Fachhochschule Südwestfalen
University of Applied Sciences

Statistik

CH.9 - Spezielle Verteilungen

SS 2022 || Prof. Dr. Buchwitz, Sommer, Henke

Wir geben Impulse

Lernziele

- Ziel 1
- Ziel 2
- Ziel 3

Spezielle Verteilungen

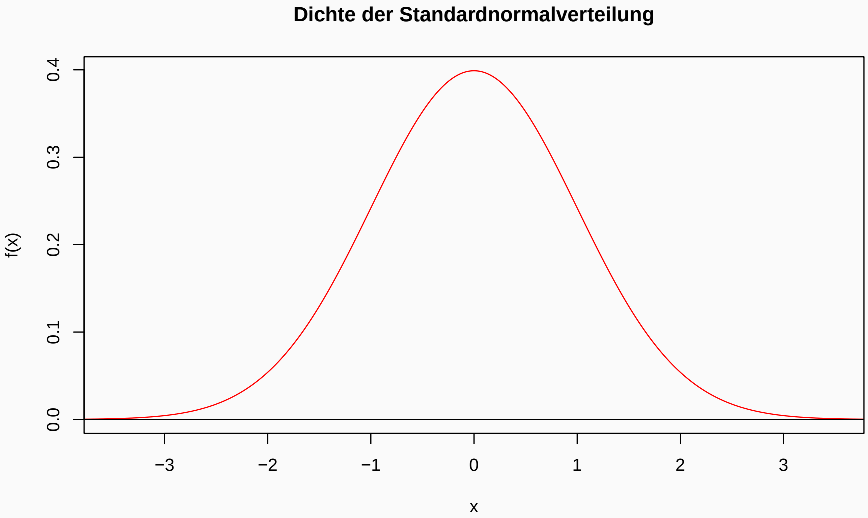
- Im Folgenden werden **stetige** und **diskrete** Verteilungen diskutiert, die vielfach in der Statistik angewandt werden.

Inhaltsübersicht

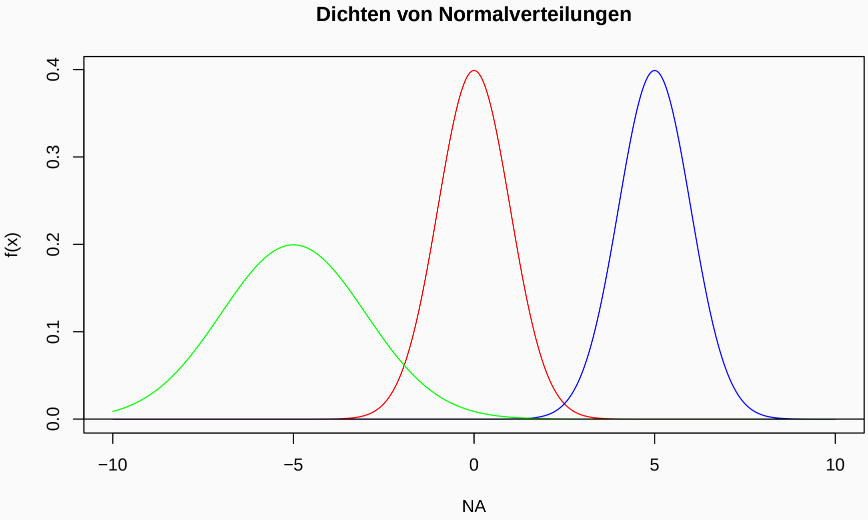
Normalverteilung

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

Normalverteilung



Normalverteilung



Normalverteilung

Definition: Normalverteilung

Eine stetige Zufallsvariable X heißt **normalverteilt**, wenn ihre Dichtefunktion mit den beiden Parameter Erwartungswert μ und Standardabweichung σ gegeben ist durch

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

- Ist die am meisten verwendete stetige Verteilung.
- Kann in der Praxis in vielen Situationen (näherungsweise) beobachtet werden.
- Stellt oft eine hinreichend gute Approximation für andere Verteilungen dar.
- **R-Funktion:** `dnorm()`, `pnorm()`, `qnorm()` und `rnorm()`

Normalverteilung

- Parameter der Normalverteilung sind Erwartungswert μ und Standardabweichung σ.
- Die Kurzschreibweise für eine normalverteilte Zufallvariale X mit Erwartungswert μ und Varianz σ^2 lautet: $X \sim N(\mu = 0, \sigma^2 = 1)$.
- Der Graph der Dichtefunktion der Normalverteilung ist **symmetrisch** an der Stelle μ und besitzt an dieser Stelle ein Maximum.
- Der Wert an der Stelle μ ist **umso größer** (die Dichtefunktion also umso höher und steiler), **je kleiner** der Wert der Standardabweichung ist.

Normalverteilung

- Die Wahrscheinlichkeit, dass der Wert einer normalverteilten Zufallsvariablen um den Betrag *einer Standardabweichung* σ vom Erwartungswert μ abweicht, beträgt etwa 68%.

$$P(\mu - \sigma \leq x \leq \mu + \sigma) \approx 0.6827$$

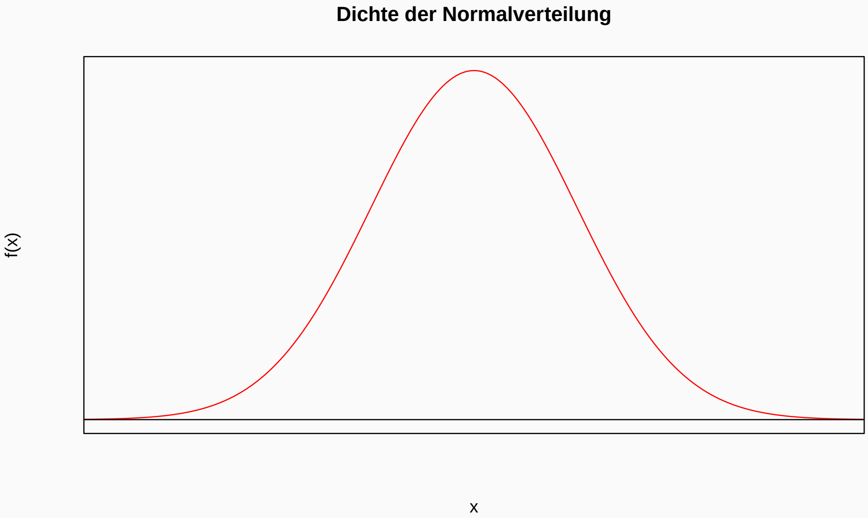
- Die Wahrscheinlichkeit, dass der Wert einer normalverteilten Zufallsvariablen um den Betrag von *zwei Standardabweichungen* σ vom Erwartungswert μ abweicht, beträgt etwa 95%.

$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) \approx 0.9545$$

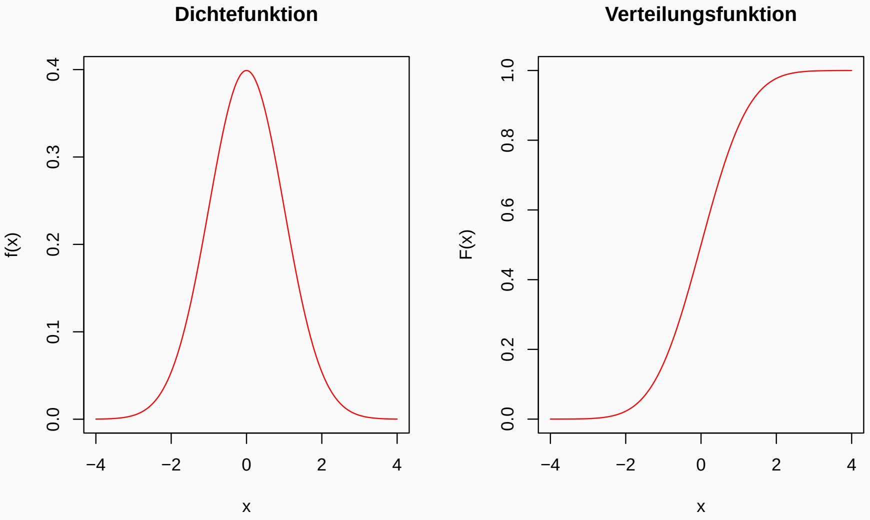
- Die Wahrscheinlichkeit, dass der Wert einer normalverteilten Zufallsvariablen um den Betrag von *drei Standardabweichungen* σ vom Erwartungswert μ abweicht, beträgt über 99%.

$$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) \approx 0.9973$$

Normalverteilung



Standardnormalverteilung

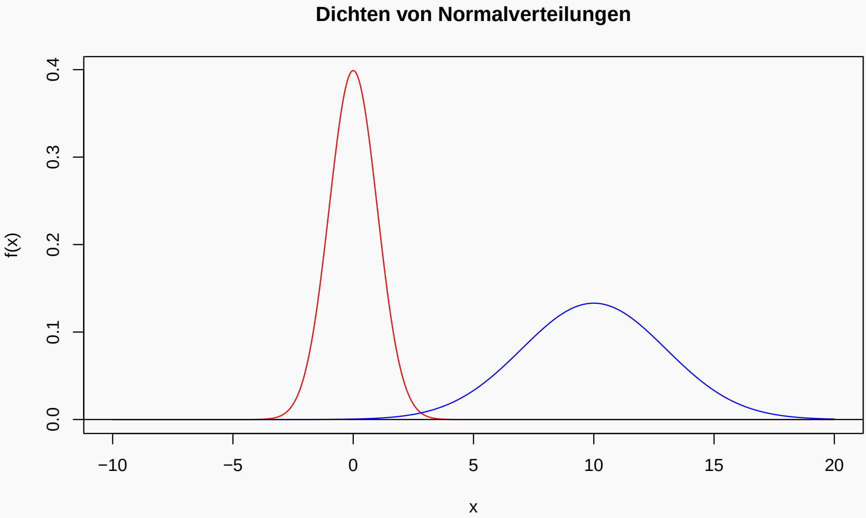


Standardnormalverteilung

- Als **Standardnormalverteilung** wird die Normalverteilung mit den Parametern $\mu = 0$ und $\sigma = 1$ bezeichnet.
- Durch Transformation lassen sich normalverteilte Daten in **standardnormalverteilte Daten** übertragen.
- Die Transformation zur Standardisierung von normalverteilten zu standardnormalverteilten Zufallsvariablen heißt **z-Transformation**.

$$z = \frac{x - \mu}{\sigma}$$

Standardnormalverteilung



Standardnormalverteilung

Die Verteilungsfunktion der Normalverteilung wird per Definition durch Integration der Dichtefunktion zwischen $-\infty$ und x bestimmt:

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- Das erforderliche Integral ist nicht einfach lösbar und die Variantenvielfalt möglicher Normalverteilungen sehr groß. Tabelliert wird daher üblicherweise **nur die Standardnormalverteilung**.
- Die Verteilungsfunktion der Standardnormalverteilung wird üblicherweise zur einfachen Unterscheidung mit $\Phi(z)$ bezeichnet.

Standardnormalverteilung

Ablesen aus der Tabelle:

- Die Zeilen geben den z-Wert bis zur **ersten Nachkommastelle** an.
- Die Spalten ergänzen die **zweite Nachkommastelle** des z-Wertes.
- Innerhalb der Tabelle befindet sich der Flächeninhalt unter der Dichtefunktion von $-\infty$ bis zur Stelle z.

$$\Phi(z) = F(z) = P(Z \leq z)$$

- In den Tabellen sind aufgrund der Symmetrie nur Werte für positive z angegeben. Es gilt $\Phi(-z) = F(-z) = 1 - F(z)$.

Die Tabellen befinden sich im Studienbuch auf S. 352 ff

Verteilungsfunktion $F(z)$ für $z \sim N(\mu = 0, \sigma^2 = 1)$

Verteilungsfunktion $F(z)$ der Standardnormalverteilung $N(0, 1)$

Beispiel: $F(z) = P(z \leq 1.96) = 0.9750$

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981

Normalverteilung (R-Funktionen)

```r
# Dichtefunktion (Density)
dnorm(0, mean=0, sd=1)
```

```
## [1] 0.3989
```

```r
# Verteilungsfunktion (Probabilitydistribution)
pnorm(q=0, mean=0, sd=1)
```

```
## [1] 0.5
```

```r
# Quantilsfunktion
qnorm(p=0.5, mean=0, sd=1)
```

```
## [1] 0
```

```r
# Zufallszahlengenerator (Random Number Generator)
set.seed(100)
rnorm(n=1, mean=0, sd=1)
```

```
## [1] -0.5022
```

Beispiel: Normalverteilung

Die Gleichspannung von Batterien, die zum Betrieb eines Sensors verwendet werden, lässt sich als normalverteilte Zufallsgröße mit $\mu = 12V$ und $\sigma = 0.6V$ modellieren. Für den Betrieb des Sensors sind mehr als 11.3V notwendig.

- Wie groß ist die Wahrscheinlichkeit, dass eine zufällig ausgewählte Batterie eine Spannung von mehr als 12V liefert?
- Wie groß ist die Wahrscheinlichkeit, dass ein Sensor mit einer zufällig ausgewählten Batterie nicht betrieben werden kann?

Normalverteilung

Inhaltsübersicht

Binomialverteilung (diskret)

Experiment: n-fache stochastisch unabhängige Wiederholung eines Bernoulli-Experiments mit identischer Erfolgswahrscheinlichkeit p und den Zufallsvariablen X_i mit $i = 1, \ldots, n$ und $X_i \in 0, 1$.

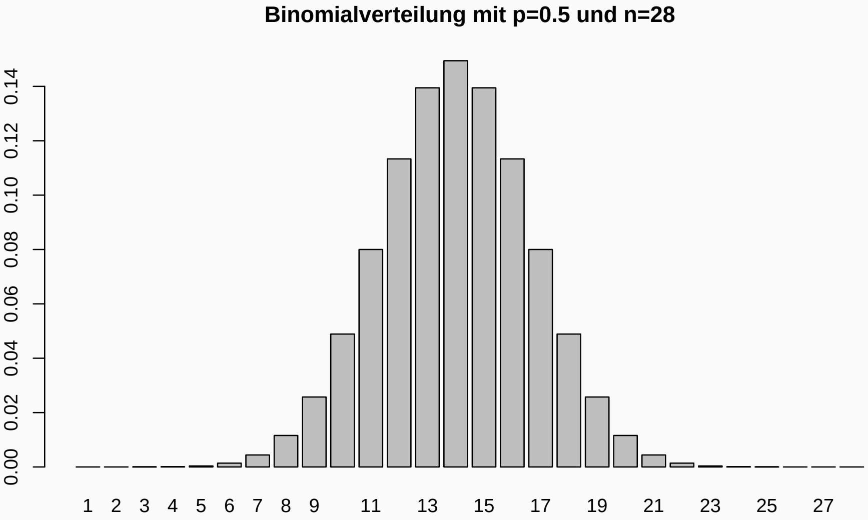
- **Wahrscheinlichkeitsfunktion:** $f(x) = \binom{n}{k} p^x (1-p)^{n-x}$
- **Erwartungswert:** $\mu = n \cdot p$
- **Varianz:** $\sigma^2 = n \cdot p \cdot (1-p)$

Anwendungsfall

Qualitätssicherung bei großen Produktionsmengen: n Proben werden zur Prüfung ausgewählt, p ist die Wahrscheinlichkeit für einen Defekt.

- **R-Funktion:** `dbinom()`, `pbinom()`, `qbinom()` und `rbinom()`

Binomialverteilung (diskret)



Inhaltsübersicht

Poissonverteilung (diskret)

Experiment: Modellierung seltener Ereignisse, die (theoretisch) in unbegrenzter Anzahl auftreten können.

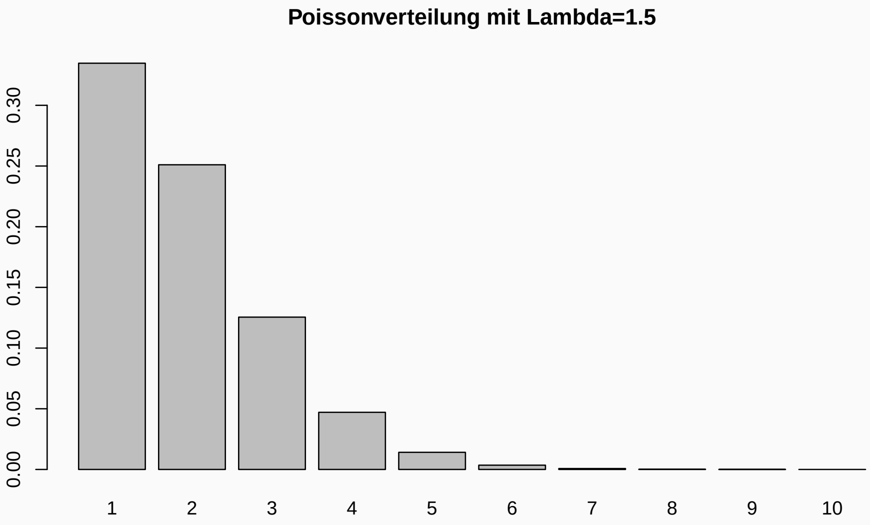
- **Wahrscheinlichkeitsfunktion:** $f(x) = \frac{\lambda^x}{x!} e^{-\lambda}$
- **Erwartungswert:** $\mu = \lambda$
- **Varianz:** $\sigma^2 = \lambda$

Anwendungsfall

z.B. Anzahl Großbrände in einem Bezirk in einem Monat; Anzahl Personen, die sich in einem bestimmten Zeitintervall in eine Warteschlange stellen.

- **R-Funktion:** `dpois()`, `ppois()`, `qpois()` und `rpois()`

Poissonverteilung (diskret)



Inhaltsübersicht

Hypergeometrische Verteilung (diskret)

Experiment: Eine Urne enthält N Kugeln, von denen $N \cdot p$ weiß und $N \cdot (1-p)$ schwarz sind. Aus der Urne werden n Kugeln ohne Zurücklegen gezogen. Die Zufallsvariable X gibt die Anzahl der weißen Kugeln an.

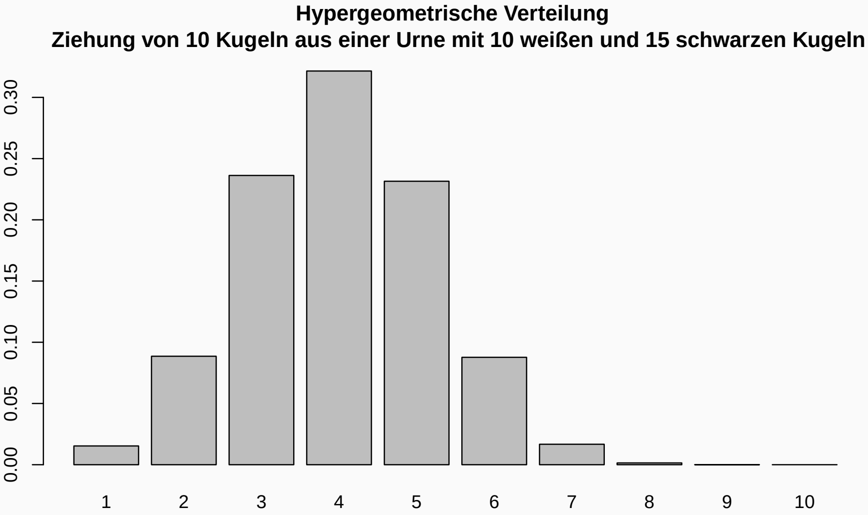
- **Wahrscheinlichkeitsfunktion:** $f(x) = \frac{\binom{N \cdot p}{x} \cdot \binom{N \cdot (1-p)}{n-x}}{\binom{N}{n}}$
- **Erwartungswert:** $\mu = n \cdot p$
- **Varianz:** $\sigma^2 = n \cdot p \cdot (1-p) \cdot \frac{N-n}{N-1}$

Anwendungsfall

Ziehen ohne Zurücklegen, z.B. in der Qualitätskontrolle bei kleinen Stückzahlen (n klein, N nicht riesig).

- **R-Funktion:** `dhyper()`, `phyper()`, `qhyper()` und `rhyper()`

Hypergeometrische Verteilung (diskret)



Inhaltsübersicht

Exponentialverteilung (stetig)

Experiment: Modellierung der Zeit zwischen dem Auftreten zwei aufeinander folgender poisson-verteilter Ereignisse.

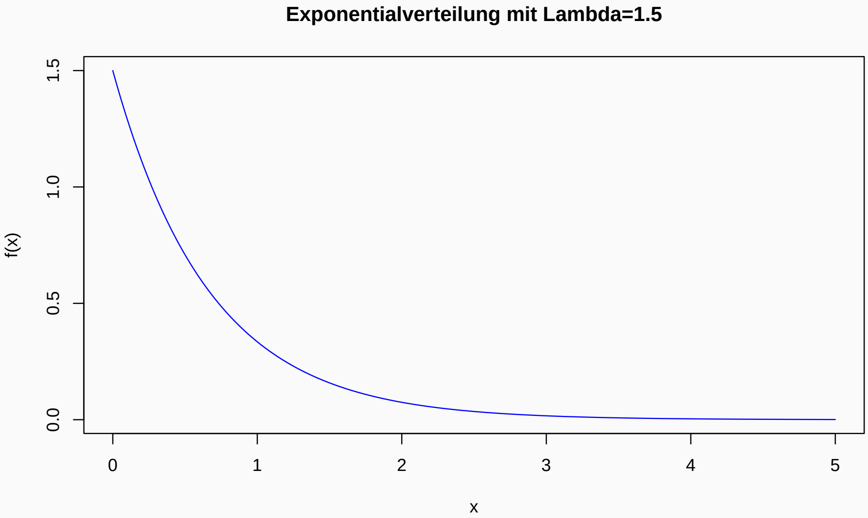
- **Dichtefunktion:** $f(x) = \begin{cases} 0, & \text{für } x < 0 \\ \lambda \cdot e^{-\lambda x}, & \text{für } x \geq 0 \end{cases}$
- **Erwartungswert:** $\mu = \frac{1}{\lambda}$
- **Varianz:** $\sigma^2 = \frac{1}{\lambda^2}$

Anwendungsfall

Verteilung der Ankunftszeiten in Warteschlangen.

- **R-Funktion:** `dexp()`, `pexp()`, `qexp()` und `rexp()`

Exponentialverteilung (stetig)



Approximation

Weitere Verteilungen

- Im Rahmen der Inferenzstatistik werden wir noch weitere Verteilungen kennenlernen, die zur Berechnung von Teststatistiken und Konfidenzintervallen benötigt werden.
- Die **wichtigsten Verteilungen** sind im Studienbuch tabelliert.
- Weitere Verteilungen sind:
  - χ^2-Verteilung
  - t-Verteilung
  - F-Verteilung

Verständnisfragen

1. Wieso ist es nützlich, die Standardnormalverteilung zu kennen?
2. Wie unterscheiden sich die Dichten $N(0, 5)$, $N(0, 0.5)$ und $N(5, 1)$ von der Gestalt der Standardnormalverteilung?
3. Wie kann eine normalverteilte Zufallsvariable in eine standardnormalverteilte Zufallsvariable überführt werden?