Kapitel 3

Zufallsvariable

Lernziele

- Diskrete und stetige Zufallsvariable
- Wahrscheinlichkeitsfunktion, Dichte und Verteilungsfunktion
- Erwartungswert, Varianz und Standardabweichung
- Binomial- und Poissonverteilung
- Normal- und Exponentialverteilung

Problem

Sie möchten einen Multiple-Choice-Test mit 20 Fragen bestehen. Für jede Frage gibt es 5 Antwortmöglichkeiten, wobei immer nur (genau) eine richtig ist.

Die erste Frage können Sie nicht beantworten und müssen raten.
Wie groß ist die Wahrscheinlichkeit, dass Sie richtig raten?

Wenn Sie keine einzige Frage beantworten können und daher bei jeder Antwort raten, wie groß ist dann die Wahrscheinlichkeit, dass Sie den Test bestehen?
Mindestens 10 Antworten müssen richtig sein.

Zufallsvariable

Eine **Zufallsvariable** ist eine Abbildung, die jedem Ergebnis eines Zufallsexperiments eine reelle Zahl zuordnet.

- Experiment: Frage nach Anzahl PKW im Haushalt
  Die ZV ordnet jedem Haushalt die Anzahl PKW zu.
  Werte: $0, 1, 2, \ldots$ – **diskrete** ZV
- Experiment: Gewichtsbestimmung von Äpfeln
  Gewicht eines Apfels(g): $123, 245, 301, \ldots$ – **stetige** ZV

Zufallsvariable werden mit Großbuchstaben bezeichnet, z.B. X, mögliche Realisationen mit Kleinbuchstaben, hier x.

Diskrete Zufallsvariable

- Nur ganze Zahlen sind als Ergebnisse möglich.
  Z.B.: 0, 1, 2,. . . .
- Tritt als Ergebnis von Zählexperimenten auf.
- Hat meistens nur eine endliche Anzahl an Werten.
- Wird durch die **Wahrscheinlichkeitsfunktion** oder **Verteilungsfunktion** beschrieben.

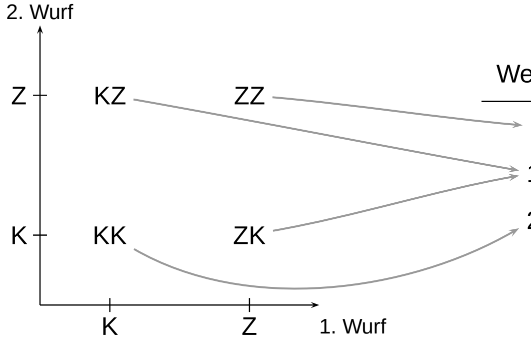
Wahrscheinlichkeitsfunktion

- Liste **aller** Paare $(x, f(x))$
  x . . . Wert der Zufallsvariablen X
  $f(x)$ . . . Wahrscheinlichkeit, dass Wert x eintritt.
  $$f(x) = \mathrm{P}(X = x)$$
- Jeder x Wert kommt nur einmal vor. (Funktion!)
- $0 \le f(x) \le 1$
- $\sum_{x \in S} f(x) = 1$

Wahrscheinlichkeitsfunktion / Beispiel

Experiment: Werfe zwei Münzen; Zähle Anzahl „Kopf“.

Zufallsvariable heißt: „Anzahl Köpfe“



Wert x	$f(x)$
0	$\frac{1}{4} = 0.25$
1	$\frac{2}{4} = 0.50$
2	$\frac{1}{4} = 0.25$

Wahrscheinlichkeitsfunktion / Darstellung

- **Liste**

  $\{(0, 0.25), (1, 0.50), (2, 0.25)\}$

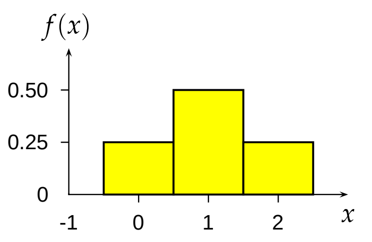
- **Tabelle**

x	Anzahl	$f(x)$
0	1	$\frac{1}{4} = 0.25$
1	2	$\frac{2}{4} = 0.50$
2	1	$\frac{1}{4} = 0.25$

- **Formel**

  $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$

- **Grafik**



Verteilungsfunktion

Gibt die Wahrscheinlichkeit an, dass die diskrete Zufallsvariable nicht größer als ein vorgegebener Wert ist:

$$F(x) = \mathrm{P}(X \leq x) = \sum_{y \leq x} f(y)$$

Die Verteilungsfunktion ist monoton in x.

Erwartungswert

Der **Erwartungswert** (*expectation*) einer diskreten Zufallsvariable ist die gewichtete Summe der möglichen Realisationen.

$$\mathrm{E}(X) = \mu_x = \sum_x x f(x) = \sum_x x\,\mathrm{P}(X = x)$$

Achtung!
Der Erwartungswert muss nicht immer existieren.

Erwartungswert / Beispiel

Angenommen sie arbeiten für eine Versicherung, und verkaufen Lebensversicherungen mit einer Vertragssumme von 10 000 €.
Die Jahresprämie dafür beträgt 290 €.

Sterbetafeln geben für einen Kunden in Abhängigkeit von Alter, Geschlecht, Gesundheitszustand, etc. eine Wahrscheinlichkeit von 0.001 an, dass er in diesem Jahr verstirbt.

Was ist der erwartete jährliche Gewinn für Polizzen dieser Art?

Die ZV Nettogewinn, Einzahlung minus Auszahlung, bezeichnen wir mit X. Wahrscheinlichkeitsfunktion:

Gewinn, x	Ereignis	Wahrscheinlichkeit
290 €	Kunde lebt	0.999
-9710 €	Kunde stirbt	0.001

Erwartungswert / Beispiel – Fortsetzung

Wenn die Versicherung 1000 Polizzen davon verkauft, so wird – *vereinfacht* gesprochen – in 999 Fällen von den 1000 der Kunde das Jahresende erleben, in einem von den 1000 Fällen wird er innerhalb dieses Jahres sterben.

290 € wird die Versicherung in 999 Fällen ohne monetäre Gegenleistung erhalten, also einen Nettoertrag von 290 €. In einem Fall von 1000 hat sie einen Nettoabgang von 9710 € ($= 290 - 10000$).

Der durchschnittliche Gewinn ist daher 280 €:

$$\frac{999}{1000}\,290 + \frac{1}{1000}\,(-9710) = 0.999 \cdot 290 + 0.001 \cdot (-9710) = 280$$

Erwartungswert / Beispiel – Fortsetzung

Die Rechenschritte, die in der Formel

$$\mathrm{E}(X) = \sum_x x\,\mathrm{P}(X = x) = \mu_x$$

angegeben sind, kann man in die Tabelle der Wahrscheinlichkeitsverteilung leicht integrieren.

x	$\mathrm{P}(X = x)$	$x\,\mathrm{P}(X = x)$	
290	0.999	$290 \cdot 0.999 =$	289.71
−9710	0.001	$-9710 \cdot 0.001 =$	−9.71
Summe	1.000		280.00
			$= \mathrm{E}(X)$

Erwartung einer Funktion einer ZV

Der Erwartungswert einer Funktion g einer diskreten Zufallsvariable X, $g(X)$, ist

$$\boxed{\mathrm{E}[g(X)] = \sum_x g(x) f(x) = \sum_x g(x)\,\mathrm{P}(X = x)}$$

Angenommen Sie erhalten 10% Gewinnbeteiligung obiger Versicherung abzüglich 10 € Bearbeitungsgebühr. Ihr erwarteter Gewinn beträgt dann

$$\begin{aligned}&\mathrm{E}(G) = \mathrm{E}[g(X)] = \mathrm{E}[0.1\,X - 10] = \\ &= (0.1 \cdot 290 - 10) \cdot 0.999 + (-0.1 \cdot 9710 - 10) \cdot 0.001 = 18\end{aligned}$$

Erwartungswert / Rechenregeln

Sei $Y = a + b\,X$, dann gilt allgemein für den Erwartungswert von Y

$$\boxed{\mathrm{E}(Y) = \mathrm{E}(a + b\,X) = a + b\,\mathrm{E}(X)}$$

Die Erwartung einer Linearkombination von ZVen ist gleich die Linearkombination der Erwartungswerte der einzelnen ZVen.

Bemerkung: Der Erwartungswert von X muss existieren.

Achtung!
Ist f nicht linear, kann Erwartung und Funktion nicht vertauscht werden. Im allgemeinen gilt

$$\mathrm{E}[f(X)] \neq f(\mathrm{E}[X])$$

Varianz

Die **Varianz** einer ZV X, $\mathrm{V}(X) = \sigma_x^2$, ist definiert als die erwartete quadratische Abweichung vom Erwartungswert (mittlere quadratische Abweichung vom Mittel).

$$\mathrm{V}(X) = \mathrm{E}([X - \mathrm{E}(X)]^2) = \mathrm{E}([X - \mu]^2) = \sigma_x^2$$

Die **Standardabweichung** ist ein Maß für die Streuung:

$$\sqrt{\mathrm{V}(X)} = \sqrt{\sigma_x^2} = \sigma_x$$

Die Standardabweichung wird oft als Maß für die Unsicherheit über den Ausgang eines Experiments verwendet.

Varianz / Verschiebungssatz

Der **Verschiebungssatz** erleichtert die Berechnung der Varianz. Er lautet mit $\mathrm{E}(X) = \mu$:

$$\mathrm{V}(X) = \mathrm{E}([X - \mu]^2) = \mathrm{E}(X^2) - \mu^2$$

bzw.

$$\mathrm{V}(X) = \sum_x ([x - \mu]^2)\, \mathrm{P}(X = x) = \sum_x x^2\, \mathrm{P}(X = x) - \mu^2$$

$$\mathrm{V}(X) = \sum_x ([x - \mu]^2) f(x) = \sum_x x^2 f(x) - \mu^2$$

Varianz / Rechenregeln

Sei $Y = a + b\,X$, dann gilt allgemein für die Varianz von Y

$$\mathrm{V}(Y) = \mathrm{V}(a + b\,X) = b^2\,\mathrm{V}(X)$$

Die Varianz ist verschiebungsinvariant. Sie hängt nicht von der Konstanten a ab.

- Die Varianz von Y steigt linear mit b^2,
- die Standardabweichung von Y steigt linear mit b.

Bemerkung: Erwartungswert und Varianz von X müssen existieren.

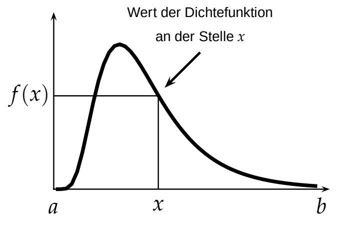
Stetige Zufallsvariable

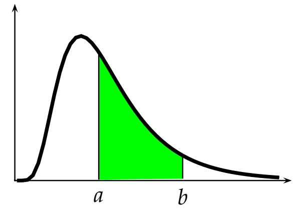
- Ganze Zahlen (gerundet), Bruchzahlen oder reelle Zahlen.
- Erhält man bei Messungen.
- Unendlich viele mögliche Werte in einem Intervall.
  Zu viele, um sie wie bei diskreten Zufallsvariablen aufschreiben zu können.
- Wird durch die **Dichtefunktion** oder **Verteilungsfunktion** beschrieben.

Dichtefunktion

- Ein Graph, der alle x die „Häufigkeit“ $f(x)$ zeigt.
  $f(x)$ ist keine Wahrscheinlichkeit.
- Eigenschaften:
  - $\int_{-\infty}^{\infty} f(x)\,dx = 1.$
    Die Fläche unter der Kurve ist 1.
  - $f(x) \geq 0.$
- Wahrscheinlichkeiten können nur für Intervalle angegeben werden:

$$\mathrm{P}(a < X \leq b) = \int_a^b f(x)\,dx$$





Verteilungsfunktion

Gibt die Wahrscheinlichkeit an, dass die Zufallsvariable nicht größer als ein vorgegebener Wert ist:

$$F(x) = \mathrm{P}(X \leq x) = \int_{-\infty}^{x} f(y)\,dy$$

Die Verteilungsfunktion ist monoton in x.

Mit Hilfe der Verteilungsfunktions lassen sich Wahrscheinlichkeiten ausrechnen:

$$\mathrm{P}(a < X \leq b) = F(b) - F(a)$$

Erwartungswert und Varianz

- **Erwartungswert**

$$\mathrm{E}(X) = \mu_x = \int_{-\infty}^{\infty} x f(x)\, dy$$

- **Varianz**

$$\mathrm{V}(X) = \sigma_x = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)\, dy = \int_{-\infty}^{\infty} x^2 f(x)\, dy - \mu_x^2$$

- **Standardabweichung**

$$\sigma_x = \sqrt{\mathrm{V}(X)}$$

Spezielle Verteilungen

Viele Merkmale folgen der gleichen (oder ähnlichen) Verteilungen. Diese sind daher besonders wichtig und gut untersucht.

- Jede Verteilung beschreibt ein zugrundeliegendes Phänomen (**Modell**).
- Für Varianten derselben Fragestellung sind nur die Parameter der Verteilung anzupassen.
- Für bestimmte Fragestellungen können die Verteilungen – die mathematische Formeln – explizit angegeben werden.
- Wahrscheinlichkeiten, Erwartungswert, Varianz, ..., können dann exakt berechnet werden.

Binomialverteilung

Beschreibt die Zahl der „Erfolge“ in einer Stichprobe von Umfang n

- Zufallsvariable heißt: Zahl der „Erfolge“ unter n (unabhängigen) Beobachtungen (Versuchen).

Beispiele:

- Anzahl „Kopf“ bei zehnmaligem Werfen einer Münze.
- Anzahl der richtigen Antworten bei MC-Test mit 20 Fragen.
- Anzahl defekter Teile bei Qualitätskontrolle in einer Kiste mit 50 Stück.
- Anzahl der erfolgreichen Verkaufsgespräche bei 100 geführten.

Binomialverteilung / Modellvoraussetzungen

- Folge von n **identischen**, **unabhängigen** Versuchen.
- Jeder Versuch hat genau zwei mögliche Ausgänge: **„Erfolg"** oder **„Mißerfolg"**.
- Die Wahrscheinlichkeit für „Erfolg", bzw. „Mißerfolg", ist für alle Versuche gleich.
- Zwei Stichprobenauswahlverfahren dazu:
  - **Unendliche** Grundgesamtheit **ohne** Zurücklegen, oder
  - **Endliche** Grundgesamtheit **mit** Zurücklegen

Binomialverteilung / Wahrscheinlichkeitsfunktion

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}$$

$f(x)$ ... Wahrscheinlichkeit für x „Erfolge", $\mathrm{P}(X = x) = f(x)$

n ... Anzahl der Wiederholungen (Stichprobengröße)

p ... Wahrscheinlichkeit für „Erfolg"

x ... Anzahl an „Erfolgen" ($x = 0, 1, 2, \dots, n$)

$\binom{n}{x}$ ... Binomialkoeffizient. Man sagt: „n über x".

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!}$$

Binomialverteilung / Erwartungswert

- **Erwartungswert**

$$\mu = \mathrm{E}(X) = n\,p$$

- **Standardabweichung**

$$\sigma = \sqrt{n\,p\,(1-p)}$$

Binomialverteilung / Beispiel

Experiment: Münze wird 5 mal geworfen. Beobachte Anzahl „Kopf“.

Wie groß ist die Wahrscheinlichkeit für 3 mal „Kopf“?

Antwort: $n = 5$, $p = 0.5$, $x = 3$, $\mathrm{P}(X = 3) = f(3)$,

$$\begin{aligned} f(x) &= \frac{n!}{x!\,(n-x)!}\, p^x\,(1-p)^{n-x} \\ f(3) &= \frac{5!}{3!\,(5-3)!}\, 0.5^3\,(1-0.5)^{5-3} \\ &= 10 \cdot 0.5^3 \cdot 0.5^2 \\ &= 0.3125 \end{aligned}$$

Binomialverteilung / Beispiel

Bei einem Multiple-Choice-Test gibt es 20 Fragen mit je 5 Antwortmöglichkeiten, wobei immer nur (genau) eine richtig ist.

Wenn Sie keine einzige Frage beantworten können und daher bei jeder Antwort raten, wie groß ist dann die Wahrscheinlichkeit, dass Sie den Test bestehen? Mindestens 10 Antworten müssen richtig sein.

Wie lauten Erwartungswert und Standardabweichung für die Anzahl an richtigen Antworten?

Binomialverteilung / Lösung

Die Zufallsvariable, X, heißt: „Anzahl der Erfolge unter 20“. Prüfung bestanden heißt: 10, 11, 12, . . . , 19, oder 20 Antworten sind richtig. Die Anzahl der richtigen Antworten ist binomial verteilt mit $n = 20$ und $p = 0.2$. Daher

$$\begin{aligned} \mathrm{P}(\text{bestanden}) &= \mathrm{P}(X \geq 10) = f(10) + f(11) + \cdots + f(20) \\ &\approx 2.03 \cdot 10^{-3} + 4.62 \cdot 10^{-4} + \cdots + 1.05 \cdot 10^{-14} \\ &\approx 0.0026 \end{aligned}$$

Erwartungswert und Standardabweichung

$$\begin{aligned} \mu &= n\,p = 20 \cdot 0.2 = 4 \\ \sigma &= \sqrt{n\,p\,(1-p)} = \sqrt{20 \cdot 0.2 \cdot (1-0.2)} = 1.79 \end{aligned}$$

Poissonverteilung

Beschreibt die Anzahl von Ereignissen innerhalb eines Intervalls.

- Zufallsvariable heißt: Anzahl von Ereignissen je Einheit (Zeit, Länge, Fläche, Raum, ...).

Beispiele:

- Anzahl von Kunden, die innerhalb von 20 Minuten eintreffen.
- Anzahl von Streiks pro Jahr.
- Anzahl an Schlaglöchern pro Straßenkilometer.

Poissonverteilung / Modellvoraussetzungen

- Die Wahrscheinlichkeit für das Eintreten eines Ereignisses ist **konstant**.
  - Angenommen pro Stunde kommen im Durchschnitt 120 Kunden. So bedeutet dies, dass in jeder der 60 Minuten im Durchschnitt 2 Kunden eintreffen.
- Das Eintreten der einzelnen Ereignisse ist **unabhängig**.
  - Das Eintreffen eines Kunden beeinflusst nicht das Eintreffen eines anderen Kunden.
  - Das Auftreten von Schlaglöchern in einem Strassen-abschnitt beeinflusst nicht andere Strassenabschnitte.
- Ein Ereignis pro Einheit.
  - Es können nicht 2 Kunden zur selben Zeit eintreffen.

Poissonverteilung / Wahrscheinlichkeitsfunktion

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$f(x)$ ... Wahrscheinlichkeit für x „Erfolge“, $f(x) = \mathrm{P}(X = x)$

λ ... Erwartete (durchschnittliche) Anzahl von „Erfolgen“

e ... Eulersche Zahl ($2.71828\ldots$)

x ... Anzahl an „Erfolgen“ pro Einheit ($x = 0, 1, 2, \ldots, \infty$)

$x!$ ... x-Faktorielle. $x! = x \cdot (x-1) \cdot (x-2) \cdot \ldots \cdot 2 \cdot 1$,

Poissonverteilung / Erwartungswert

- **Erwartungswert**

$$\mu = \mathsf{E}(X) = \lambda$$

- **Standardabweichung**

$$\sigma = \sqrt{\lambda}$$

Poissonverteilung / Beispiel

In einer Bankfiliale treffen pro Stunde durchschnittlich 22 Kunden pro Stunde ein.

Wie groß ist die Wahrscheinlichkeit, dass innerhalb von 15 Minuten genau 4 Kunden eintreffen?

Wie groß sind Erwartungswert und Standardabweichung für die Anzahl der Kunden in einem 15-Minuten-Intervall?

Poissonverteilung / Lösung

Anzahl der Kunden pro 15-Minuten-Intervall ist poisson verteilt mit

$$\lambda = \tfrac{22}{4} = 5.5.$$

(22 Kunden pro Stunde = 5.5 Kunden pro 15 Minuten).

Wahrscheinlichkeit, dass genau 4 Kunden eintreffen:

$$\begin{aligned} f(x) &= \mathrm{P}(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \\ f(4) &= \mathrm{P}(X = 4) = \frac{5.5^4 e^{-5.5}}{4!} = 0.156 \end{aligned}$$

Erwartungswert und Standardabweichung

$$\begin{aligned} \mu &= \lambda = 5.5 \\ \sigma &= \sqrt{\lambda} = 2.345 \end{aligned}$$

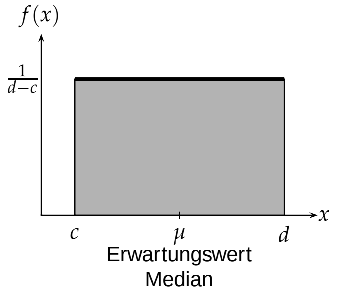
Stetige Gleichverteilung

- Gleichwahrscheinliche Realisationen (Ergebnisse).
- **Dichtefunktion**, $c < x < d$:

$$f(x) = \frac{1}{d-c}$$

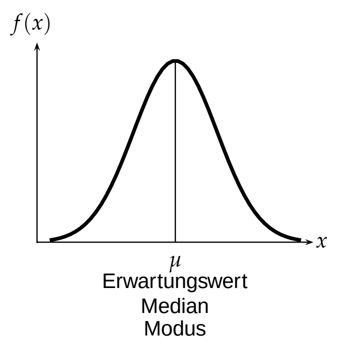
- **Erwartungswert** und **Standardabweichung**

$$\mu = \frac{c+d}{2}, \qquad \sigma = \frac{d-c}{\sqrt{12}}$$



Normalverteilung

- Einfache mathematische Eigenschaften.
- Beschreibt gut viele stochastische (zufällige) Prozesse und stetige Phänomene.
- Approximiert gut andere (schwierig zu berechnende) Verteilungen.
- Ist Basis für klassische statistische Inferenz (statistisches Schließen).



Normalverteilung / Dichtefunktion

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$f(x)$ … Wert der Dichte der Zufallsvariablen X an der Stelle x

μ … Erwartungswert, Mittel der Grundgesamtheit

σ … Standardabweichung in der Grundgesamtheit

π … $3.14159\ldots$

e … Eulersche Zahl ($2.71828\ldots$), $\exp(x) = e^x$

x … Wert der Zufallsvariablen X, $(-\infty < x < \infty)$

Die Normalverteilung hat 2 Parameter: μ und σ.

Approximation der Binomialverteilung

Die Binomialverteilung mit Parametern n (Wiederholungen) und p (Wahrscheinlichkeit für „Erfolg") kann durch die Normalverteilung approximiert werden, falls gilt (Fausregel!)

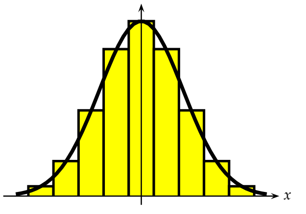
$$\boxed{n\,p\,(1-p) \geq 9}$$

- Erwartungswert

  $$\mu_p = n\,p$$

- Standardabweichung

  $$\sigma_p = \sqrt{\frac{p\,(1-p)}{n}}$$



Zentraler Grenzwertsatz

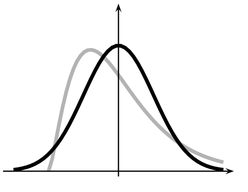
Seien X_i ($i = 1, \ldots, n$) unabhängige stetige Zufallsvariable mit Mittelwert μ und Standardabweichung σ.
Wir betrachten die Zufallsvariable (arithmetisches Mittel)

$$\bar{X} = \frac{1}{n} \sum X_i$$

Bei hinreichend großem n (Stichprobengröße) ist $\bar{X}$ annäherend normalverteilt mit

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

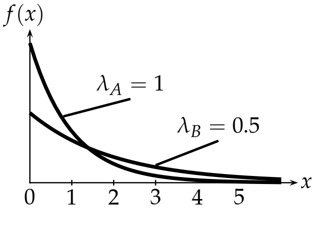
Faustregel: $n \gtrsim 30$.



Exponentialverteilung

- Parameter: λ
- **Dichtefunktion**, $x \geq 0$:

  $$\boxed{f(x) = \lambda\, \exp(-\lambda x)}$$



- **Verteilungsfunktion**:

  $$\boxed{F(x) = P(X \leq x) = 1 - \exp(-\lambda x)}$$

- **Erwartungswert** und **Standardabweichung**:

  $$\boxed{\mu = \frac{1}{\lambda} \qquad \sigma = \frac{1}{\lambda}}$$

Exponentialverteilung

Beschreibt Zeit oder Distanz zwischen Ereignissen.

- Wird in Warteschlangemodellen verwendet.