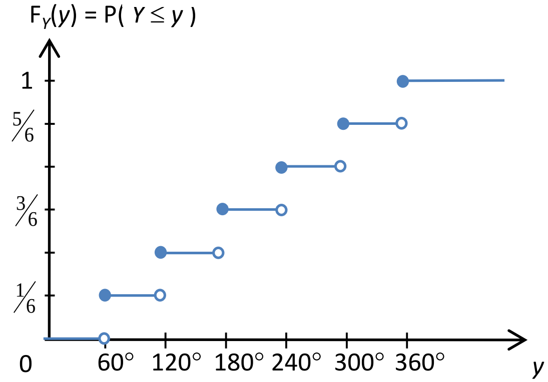
Flaschendrehen

Kumulierte Verteilungsfunktion von *Y*

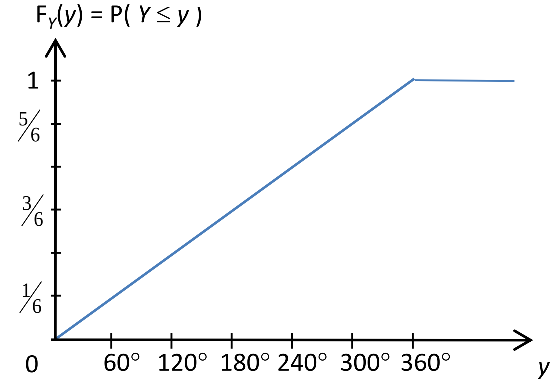
***Y* gerundet auf** 120°



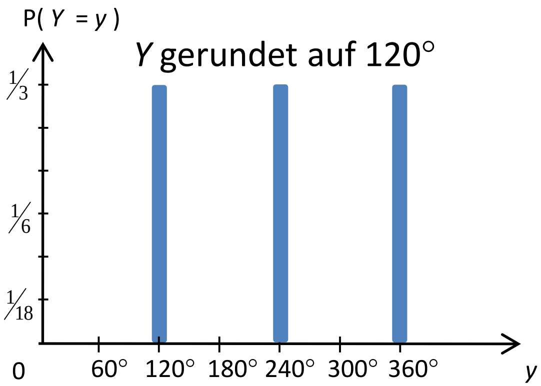
***Y* gerundet auf** 60°

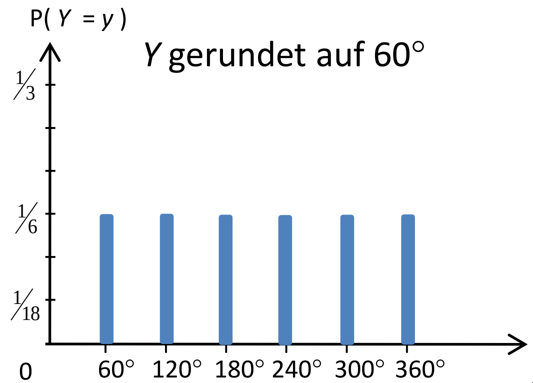


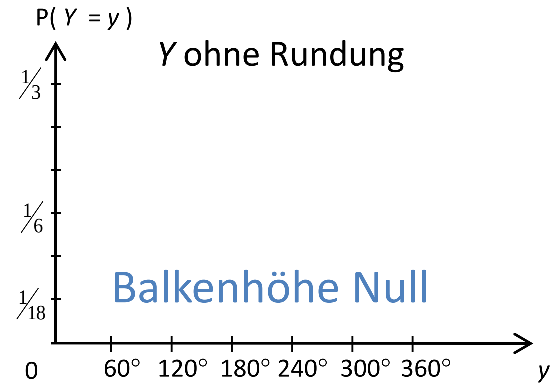
***Y* ohne Rundung**



Wahrscheinlichkeitsfunktion von *Y*







Kum. Verteilungsfunktion glatter, Säulen niedriger, je kleiner die Schrittweite

→***Säulenhöhe*** *geteilt durch* ***Schrittweite*** bleibt konstant

Stetige Zufallsvariable

Definition:

Eine ***Zufallsvariable*** X heißt ***stetig***, wenn ihre kumulierte Verteilungsfunktion stetig ist. Dies impliziert, dass $P(X = x) = 0$ für alle x.

Erinnerung: Eine Zufallsvariable X heißt ***diskret***, wenn sie nur endlich viele oder abzählbar unendlich viele Werte annehmen kann.

Bemerkung:

- Bei diskreten Zufallsvariablen ist in der Regel für jeden möglichen Wert x_i die Wahrscheinlichkeit, dass er exakt angenommen wird, größer als Null.
- Ein unmögliches Ereignis hat Wahrscheinlichkeit Null.
- Ein Ereignis, das Wahrscheinlichkeit Null hat, muss nicht unmöglich sein.

Diskret oder stetig?

- Augenzahl beim Würfeln: *diskret*
- Anzahl Anrufe, die in einem Callcenter in einer Stunde eingehen: *diskret*
- Winkel beim Flaschendrehen (ohne Rundung): *stetig*
- Laufzeit eines zufällig ausgewählten Datenpaketes: *stetig*

wichtig

Eigenschaften stetiger Z-Variablen

Satz: Für jede ***stetige*** Zufallsvariable X gilt für alle Werte x, a, b:

$$P(X = x) = 0,$$

$$P(X < x) = P(X \leq x), \text{ und}$$

$$\mathbf{P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = F_X(b) - F_X(a)}$$

ACHTUNG: Das gilt nur für stetige Zufallsvariablen. Für diskrete Zufallsvariablen, z.B. beim Würfeln, gibt es sehr wohl einen Unterschied: $P(\textit{Augensumme} \leq 3) \neq P(\textit{Augensumme} < 3) = P(\text{Augensumme} \leq 2)$

Bei stetigen Zufallsvariablen macht es keinen Sinn zu fragen, wie wahrscheinlich sie einen bestimmten Wert exakt annehmen. Es macht nur Sinn zu fragen, wie wahrscheinlich Sie in einem bestimmten Intervall liegen.

Motivation von Wahrscheinlichkeits***dichten***

Beispiel:

X = Winkel beim Flaschendrehen

Beobachtung:

Die Wahrscheinlichkeit, dass X in einem (kleinen) Bereich liegt ist proportional zur Breite des Bereiches.

$$P(180° < X \leq 190°) = F_X(190) - F_X(180) = 10/360 \approx 2.8\%$$

$$P(180° < X \leq 185°) = F_X(185) - F_X(180) = 5/360 \approx 1.4\%$$

$$P(180° < X \leq 182.5°) = F_X(182.5) - F_X(180) = 2.5/360 \approx 0.7\%$$

Dies motiviert die folgende Definition der

Dichtefunktion f einer stetigen Zufallsvariablen X:

$$f(x) := \lim_{dx \to 0} \frac{P(x < X \leq x + dx)}{dx} = \lim_{dx \to 0} \frac{F_X(x+dx) - F_X(x)}{dx} = F_X'(x)$$

Dichtefunktion

Definition 27.7 Sei *X* eine stetige Zufallsvariable mit differenzierbarer kumulierter Verteilungsfunktion $F_X(x)$. (Erinnerung: $F_X(x) := P(X \leq x)$)
Die Ableitung

$$f_X(x) := F_X'(x)$$

wird (*Wahrscheinlichkeits-*)***Dichtefunktion*** genannt. Umgekehrt erhält man die kumulierte Verteilungsfunktion durch Integration der Dichtefunktion ab -∞:

$$F_X(x) := \int_{-\infty}^{x} f_X(t)\,dt$$

Satz: Für jede Dichtefunktion f gilt:

- $f(x) \geq 0$ für alle x
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$

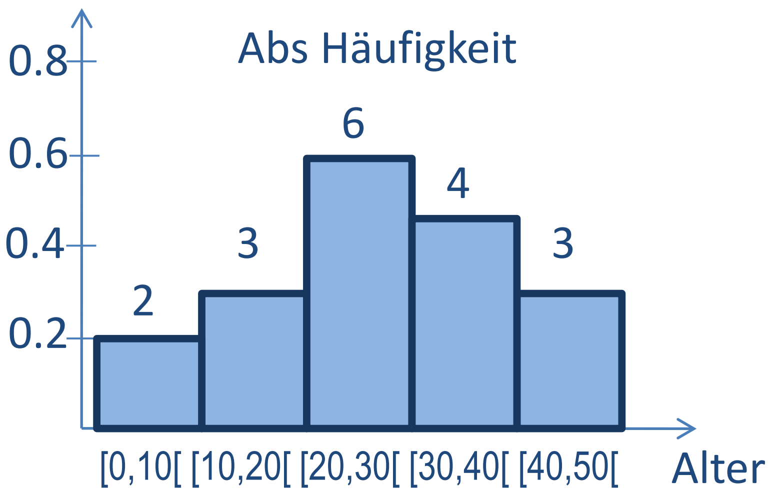
wichtig

Zum Vergleich: Histogramme

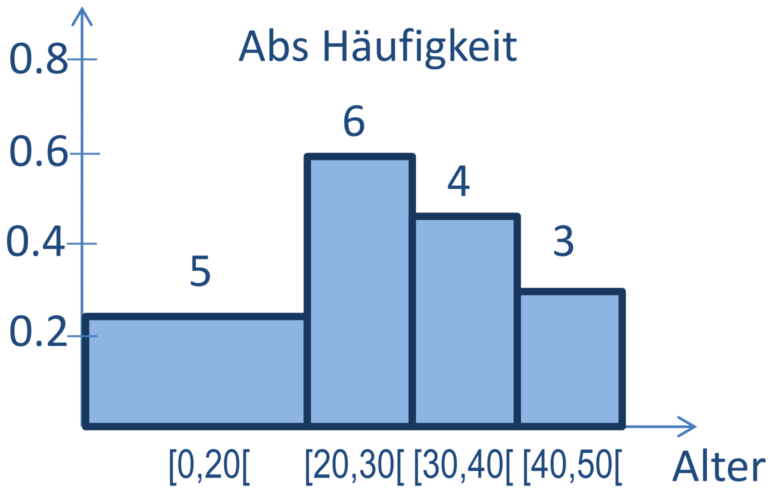
Im Unterschied zum Stabdiagramm gibt bei Histogrammen die Balkenbreite die Klassenbreite wieder, und die Balkenfläche, nicht die Höhe gibt die relative oder absolute Häufigkeit an.

Die Balkenhöhe ist deshalb **Häufigkeit/Balkenbreite.**

Normale Klasseneinteilung:



Verschieden große Klassen:



Eine Dichtefunktion entspricht einem Histogramm mit „unendlich kleiner Klassenbreite“.

Anwenden von Dichte- und Verteilungsfunktion

Satz: Für jede stetige Zufallsvariable X mit Dichtefunktion f und kumulierter Verteilungsfunktion F gilt:

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) \quad = F_X(b) - F_X(a) = \int_a^b f_X(x)\, dx$$

Die Wahrscheinlichkeit, dass X im Intervall [a;b] liegt entspricht also der **Differenz der Funktionswerte der kumulierten Verteilungsfunktion an den Stellen** *b* und *a*

oder der **Fläche** zwischen *a* und *b* **unter der Dichtefunktion** .

Anschauliche Interpretation von Dichten:

Folgerung: Bestimmt man das Verhältnis der Werte der Dichtefunktion einer Zufallsvariablen an verschiedenen Stellen *x* und *y*, so entspricht es dem Verhältnis der ***Wahrscheinlichkeit***en, dass die Zufallsvariable einen ***Wert nahe x*** bzw. nahe *y* annimmt.
(Unter „nahe“ sollen dabei genügend kleine Intervalle gleicher Größe um x bzw. y verstanden werden)

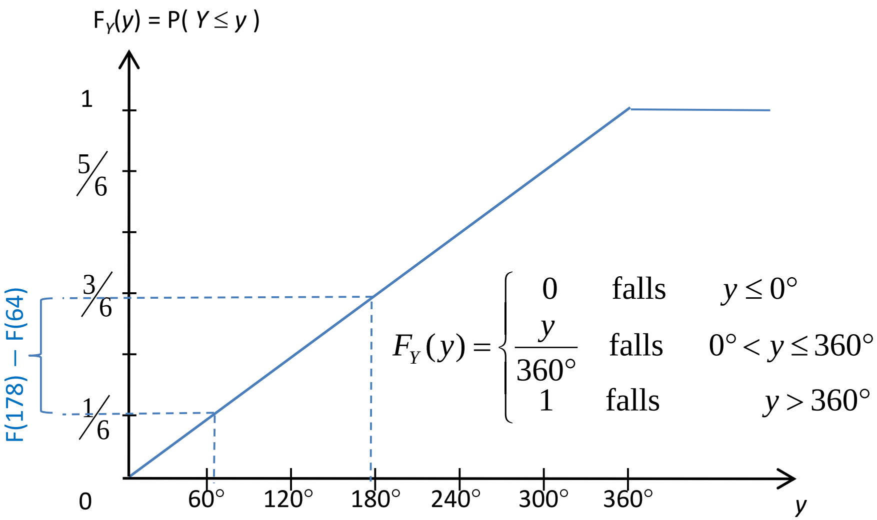
Ist f_X die Dichtefunktion der Zufallsvariablen X und $f_X(x) = 2 \cdot f_X(y)$, so sind also Werte nahe an x doppelt so wahrscheinlich, wie Werte nahe an y.

wichtig

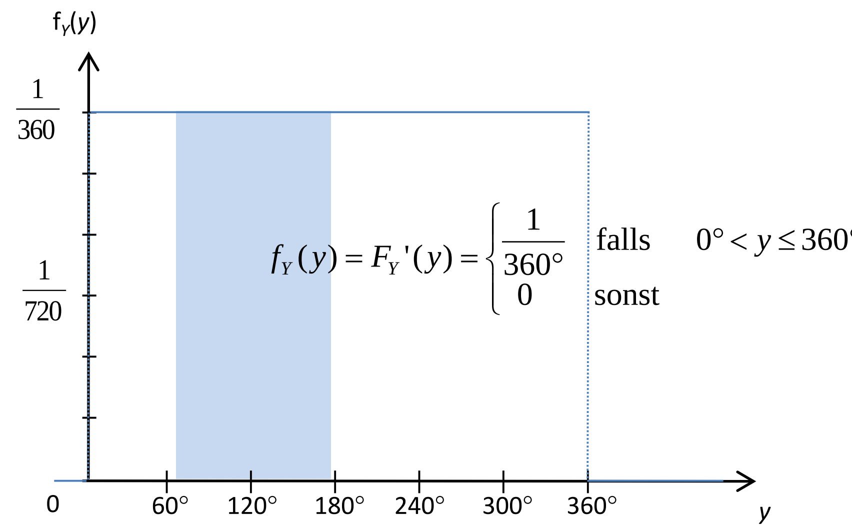
Dichtefunktion

Fortsetzung des Flaschendrehbeispiels *ohne* Rundung: *Y* ist stetige Zufallsvariable

a) Bestimmen Sie kumulierte Verteilungsfunktion und Dichtefunktion



$$F_Y(y) = \begin{cases} 0 & \text{falls} \quad y \le 0° \\ \dfrac{y}{360°} & \text{falls} \quad 0° < y \le 360° \\ 1 & \text{falls} \quad y > 360° \end{cases}$$



$$f_Y(y) = F_Y'(y) = \begin{cases} \dfrac{1}{360°} & \text{falls} \quad 0° < y \le 360° \\ 0 & \text{sonst} \end{cases}$$

b) Berechnen Sie P(64° < Y ≤ 178°):

$$P(64 < Y \le 178) = F_Y(178) - F_Y(64) = \frac{178-64}{360} \approx 31.7\%$$

oder

$$P(64 < Y \le 178) = \int_{-\infty}^{178} f_Y(t)\,dt - \int_{-\infty}^{64} f_Y(t)\,dt = \int_{64}^{178} f_Y(t)\,dt = F(178) - F(64) = \frac{178-64}{360} \approx 31.7\%$$

Stetige Gleichverteilung

Definition (**stetige Gleichverteilung**)
Man nennt eine stetige Zufallsvariable *X* *gleichverteilt auf dem Intervall [a; b]* wenn alle möglichen Realisationen gleiche Wahrscheinlichkeitsdichte haben. Für die Dichtefunktion *f* und die kumulierte Verteilungsfunktion *F* gilt dann:

$$f_X(x_0) = \begin{cases} \frac{1}{b-a} & \text{für} \quad x_0 \in [a;b] \\ 0 & \text{für} \quad x_0 \notin [a;b] \end{cases} \qquad F_X(x_0) = \begin{cases} 0 & \text{für} \quad x_0 < a \\ \frac{x_0-a}{b-a} & \text{für} \quad x_0 \in [a;b] \\ 1 & \text{für} \quad x_0 > a \end{cases}$$

Beispiel:
Der Winkel beim Flaschendrehen war gleichverteilt auf]0; 360].

Satz (Erwartungswert und Std-Abw. der Gleichverteilung)

Für eine *auf dem Intervall [a; b] gleichverteilte stetige Zufallsvariable X* gilt:

$$E(X) = \frac{a+b}{2}, \qquad \sigma(X) = \frac{1}{6}\sqrt{3}\,(b-a)$$

wichtig

Gleichverteilung einer stetigen Zufallsvariable

Beispiel 27.11 Gleichverteilung
Angenommen, eine Straßenbahn fährt pünktlich alle 10 Minuten. Wenn man zufällig zur Haltestelle kommt, dann ist die Wartezeit X eine Zufallsvariable, die kontinuierlich alle Werte von 0 bis 10 annehmen kann, wobei jede Wartezeit gleich wahrscheinlich ist (wenn wir einfachheitshalber das Gesetz von Murphy vernachlässigen;-). Die zugehörige Wahrscheinlichkeitsdichte ist daher

$$f(x) = \begin{cases} k, & \text{für } 0 < x < 10 \\ 0, & \text{sonst} \end{cases}$$

wobei k eine Konstante ist.

a) Bestimmen Sie k.

b) Geben Sie die Verteilungsfunktion F an.

c) Wie groß ist die Wahrscheinlichkeit, höchstens 3 Minuten zu warten?

d) Wie groß ist die Wahrscheinlichkeit, mindestens 2 Minuten zu warten?

e) Wie groß ist die Wahrscheinlichkeit, zwischen 5 und 9 Minuten zu warten?

Lösung zu 27.11

a) Die Konstante muss so gewählt werden, dass die Gesamtfläche unter dem Graphen von f gleich 1 ist, also

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^{10} k\,dx = 10k.$$

Daher muss $k = 0.1$ sein.

b) Für $x < 0$ ist auch $F(x) = 0$. Für x zwischen 0 und 10 gilt $F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x 0.1\,dt = 0.1 \cdot x$. Für $x > 10$ ist $F(x) = \int_{-\infty}^{x} f(t)dt = \int_0^{10} f(t)dt + \int_{10}^{x} f(t)dt = 1 + 0 = 1$. Insgesamt:

$$F(x) = \begin{cases} 0\;, & \text{für } x \leq 0, \\ 0.1x, & \text{für } 0 < x < 10 \\ 1\;, & \text{für } x \geq 10 \end{cases}$$

c) Wir beantworten diese Frage, indem wir in die Verteilungsfunktion $F(x) = 0.1 \cdot x$ einsetzen: $P(X \leq 3) = F(3) = 0.1 \cdot 3 = 30\%$. Die Wahrscheinlichkeit, höchstens 3 Minuten zu warten, ist in Abbildung 27.6 dargestellt.

d) Wieder drücken wir die gesuchte Wahrscheinlichkeit mithilfe der Verteilungsfunktion aus: $P(X \geq 2) = 1 - P(X < 2) = 1 - F(2) = 1 - 0.1 \cdot 2 = 80\%$.

e) $P(5 < X < 9) = F(9) - F(5) = 0.1 \cdot 9 - 0.1 \cdot 5 = 40\%$. ■

Lösung zu 27.11 (Forts.)

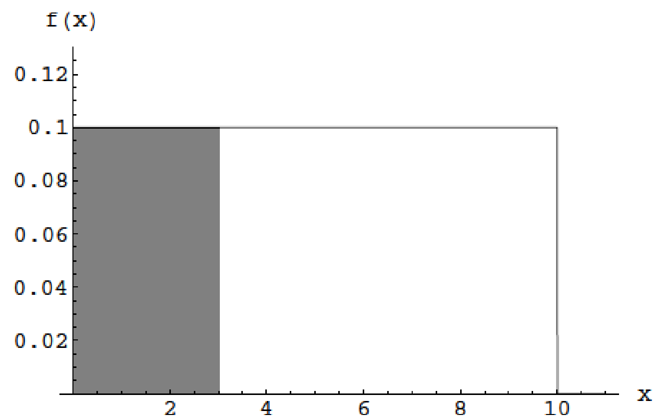


Abbildung 27.6. Schattierte Fläche = Wahrscheinlichkeit, höchstens 3 Minuten zu warten.

Zusammenfassung zu stetigen Zufallsvariablen

Diskrete Zufallsvariable:

Die möglichen Werte sind aufzählbar.

Wahrscheinlichkeitsfunktion: $x \mapsto P(X = x)$

Kumulierte Verteilungsfunktion $F(x)$: $x \mapsto P(X \leq x)$

Es gilt: $P(a < X \leq b) = F(b) - F(a) = \sum_{x:\, a<x\leq b} P(X = x)$

Stetige Zufallsvariable:

Zusammenhängender, kontinuierlicher Wertebereich

Kumulierte **Verteilungsfunktion** $F(x)$: $x \mapsto P(X \leq x)$

Dichtefunktion $f(x)$: $x \mapsto F'(x)$

Es gilt: $P(a < X \leq b) = P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = \mathbf{F(b) - F(a)} = \int_a^b f(x)\, dx$

Jeder einzelne Wert hat Wahrscheinlichkeit Null,
also gilt für jedes x: $P(X=x) = 0$ und $P(X < x) = P(X \leq x)$

Kennzahlen stetiger Verteilungen

Definition

Sei X eine **stetige Zufallsvariable** mit Verteilungsfunktion F(x) und Dichtefunktion f(x).

Überträgt man die Kennzahldefinitionen aus dem diskreten Fall, ändert sich nichts Wesentliches, außer dass der Summen jeweils Integrale und statt Wahrscheinlichkeiten Wahrscheinlichkeitsdichten stehen:

Erwartungswert: $$\mu = E(X) := \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$ (Def. 27.19)

Varianz: $$\sigma^2 = V(X) := \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x)\, dx$$ (Def. 27.30)

Covarianz: $$\sigma_{X,Y} = Cov(X,Y) := \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)\cdot(y-\mu_Y)\cdot f(x,y)\, dx\, dy$$ (Def. 27.37)

Die Rechenregeln für Erwartungswert und Varianz gelten unverändert weiter.

Die Quantilsbestimmung vereinfacht sich sogar gegenüber dem diskreten Fall:
Das ***p*-Quantil** von X ist derjenige Wert q_p, für den gilt:

$$F(q_p) = p \quad \textit{also} \quad P(X \leq q_p) = p\,.$$

Im stetigen gibt es anders als im diskreten Fall immer eine eindeutige Lösung q_p, weil F eine stetige Funktion ist.

Kennzahlen stetiger Verteilungen (2)

Beispiel (Flaschendrehen)

Sei X der Winkel beim Flaschendreh-Experiment. Dann gilt:

f(x) = 1/360° für 0° < x < 360° (0 außerhalb des Bereichs)

F(x) = x/360° für 0° < x < 360° (0 für $x \leq 0$; 1 für $x \geq 360$)

Der Erwartungswert von X ist:

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx = \int_{0}^{360} x \cdot \tfrac{1}{360} \cdot dx = \left[\frac{x^2}{2 \cdot 360} \right]_0^{360} = \frac{360^2}{2 \cdot 360} - 0 = 180$$

Die Varianz von X ist:

$$V(X) = \int_{0}^{360} (x-180)^2 \cdot \tfrac{1}{360} \cdot dx = \left[\frac{(x-180)^3}{3 \cdot 360} \right]_0^{360} = \frac{180^3}{3 \cdot 360} - \frac{-180^3}{3 \cdot 360} = 5400$$

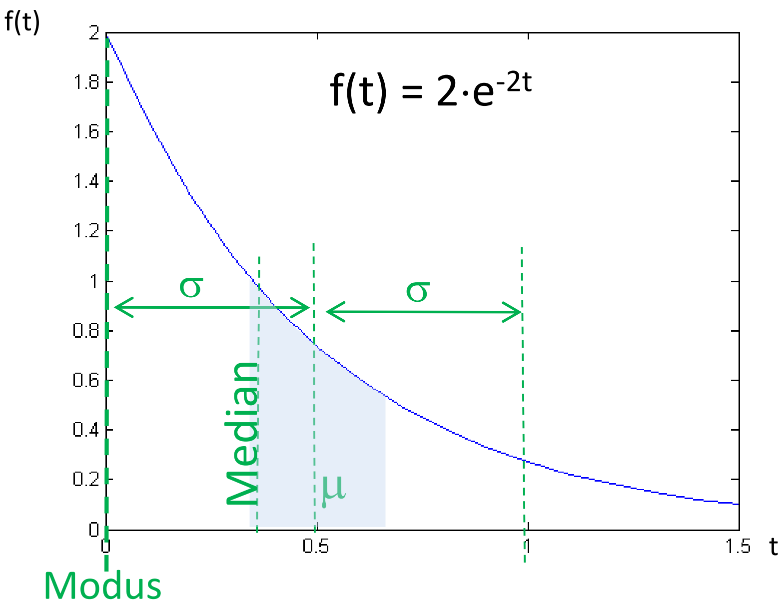
Die Standardabweichung ist:

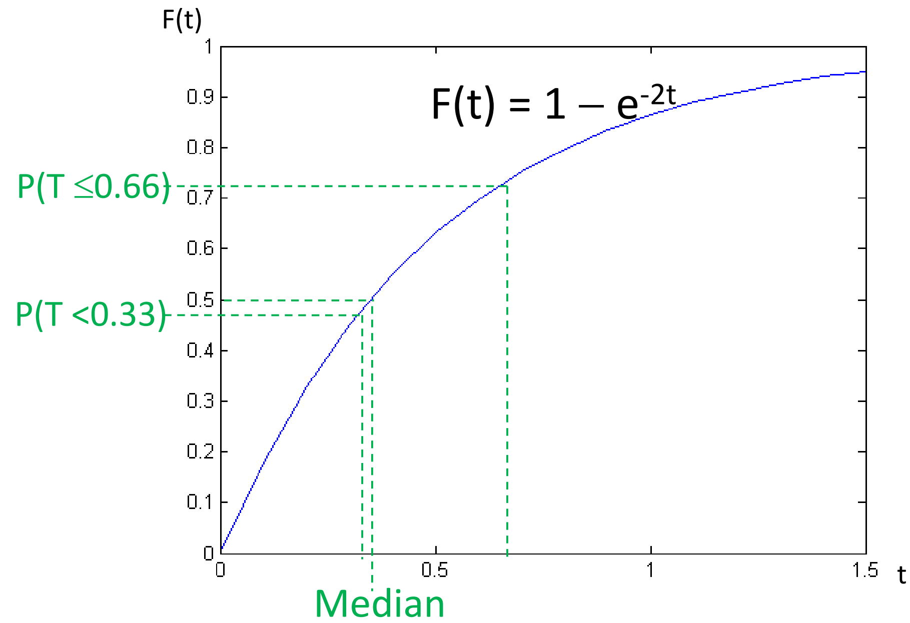
$$\sigma(X) = \sqrt{V(X)} = \sqrt{5400} \approx 73.5$$

Das 10% - Quantil:

$$F(x_{0.1}) = x_{0.1} / 360° \overset{!}{=} 10\,\% \quad \Leftrightarrow x_{0.1} = 360° \cdot 0.1 = \underline{\mathbf{36°}}$$

Beispiel: Der Zeitabstand *T* in Minuten zwischen aufeinanderfolgenden Telefonanrufen in einem Callcenter besitze folgende Dichte bzw. kumulierte Verteilungsfunktion:





Berechnen Sie P(1/3 min < *T* ≤ 2/3 min).

$P(1/3 \text{ min} < T \leq 2/3 \text{ min}) = F(2/3 \text{ min}) - F(1/3 \text{ min}) = e^{-2/3} - e^{-4/3} \approx 0.25$

Wie kann man P(1/3 min< *T*≤ 2/3 min) aus dem Graphen von *f* ablesen, wie aus dem von *F*?

Fläche unter *f* für 1/3<t ≤ 2/3. Oder: *F*(2/3) minus *F*(1/3) ablesen.

Bestimmen Sie den Median graphisch und exakt.

$F(m) = 1 - e^{-2m} = 0.5 \Rightarrow m = -\ln(0.5)/2 \approx 0.35$

Was Sie gelernt haben sollten

- Histogramme
- Unterschied zwischen diskreten und stetigen Zufallsvariablen
- Interpretation von Wahrscheinlichkeitsdichten
- Beziehung zwischen kumulierter Verteilungsfunktion einerseits, und Dichtefunktion bzw. Wahrscheinlichkeitsfunktion andererseits.
- Aus jeder dieser Funktionen $P(X<a)$, $P(X>a)$, $P(X\leq a)$, $P(X\geq a)$, $P(a\leq X\leq b)$ bestimmen.
- Aus angegebenen Informationen unbekannte Parameter eine W-Verteilung bestimmen.