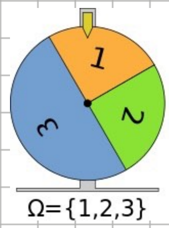
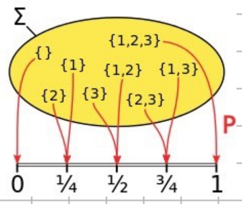
Grundlegende Begriffe

Symbole und Begriffe





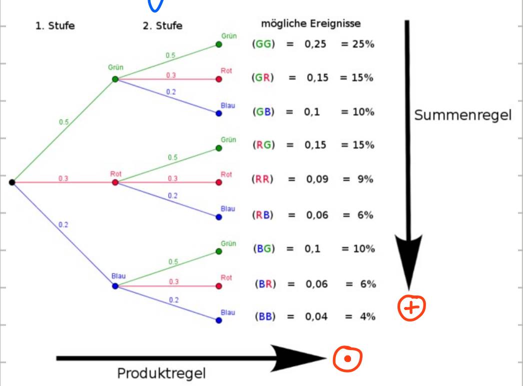
ω = Ergebnis:	Ergebnis eines Zufallsexperiment
Ω = Ergebnismenge: ↳ auch Ergebnisraum	Menge aller möglichen Ergebnisse eines Zufallsexperiment
$\lvert\Omega\rvert$ = Mächtigkeit von Ω: ↳ auch Kardinalität	Anzahl aller möglichen Ergebnissen (Anz. Elemente in Ω)
E = Ereignis:	Teilmenge der Ergebnismenge ($E \subseteq \Omega$)
$\{\omega\}$ = Elementarereignis:	Ereignis mit nur einem Ergebnis ($\{\omega\} \subseteq \Omega$)
Σ = Ereignisalgebra:	Menge aller in einem Zufallsexperiment zu betrachtenden Ereignisse
P = Wahrscheinlichkeit: ↳ kurz W'keit	Wahrscheinlichkeit(smass) ist Funktion, welche jedem Ereignis seine Wahrscheinlichkeit zuordnet: $P: \Sigma \longrightarrow [0,1]$, $E \mapsto P(E)$

Wahrscheinlichkeitsraum $:= (\Omega, \Sigma, P)$

→ diskret, falls Ω endlich oder abzählbar ist, zB $\Omega = \mathbb{N}$

oft hilfreich:

Baumdiagramm



Wahrscheinlichkeit und Mengenlehre

$\bar{A}$ $(= A^c)$	Komplement von A / kein Elementarereignis aus A findet statt		
$A \cap B$	Schnittmenge von A und B	⇨	ein Elementarereignis aus A **und** B findet statt
$A \cup B$	Vereinigung von A und B	⇨	ein Elementarereignis aus A **oder** B findet statt
$A \setminus B$	A ohne B	⇨	ein Elementarereignis aus A tritt ein, aber nicht aus B.
$A \subset B$	A ist eine Teilmege von B	⇨	wenn ein Elementarereignis aus A stattfindet, dann immer auch ein Elementarereignis aus B

klassische Wahrscheinlichkeit

Die klassische Wahrscheinlichkeit tritt dann auf, wenn die Anzahl der Elementarereignisse endlich ist und alle Elementarereignisse gleich wahrscheinlich sind. (Auch Laplace Wahrscheinlichkeit genannt)

$$\boxed{p = \frac{\text{Anzahl günstige Fälle}}{\text{Anzahl mögliche Fälle}} = \frac{|E|}{|\Omega|}}$$

Vorgehen:
- Ereignisraum Ω bestimmen für alle möglichen Fälle und Ereignis E bestimmen für alle günstigen Fälle.
- Die Kardinalität $|\Omega|$ bzw. $|E|$ bestimmen, um die Anzahl möglichen Fälle und die Anzahl günstigen Fälle zu erhalten. oft mit Kombinatorik
- Formel für klassische Wahrscheinlichkeit anwenden.

geometrische Wahrscheinlichkeit

Die Wahrscheinlichkeit ist proportional zur Masszahl eines Flächeninhalts oder ein anderes geometrisches Objekt.

$$\boxed{p = \frac{\text{günstige Fläche}}{\text{mögliche Fläche}}}$$

Vorgehen:
- Geometrisches Objekt zeichnen.
- Mögliche Fläche markieren und berechnen (Bruchteil, der die mögliche Fläche des Ganzen ausmacht).
- Günstige Fläche markieren und berechnen (Bruchteil, der die günstige Fläche des Ganzen ausmacht).
- Formel für geometrische Wahrscheinlichkeit anwenden.

Rechenregeln Wahrscheinlichkeit

$0 \le P[E] \le 1$ für alle E.

$P(\Omega) = 1$ sicheres Ereignis

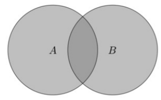
$P(\phi) = 0$ unmögliches Ereignis

$\boxed{P(\bar{E}) = 1 - P(E)}$ Komplementärereignis (oft hilfreich)

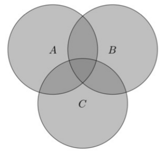
$\boxed{P(E \cup F) = P(E) + P(F) - P(E \cap F)}$ Vereinigung

$P(E \cup F) = P(E) + P(F) \quad falls\ E \cap F = \phi\ (unvereinbar = disjunkt)$

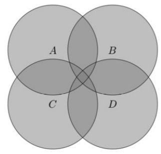
$|A \cup B| = |A| + |B| - |A \cap B|$



$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$



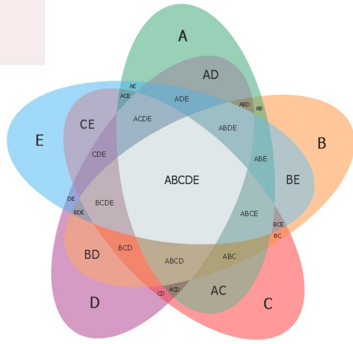
$$|A \cup B \cup C \cup D| = |A| + |B| + |C| + |D| - |A \cap B| - |A \cap C| - |A \cap D| - |B \cap C| - |B \cap D| - |C \cap D| + |A \cap B \cap C| + |A \cap B \cap D| + |A \cap C \cap D| + |B \cap C \cap D| - |A \cap B \cap C \cap D|$$



für mehr als 2 Ereignisse: Siebformel

$$P(E_1 \cup \cdots \cup E_n) = \sum_{1=j_1}^{n} P(E_{j_1}) - \sum_{1=j_1<j_2}^{n} P(E_{j_1} \cap E_{j_2}) + \sum_{1=j_1<j_2<j_3}^{n} P(E_{j_1} \cap E_{j_2} \cap E_{j_3}) + \cdots + (-1)^{n+1} P(E_1 \cap \cdots \cap E_n)$$

$$= \sum_{k=1}^{n} (-1)^{k+1} \sum_{1=j_1<j_2<\cdots<j_k}^{n} P(E_{j_1} \cap E_{j_2} \cap \cdots \cap E_{j_k})$$ → S3A8



falls W'keiten der Schnittmengen mit gleicher Anzahl Teilmengen gleich sind "vereinfacht" sich die Formel zu:

$$nP(A_1) - \binom{n}{2} P(A_1 \cap A_2) + \binom{n}{3} P(A_1 \cap A_2 \cap A_3) - \binom{n}{4} P(A_1 \cap A_2 \cap A_3 \cap A_4)$$
$$+ \binom{n}{5} P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) - \cdots + \cdots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n),$$

—> S3A8

Bonferroni - Ungleichungen

In der Siebformel werden die grösseren Durchschnitte immer kleiner
=> Wir erhalten eine obere Grenze für die Wahrscheinlichkeit der Vereinigung, wenn wir in der Formel nach einer positiven Summe abbrechen und eine untere Grenze wenn wir nach einer negativen Summe abbrechen:

$$P(E_1 \cup \cdots \cup E_n) \le \sum_{1=j_1}^{n} P(E_{j_1})$$ → S4A1

$$P(E_1 \cup \cdots \cup E_n) \ge \sum_{1=j_1}^{n} P(E_{j_1}) - \sum_{1=j_1<j_2}^{n} P(E_{j_1} \cap E_{j_2})$$

$$P(E_1 \cup \cdots \cup E_n) \le \sum_{1=j_1}^{n} P(E_{j_1}) - \sum_{1=j_1<j_2}^{n} P(E_{j_1} \cap E_{j_2}) + \sum_{1=j_1<j_2<j_3}^{n} P(E_{j_1} \cap E_{j_2} \cap E_{j_3})$$

$\vdots$

genauer ↓ einfacher ↑

Unabhängigkeit von Ereignissen

A und B sind (paarweise) unabhängig voneinander und beeinflussen sich somit nicht, wenn gilt:

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$ Schnittmenge

$$P(A|B) = P(A)$$

Dabei ist zu beachten, dass unabhängig nicht das Gleiche ist wie disjunkt!

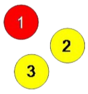
$$P(A \cap B) = P(A) \cdot P(B) \quad \neq \quad P(A \cap B) = 0$$

unabhängig ≠ disjunkt
↳ schliessen sich gegenseitig aus

Kombinatorik

Grundmenge

alle Elemente | Auswahl

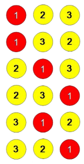


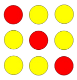
immer mit Reihenfolge

Variation oder Kombination?

mit Reihenfolge | ohne Reihenfolge

Permutation | **Variation** | **Kombination**





	Permutation		Variation		Kombination	
	Treten Elemente mehrfach auf?		Treten Elemente mehrfach auf?		Treten Elemente mehrfach auf?	
	nein	ja	nein	ja	nein	ja
	ohne Wdh.	mit Wdh.	ohne Wdh.	mit Wdh.	ohne Wdh.	mit Wdh.
Formeln:	$n!$	$\frac{n!}{m_1! \cdot m_2! \cdot \ldots \cdot m_n!}$	$\frac{n!}{(n-k)!}$	n^k	$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$	$\binom{n+k-1}{k} = \frac{(n+k-1)!}{(n-1)! \cdot k!}$

Urne 6 Kugeln 2 Ziehungen

	Ziehen ohne Zurücklegen (ZOZ)	Ziehen mit Zurücklegen (ZMZ)
ungeordnet	Kombination ohne W. $\binom{n}{k} = \frac{n!}{k! * (n-k)!}$ TR: nCr — 15 Möglichkeiten	Kombination mit W. $\binom{n+k-1}{k} = \frac{(n+k-1)!}{k! * (n-1)!}$ — 21 Möglichkeiten
geordnet	Variation ohne W. $\frac{n!}{(n-k)!}$ — 30 Möglichkeiten	Variation mit W. n^k — 36 Möglichkeiten
	k <u>kann nicht</u> größer sein als n	k <u>kann</u> größer sein als n

Kombination ohne W.:

	1	2	3	4	5	6
1		X	X	X	X	X
2			X	X	X	X
3				X	X	X
4					X	X
5						X
6						

15 Möglichkeiten

Kombination mit W.:

	1	2	3	4	5	6
1	X	X	X	X	X	X
2		X	X	X	X	X
3			X	X	X	X
4				X	X	X
5					X	X
6						X

21 Möglichkeiten

Variation ohne W.:

	1	2	3	4	5	6
1		X	X	X	X	X
2	X		X	X	X	X
3	X	X		X	X	X
4	X	X	X		X	X
5	X	X	X	X		X
6	X	X	X	X	X	

30 Möglichkeiten

Variation mit W.:

	1	2	3	4	5	6
1	X	X	X	X	X	X
2	X	X	X	X	X	X
3	X	X	X	X	X	X
4	X	X	X	X	X	X
5	X	X	X	X	X	X
6	X	X	X	X	X	X

36 Möglichkeiten

Auswahl : k Objekte aus n auswählen

Reihenfolge: ABC ist nicht das gleiche wie CBA

Wiederholung: Objekte dürfen mehrmals benutzt / gezogen werden

Anzahl k-te partielle Ableitungen von $f: \mathbb{R}^n \to \mathbb{R}$ (siehe Satz von Schwarz, AN3)

Anzahl Ableitungen ohne Satz von Schwarz: n^k

Anzahl Möglichkeiten k Variablen aus n abzuleiten
wobei es auf die Reihenfolge ankommt
und die Variablen mehrmals gewählt werden dürfen
$\to$ Variation mit Wdh.

Auswahl
mit Reihenfolge
mit Wdh.

	1	2	3	4	5	6
1	X	X	X	X	X	X
2	X	X	X	X	X	X
3	X	X	X	X	X	X
4	X	X	X	X	X	X
5	X	X	X	X	X	X
6	X	X	X	X	X	X

36 Möglichkeiten

Anzahl zu berechnender Ableitungen für Hesse-Matrix (bei $n=6$)

Anzahl Ableitungen mit Satz von Schwarz: $\binom{k+n-1}{k}$

Anzahl Möglichkeiten k Variablen aus n abzuleiten
wobei es nicht auf die Reihenfolge ankommt
und die Variablen mehrmals gewählt werden dürfen
$\to$ Kombination mit Wdh.

Auswahl
ohne Reihenfolge
mit Wdh.

	1	2	3	4	5	6
1	X	X	X	X	X	X
2		X	X	X	X	X
3			X	X	X	X
4				X	X	X
5					X	X
6						X

21 Möglichkeiten

Anzahl zu berechnender Ableitungen für Hesse-Matrix (bei $n=6$)

Binomialkoeffizient

$$\boxed{\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}}$$

«n tief k»
«n choose k»
$\to$ TR: nCr(n,r)

Anzahl Möglichkeiten k Objekte aus n auszuwählen (Komb. ohne Wdh.)

wichtige Spezialfälle:
(≙ Rand d. pascalschen Dreiecks)

$$\binom{n}{0} = \binom{n}{n} = 1, \quad \binom{n}{1} = \binom{n}{n-1} = n$$

pascalsches Dreieck

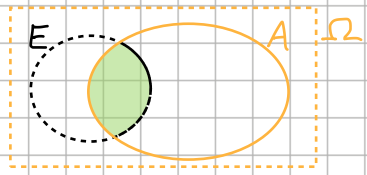
n											k
0						1					0
1					1		1				1
2				1		2		1			2
3			1		3		3		1		3
4		1		4		6		4		1	4
	1		5		10		10		5		1

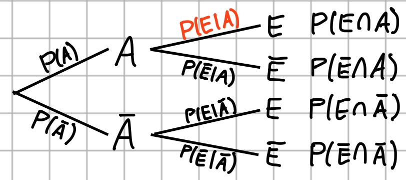
Bedingte Wahrscheinlichkeit

$$\boxed{P(E|A) = \frac{P(E\cap A)}{P(A)}}$$

auch $P_A(E)$

«E gegeben A»





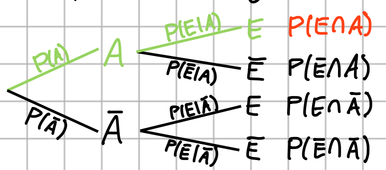
Wahrscheinlichkeit, dass E eintritt gegeben dass A eingetreten ist

Multiplikationsregel: Schnittmenge mithilfe bedingter W'keit

2 Ereignisse:

$$\boxed{P(E\cap A) = P(E|A)\cdot P(A)}$$

Multiplikation entlang Ast:

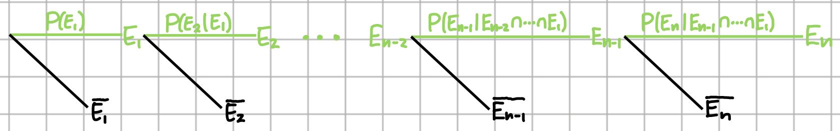


mehr als 2 Ereignisse:

$$P(E_n\cap E_{n-1}\cap\cdots\cap E_1)$$
$$= P(E_n|E_{n-1}\cap E_{n-2}\cap\cdots\cap E_1)\cdot P(E_{n-1}|E_{n-2}\cap E_{n-3}\cap\cdots\cap E_1)\cdot$$
$$\cdots\cdot P(E_2|E_1)\cdot P(E_1).$$

Herleitung:

$$P(E_n\cap E_{n-1}\cap\cdots\cap E_1)$$
$$= \frac{P(E_n\cap E_{n-1}\cap\cdots\cap E_1)}{P(E_{n-1}\cap E_{n-2}\cap\cdots\cap E_1)}\cdot\frac{P(E_{n-1}\cap E_{n-2}\cap\cdots\cap E_1)}{P(E_{n-2}\cap E_{n-3}\cap\cdots\cap E_1)}\cdot$$
$$\cdots\cdot\frac{P(E_2\cap E_1)}{P(E_1)}\cdot P(E_1)$$

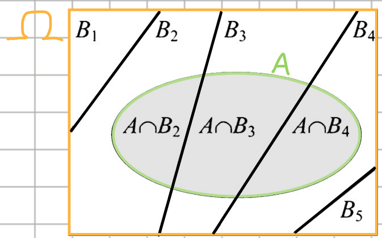


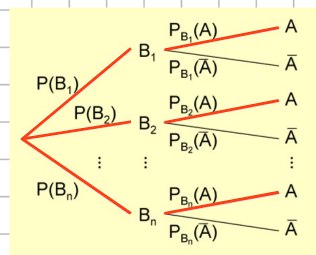
$P(E_n\cap\cdots\cap E_1)$

= Multiplikation des ganzen Astes

Gesetz d. totalen Wahrscheinlichkeit: W'keit mithilfe bed. W'keit

$$\boxed{P(E) = \sum_{k=1}^{n} P(E|B_k)\,P(B_k)}$$





Ω in disjunkte $B_1, \dots, B_n$ aufteilen

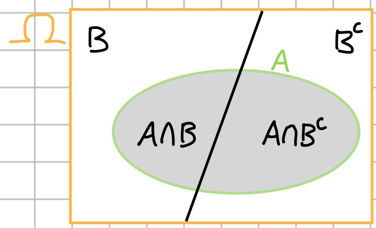
⇒ $B_1, \dots, B_n$ müssen 2 Bedingungen erfüllen:

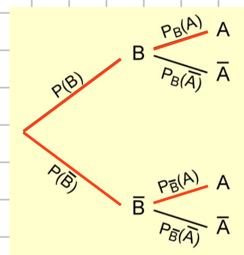
- müssen paarweise disjunkt sein
- $B_1 \cup \dots \cup B_n = \Omega$ muss gelten

2 Teilmengen:

$$P(E) = P(E\cap B) + P(E\cap B^c)$$
$$= P(E|B)\cdot P(B) + P(E|B^c)\cdot P(B^c)$$

hier wird Ω in B und $\bar{B}$ aufgeteilt





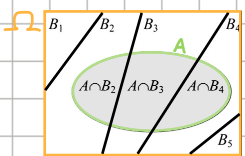
Satz von Bayes: P(B|A) mithilfe von P(A|B) berechnen

$$\boxed{P(B|A) = \frac{P(B\cap A)}{P(A)} = \frac{P(A|B)\,P(B)}{P(A)}}$$

Herleitung: Multiplikationsregel auf P(B∩A) anwenden

mit totaler W'keit:

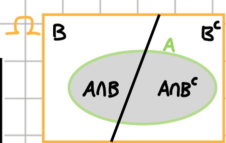
$$P(B_i|A) = \frac{P(B_i\cap A)}{P(A)} = \frac{P(A|B_i)\,P(B_i)}{\sum_{k=1}^{n} P(A|B_k)\,P(B_k)}$$



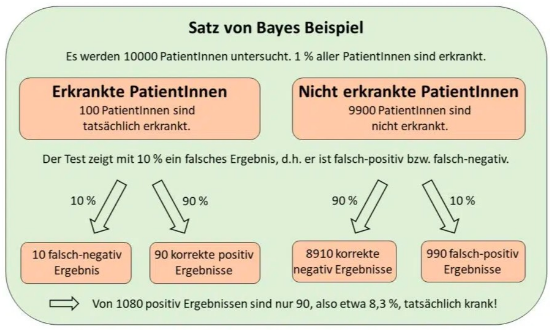
Herleitung: Gesetz d. tot. W'keit auf P(A) anwenden

2 Teilmengen:

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)\,P(B) + P(A|B^c)\,P(B^c)}$$



Bsp:



$$P(krank|positiv) = \frac{P(positv|krank)\cdot P(krank)}{P(positiv)}$$

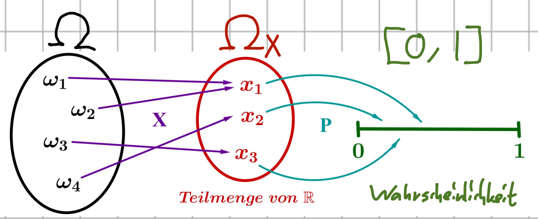
$$P(krank|positv) = \frac{0{,}9\cdot 0{,}01}{1080/10000} = \frac{0{,}9\cdot 0{,}01}{0{,}1089} \sim 0{,}083 = 8{,}3\,\%$$

Diskrete Zufallsvariablen

Zufallsvariable

Funktion welche jedem Ergebnis $\omega \in \Omega$ eine Zahl $x \in \mathbb{R}$ zuweist
→ manchmal auch $\mathbb{R}^n$ anstatt $\mathbb{R}$, resp $\vec{x}$ anstatt x!

$$\boxed{\begin{aligned} X: \Omega &\to \mathbb{R} \\ \omega &\mapsto x \end{aligned}}$$



salopp: Grösse, deren Wert vom Zufall abhängt

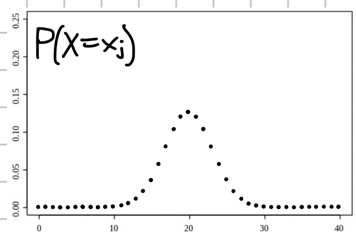
Wahrscheinlichkeitsfunktion

EN: probability mass function (PMF)

$$\boxed{\begin{aligned} p_X: \Omega_X &\to [0,1] \\ x_j &\mapsto p_X(x_j) = P(X = x_j) \end{aligned}}$$

Funktion welche allen $x_j \in \Omega_X$ die Wahrscheinlichkeit zuordnet dass die diskrete Zufallsvariable X diesen Wert x_j annimmt

x_j	x_1	x_2	...	x_{m-1}	x_m
$p_X(x_j)$	$P(X=x_1)$	$P(X=x_2)$	...	$P(X=x_{m-1})$	$P(X=x_m)$



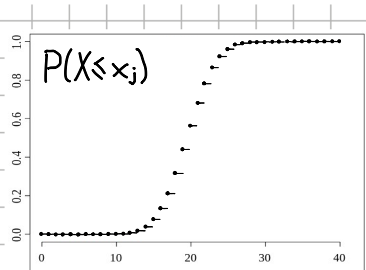
kumulierte Verteilungsfunktion

EN: cumulative distribution function (CDF)

$$\boxed{P(X \leq x_j) = \sum_{k=1}^{j} p_X(x_k)}$$

W'keit, dass X einen Wert kleiner gleich x_j annimmt

x_j	x_1	x_2	...	x_{m-1}	x_m
$\sum_{k=1}^{j} p_X(x_k)$	$P(X \leq x_1)$	$P(X \leq x_2)$	...	$P(X \leq x_{m-1})$	$P(X \leq x_m) = 1$



Unabhängigkeit

$X_1, \cdots, X_n$ sind unabhängig, falls

$$\begin{aligned} p_{\vec{X}}(x_1, \cdots, x_n) &= P(X_1 = x_1 \text{ und } \cdots \text{ und } X_n = x_n) \\ &= P(X_1 = x_1) \cdots P(X_n = x_n) \\ &= p_{X_1}(x_1) \cdots p_{X_n}(x_n) \end{aligned}$$

für alle $\vec{x} = (x_1, \cdots, x_n) \in \Omega_{\vec{X}} = \Omega_{X_1} \times \cdots \times \Omega_{X_n}$ gilt

Jede der Koordinatenfunktionen $X_1, \cdots, X_n$ einer diskreten, vektorwertigen Zufallsvariable

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^n$$

können wir als diskrete, reellwertige Zufallsvariable auffassen. Die Verteilung $p_{\mathbf{X}}$ der diskreten, vektorwertigen Zufallsvariable $\mathbf{X}$ enthält Zusammenhänge zwischen den Zufallsvariablen $X_1, \cdots, X_n$ und diese Zusammenhänge verlieren wir, wenn wir nur die Verteilungen $p_{X_1}, \cdots, p_{X_n}$ der einzelnen reellwertigen Zufallsvariablen betrachten. Wir werden diese Thematik später genauer studieren und uns an dieser Stelle auf einen wichtigen Spezialfall beschränken. Wenn keine zusätzliche Information in der Verteilung der vektorwertigen Zufallsvariable $\mathbf{X}$ enthalten ist, nennen wir die Zufallsvariablen $X_1, \cdots, X_n$ unabhängig.

Erwartungswert μ (Mü)

$$\boxed{\mu = E[X] = \sum_{j=1}^{m} x_j \cdot p_X(x_j)}$$

gewichtete Summe von allen möglichen $x_j \in \Omega_X$, wobei jeder Wert x_j mit seiner Wahrscheinlichkeit $p_X(x_j)$ gewichtet wird

$$\boxed{E[a_1 \cdot X_1 + a_2 \cdot X_2] = a_1 \cdot E[X_1] + a_2 \cdot E[X_2]}$$ Linearität

$$\boxed{E[f(X)] = \sum_{j=1}^{m} f(x_j) \cdot p_X(x_j)}$$

$$\boxed{E[X^2] = \sum_{j=1}^{m} x_j^2 \cdot p_X(x_j)}$$

Varianz σ^2

$$\boxed{\sigma^2 = V[X] = \sum_{j=1}^{m} (x_j - \mu)^2 \cdot p_X(x_j)}$$

gewichtete Summe der quadrierten Abweichungen vom Erwartungswert

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

oft hilfreich

$$\boxed{V[aX+b] = a^2 \cdot V[X]}$$

Standardabweichung σ (Sigma)

$$\boxed{\sigma = \sqrt{V[X]} = \sqrt{\sum_{j=1}^{m} (x_j - \mu)^2 \cdot p_X(x_j)}}$$

$$\boxed{\sqrt{V[aX+b]} = |a| \cdot \sqrt{V[X]}}$$

W'keit einer Abweichung von mehr als $d > 0$ vom Erwartungswert

$$\boxed{P(|X - E[X]| \geq d) \leq \frac{V[X]}{d^2}}$$

Chebyshev-Ungleichung
selten hilfreich

Rundengewinn → S6A6

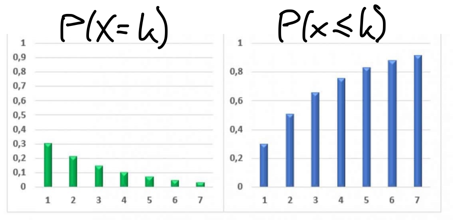
$$P(\text{zuerst } n \text{ Runden}) = \sum_{k=0}^{n-1} \binom{n+k-1}{n-1} p^n (1-p)^k$$

Wer zuerst n Runden gewinnt, gewinnt das Turnier
p = W'keit eine einzelne Runde zu gewinnen

Diskrete Verteilungen

Geometrische Verteilung

Anzahl Versuche bis zum ersten Erfolg bei konstanter Erfolgswahrscheinlichkeit p. Beispiel: Kugeln ziehen *mit* zurücklegen, von den Kugeln ist ein Anteil p rot, Erfolg = eine rote.



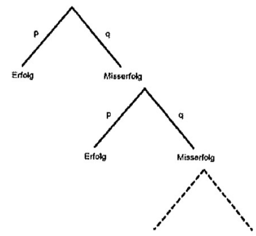
Parameter

- p: Erfolgswahrscheinlichkeit

erster Erfolg beim k-ten Versuch: $P(X = k) = p \cdot (1-p)^{k-1}$

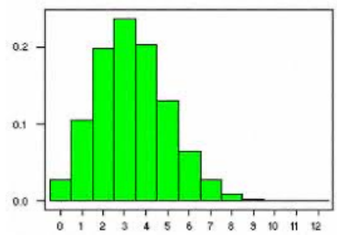
erster Erfolg nach nicht mehr als k Versuchen: $P(X \leq k) = 1 - (1-p)^k$

$$E(X) = 1/p$$
$$\sigma^2(X) = (1-p)/p^2$$



Binomische Verteilung

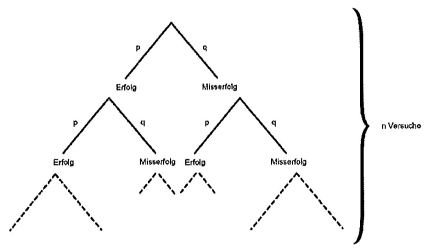
Anzahl Erfolge in n Versuchen mit konstanter Wahrscheinlichkeit p. Beispiel: Kugeln ziehen *mit* zurücklegen, von den Kugeln ist ein Anteil p rot, Erfolg = eine rote



Parameter

- p: Erfolgswahrscheinlichkeit
- n: Anzahl Versuche

$$\begin{aligned} P(X = k) &= \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k} \\ E(X) &= n \cdot p \\ \sigma^2(X) &= n \cdot p \cdot (1-p) \end{aligned}$$

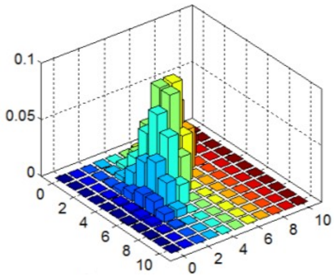


$$P(X \leq l) = \sum_{k=0}^{l} \binom{n}{k} p^k \cdot (1-p)^{n-k}$$ → S6A3 (Bakterien), S6A5 (Plantage)

$$P(X \geq l) = \sum_{k=l}^{n} \binom{n}{k} p^k \cdot (1-p)^{n-k}$$ → S11A7

Multinomische Verteilung -> 5647

Anzahl Ergebnisse vom Typ $1, 2, \ldots, k$, welche konstante Wahrscheinlichkeiten $p_1, p_2, \ldots, p_k$ haben, n Versuchen. Beispiel: Anzahl Würfe mit Augenzahl $1, 2, \ldots, 6$ in n unabhängigen Würfen eines Würfels.



Parameter

- $p_1, \ldots p_k$: Erfolgswahrscheinlichkeiten
- n: Anzahl Versuche

$$
\begin{aligned}
P(n_1, n_2, \ldots, n_k) &= \binom{n}{n_1\, n_2 \ldots n_k} \cdot p_1^{n_1} \cdot \ldots \cdot p_k^{n_k} \\
&= \frac{n!}{n_1! \cdot n_2! \cdot \ldots \cdot n_k!} \cdot p_1^{n_1} \cdot \ldots \cdot p_k^{n_k}
\end{aligned}
$$

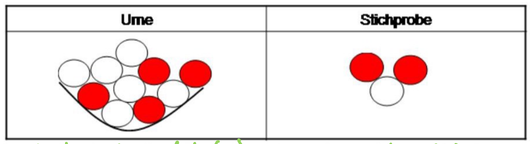
Wobei $n = n_1 + \ldots + n_k$

Hypergeometrische Verteilung → S6A8

Anzahl Erfolge in n Versuchen, *ohne* zurücklegen. Beispiel: n Kugeln ziehen aus g Kugeln, von denen h rot sind. Erfolg = eine rote

Parameter

- g: Anzahl Kugeln
- h Anzahl rote Kugeln
- n: Anzahl Versuche



g: Kugeln total (10)
h: rote Kugeln (4)
n: Versuche (3)
k: Erfolge (2)

$$
\begin{aligned}
P(X=k) &= \frac{\binom{h}{k}\cdot\binom{g-h}{n-k}}{\binom{g}{n}} \\
E(X) &= n\cdot\frac{h}{g} \\
\sigma^2(X) &= n\cdot\frac{h}{g}\cdot\left(1-\frac{h}{g}\right)\cdot\frac{g-n}{g-1}
\end{aligned}
$$

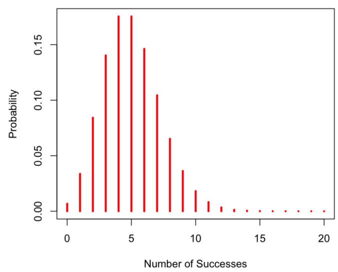
W'keit k (2) rote Kugeln zu erhalten wenn man n(3) Kugeln aus einer Urne zieht in welcher g(10) Kugeln liegen, wovon h(4) rot sind

Poisson Verteilung → S6A9, S6A10

Anzahl Ereignisse pro Zeiteinheit, wenn der Mittelwert (Erwartungswert) der Ereignisse bekannt ist und die Ereignisse unabhängig voneinander eintreten. Beispiel: Anzahl radioaktiver Zerfälle / Sekunde, Anzahl Kunden, die pro Stunde am Postschalter eintreffen

Parameter

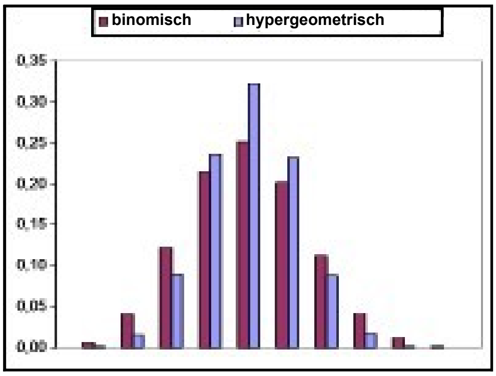
- λ Erwartungswert (= Varianz)



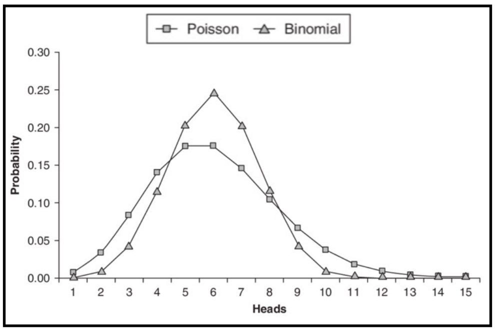
$$
\begin{aligned}
P(X=k) &= e^{-\lambda}\cdot\frac{\lambda^k}{k!} \\
E(X) &= \lambda \\
\sigma^2(X) &= \lambda
\end{aligned}
$$

Näherungen

Verteilung 1	Parameter	Verteilung 2	Parameter	Bedingung	
hypergeometrisch	g, h, n	binomisch	$n = n, p = h/g$	$h \ll g$	→ S6A8
binomisch	n, p	Poisson	$\lambda = n \cdot p$	$p \ll 1$	→ S6A9, S6A10

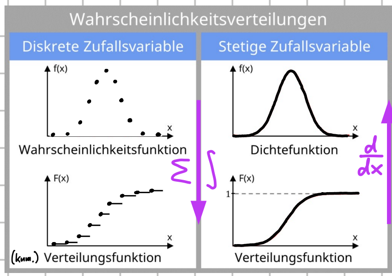


Aus einer Urne mit roten und blauen Kugeln werden 10 Kugeln gezogen. Zufallsgröße X ist die Zahl der gezogenen roten Kugeln.
a) Ziehen mit Zurücklegen. Die Grundwahrscheinlichkeit ändert sich nicht. Die Wahrscheinlichkeiten sind binomial verteilt.
b) Ziehen ohne Zurücklegen. Die Grundwahrscheinlichkeit ändert sich. Die Wahrscheinlichkeiten sind hypergeometrisch verteilt.



Poisson versus binomial distribution, from number of heads in a coin toss. The Poisson distribution shown assumes an average of 5 heads per toss (λ = 5); the binomial distribution shown assumes 10 coins were tossed, and the probability of a head is 0.5.

Stetige Zufallsvariablen



Dichtefunktion

EN: probability density function (PDF)

analog zu Wahrscheinlichkeitsfunktion bei diskret

$$f(x) = \frac{d}{dx} F(x)$$

→ S7A1, S7A2

Verteilungsfunktion

EN: cumulative distribution function (CDF)

analog zu kumulierten Verteilung bei diskret

$$\boxed{F(x) = \int_{-\infty}^{x} f(t)\,dt = P(X \leq x)}$$

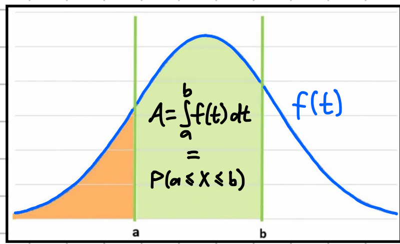
W'keit, dass X einen Wert kleiner gleich x annimmt

$$\int_{-\infty}^{\infty} f(t)\,dt = 1$$ → S7A3

$$P(X = x) = \int_x^x f(t)\,dt = 0$$

$$\boxed{P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int_a^b f(t)\,dt}$$

$$P(X > x) = 1 - P(X \leq x) = 1 - F(x)$$



Quantile

p-Quantil x_p: $\quad x_p = F^{-1}(p)$

Median

Median $a = x_{\frac{1}{2}}$ (0.5-Quantil): $\quad a = F^{-1}(0.5) \neq E[X]$

Erwartungswert μ (Mü)

$$\boxed{\mu = E[X] = \int_{-\infty}^{\infty} t \cdot f(t)\,dt}$$

$$\boxed{E[a_1 \cdot X_1 + a_2 \cdot X_2] = a_1 \cdot E[X_1] + a_2 \cdot E[X_2]}$$ Linearität

$$\boxed{E[r(X)] = \int_{-\infty}^{\infty} r(t) \cdot f(t)\,dt}$$ μ einer transformierten Zufallsvariablen

$$\boxed{E[X^2] = \int_{-\infty}^{\infty} t^2 \cdot f(t)\,dt}$$

Varianz σ^2

$$\boxed{\sigma^2 = V[X] = \int_{\infty}^{\infty} (t-\mu)^2 \cdot f(t)\,dt}$$

$$\boxed{V[X] = E[X^2] - (E[X])^2}$$ oft hilfreich!

$$\boxed{V[aX + b] = a^2 \cdot V[X]}$$

Standardabweichung σ

$$\boxed{\sigma = \sqrt{V[X]} = \sqrt{\int_{\infty}^{\infty} (t-\mu)^2 \cdot f(t)\,dt}}$$

$$\boxed{\sqrt{V[aX+b]} = |a| \cdot \sqrt{V[X]}}$$

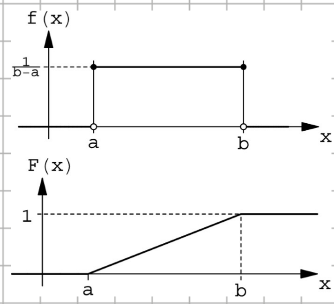
W'keit einer Abweichung von mehr als $d > 0$ vom Erwartungswert

$$\boxed{P(|X - E[X]| \geq d) \leq \frac{V[X]}{d^2}}$$

Chebyshev-Ungleichung
selten hilfreich

Stetige Verteilungen

Gleichverteilung/Uniformverteilung



Eine **auf dem Intervall** (a, b) **gleichverteilte** Zufallsvariable X ist absolut stetig und hat die Verteilungsfunktion und Dichte

$$\boxed{F_X(x) = \begin{cases} 0 & \text{falls } x \le a \\ \frac{x-a}{b-a} & \text{falls } a < x \le b \\ 1 & \text{falls } b < x \end{cases} \quad \text{und } f(x) = \begin{cases} 0 & \text{falls } x \le a \\ \frac{1}{b-a} & \text{falls } a < x \le b \\ 0 & \text{falls } b < x \end{cases}}$$

Wir haben den Erwartungswert und die Varianz

$$\boxed{E[X] = \frac{a+b}{2} \text{ und } \operatorname{var}(X) = \frac{(b-a)^2}{12}}$$

berechnet.

W'keit berechnen

Lösen mit Dichtefunktion

$P[u \le X \le v] = \int_u^v f(x)\,dx$

$P[X \ge u] = \int_u^\infty f(x)\,dx$

$P[X \le v] = \int_{-\infty}^v f(x)\,dx$

Zu beachten: $f(x)$ ist nicht für alle x gleich!!!
Ausserhalb des Intervalls [a, b] gilt $f(x) = 0$

Lösen mit Verteilungsfunktion

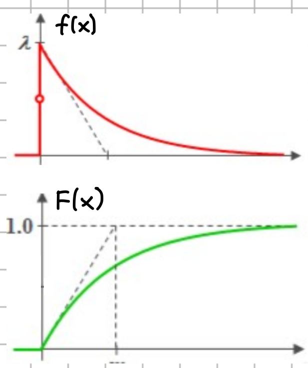
$P[u \le X \le v] = F(v) - F(u)$

$P[X \ge u] = 1 - F(u)$

$P[X \le v] = F(v)$

Zu beachten: $F(x)$ ist nicht für alle x gleich!!!
Für $x < a$ gilt $F(x) = 0$, für $x > b$ gilt $F(x) = 1$

Exponentialverteilung



Die Exponentialverteilung wird häufig verwendet, um die Wartezeiten zwischen Ereignissen zu modellieren (Zeit zwischen Anrufen auf Hotline, Zeit zwischen ankommenden Kunden, Zeit zwischen dem Ausfall von Glühbirnen eines Gebäudes, ...) und kann als nicht diskrete Version der geometrischen Verteilung aufgefasst werden.[20]

Eine **exponentiell mit der Rate** $\lambda > 0$ **verteilte** Zufallsvariable hat die Verteilungsfunktion und Dichte

↳ TR: *Expr*|**λ>0**

$$\boxed{F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{falls } x > 0 \\ 0 & \text{sonst} \end{cases} \quad \text{und } f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{falls } x > 0 \\ 0 & \text{sonst} \end{cases}}$$

Mittels partieller Integration können der Erwartungswert und die Varianz

$$\boxed{E[X] = \frac{1}{\lambda} \text{ und } \operatorname{var}(X) = \frac{1}{\lambda^2}}$$

berechnet werden. (Darstellung [7])

Eine wichtige Eigenschaft der Exponentialverteilung ist die sogenannte **Gedächtnislosigkeit**. Für jede exponentiell mit Rate λ verteilten Zufallsvariable X und beliebige Werte x, $d > 0$ gilt die Gleichung

$$\boxed{P(X > x + d \,|\, X > x) = P(X > d)}$$ →S8A1, S8A2

In Worten: Wenn wir schon x Zeiteinheiten gewartet haben, dann ist die Wahrscheinlichkeit, noch einmal mindestens d Zeiteinheiten warten zu müssen die gleiche, wie wenn wir überhaupt noch nicht gewartet hätten.

W'keit berechnen

Lösen mit Dichtefunktion

$P[a \le X \le b] = \int_a^b f(x)\,dx$

$= \int_a^b \lambda e^{-\lambda x}\,dx = \left[-e^{-\lambda x}\right]_a^b$

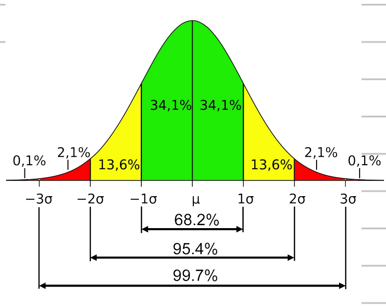
Lösen mit Verteilungsfunktion

$P[a \le X \le b] = F(b) - F(a)$

$= (1 - e^{-\lambda b}) - (1 - e^{-\lambda a})$

$= e^{-\lambda a} - e^{-\lambda b}$

Normalverteilung/Gaussverteilung



Wir betrachten in diesem Abschnitt die sogenannte Normal- oder Gaussverteilung, welche die wahrscheinlich wichtigste aller Verteilungen ist. Eine **normalverteilte Zufallsvariable** X besitzt die Dichte

$$\boxed{f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \mathcal{N}_{\mu,\sigma}(x)}$$

TR: **normPdf(x,μ,σ)**

mit dem Mittelwert $\mu \in \mathbb{R}$ und der Standardabweichung $\sigma > 0$. Die Bezeichnung der Parameter ist gerechtfertigt, denn mittels Rechnung erhalten wir den Erwartungswert und die Varianz

$$E[X] = \mu \text{ und } \operatorname{var}(X) = \sigma^2.$$

Die Verteilungsfunktion

$$\boxed{F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t = \Phi_{\mu,\sigma}(x)}$$

TR: **normCdf(-∞,x,μ,σ)**

kann leider nicht geschlossen integriert werden[21]. Eine Darstellung als Potenzreihe ist jedoch möglich und für die numerische Berechnung der Werte werden geeignete Approximationstechniken verwendet.

Da die Dichte $f_X(x)$ der Normalverteilung symmetrisch um den Mittelwert μ liegt, erhalten wir für die Verteilungsfunktion die Beziehung

$$\boxed{F_X(\mu - x) = 1 - F_X(\mu + x)}$$

Die Bedeutung der Normalverteilung ergibt sich aus dem zentralen Grenzwertsatz, welcher vereinfachend formuliert besagt[22], dass jede Summe von vielen kleinen und untereinander unabhängigen Beiträgen annähernd normalverteilt ist.[23] Wir werden diesen Grenzwertsatz im Kapitel 5 studieren.

In vielen Quellen finden wir die Werte der Verteilungsfunktion

$$\boxed{\Phi(x) = \Phi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,\mathrm{d}t}$$

der sogenannten **Standard-Normalverteilung**, also der Normalverteilung mit den Parametern $\mu = 0$ und $\sigma = 1$, in Tabellenform.[24] Daraus können wir bei Bedarf die Werte der Verteilungsfunktion der allgemeinen Normalverteilung berechnen, denn für jede normalverteilte Zufallsvariable X und jede Zahl $x \in \mathbb{R}$ erhalten wir mittels der Integraltransformation $t = \sigma s + \mu = g(s)$ die Verteilungsfunktion

$$\boxed{\begin{aligned} \Phi_{\mu,\sigma}(x) = F_X(x) = P(X \leq x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,\mathrm{d}t \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-\frac{s^2}{2}}\,\mathrm{d}s = \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned}}$$

→ S10A11

Für eine beliebige normalverteilte Zufallsvariable X und einen Wert $k \in \mathbb{R}$ ist die Wahrscheinlichkeit, dass der Wert der Zufallsvariable X um weniger als k Standardabweichungen σ vom Mittelwert μ abweicht

$$\boxed{\begin{aligned} P(\mu - k\sigma \leq X \leq \mu + k\sigma) &= F_X(\mu + k\sigma) - F_X(\mu - k\sigma) \\ &= 2F_X(\mu + k\sigma) - 1 = 2\Phi(k) - 1, \end{aligned}}$$

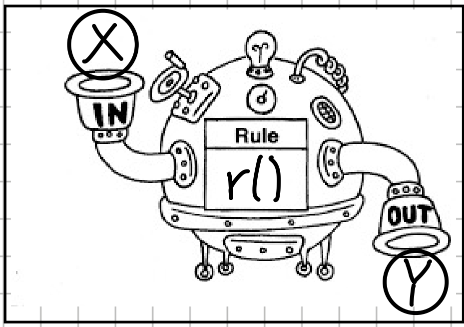
→ S8A5

immer die gleiche.

$$\begin{aligned} P(\mu - \sigma \leq X \leq \mu + \sigma) &\cong 0.683 \\ P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) &\cong 0.955 \\ P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) &\cong 0.997 \\ &\vdots \end{aligned}$$

Bei der Berechnung der p-Quantilen x_p kommt die inverse Funktion der Verteilungsfunktion F_X oder die inverse der Standard-Normal-Verteilungsfunktion Φ zum Einsatz.

Transformation von Zufallsvariablen



Notation & Definition:

Y ist die transformierte Zufallsvariable, die aus X durch Anwenden der Funktion $r: \mathbb{R} \to \mathbb{R}$ entsteht:

$$\boxed{r(X) = Y}$$

Dichtefunktion von Y

falls $P(a < X < b) = 1$ und $r: (a,b) \to (c,d)$ dann hat $Y = r(X)$ die Dichte

$$\boxed{f_Y(y) = \begin{cases} f_X(s(y)) \left|\frac{d}{dy} s(y)\right| & c < y < d \\ 0 & \text{sonst} \end{cases}}$$

$s(y) = r^{-1}(y)$ Umkehrfunktion

$\to r(x) = y$ nach x auflösen

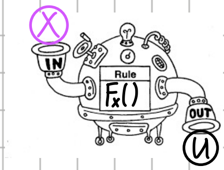
→ S9A2-6

2 nützliche Spezialfälle

Wenn man eine beliebige Zufallsvariable X mit ihrer Verteilungsfunktion F_X transformiert, ist die daraus entstehende Zufallsvariable $Y = F_X(X)$ gleichverteilt auf dem Intervall $(0,1)$:

$$\boxed{F_X(X) = U}$$

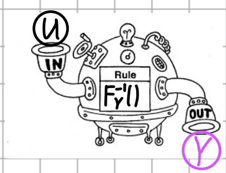
X beliebig
$r = F_X$
$Y = U$



Wenn man umgekehrt eine auf $(0,1)$ gleichverteilte Zufallsvariable X mit der Umkehrfunktion F_Y^{-1} einer beliebigen Verteilungsfunktion F_Y transformiert, hat die daraus entstehende Zufallsvariable $Y = F_Y^{-1}(X)$ die Verteilungsfunktion F_Y

$$\boxed{F_Y^{-1}(U) = Y}$$

$X = U$
$r = F_Y^{-1}$
Y beliebig

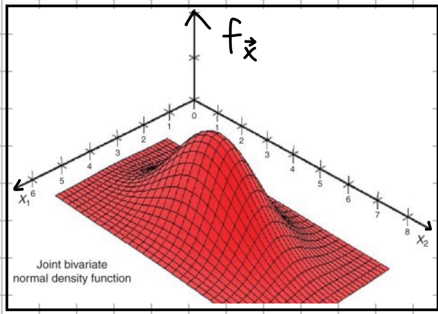


→ Zufallszahlengenerator mit beliebiger Verteilung

→ S9A1

Mehrdimensionale Zufallsvariablen

vektorwertige Zufallsvariable: $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$



gemeinsame Dichte

$$\boxed{\begin{aligned} f_{\vec{X}} : \mathbb{R}^n &\to [0,\infty) \\ \vec{t} &\mapsto f_{\vec{X}}(\vec{t}) = f_{\vec{X}}(t_1,\dots,t_n) = \frac{\partial^n F(x_1,\dots,x_n)}{\partial x_1 \cdots \partial x_n} \end{aligned}}$$

$\uparrow \frac{\partial^n}{\partial x_1 \cdots \partial x_n}$ $\quad \downarrow \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n}$

gemeinsame Verteilung

$$\boxed{\begin{aligned} F_{\vec{X}} : \mathbb{R}^n &\to [0,1] \\ \vec{x} &\mapsto F_{\vec{X}}(\vec{x}) = F_{\vec{X}}(x_1,\dots,x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\vec{X}}(t_1,\dots,t_n)\, dt_1 \cdots dt_n \end{aligned}}$$

$$\boxed{P(X_1 \leq x_1 \wedge \dots \wedge X_n \leq x_n) = F_{\vec{X}}(x_1,\dots,x_n)}$$

→ S10A2

Wahrsheinlichkeit für Ereignis $A \subset \mathbb{R}^n$

$$\boxed{P(\vec{X} \in A) = \int_A f_{\vec{X}}(\vec{t})\, d\vec{t}}$$ → S9A7-9

Normierung:

$$\boxed{P(\vec{X} \in \mathbb{R}^n) = \int_{\mathbb{R}^n} f_{\vec{X}}(\vec{t})\, d\vec{t} \overset{!}{=} 1}$$ → S11A2

Annäherung für kleines, an Stelle $\vec{x}_0$ zentriertes $A_{\vec{x}_0}$:

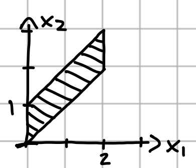
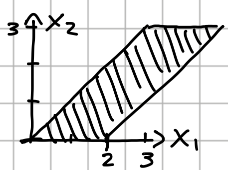
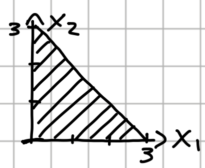
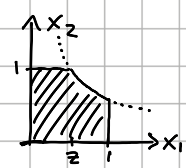
$$\boxed{P(\vec{X} \in A_{\vec{x}_0}) = \int_{A_{\vec{x}_0}} f_{\vec{X}}(\vec{t})\, d\vec{t} \approx \mathrm{Vol}_n(A_{\vec{x}_0}) \cdot f_{\vec{X}}(\vec{x}_0)}$$

$\hat{=}$ «Grundfläche · Höhe» falls $n=2$

Integralgrenzen finden

1. Grenzen des Ereignisses A bestimmen/skizzieren
2. Bereich finden/skizzieren wo $f_{\vec{X}}(\vec{x}) \neq 0$ ist
   → überall wo $f_{\vec{X}}(\vec{x}) = 0$ ist, ist auch Integral davon 0
3. Schnittmenge davon bilden (Bedingungen aus 1. und 2. kombinieren)
4. Integrationsreihenfolge bestimmen:
   - 4.1 äussere Integralgrenzen müssen konstant sein
   - 4.2 innere Grenzen dürfen Funktionen der äusseren Variabeln sein
     → bei Ecken/Sprüngen Integral aufteilen mit verschiedenen Funktionen → S9A9

häufige Situationen:

• $0 < x_1 < 2$ • $x_1 < x_2 < x_1 + 1$	• $0 < x_2 < 3$ • $x_2 < x_1 < x_2 + 2$	• $0 < x_1$ • $0 < x_2$ • $x_1 + x_2 < 3$	• $0 < x_1 < 1$ • $0 < x_2 < 1$ • $x_2 x_1 < z \Leftrightarrow x_2 < z/x_1$
			
$\int_0^2 \int_{x_1}^{x_1+1} \cdots \, dx_2\, dx_1$	$\int_0^3 \int_{x_2}^{x_2+2} \cdots \, dx_1\, dx_2$	$\int_0^3 \int_0^{-x_1+3} \cdots \, dx_2\, dx_1$	$\int_0^1 \int_0^{\min(1,\frac{z}{x_1})} \cdots \, dx_2\, dx_1 = \int_0^z \int_0^1 \cdots \, dx_2\, dx_1 + \int_z^1 \int_0^{\frac{z}{x_1}} \cdots \, dx_2\, dx_1$

Randverteilung

gegeben: Zufallsvektor $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \\ \vdots \\ X_n \end{pmatrix}$ mit gemeinsamer Dichte $f_{\vec{X}}(\vec{t}) = f_{\vec{X}}(t_1, \dots, t_n)$

Rand-Dichtefunktion → S9A10/11

$$\boxed{f_{X_k}(x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\vec{X}}(t_1, \dots, t_{k-1}, x_k, t_{k+1}, \dots, t_n)\, dt_1 \cdots dt_{k-1}\, dt_{k+1} \cdots dt_n}$$

Integral über alle anderen Komponenten

→ Grenzen dürfen x_k enthalten

Rand-Verteilungsfunktion

$$\boxed{F_{X_k}(x_k) = \int_{-\infty}^{x_k} f_{X_k}(t_k)\, dt_k}$$

Erwartungswert

$$E[\vec{X}] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix} = \begin{pmatrix} \int_{-\infty}^{\infty} t_1 f_{X_1}(t_1)\, dt_1 \\ \vdots \\ \int_{-\infty}^{\infty} t_n f_{X_n}(t_n)\, dt_n \end{pmatrix}$$

Erwartungswerte komponentenweise mithilfe der Rand-Dichte berechnen

Unabhängigkeit

Definition:

$\boxed{X_1, \dots, X_n \text{ sind unabhängig falls } f_{\vec{X}}(\vec{x}) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)}$

gemeinsame Dichte = Produkt der Rand-Dichten

→ S10A2-6

gleiches gilt auch für Verteilung ⇒ falls $X_1, \dots X_n$ unabhängig, gilt also:

$$\boxed{\begin{aligned} f_{\vec{X}}(\vec{x}) &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ F_{\vec{X}}(\vec{x}) &= F_{X_1}(x_1) \cdots F_{X_n}(x_n) \end{aligned}}$$

Verteilung von $X_{max} = max(X_1, \dots X_n)$ und $X_{min} = min(X_1, \dots X_n)$ von n unabhängigen und gleich verteilten (nicht im Sinne von Uniformverteilung!) Zufallsvariabeln $X_1, \dots, X_n$ mit der Verteilungsfunktion F_X:

$$\begin{aligned} F_{X_{max}}(x) &= P(max(X_1, \dots, X_n) \leq x) = P(X_1 \leq x \wedge \dots \wedge X_n \leq x) \\ &= P(X_1 \leq x) \cdots P(X_n \leq x) \\ &= (F_X(x))^n \end{aligned}$$

$$\begin{aligned} F_{X_{min}}(x) &= P(min(X_1, \dots, X_n) \leq x) = 1 - P(min(X_1, \dots, X_n) > x) = 1 - P(X_1 > x \wedge \dots \wedge X_n > x) \\ &= 1 - P(X_1 > x) \cdots P(X_n > x) \\ &= 1 - (1 - F_X(x))^n \end{aligned}$$

→ S10A2/A9.2, S13A1.3

bedingte Dichte von X_1 gegeben X_2 → S10A7/8

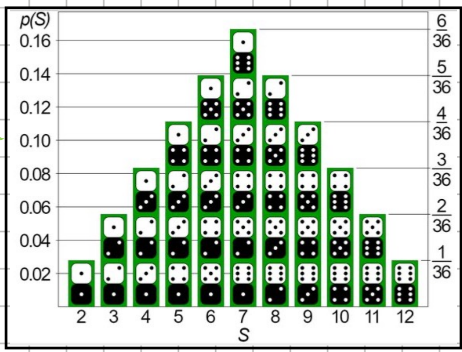
$$\boxed{f_{X_1}(x_1 \mid X_2 = x_2) = \frac{f_{\vec{X}}(x_1, x_2)}{f_{X_2}(x_2)} = \frac{f_{\vec{X}}(x_1, x_2)}{\int_{-\infty}^{\infty} f_{\vec{X}}(t_1, x_2)\, dt_1}}$$

Summen unabhängiger Zufallsvariabeln

gegeben: Summe $Y = X_1 + X_2$ von 2 unabh. Zufallsvariabeln

Wahrscheinlichkeitsfunktion (diskrete X_1, X_2)

$$\boxed{p_Y(y) = \sum_x p_{X_1}(x) \cdot p_{X_2}(y-x)}$$ diskrete Faltung



Dichtefunktion (stetige X_1, X_2)

$$\boxed{f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(x) \cdot f_{X_2}(y-x)\, dx}$$ Faltung

Exponentialverteilung

Summe von n unabhängigen, mit dem Parameter λ exponentiell verteilten Zufallsvariabeln:

$$\boxed{f_X(x) = \begin{cases} \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} & \text{falls } x > 0 \\ 0 & \text{sonst} \end{cases}}$$ gammaverteilt mit Parameter (n, λ)

-> S11A1

Normalverteilung

Summe von n unabhängigen, mit den Parametern (μ_i, σ_i) normalverteilten Zufallsvariabeln $X_1, \dots, X_n$ ist wieder normalverteilt mit Parametern $\tilde{\mu}$ und $\tilde{\sigma}$:

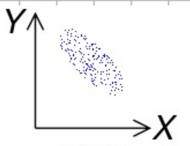
$$\boxed{Y = \sum_{i=1}^{n} X_i \sim \text{Normal}(\tilde{\mu}, \tilde{\sigma}) \quad \text{mit} \quad \tilde{\mu} = \sum_{i=1}^{n} \mu_i \quad \text{und} \quad \tilde{\sigma} = \sqrt{\sum_{i=1}^{n} \sigma_i^2}}$$

-> S10A11

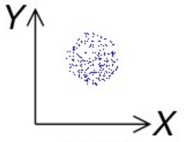
Kovarianz, Varianz, Korrelation

Kovarianz

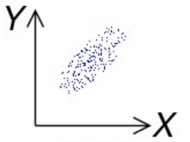
$$\boxed{cov(X_1, X_2) = \begin{cases} \sum_{x_1} \sum_{x_2} (x_1 - \mu_1)(x_2 - \mu_2)\, p_{\vec{X}}(x_1, x_2) & \text{diskret} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} (x_1 - \mu_1)(x_2 - \mu_2)\, f_{\vec{X}}(x_1, x_2)\, dx_1 dx_2 & \text{stetig} \end{cases}}$$



$cov(X,Y) < 0$



$cov(X,Y) \approx 0$



$cov(X,Y) > 0$

$$cov(X_1, X_2) = E[X_1 X_2] - E[X_1]E[X_2]$$ oft hilfreich → S11A2/4

$$cov(X_1, X_2) = cov(X_2, X_1)$$

$$cov(aX_1 + b\tilde{X}_1, X_2) = a\,cov(X_1, X_2) + b\,cov(\tilde{X}_1, X_2)$$ → S11A3/5

Korrelationskoeffizient

$$\boxed{\rho_{X_1X_2} = cov\left(\frac{X_1}{\sigma_1}, \frac{X_2}{\sigma_2}\right) = \frac{cov(X_1, X_2)}{\sigma_1 \sigma_2} \in (-1, 1)}$$ → S11A2

$$\boxed{X_1 \text{ und } X_2 \text{ unabhängig} \overset{\not\Leftarrow}{\Rightarrow} cov(X_1, X_2) = 0, \quad \rho_{X_1X_2} = 0}$$

unabhängig => unkorreliert, aber nicht umgekehrt

Varianz

$$\boxed{var(X) = cov(X, X)}$$

Summe $X_1, \ldots, X_n$ von n Zufallsvariabeln:

$$\boxed{var\left(\sum_{k=1}^{n} X_k\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} cov(X_i, X_j)}$$ → S11A5/6

bei paarweiser Unkorreliertheit vereinfacht sich die Summe zu:

$$\boxed{var\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} var(X_k)}$$

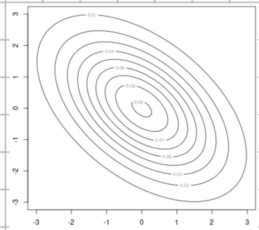
Summe von 2 Zufallsvariabeln:

$$\boxed{var(X_1 + X_2) = var(X_1) + 2cov(X_1, X_2) + var(X_2)}$$

Kovarianzmatrix

$$\underline{\Sigma} = \begin{pmatrix} cov(X_1, X_1) & \cdots & cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ cov(X_n, X_1) & \cdots & cov(X_n, X_n) \end{pmatrix}$$

mehrdimensionale Normalverteilung



tritt im Kontext von mehrdimensionalen Systemen sehr häufig auf

auch in n Dimensionen ist diese Verteilung vollständig durch Mittelwert und symmetrische Kovarianzmatrix bestimmt

ein n-dimensionaler normalverteilter Zufallsvektor $\vec{X}$ hat eine Dichte der Form

$$\boxed{f_{\vec{X}}(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}} \frac{1}{(\det(\underline{\Sigma}))^{\frac{1}{2}}} \exp\left(\frac{-1}{2}(\vec{x}-\vec{\mu})\underline{\Sigma}^{-1}(\vec{x}-\vec{\mu})\right) = \mathcal{N}_{\vec{\mu},\underline{\Sigma}}(\vec{x})}$$

für die Situation $n=1$ erhält man mit der Varianz $\underline{\Sigma} = \sigma^2$ die bekannte Gauss-Dichte

$$f_X(x) = \frac{1}{(2\pi)^{\frac{1}{2}}} \frac{1}{(\sigma^2)^{\frac{1}{2}}} \exp\left(\frac{-1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \mathcal{N}_{\mu,\sigma}(x)$$

der Ausdruck $(\vec{x}-\vec{\mu})\underline{\Sigma}^{-1}(\vec{x}-\vec{\mu})$ ist eine quadratische Form und misst das Quadrat der durch $\underline{\Sigma}$ definierten Distanz zwischen zwischen $\vec{x}$ und $\vec{\mu}$, wobei für $\underline{\Sigma} = \underline{E}_n$ genau die euklidische Distanz resultiert

somit ist die Dichte $f_{\vec{X}}(\vec{x})$ konstant auf der Menge aller Punkte $S_d = \{\vec{x} \in \mathbb{R}^n \mid (\vec{x}-\vec{\mu})\underline{\Sigma}^{-1}(\vec{x}-\vec{\mu}) = d^2\}$, welche die Distanz d zum Mittelwert haben

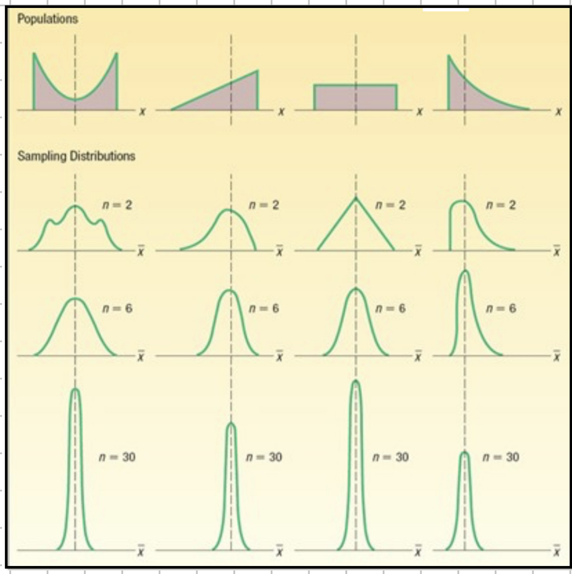
diese Mengen S_d sind Ellipsoide mit denselben im Mittelwert $\vec{\mu}$ zentrierten Hauptachsen

Grenzwertsätze

i.i.d. (independent and identically distributed):
Zufallsvariabeln $X_1, X_2, \ldots, X_n$, welche unabhängig sind und dieselbe Verteilung haben

Summe von i.i.d. Zufallsvariabeln S_n

$$\boxed{S_n = X_1 + \cdots + X_n}$$ neue Zufallsvariable



Mittelwert einer i.i.d Folge

$$\boxed{\overline{X}_n = \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}}$$ Zufallsvariable, keine Zahl!

Erwartungswert

$$\boxed{E[\overline{X}_n] = E[X_k]}$$ → S12A4

Varianz

$$\boxed{var(\overline{X}_n) = \frac{var(X_k)}{n} \neq var(X_k)}$$ → S12A4

Gesetz d. grossen Zahlen

$$\boxed{P(|\overline{X}_n - E[X_k]| \geq \epsilon) \leq \frac{var(X_k)}{n\epsilon^2}}$$

starkes Gesetz d. grossen Zahlen

$$\boxed{P(\lim_{n\to\infty} \overline{X}_n = E[X_k]) = 1}$$

zentraler Grenzwertsatz

$$\boxed{P\left(\frac{S_n - n \cdot E[X_k]}{\sqrt{n \cdot var(X_k)}} \leq x\right) \xrightarrow[n\to\infty]{} \Phi_{0,1}(x)}$$

(normalisierte) Summe von i.i.d. X_k konvergiert gegen Standardnormalverteilung

Näherungsformeln allgemein:

$$P(S_n \leq x) \approxeq \Phi_{n\mu, \sqrt{n}\sigma}(x)$$
$$P(\overline{X}_n \leq x) \approxeq \Phi_{\mu, \frac{\sigma}{\sqrt{n}}}(x)$$

→ S11A10

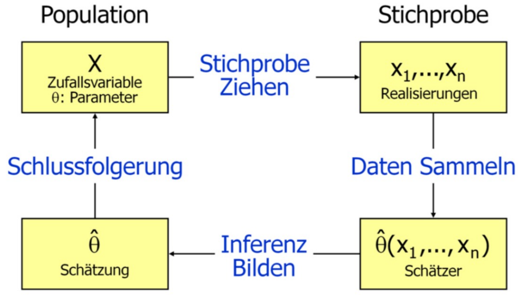
Näherung für binomisch n, p verteilte Zufallsvariable X (mit Histogrammkorrektur):

$$P(X \leq m) \approxeq \Phi_{0,1}\left(\frac{m + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

$$P(X \geq m) \approxeq 1 - \Phi_{0,1}\left(\frac{m - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$ → S11A7/8

$$P(X = m) \approxeq 2\Phi_{0,1}\left(\frac{m + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - 1$$ → S11A7/9

Schätzen von Parametern



Stichprobe

Folge $X_1, \dots, X_n$ von n i.i.d Zufallsvariabeln

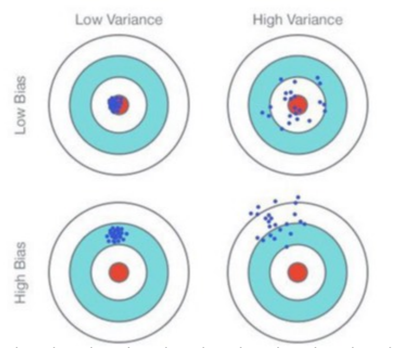
Wenn man eine Stichprobe zieht / realisiert, erhält man n unabhängige Realisierungen $x_1, \dots, x_n$ der Zufallsvariable X

Punktschätzer

Vorschrift T, mit welcher man aus einer Stichprobe $X_1, \dots, X_n$ eine Schätzung für einen Parameter θ berechnen kann

bias

$$\boxed{\text{bias}(T) = E[T] - \theta}$$ → S12A4



- misst erwartete Verzerrung des Schätzers T
- falls $\text{bias}(T) = 0$ nennt man T erwartungstreu

mse (mean squared error)

$$\boxed{\text{mse}(T) = E\left[(T-\theta)^2\right] = \text{var}(T) + (\text{bias}(T))^2}$$

- misst erwartatete quadratische Abweichung des Schätzers T
  → Mass für Güte eines Schätzers → S12A2
- entspricht bei erwartungstreuen Schätzern der Varianz des Schätzers → S12A1

wichtige Formel Varianz (Repetition)

$$\boxed{\begin{aligned} \text{var}(X) &= E[X^2] - (E[X])^2 \\ E[X^2] &= \text{var}(X) + (E[X])^2 \end{aligned}}$$

→ S12A3

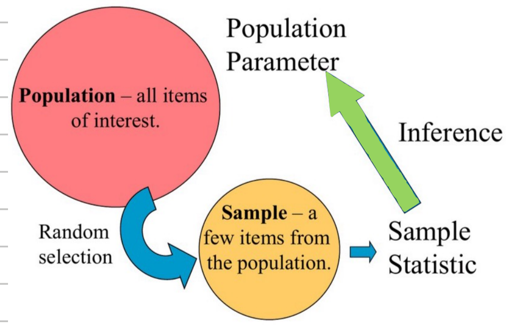
→ S12A4

Schätzer für μ, σ^2, σ

	Population		Sample
# of subjects	N		n
Mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	←	$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$	←	$S^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$
Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}$	←	$S = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}$

Note: S^2 is the formula for unbiased sample variance, since we're dividing by $n-1$. → S12A3

Note: Finding S by taking $\sqrt{S^2}$ reintroduces bias. → S12A3



Schätzer bestimmen

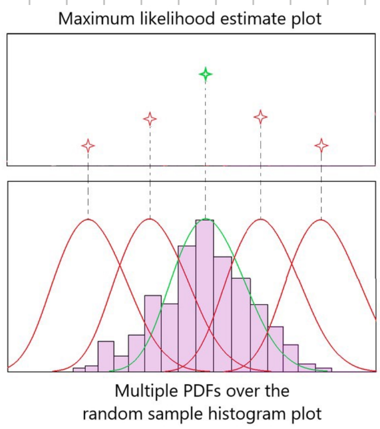
gegeben: n Werte $x_1, \ldots, x_n$ (Stichprobe)

vorgegeben: W'keits(dichte)-Funktion $f_X(x \mid \theta)$ mit unbekanntem Parameter θ

gesucht: Parameter θ, damit vorgegebene Funktion $f_X(x)$ bestmöglich zu Stichprobe passt

Maximum-Likelihood-Methode

1. $L(\theta) = f_X(x_1 \mid \theta) \cdots f_X(x_n \mid \theta) = \prod_{k=1}^{n} f_X(x_k \mid \theta)$



Multiple PDFs over the random sample histogram plot

2. evt: $\ln(L(\theta))$ falls vereinfacht

3. Kandidaten für θ

3.1 Nullstelle der Ableitung -> S12A5-7.1
   - a. $\frac{d}{d\theta}$ (ableiten nach θ)
   - b. Ableitung $\stackrel{!}{=} 0$ setzen
   - c. nach θ auflösen
   - d. evt: falls mehrere Nullstellen:
     -> einsetzen und θ so wählen, dass $L(\theta)$ möglichst gross

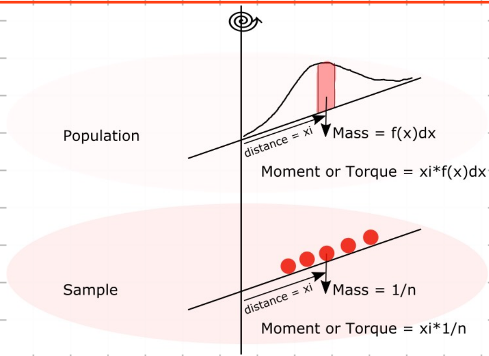
3.2 kleinster und grösster Wert der Stichprobe -> S12A5, S13A1.2
   - a. $\theta = \max(x_1, \ldots, x_n)$ und $\theta = \min(x_1, \ldots, x_n)$ setzen
   - b. prüfen, ob dies $L(\theta)$ maximiert

4. θ so wählen, dass $L(\theta)$ maximal wird

Momenten-Methode -> S12A7.2, S13A1.1

1. Mittelwert der Zufallsvariable X

$$\mu(\theta) = E[X] = \int_{-\infty}^{\infty} x \cdot f(x, \theta)\, dx$$



2. Stichprobenmittelwert

$$\overline{x} = \frac{1}{n}(x_1 + \cdots + x_n) = \frac{1}{n} \sum_{j=1}^{n} x_j$$

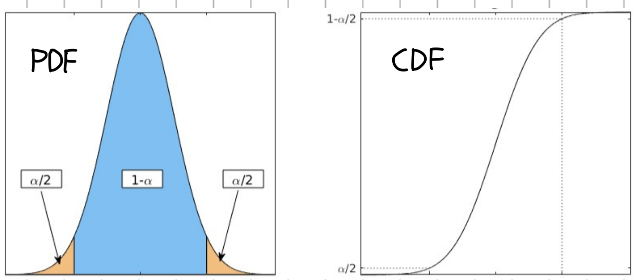
3. Formeln aus Schritt 1. und 2. gleichsetzen

$$\mu(\theta) \stackrel{!}{=} \overline{x}$$

4. Gleichung nach θ auflösen

Konfidenzintervall für Mittelwert μ

$P(\mu \in KI) = 1-\alpha$



Konfidenzintervall bestimmen

gegeben:

$\bar{x}$ empirischer Mittelwert
σ Standard-Abweichung
n Umfang der Stichprobe
α Fehlerwahrscheinlichkeit

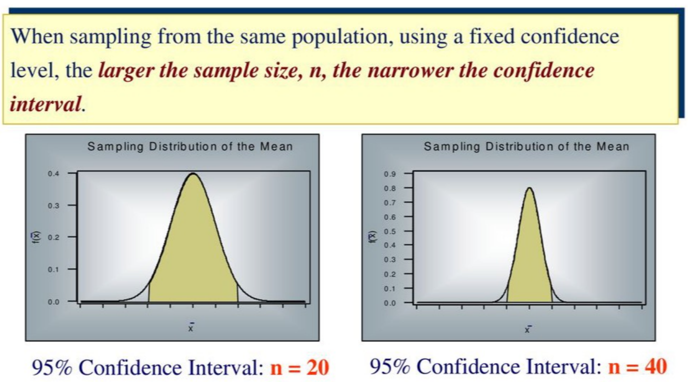
gesucht:

KI Konfidenzintervall für μ zu $1-\alpha$ –> S13A2

$z_{1-\frac{\alpha}{2}} = \Phi_{0,1}^{-1}\left(1-\frac{\alpha}{2}\right) \stackrel{TR}{=}$ **solve(normCdf(-∞,x,0,1)=1-α/2,x)**

$KI = \left[\bar{x} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$

Umfang n der Stichprobe bestimmen



gegeben:

σ Standard-Abweichung
L Länge des Konfidenzintervalls
α Fehlerwahrscheinlichkeit

gesucht:

n Umfang der Stichprobe –> S13A2.4

$z_{1-\frac{\alpha}{2}} = \Phi_{0,1}^{-1}\left(1-\frac{\alpha}{2}\right) \stackrel{TR}{=}$ **solve(normCdf(-∞,x,0,1)=1-α/2,x)**

$n = 4\left(\frac{\sigma}{L} z_{1-\frac{\alpha}{2}}\right)^2$