

# Kontinuierliche Zufallsvariablen

Statistik (Biol./Pharm.) – Herbst 2012



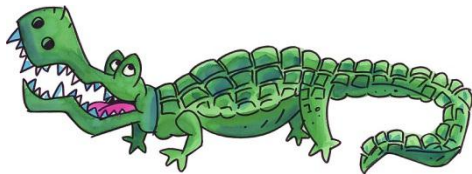
# Wa.verteilung bei kontinuierlichen Werten

- ZV  $X_0$  uniform auf  $W_0 = \{0,1, \dots, 9\} \rightarrow P(X_0 = x) = \frac{1}{10}$
- ZV  $X_1$  uniform auf  $W_1 = \{0.0,0.1, \dots, 9.9\} \rightarrow P(X_1 = x) = \frac{1}{100}$
- ZV  $X_2$  uniform auf  $W_2 = \{0.00,0.01, \dots, 9.99\} \rightarrow P(X_2 = x) = \frac{1}{1000}$
- ■ ■
- ZV  $X_i$  uniform auf  $W_i \rightarrow P(X_i = x) = \frac{1}{10^{i+1}}$
- ZV  $X_\infty$  uniform auf  $W_\infty = [0,10] \rightarrow P(X_\infty = x) = 0$

**Wa. ist nutzlos  
bei kontinuierlichen Zufallsvariablen !**

# Verteilungs-Zoo: Kontinuierliche Zufallsvariablen

Uniform



Exponential



Zugferd  
der  
Statistik



Normal



# Uniforme Verteilung

- **Situation:** Jeder Wert im Intervall  $[a,b]$  ist gleich wa.
- ZV  $X$ : Ein Wert aus  $[a,b]$

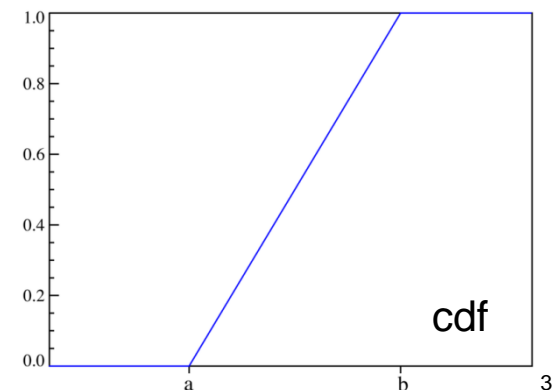
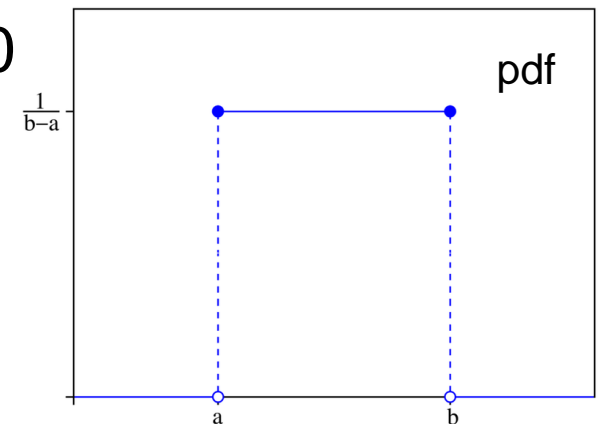
- $X \sim Unif(a, b)$

“ $X$  ist uniform verteilt auf dem Intervall  $[a,b]$ ”

- pdf:  $f(x) = \frac{1}{b-a}$  falls  $a \leq x \leq b$ , sonst 0

- cdf: 
$$F(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x < b \\ 1 & \text{for } x \geq b \end{cases}$$

- $E(X) = \frac{b+a}{2}$  ;  $Var(X) = \frac{(b-a)^2}{12}$



# Beispiel: Haltestelle

- In Zürich fahren die Trams alle 7 Minuten. Angenommen, Sie kommen zu einer zufälligen Zeit an eine Haltestelle, an der ein Tram fährt. Wie wahrscheinlich ist es, dass Sie höchstens eine Minute warten müssen?
- $X$ : Wartezeit in Minuten

$$X \sim Unif(0,7)$$

- $P(X \leq 1) = F(1) = \frac{1-0}{7-0} = \frac{1}{7}$



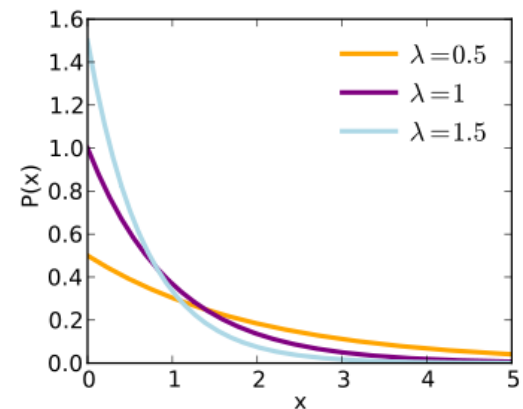
# Exponentialverteilung

- **Situation:** Wartezeit “ohne Gedächtnis”
- ZV  $X$ : Ein Wert aus  $[0, \infty[$
- $X \sim \text{Exp}(\lambda)$   
‘ $X$  ist exponentialverteilt mit Parameter  $\lambda$ ’

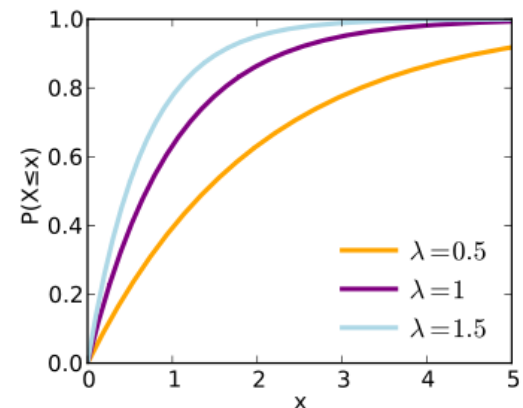
- pdf: 
$$f(x) = \begin{cases} \lambda \exp(-\lambda x), & \text{falls } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

- cdf: 
$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{falls } x \geq 0 \\ 0 & \text{falls } x < 0 \end{cases}$$

- $E(X) = \frac{1}{\lambda}, \text{Var}(X) = \frac{1}{\lambda^2}$



pdf



cdf

# Exponentialverteilung: Kein Gedächtnis

- $P(T > t) = 1 - P(T \leq t) = e^{-\lambda t}$
- $$P(T > t + s | T > s) = \frac{P(T > t+s \text{ und } T > s)}{P(T > s)} = \frac{P(T > t+s)}{P(T > s)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = P(T > t)$$

“Es spielt keine Rolle, ob man schon s Sekunden gewartet hat”
- Gut für: Radioaktiver Zerfall, manche Ionenkanäle
- Schlecht für: Lebenszeit bei Menschen, Wartezeit im Supermarkt



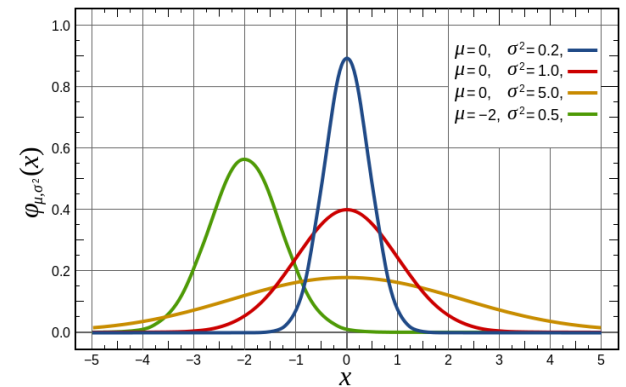
# Normalverteilung

- **Situation:** Beliebige kontinuierliche Werte; meist um einen Wert konzentriert; starke Ausreisser selten
- ZV  $X$ : Ein Wert aus  $] -\infty; \infty[$
- $X \sim N(\mu, \sigma^2)$   
' $X$  ist normalverteilt mit Erwartungswert  $\mu$  und Varianz  $\sigma^2$ '

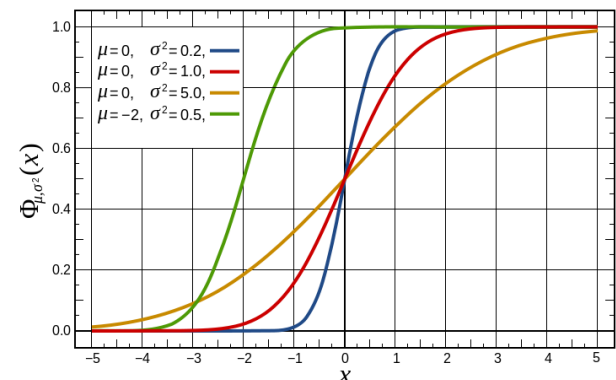
- pdf: 
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

- cdf: ???  
(Standardisieren und Tabelle oder numerisch integrieren)

- $E(X) = \mu, \text{Var}(X) = \sigma^2$
- **Summe von N's ist wieder N**



pdf



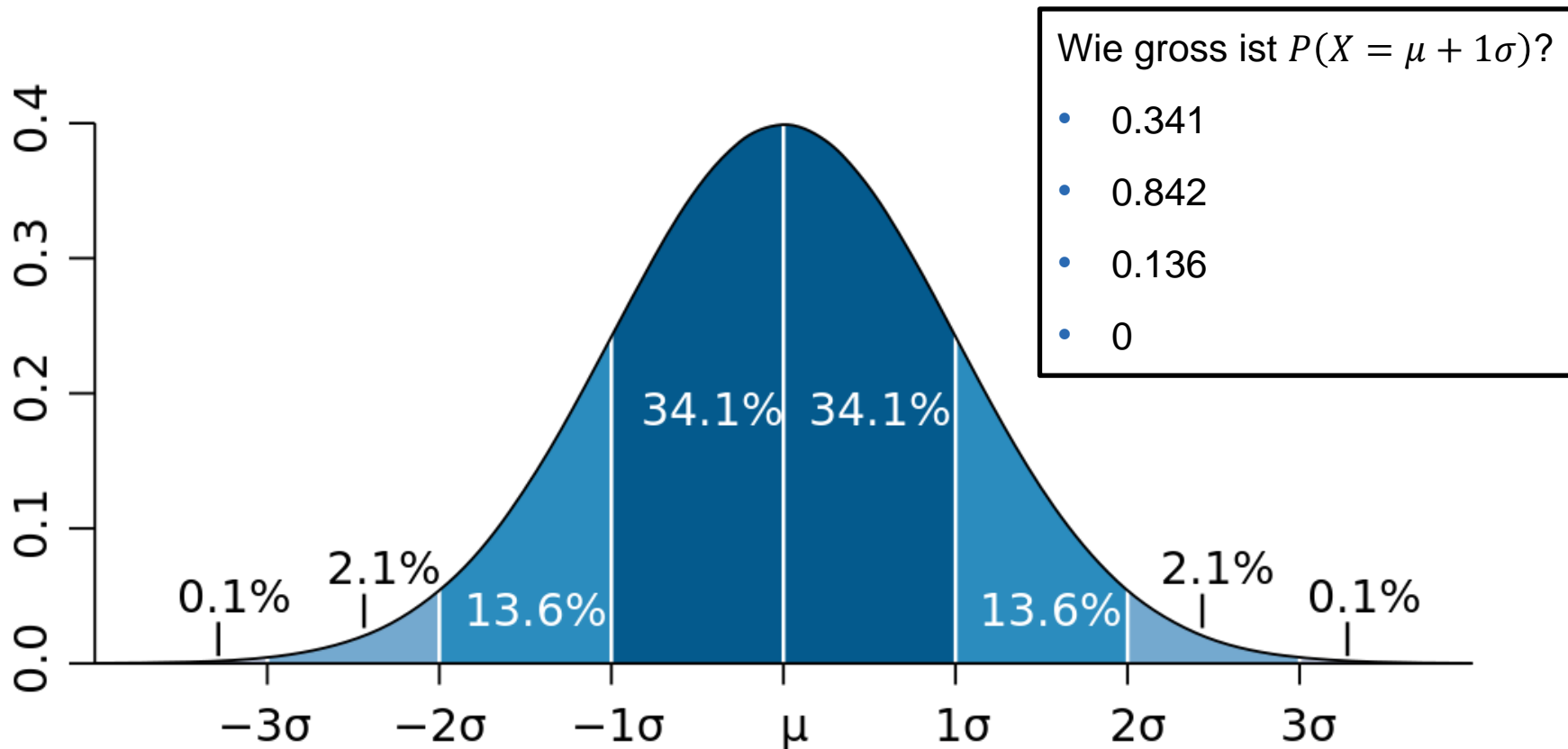
cdf



# Normalverteilung: Messfehler



- Messfehler werden meist mit der Normalverteilung modelliert  
(Begründung: Zentraler Grenzwertsatz, siehe später)

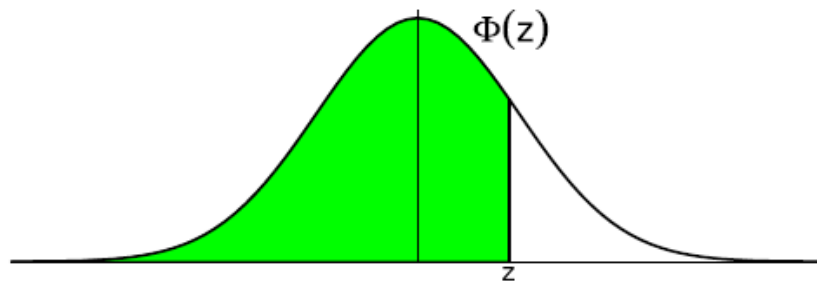


# Standardnormalverteilung

- $Z \sim N(0,1)$
- pdf mit 'φ' bezeichnet:  $\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$
- cdf mit 'Φ' bezeichnet:  $\Phi(x) = \int_{-\infty}^x \varphi(y) dy$

Analytisch nicht lösbar, daher tabelliert

- Bsp:  $P(Z < 1.64) = \Phi(1.64) = 0.9495$



Bsp.:  $P [Z \leq 1.96] = 0.975$

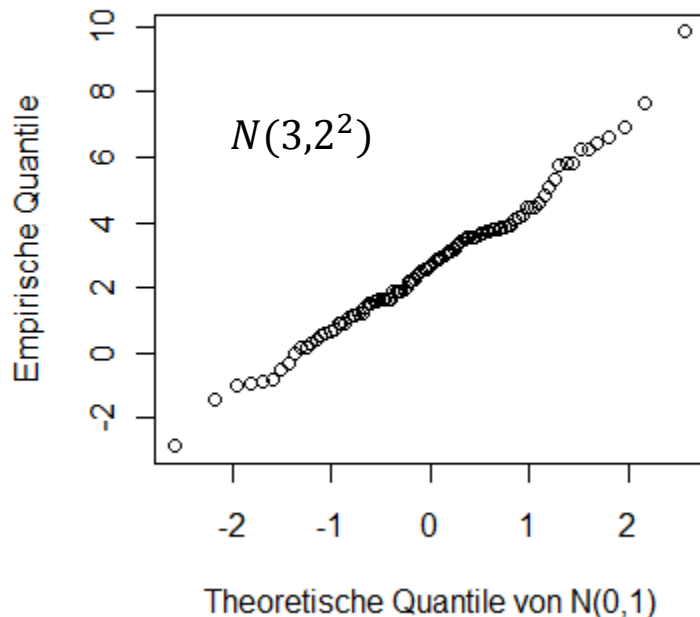
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
.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

# Wie prüft man, ob eine Normalverteilung vorliegt?

- Histogramm der Daten mit pdf vergleichen  
Schwierig kleine Abweichungen zu erkennen
- Einfacher: QQ-Plot – Theoretische Quantile gegen Empirische Quantile

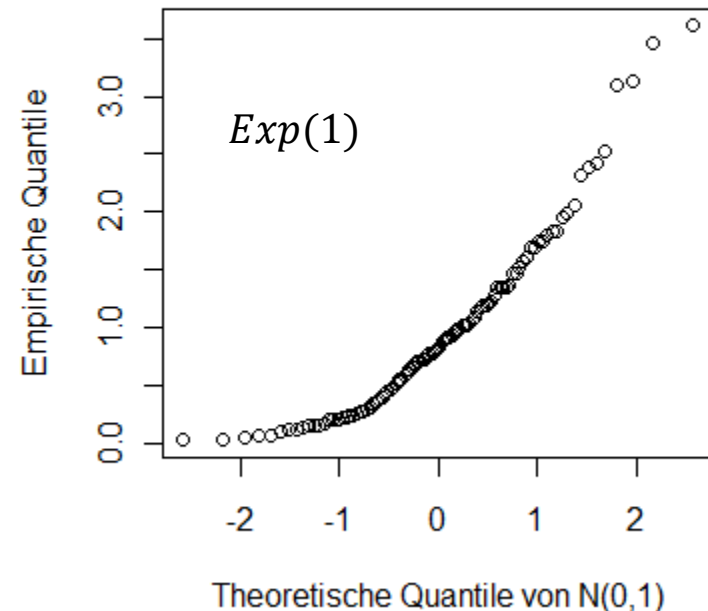
Gerade:

Normalverteilung OK  
Normal Q-Q Plot



Krümmung:

Keine Normalverteilung  
Normal Q-Q Plot



# Standardisieren

- IQ:  $X \sim N(\mu, \sigma^2)$
- Per Konvention:  $\mu = 100$ ,  $\sigma = 15$
- $P(X > 120)$ ?
- $$\begin{aligned} P(X > 120) &= 1 - P(X \leq 120) = 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{120 - \mu}{\sigma}\right) = \\ &= 1 - P\left(Z \leq \frac{120 - 100}{15}\right) = 1 - \Phi(1.33) = 1 - 0.9082 \\ &= 0.0918 \end{aligned}$$

ZV: Gewinn X	P(X=x)
-10	1/6
0	1/2
6	1/3

## ZGS: Beispiel

- n=1000 Spiele
- $E(X_i) = \frac{1}{3}$ ;  $Var(X_i) = 28.6$
- ZGS: Totaler Gewinn

$$S_n \sim N\left(1000 * \frac{1}{3}, 1000 * 28.6\right) = N(333, 28600)$$

- Mit 95% Wahrscheinlichkeit ist der totale Gewinn im Intervall

$$333 \pm 2 * \sqrt{28600} \rightarrow [-5; 671]$$

# ZGS: Normalapproximation des Binomialtests

1. Modell:  $n$  Lose ziehen, gleiche Gewinnwa., unabhängig  
Jedes Los  $X_i$ : 1 mit Wa.  $\pi$ , 0 mit Wa.  $1 - \pi$

$$E(X_i) = \pi, \text{Var}(X_i) = \pi(1 - \pi)$$

$X$ : Anzahl Gewinne;  $X = X_1 + X_2 + \dots + X_n$

2.  $H_0: \pi = \pi_0$ ; z. B.  $H_A: \pi < \pi_0$

3. Teststatistik  $T=X$

ZGS:  $T \sim N(n\pi_0, n\pi_0(1 - \pi_0))$

4.  $\alpha = 0.05$

5. Verwerfungsbereich:  $K = [0, c]$

Finde  $c$ , sodass  $P(T \leq c) = 0.05$  (mit Computer oder:)

Standardisieren & Tabelle:  $P(T \leq c) = P(Z \leq \tilde{c}) = 0.05$

mit  $\tilde{c} = \frac{c - n\pi_0}{\sqrt{n\pi_0(1 - \pi_0)}}$ ;

aus Tabelle:  $\tilde{c} = -1.64$

nach  $c$  auflösen:  $c = n\pi_0 - 1.64\sqrt{n\pi_0(1 - \pi_0)}$

6. Testentscheid