

# Zusammenfassung WAHR / R-Studio

## Zufallsexperiment

Zufallsexperiment ist definiert als Versuch/eine Situation, wo das Ergebnis nicht deterministisch vorbestimmt ist.

- Zufallsexperimente können (zumindest theoretisch) unter gleichen Bedingungen wiederholt werden.
- Um zu entscheiden, ob ein Versuch / eine Situation ein Zufallsexperiment ist, fragt man sich, ob sich bei einer Wiederholung immer exakt dasselbe Resultat einstellen würde. Ist die Antwort "nein" → Zufallsexperiment!

## Ergebnis

Wir betrachten Versuche, deren Ausgänge zufällig eintreten. Die konkreten Ausgänge werden Ergebnisse (outcome) genannt. Es tritt bei jedem Experiment **genau ein Ergebnis** ein. **Ergebnisse schliessen sich aus.**

1 x Würfelwurf: 2                                      FC Winterthur vs. FC Zürich: 1:0                                      Geburtsmonat: März

## Ergebnisraum / Wahrscheinlichkeitsraum $\Omega$

Die **Menge** aller möglichen Ergebnisse heisst Ergebnisraum und wird mit dem Symbol  $\Omega$  dargestellt. Die Ergebnisse sind Elemente von  $\Omega$ .

1 x Würfelwurf:  $\Omega = \{1-6\}$                                       Fussballmatch:  $\Omega = \{0:0, 0:1, \dots\}$                                       Geburtsmonat:  $\Omega = \{\text{Jan} - \text{Dez}\}$

## Mächtigkeit (Anzahl der Elemente) $|\Omega|$

Mit  $|\Omega|$  bezeichnet man die Mächtigkeit (Anzahl der Elemente) einer Menge, konkret die Mächtigkeit des Ergebnisraums  $\Omega$ . Die Ergebnismenge kann **endlich**, **abzählbar-unendlich** und **überabzählbar unendlich** sein.

1 x Würfelwurf:  $|\Omega| = 6$                                       Fussballmatch:  $|\Omega| = \infty$                                       Geburtsmonat:  $|\Omega| = 12$

## Ereignis

Als Ereignis (event) bezeichnet man eine Konstellation, die unter Umständen von mehreren Ergebnissen erfüllt wird. Es handelt jeweils um eine Teilmenge A von  $\Omega$  ( $A \subseteq \Omega$ ), die eines oder mehrere Elemente enthält.

Ein **Ereignis** gilt als eingetroffen, wenn ein passendes **Ergebnis** aus dem Zufallsexperiment resultiert.

- 1 x Würfeln:  $A = \text{"eine gerade Zahl zu würfeln"} = \{2, 4, 6\}$
- Fussballmatch:  $A = \text{"Winterthur gewinnt"} = \{1:0, 2:0, 2:1, \dots\}$

Ein **Elementarereignis** ist ein Ereignis, welches nur ein einziges Ergebnis enthält. Würfelwurf → einzige Augenzahl

## Ereignis vs. Ergebnis

Fussballmatch: FCW - FCZ ist 0:2 (Ergebnis)                                      Fussballmatch: FCW gewinnt gegen FCZ (Ereignis)

Ziel: Angabe von einer Wahrscheinlichkeit für das Eintreffen eines **Ereignisses** (und nicht eines Ergebnisses).

## Mengenoperationen

### Spezielle Ereignisse

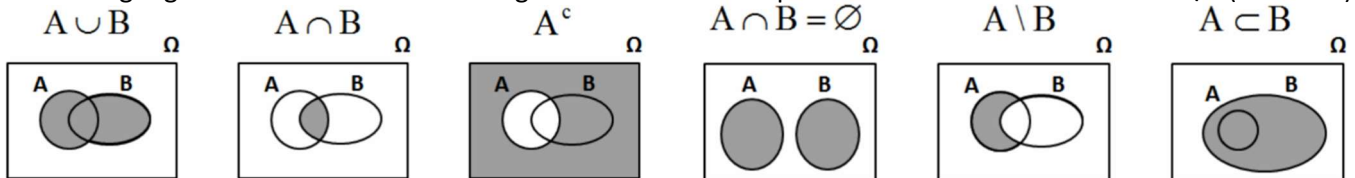
- Elementarereignisse:  $\{\omega_1, \omega_2, \dots\}$  enthalten nur ein Ereignis
- Gegenereignis von A:  $A^c$  mit  $A \cup A^c = \Omega$
- Sicheres Ereignis  $A = \Omega$
- Unmögliches Ereignis  $A = \emptyset$

Beispiele:

- Elementarereignisse bei 1 x Würfeln:  $\omega_1 = \{1\}$ ,  $\omega_2 = \{2\}$ ,  $\omega_3 = \{3\}$ ,  $\omega_4 = \{4\}$ ,  $\omega_5 = \{5\}$  und  $\omega_6 = \{6\}$
- Sicheres Ergebnis vom Treibstoffverbrauch:  $A = \text{"Es wird mehr als 0t verbraucht."}$
- Unmögliches Ereignis beim Würfeln:  $A = \text{"Augenzahl ist 7"}$

Mengenoperationen für die Ereignisse A und B

- Vereinigung:  $A \cup B$
- Schnittmenge:  $A \cap B$
- Komplement:  $A^c$
- Differenz:  $A \setminus B$  (A ohne B)



Beispiel Würfeln A = "Gerade Augenzahl"  $\{2,4,6\}$  und B = "Augenzahl kleiner 4"  $\{1,2,3\}$

Schnittmenge  $A \cap B = \{2\}$ , Vereinigung  $A \cup B = \{1, 2, 3, 4, 6\}$ , Komplement  $A^c = \{1, 3, 5\}$  und Differenz  $A \setminus B = \{4, 6\}$

Ergebnisse A und B heissen gegenseitig ausschliessend oder **disjunkt**, wenn ihre Schnittmenge leer ist:  $A \cap B = \emptyset$

Ergebnis A heisst Teilmenge von B, wenn alle Ergebnisse von A auch in B enthalten sind.

## Wahrscheinlichkeiten

Wir müssen jedem Ereignis eine Wahrscheinlichkeit zuordnen. Dafür gibt es die folgenden Möglichkeiten:

1. Modell (Theorie) basiert
2. Daten basiert (frequentistisch)
3. Mischform Modell und Daten
4. Subjektive Einschätzung

### Modell basiert: Laplace Wahrscheinlichkeiten

Einfache Möglichkeit, wenn man die Zufallsmechanismus kennt, bzw. (genügend) genau versteht. Wenn **alle Elementarereignisse gleich wahrscheinlich sind** und der Ergebnisraum endlich ist, so können wir die Laplace-Formel verwenden:  $P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl "günstige" Fälle}}{\text{Anzahl "mögliche" Fälle}}$  **Beispiel:** Würfel  $P(A = \text{geraden Zahlen}) = \frac{|2,4,6|}{|1,2,3,4,5,6|} = \frac{3}{6} = 0.5$

### Daten basiert: Frequentistische Wahrscheinlichkeit

Für viele Situationen ist kein Modell über den Zufallsmechanismus verfügbar. Wir können die Wahrscheinlichkeit aber aufgrund vergangener Beobachtungen abschätzen. Dazu führen wir das Zufallsexperiment n-Mal durch und zählen wir oft das Ereignis A Eintritt:  $P(A) = r_n(A) = \frac{n_A}{n}$  wobei  $r_n(A) = \text{relative Häufigkeit}$   
Wir sind nicht mehr auf Elementarereignisse mit gleicher Wahrscheinlichkeit reduziert.

**Beispiel:** Würfeln → Wie häufig erhält man mit 2 Würfeln eine Augenzahl grösser 10? Ereignis:  $A = \{11, 12\}$   
Berechnung der Wahrscheinlichkeit:  $r_n(\{11,12\}) = \frac{n_{A=\{11\}} \cup n_{A=\{12\}}}{n} = \frac{n_{A=\{11\}} + n_{A=\{12\}}}{n} = r_n(\{11\}) + r_n(\{12\})$   
Dabei haben wir ausgenutzt, dass die Augensumme entweder 11 oder 12 ist. Allgemein gilt, falls sich die Ereignismengen  $A_1$  und  $A_2$  gegenseitig ausschliessen (d.h.  $A_1 \cap A_2 = \emptyset$ ):  $r_n(A_1 \cup A_2) = r_n(A_1) + r_n(A_2)$

Frequentistische Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit:  $P(A) = \lim_{n \rightarrow \infty} r_n(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$   
 $n$  = Anzahl Versuche (Zufallsexperimente)  $r_n$  = relative Häufigkeit  
 $n_A$  = Anzahl Versuche in denen Ereignis A eintritt

### Mischform: Daten und Modell basiert

Vielmals verwendet man eine Mischform aus Modell und Daten.

1. Plausibles Modell, das bis auf einige Parameter bestimmt ist
2. Parameter aus den empirischen Daten abgeschätzt

Wir machen Annahmen über die Verteilung und schätzen die Parameter aus den Daten.

### Subjektive Wahrscheinlichkeit

Wahrscheinlichkeit = Mass des persönlichen Glaubens (Bayesianische Interpretation).

### Axiome von Kolmogorov - Eigenschaften der Wahrscheinlichkeit

Nach Kolmogorov gelten für Wahrscheinlichkeiten 3 Axiome. Jedem Ereignis  $A \subset \Omega$  wird eine Zahl  $P(A)$  mit folgenden Eigenschaften zugeordnet:

- Axiom 1:  $P(A) \geq 0$  für jedes Ereignis (A) (Nichtnegativität)
- Axiom 2:  $P(\Omega) = 1$  (Normierung)
- Axiom 3: Falls  $A \cap B = \emptyset$ , dann gilt  $P(A \cup B) = P(A) + P(B)$  (Additivität)

Frequentistische, Laplace-Definition und Bayesianische Interpretation erfüllen diese Axiome.

Ereignisse A mit  $P(A) = 0$ , die aber dennoch eintreten können, nennt man **fast unmögliche Ereignisse**.  
Ereignisse A mit  $P(A) = 1$ , aber endlich viele Ausnahmen haben, nennt man **fast sicheren Ereignissen**.

### Laplace

- Axiom 1:  $P(A) = \frac{|A|}{|\Omega|}$  da  $|A| \geq 0$  und  $|\Omega| > 0$  (mindestens ein Ergebnis), ist  $P(A) \geq 0$ .
- Axiom 2:  $P(\Omega) = \frac{|\Omega|}{|\Omega|} = 1$  Axiom 3: Falls  $A \cap B = \emptyset$ , dann gilt  $P(A \cup B) = P(A) + P(B)$

### Relative Häufigkeiten

- Axiom 1:  $P(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$ . Es ist  $n_A \geq 0$  und  $n > 0$ , somit  $P(A) \geq 0$ .
- Axiom 2:  $P(\Omega) = \lim_{n \rightarrow \infty} \frac{n_\Omega}{n} = \lim_{n \rightarrow \infty} \frac{n}{n} = 1$  Axiom 3:  $P(A \cup B) = \lim_{n \rightarrow \infty} \frac{n_{A \cup B}}{n} = \lim_{n \rightarrow \infty} \frac{n_A + n_B}{n} = P(A) + P(B)$

Aus den Axiomen lassen sich folgende Rechenregeln der Wahrscheinlichkeit herleiten:

- |   |  |
|---|--|
| <ol style="list-style-type: none"><li>1. <math>P(A) + P(A^c) = 1</math> (Ereignis + Gegenereignis = 1)</li><li>2. <math>P(\emptyset) = 1 - P(\Omega) = 0</math></li><li>3. <math>P(A \cup B) = P(A) + P(B) - P(A \cap B)</math> (Additionssatz)</li><li>4. Wenn <math>A \subseteq B</math>, so gilt <math>P(A) \leq P(B)</math> (A liegt innerhalb von B oder gleich B)</li></ol> | <ol style="list-style-type: none"><li>5. Für jedes Ereignis A gilt <math>0 \leq P(A) \leq 1</math></li><li>6. <math>P(A_1 \cup A_2 \cup \dots \cup A_k) = P(A_1) + P(A_2) + \dots + P(A_k)</math>, falls <math>A_i \cap A_j = \emptyset</math> für alle i, j, paarweise disjunkt</li></ol> |
|---|--|

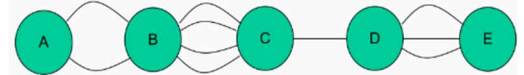
Kombinatorik

Gesucht ist die Anzahl Möglichkeiten **eine Auswahl von k Objekten** (Stichprobe vom Umfang k) aus **insgesamt n Objekten** anzuordnen bzw. auszuwählen. **R**:  $n! = \text{factorial}(n)$  und  $\binom{n}{k} = \text{choose}(n, k)$

	Ohne Wiederholung/ Zurücklegen	Mit Wiederholung/Zurücklegen
Auswahl: k = n Reihenfolge wichtig	Permutation <b>ohne</b> Wiederholung $z = n!$	Permutation <b>mit</b> Wiederholung $z = \frac{n!}{n_1! * \dots * n_s!}$
Auswahl: k Reihenfolge wichtig	geordnete Stichprobe <b>ohne</b> Zurücklegen (Es gilt: $k < n$ ) $z = \frac{n!}{(n-k)!}$	geordnete Stichprobe <b>mit</b> Zurücklegen (k beliebig) $z = n^k$
Auswahl: k Reihenfolge unwichtig	ungeordnete Stichprobe <b>ohne</b> Zurücklegen (Es gilt: $k < n$ ) $z = \binom{n}{k}$	ungeordnete Stichprobe <b>mit</b> Zurücklegen (k beliebig) $z = \binom{n+k-1}{k}$

Produktregel - `epand.grid()`

Wie viele Wege gibt es von A nach E?



In jeder Stufe ist die Anzahl der Wegstücke unabhängig von der jeweiligen Wahl in den anderen Stufen.

Gesamtanzahl an möglichen Wegen:  $n_{AB} * n_{BC} * n_{CD} * n_{DE} = 2 * 4 * 1 * 3 = 24$

Die Anzahl Möglichkeiten berechnet sich als Produkt der Möglichkeiten pro Stufe.

Es gibt 6 Chauffeure, 4 Zugfahrzeuge und 8 Anhänger. Wie viele Kombinationen gibt es?  $z = 6 * 4 * 8 = 192$

Permutationen

Die Anzahl der Elemente ist vorgegeben und vorhanden. Fragestellung: Auf wie viele Arten kann man die Elemente vertauschen. Die Reihenfolge der eindeutig identifizierbaren Elemente ist wichtig. Gesucht: Anzahl Anordnungen.

Permutation **ohne** Wiederholung - `library(combinat) → permn()`

Es gibt 4 Flugzeuge und 4 Gates. Wie viele Möglichkeiten gibt es, die Flugzeuge an die Gates zu stellen?

Das 1. Flugzeug hat 4 Möglichkeiten. Das 2. Flugzeug hat 3 Möglichkeiten. Das 3. Flugzeug hat 2 Möglichkeiten.

Das 4. Flugzeug muss das übrig gebliebene Gate nehmen:  $z = 4 * 3 * 2 * 1$ .

Allgemein: Anzahl Möglichkeiten n verschiedene Objekte anzuordnen:

$z = n!$  Es gilt  $1! = 1$  und  $0! = 1$ . (Spezialfall Produktregel: Objekt j hat  $n_j = n - j + 1$  Optionen)

Permutation **mit** Wiederholung - `library(combinat) → permn() + unique()`

Permutationen von ERBSE. Wie viele Möglichkeiten gibt es, die Buchstaben anzuordnen?

- Wären es 5 verschiedene Buchstaben gäbe  $n! = 5! = 120$  Möglichkeiten.
- Der Buchstaben E kommt zweimal vor. Entsprechende Kombinationen sind identisch. Deshalb teilt man durch die nicht interessierenden Permutationen  $\rightarrow 2! = 2$  "identische" Möglichkeiten für E.  $z = \frac{5!}{2!} = \frac{120}{2} = 60$

Allgemein: Anzahl Möglichkeiten n Objekte anzuordnen, wobei je  $n_1, \dots, n_s$  identisch sind:  $z = \frac{n!}{n_1! * \dots * n_s!} = \frac{n!}{n_1! * \dots * n_s!}$

Ein Zug besteht aus 4 Wagen der 1. Klasse, 7 Wagen der 2. Klasse, 1 Speisewagen, 2 Gepäckwagen.

- Wie viele unterschiedliche Wagenfolgen sind möglich?  $\frac{14!}{4! * 7! * 1! * 2!} = 360'360$
- Wie viele unterschiedliche Wagenfolgen sind möglich, wenn die 1. Klass-Wagen nicht getrennt werden dürfen?  
Wir betrachten die 4 Wagen der 1. Klasse als 1 Element:  $\frac{11!}{1! * 7! * 1! * 2!} = 3'960$
- Was ist die Wahrscheinlichkeit bei einer zufälligen Anordnung der Wagen, dass alle Wagen der 1. Klasse zusammen sind?  $P(\text{"1. Klasse zusammenhängend"}) = \frac{3'960}{360'360} = 0.011$

Geordnete Stichprobe **ohne** Zurücklegen - `factorial()`

In einer Urne befinden sich  $n = 6$  verschieden farbige Smarties. Wir ziehen 3 Mal ein Smartie, notieren die Farbe und legen es nicht zurück. Die Farbenreihenfolge ist uns wichtig. Wie viele unterschiedliche Ziehungen gibt es? Beim ersten Zug gibt es 6 Möglichkeiten. Bei jedem weiteren Zug verkleinert sich die Anzahl möglicher Smarties, um 1:  $z = 6 * 5 * 4 = 120$

$z = \frac{n!}{(n-k)!}$  (Spezialfall Produktregel: Objekt j hat  $n_j = n - j + 1$  Optionen.)

An der Schweizer Meisterschaft nehmen  $n = 23$  Fahrer teil. Wie viele Möglichkeiten gibt es, das Podest, d.h. die ersten  $k = 3$  Plätze, zu besetzen?  $z = \frac{23!}{(23-3)!} = \frac{23!}{20!} = 23 * 22 * 21 = 10'626$

Geordnete Stichprobe **mit** Zurücklegen - `sample()`

In einer Urne sind  $n = 6$  verschieden farbige Smarties. Wir ziehen  $k = 3$  Mal ein Smartie, notieren die Farbe und legen das Smartie jedes Mal wieder zurück. Die Farbenreihenfolge ist uns wichtig. Wie viele unterschiedliche Ziehungen gibt es? Mit der Produktregel erhalten wir:  $z = 6 * 6 * 6 = 6^3 = 216$

Beispiel: Bei einem Kombinationsschloss sind die einzelnen Einstellungen durch 3-ziffrige Zahlen mit Ziffern aus 0 bis 9 möglich. Wie viele mögliche Einstellungen gibt es?  $z = 10^3 = 1'000$

Wie gross ist die Wahrscheinlichkeit, dass jemand beim ersten Versuch die richtige Kombination findet?

$$\Rightarrow P(A) = P(\text{"1x Erfolgreich"}) = \frac{1}{1'000}$$

Wie gross ist die Wahrscheinlichkeit, dass jemand (ohne Gedächtnis) durch 3x Probieren die Kombination findet?

Hier ist es leichter die Gegenwahrscheinlichkeit zu bestimmen:

Anzahl günstige Fälle (d.h. 3x Nicht Finden):  $999^3$  und Anzahl mögliche Fälle:  $1'000^3$

$$\Rightarrow P(B) = 1 - P(\text{"3x Nichtfinden"}) = 1 - \left(\frac{999}{1'000}\right)^3 \approx 0.003$$

Wahrscheinlichkeit, dass Schloss in 1000 Versuchen zu öffnen  $\Rightarrow P(C) = 1 - \left(\frac{999}{1000}\right)^{1000} \approx 0.632$  und nicht 1!

Wie gross ist die Wahrscheinlichkeit, dass jemand (mit Gedächtnis) durch 3x Probieren die Kombination eines Zahlenschloss findet?

Anzahl günstiger Fälle (3x Nicht Finden):  $999 * 998 * 997$  und Anzahl mögliche Fälle:  $1'000 * 999 * 998$

$$\Rightarrow P(D) = 1 - P(\text{"3x Nichtfinden"}) = 1 - \frac{999*998*997}{1'000*999*998} = 1 - \frac{997}{1'000} \approx 0.003$$

Ungeordnete Stichprobe **ohne** Zurücklegen - `choose()` / `combn()`

In einer Urne sind  $n = 6$  verschieden farbige Smarties. Wir ziehen  $k = 3$  Mal ein Smartie, notieren die Farbe und legen das Smartie nicht zurück. Uns ist wichtig, welche Farben gezogen wurden, aber nicht in welcher Reihenfolge. Wie viele unterschiedliche Ziehungsergebnisse gibt es? Es gäbe  $z = 6 * 5 * 4 = 120$  Ziehungsmöglichkeiten. Jedoch müssen identische Ziehungen eliminiert werden. Identisch = rot, blau, gelb = rot, gelb, blau = blau, rot gelb.  $\rightarrow 3!$

$$z = \frac{6*5*4}{3!} = 20 \quad z = \frac{n!}{(n-k)! k!} = \binom{n}{k} \text{ heisst Binomialkoeffizient und «n tief k» oder «k aus n» gesprochen}$$

Aus Schachtel mit 9 verschiedenen Pralines entnimmt man 3 Stück zum Testen:  $z = \frac{9!}{(9-3)!3!} = \frac{9!}{6!3!} = \binom{9}{3} = 84$

Ungeordnete Stichprobe **mit** Zurücklegen – `library(partitions)`  $\rightarrow$  `compositions()`

In einer Urne sind  $n = 6$  verschieden farbige Smarties. Wir ziehen  $k = 3$  Mal ein Smartie, notieren die Farbe und legen das Smartie wieder zurück. Uns ist wichtig, welche Farben gezogen wurden, aber nicht in welcher Reihenfolge. Wie viele unterschiedliche Ziehungsergebnisse gibt es?  $z = \frac{(n+k-1)!}{(n-1)!k!} = \binom{n+k-1}{k}$

Wie viele Möglichkeiten gibt es, aus 6 Früchtesorten einen Korb mit 20 Früchten zusammenstellen? Hier handelt es sich um eine Auswahl von  $k = 20$  Früchten aus  $n = 6$  Sorten  $z = \frac{(6+20-1)!}{(6-1)!20!} = \frac{(25)!}{5!20!} = \binom{6+20-1}{20} = 53'130$

## Ereignisbäume (Baumdiagramm)

Hilft bei mehrstufigen Zufallsexperimenten. Ein Pfad des Ereignisbaumes, d.h. Durchlaufen der Kanten von der Wurzel bis zu einem Endknoten, entspricht einem Elementarereignis des mehrstufigen Zufallsexperimentes. Wahrscheinlichkeit an den Endknoten: Multiplikation der Wahrscheinlichkeiten entlang des Pfades (Produktregel).

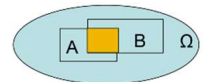
## Bedingte Wahrscheinlichkeit

Interesse für Wahrscheinlichkeit Ereignis A, wenn schon bekannt ist, dass ein anderes Ereignis B eingetreten ist.

• A = Herzinfarkt, wenn B = Bluthochdruck

Seien A und B zwei Ereignisse mit  $P(B) \neq 0$  dann ist die bedingte Wahrscheinlichkeit (conditional probability) von A

gegeben B definiert als:  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  Man sagt Wahrscheinlichkeit von A gegeben B.



### Interpretation

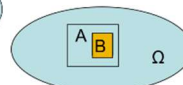
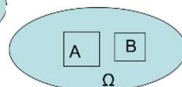
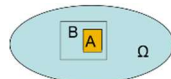
Wenn gegeben ist, dass das Ereignis B eingetreten ist, geht man vom Wahrscheinlichkeitsraum  $\Omega$  zu einem kleineren Wahrscheinlichkeitsraum B über. Dadurch werden alle Ergebnisse, die in B liegen, wahrscheinlicher und alle Ergebnisse, die nicht in B liegen, unmöglich.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{"günstige"}}{\text{"möglich"}}$$

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

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{P(B)} = 0$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$$



A = "gerade Zahl" = {2, 4, 6}

$A^C$  = "ungerade Zahl" = {1, 3, 5}

B = "geworfene Zahl  $\leq 3$ " = {1, 2, 3}

Wie gross ist die Wahrscheinlichkeit für A und  $A^C$ , wenn B eingetroffen ist?

$$A \cap B = \{2\} P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{|A \cap B|}{|B|} = \frac{1}{3}$$

$$A^C \cap B = \{1, 3\} P(A^C|B) = \frac{P(A^C \cap B)}{P(B)} = \frac{|A^C \cap B|}{|B|} = \frac{2}{3}$$

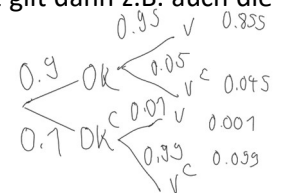
Bedingte Wahrscheinlichkeiten genügen auch den drei Wahrscheinlichkeitsaxiomen. Damit gilt dann z.B. auch die Formel:  $P(A^c|B) = 1 - P(A|B)$

Hingegen ist im Allgemeinen:  $P(A|B^c) \neq 1 - P(A|B)$

$P(\text{"Sechs geworfen"} | \text{"Augenzahl gerade"}) = 1/3$

$P(\text{"Keine Sechs geworfen"} | \text{"Augenzahl gerade"}) = 2/3$

$P(\text{"Sechs geworfen"} | \text{"Augenzahl ungerade"}) = 0$



90% der Schrauben sind ok. Von diesen werden 95% verkauft. Von den kaputten Schrauben werden 1% verkauft.

- Schraube ist verkauft und ok  $\rightarrow 0.855$
- Schraube ist verkauft  $\rightarrow 0.856$
- Eine verkaufte Schraube ist ok  $\rightarrow \frac{P(v \cap OK)}{P(v)} = \frac{0.855}{0.856}$

	OK	Nicht OK	
Verkauft	0.855	0.001	0.856
Nicht verkauft	0.045	0.099	0.144
	0.900	0.100	1.000

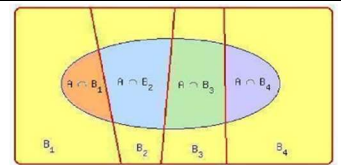
Für jedes Ereignis A gilt  $P(A \cup A^c) = P(A) + P(A^c) = 1$ . Es gilt immer:  $P(A \cap B) \leq P(A)$ .

Es gilt immer:  $P(A \cup B) + P(A \cap B) = P(A) + P(B)$ . Es gilt immer:  $P(A \cap B) \leq P(A)$ .

Es gilt immer:  $P(A) < P(B)$ , falls  $A \subset B$  ist. falsch (z.B. Würfel  $A = 1 - 6, B = 1 - 7$ )

Es gilt immer:  $P(A \cup B) = P(A) + P(B)$ . falsch (nur falls A und B disjunkt sind)

Es gilt immer:  $P(A \cap B) = P(A) \times P(B)$ . falsch ( nur falls A und B unabhängig)



### Multiplikationssatz

Die Formel  $P(A|B) = \frac{P(A \cap B)}{P(B)}$  kann man umformen, dann ist sie als allgemeiner Multiplikationssatz bekannt:

$$P(A \cap B) = P(A|B) * P(B)$$

**Bedeutung:** Wahrscheinlichkeit, dass A und B zusammen eintreten, ist das Produkt der Wahrscheinlichkeit von B und der bedingten Wahrscheinlichkeit von A gegeben B.

### Satz der totalen Wahrscheinlichkeit

Seien  $B_1, B_2, \dots, B_k$  paarweise disjunkte Ereignisse und  $B_1 \cup B_2 \cup \dots \cup B_k = \Omega \rightarrow$  d.h. genau ein  $B_i$  tritt immer ein. Die  $B_i$  ergeben eine disjunkte Zerlegung des W-Raumes  $\Omega$ .

Jedes Element von einem weiteren Ereignis A liegt so genau in einem  $B_i$  und es gilt:  $P(A) = \sum_{i=1}^k P(A|B_i) * P(B_i)$

Bei Niederschlag beträgt die Wahrscheinlichkeit, dass ein Flugzeug pünktlich ist 50%. Bei schönem Wetter ist das Flugzeug in 90% der Fälle pünktlich. Morgen wird es mit 70% Wahrscheinlichkeit schön sein. Wie gross ist die Wahrscheinlichkeit, dass morgen ein Flugzeug pünktlich eintreffen wird?

$$P(\text{pünktlich}) = P(\text{pünktlich}|\text{schön}) * P(\text{schön}) + P(\text{pünktlich}|\text{schlecht}) * P(\text{schlecht}) = 0.9 * 0.7 + 0.5 * 0.3 = 0.78$$

### Satz von Bayes

Der Satz von Bayes beschreibt die Beziehung zwischen den zwei bedingten Wahrscheinlichkeiten  $P(A|B)$  und  $P(B|A)$

und erlaubt somit das «Umkehren» der bedingten Wahrscheinlichkeiten.  $P(B_i|A) = \frac{P(A|B_i) * P(B_i)}{\sum_{j=1}^k P(A|B_j) * P(B_j)}$

Ziegenproblem  $P(G_2 | M_3) = ?$   $P(G_1) = P(G_2) = P(G_3) = \frac{1}{3}$

$P(M_3 | G_1) = \frac{1}{2}$   $P(M_3 | G_2) = 1$   $P(M_3 | G_3) = 0$

Satz der totalen Wahrscheinlichkeit:  $P(M_3) = P(M_3|G_1) * P(G_1) + P(M_3|G_2) * P(G_2) + P(M_3|G_3) * P(G_3)$   
 $= \frac{1}{2} * \frac{1}{3} + 1 * \frac{1}{3} + 0 * \frac{1}{3} = \frac{1}{2}$

Mit dem Satz von Bayes:  $P(G_2|M_3) = \frac{P(M_3|G_2) * P(G_2)}{P(M_3)} = \frac{1 * \frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$

Der Satz von Bayes erlaubt unterschiedliche sich gegenseitig ausschliessende Möglichkeiten  $B_i$  (Hypothesen) auf Grund von beobachteten Daten (A) zu vergleichen. Dafür braucht man ...

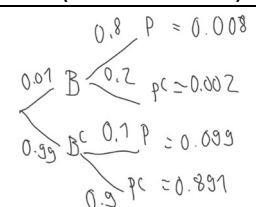
- eine a-priori Meinung über die Hypothesen  $P(B_i)$  (d.h. vor dem Eintreten von A)
- zu wissen, welche Wirkung die Hypothesen auf das Eintreten von A haben (das sind die  $P(A|B_i)$ )

Kurz: Der Satz von Bayes beschreibt wie sich unsere a-priori Meinung anhand vom Eintreten eines Ereignisses verändert, also wie man aus Erfahrung lernt.  $\Rightarrow$  Der Satz von Bayes beschreibt, wie wir denken sollen (wenn rational).

1% der Frauen haben Brustkrebs. Das Testverfahren ist zu 80% positiv, wenn er vorliegt

und zu 10% falsch positiv. Mit welcher Wahrscheinlichkeit habe ich mit positivem

Testresultat auch Brustkrebs?  $P(P|B) = \frac{P(P \cap B)}{P(P)} = \frac{0.008}{0.008 + 0.099} = 0.0747$



### Diagnostische Tests

Diagnostische Tests sind eine wichtige Anwendung des Satzes von Bayes.

Kein Test mit vernünftigen Aufwand ist perfekt, auch der ELISA-Test nicht. Es gibt zwei mögliche Fehldiagnosen:

a) Falsch positive Resultate → Eine eigentlich HIV-negative Person erhält ein positives Testresultat.

b) Falsch negative Resultate → Eine eigentlich HIV-positive Person erhält ein negatives Testresultat.

ELISA+ = ELISA Test ist positiv

ELISA- = ELISA Test ist negativ

HIV+ = HIV-positive Blutprobe

HIV- = HIV-negative Blutprobe

a) Falsch Positive: 14925 Personen

b) Falsch Negative: 15 Personen

	ELISA+	ELISA-	Total
HIV+	4985	15	5000
HIV-	14925	980075	995000
Total	19910	980090	1000000

**Sensitivität** ist der Anteil der HIV-positiven Blutproben, welche durch den ELISA-Test erkannt werden.

$$P(ELISA + | HIV +) = 99.7\%$$

**Spezifität** ist der Anteil an HIV-negativen Personen, die auch ein negatives ELISA-Testresultat erhalten.

$$P(ELISA - | HIV -) = 98.5\% \quad P(ELISA + | HIV -) = 1 - P(ELISA - | HIV -) = 1 - 0.985 = 0.015$$

**Prävalenz** bezeichnet die Wahrscheinlichkeit, dass eine zufällig herausgegriffene Blutspende HIV-positiv ist.

$$P(HIV +) = 0.5\%$$

$$P(HIV -) = 1 - P(HIV +) = 1 - 0.005 = 0.995$$

Wie gross ist die Wahrscheinlichkeit, dass eine Person von der der ELISA Test + ausfiel, tatsächlich HIV-positiv ist?

$$P(HIV + | ELISA +) = \frac{P(ELISA + | HIV +) \cdot P(HIV +)}{P(ELISA + | HIV +) \cdot P(HIV +) + P(ELISA + | HIV -) \cdot P(HIV -)} = \frac{0.997 \cdot 0.005}{0.997 \cdot 0.005 + 0.015 \cdot 0.995} = 0.2504$$

### Stochastische Unabhängigkeit

Stochastische Unabhängigkeit bedeutet, dass sich zwei Ereignisse nicht gegenseitig beeinflussen.

Zwei Ereignisse A und B heissen stochastisch unabhängig, falls  $P(A \cap B) = P(A) \cdot P(B)$

Wir werfen einen Würfel und betrachten die Ereignisse A = „die geworfene Augenzahl ist gerade“ und B = „die geworfene Augenzahl ist 3 oder grösser“. Mit  $\Omega = \{1, 2, 3, 4, 5, 6\}$  gilt  $A = \{2, 4, 6\}$   $B = \{3, 4, 5, 6\}$   $A \cap B = \{4, 6\}$

$$P(A) \cdot P(B) = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3} = P(A \cap B)$$

Drei Ereignisse A, B, C sind unabhängig, wenn

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(B \cap C) = P(B) \cdot P(C)$$

$$P(A \cap C) = P(A) \cdot P(C)$$

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C)$$

Diese Formel impliziert unter anderem, dass jedes Paar von Ereignissen  $A_i$  und  $A_j$  unabhängig sein muss. Dies allein reicht aber nicht aus, wie das folgende Beispiel zeigt.

Zweifacher Münzwurf mit Z = Zahl und K = Kopf  $\Omega = \{(Z, Z), (Z, K), (K, Z), (K, K)\}$ .

Ereignis B: „Erster Wurf Zahl“ =  $\{(Z, Z), (Z, K)\}$   $P(B) = \frac{1}{2}$

Ereignis C: „Zweiter Wurf Zahl“ =  $\{(Z, Z), (K, Z)\}$   $P(C) = \frac{1}{2}$

Ereignis D: „Beide Würfe dasselbe Ergebnis“ =  $\{(Z, Z), (K, K)\}$   $P(D) = \frac{1}{2}$

$$P(B \cap C) = P((Z, Z)) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(B) \cdot P(C) \quad P(B \cap D) = P((Z, Z)) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(B) \cdot P(D)$$

$$P(C \cap D) = P((Z, Z)) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(C) \cdot P(D)$$

Ergebnisse B, C, D sind paarweise stochastisch unabhängig. Aber

$$P(B \cap C \cap D) = P((Z, Z)) = \frac{1}{4} \neq \frac{1}{8} = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = P(B) \cdot P(C) \cdot P(D)$$

B, C, D sind **nicht** stochastisch unabhängig.

Disjunkt und unabhängig werden oft verwechselt.

- Zwei Ereignisse A, B sind disjunkt, wenn  $A \cap B = \emptyset$ , d.h. sie treten nicht gemeinsam auf.
- Falls  $P(A) > 0$  und  $P(B) > 0$ , schliessen sich die Eigenschaften disjunkt und unabhängig aus.

Wenn A und B nicht gemeinsam vorkommen können, weiss man sehr viel über A (es kommt nicht vor), wenn B eingetreten ist.

Beweis: Sind A und B disjunkt, so ist  $P(A \cap B) = 0 \neq P(A) \cdot P(B)$ , da  $P(A) > 0$  und  $P(B) > 0$ . → Also disjunkt ⇒ abhängig.

Umgekehrt: Sind A und B stochastisch unabhängig ( $P(A) > 0$  und  $P(B) > 0$ ), dann  $P(A \cap B) = P(A) \cdot P(B) > 0$ , d.h.  $P(A \cap B) \neq 0$  (stochastisch unabhängig ⇒ nicht disjunkt)

### Wahrscheinlichkeitsverteilung

Wie verteilen sich bei einem Zufallsexperiment die Wahrscheinlichkeiten auf die verschiedenen Ergebnisse?

- physikalische Modelle (z.B. Würfelwurf) Oder, und darum wird es hier gehen
- rein datengestützt (z.B. Passagiere) • Mischform Modell & empirisch geschätzte Parameter
- subjektiv (z.B. atomarer Störfall)

Gedankenexperiment:

1. Planen Sie einen Zufallsspiel (z.B. Würfelfwurf, Roulette)
2. Machen Sie sich Gedanken über den Ergebnisraum (**Zufallsvariable**)
3. Ordnen Sie jedem Elementarereignis eine Wahrscheinlichkeit zu. (**Wahrscheinlichkeitsverteilung**)
4. Wie gross sind meine Gewinnchancen? Ist das Spiel fair? (**Erwartungswert/Varianz**)

Ziel: Das Ergebnis eines Zufallsexperiments soll durch eine einzelne Zahl beschrieben werden.

- Eine Grösse, die unter bestimmten, konstanten Bedingungen unterschiedliche, durch den Zufall bedingte Werte annehmen kann, nennt man Zufallsvariable.
- Mathematisch ist es eine Funktion,  $X: \Omega \rightarrow \mathbb{R}$ , die jedem Element des Ergebnisraumes  $\Omega$  eine Zahl zuordnet.
- Quantifizierung des Konzepts von Zufallsexperimenten.

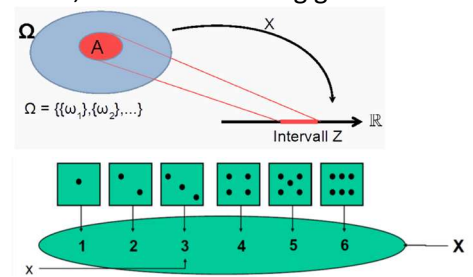
X heisst **Zufallsgrösse** oder **Zufallsvariable**, weil die Werte, die X annehmen kann, vom Zufall abhängig sind.

Zufallsvariablen

Die Zufallsvariable X ist eine Abbildung. Die Abbildung ordnet jedem Ergebnis eines Zufallsexperiments einen Wert zu:

Frage: Mit welcher Wahrscheinlichkeit liegt der Wert X in Z? Antwort:

$P(X \text{ in } Z) = P(A)$ , wobei A alle  $\omega_i$  enthält, für welche  $X(\omega_i)$  in Z liegt.



Beispiel: Würfeln von 2 Würfeln

Ereignisraum:  $\Omega = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), \dots, (6, 6)\}$

Um das Zufallsexperiment durch eine Zahl zu beschreiben, kann man hier zum Beispiel die Summe der beiden gewürfelten Augenzahlen wählen:  $X : (1, 1) \rightarrow 2$  oder  $X : (1, 2) \rightarrow 3$   $X = 4$  steht für das Ergebnis: Augensumme gleich 4. Es ist dann eines der folgenden drei Elementarereignisse eingetreten  $\{(1, 3), (2, 2), (3, 1)\}$

Zufallsvariablen beschreiben wir mit einem Grossbuchstaben.

Eine **Realisierung** der **Zufallsvariablen** wird durch den entsprechenden **Kleinbuchstaben** abgekürzt.

Beispiel:  $X =$  gewürfelte Augenzahl Realisierung von X:

- $x = 3$ , wenn die Augenzahl 3 gewürfelt wird.
- $x = 4$ , wenn die Augenzahl 4 gewürfelt wird.

Diskrete Zufallsvariablen

- X nimmt endlich oder unendlich abzählbar viele Werte an.
- Überall dort, wo gezählt wird.  $\rightarrow$  Beispiel: Anzahl Kunden, Anzahl Fahrzeuge, Anzahl Unfälle.

Stetige Zufallsvariable

- X kann jeden beliebigen Wert eines Intervalls annehmen.
- Überall dort, wo gemessen wird.  $\rightarrow$  Beispiel: Länge, Gewicht, Temperatur.

Die Wahrscheinlichkeitsverteilung gibt an, welche Werte eine Zufallsvariable mit welcher Wahrscheinlichkeit annimmt. Je nach Ausprägung der Zufallsvariablen (diskret/stetig) gibt es

Die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariable gibt zu jedem möglichen Wert (Realisierung)  $x_i$  die entsprechende Wahrscheinlichkeit  $p_i$  an. Notation:  $p_i = p(x_i) = P(X = x_i)$ .

Eigenschaften:

- Die Summe aller  $p_i$  ergibt 1, d.h.  $\sum_{i=1}^N p_i = 1$
- Die Wahrscheinlichkeitsfunktion heisst **diskrete Gleichverteilung**, wenn alle  $p_i$  gleich sind.

Diskrete Gleichverteilung

Beim Würfeln  $X =$  "gewürfelte Augenzahl" diskrete Gleichverteilung.

X	$x_1$	$x_2$	$x_3$	...	$x_N$
$p(\cdot)$	$p_1$	$p_2$	$p_3$	...	$p_N$

X	$x_1 = 1$	$x_2 = 2$	$x_3 = 3$	$x_4 = 4$	$x_5 = 5$	$x_6 = 6$
$p(\cdot)$	$p_1 = 1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

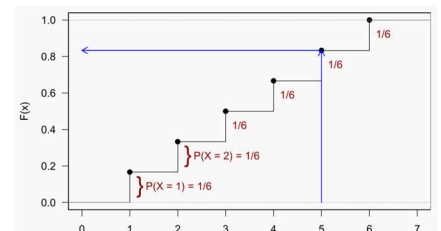
Kumulative Verteilungsfunktion

Zu jeder (diskreten) Wahrscheinlichkeitsverteilung gibt es eine kumulative Verteilungsfunktion:  $F(x) = P(X \leq x)$

Die kumulative Verteilungsfunktion gibt die Wahrscheinlichkeit an, mit welcher die Zufallsvariable X einen Wert annimmt, der kleiner / gleich einem vorgegebenen Wert ist.

Eigenschaften:

- Für  $F(x)$  gilt  $F(-\infty) = 0$  und  $F(\infty) = 1$ .
- Eine kumulative Verteilungsfunktion ist monoton wachsend, d.h. für  $u < v$  gilt  $F(u) \leq F(v)$ .
- Für diskrete Zufallsvariablen ist  $F(x)$  eine Treppenfunktion.



Beispiel: Würfeln  $X =$  "gewürfelte Augenzahl"  $\rightarrow$  Kumulative Wahrscheinlichkeit für  $x = 5$ :

$F(5) = P(X \leq 5) = P(X = 1) + P(X = 2) + \dots + P(X = 5) = 5/6$

### Erwartungswert

Der Erwartungswert  $E[X]$  einer Zufallsvariable  $X$  ist das, was man "im Schnitt" bei unendlich vielen Realisierungen von  $X$  erhält. Er ist ein Lagemass des Wahrscheinlichkeitsmodells. Wichtig: Daten spielen hier keine Rolle! Im diskreten Fall ergibt sich der Erwartungswert als gewichtete Summe der möglichen Werte, wobei die Gewichte den Wahrscheinlichkeiten entsprechen, mit denen die Werte realisiert werden.

Erwartungswert für eine diskrete Zufallsvariable  $X$  ( $E[X]$ ):  $\mu = E[X] = \sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} x_i p_i$

Beispiel: Würfeln  $E[X] = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3.5$

### Kennzahlen

Kennzahlen von Stichproben werden zu Kennzahlen von Verteilungen, wenn man die relativen Häufigkeiten durch Wahrscheinlichkeiten ersetzt. arithmetisches Mittel:  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \frac{x_1+x_2+x_3+\dots+x_n}{n}$

Erwartungswert (der Verteilung):  $E[X] = \sum_{i=1}^{\infty} x_i P(X = x_i) = \sum_{i=1}^{\infty} x_i p_i$

### Varianz

Die Varianz  $Var(X)$  ist ein Streuungsmass einer Zufallsvariablen und gibt an, wie breit die angenommenen Werte des Wahrscheinlichkeitsmodells streuen. Varianz für eine diskrete Zufallsvariable  $X$ :

$$Var(X) = E[(X - E[X])^2] = \sum_{i=1}^{\infty} (x_i - E[X])^2 p_i$$

Varianz lässt sich auf wie folgt berechnen:  $Var(X) = E[X^2] - (E[X])^2 = (\sum_{i=1}^{\infty} (x_i)^2 p_i) - (E[X])^2$

Beispiel: Würfeln

$$Var(X) = (1 - 3.5)^2 \cdot 1/6 + (2 - 3.5)^2 \cdot 1/6 + (3 - 3.5)^2 \cdot 1/6 + (4 - 3.5)^2 \cdot 1/6 + (5 - 3.5)^2 \cdot 1/6 + (6 - 3.5)^2 \cdot 1/6 = (-2.5)^2 \cdot 1/6 + (-1.5)^2 \cdot 1/6 + (-0.5)^2 \cdot 1/6 + (0.5)^2 \cdot 1/6 + (1.5)^2 \cdot 1/6 + (2.5)^2 \cdot 1/6 = 2.92$$

Alternativen Formel:  $E[X^2] = \frac{1}{6} \cdot 1^2 + \frac{1}{6} \cdot 2^2 + \frac{1}{6} \cdot 3^2 + \frac{1}{6} \cdot 4^2 + \frac{1}{6} \cdot 5^2 + \frac{1}{6} \cdot 6^2 = 15.17$

$$Var(X) = 15.17 - (3.5)^2 = 2.92$$

## Diskrete Wahrscheinlichkeitsverteilung

### Bernoulli Verteilung

Ein Bernoulli-Zufallsexperiment ist ein Experiment mit zwei möglichen Ausgängen: Erfolg (Ja) und Misserfolg (Nein). In der Regel sind die Ausgänge mit 0 (= Misserfolg) und 1 (= Erfolg) kodiert.

**Wahrscheinlichkeitsfunktion:**  $p = P(X = 1)$  ist die Erfolgswahrscheinlichkeit.

**Notation:  $X \sim \text{Bernoulli}(p)$**

**alternative Notation:  $X \sim \text{Bin}(1, p)$**

- **Erwartungswert:**  $E[X] = p$
- **Varianz:**  $Var(X) = p \cdot (1 - p)$

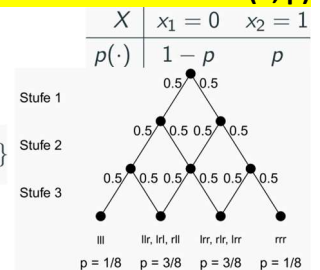
### Mehrstufige Bernoulli-Experimente

Beispiel: Galtonbrett mit  $n = 3$  Stufen

$Y$  = Anzahl Sprünge nach links

$X_i$  = Kugel springt auf Stufe  $i$  nach links

Es gilt:  $Y = \sum_{i=1}^3 X_i$ , d.h.  $Y \in \{0, 1, 2, 3\}$



### Binomialverteilung

Die Binomialverteilung beschreibt  **$n$  unabhängige Bernoulli Experimente** mit jeweils **konstanter Erfolgswahrscheinlichkeit  $p$** , die  **$n$ -mal nacheinander ausgeführt wurden**.

Die **Anzahl Erfolge  $X$**  hat dann die folgende Verteilung:  $P(X = k) = \binom{n}{k} \cdot p^k (1 - p)^{n-k}$

Hierbei sind  $n$  und  $p$  fix gegeben.

$p^k (1 - p)^{n-k} \rightarrow$  Wahrscheinlichkeit für ein **Ereignis mit  $k$  Erfolgen und  $n - k$  Misserfolgen**.

$\binom{n}{k} \rightarrow$  Anzahl Ereignisse mit  $k$  Erfolgen in  $n$  Versuchen

**Notation:  $X \sim \text{Bin}(n, p)$**

### Beispiele:

- Die Abwesenheitsrate einer Firma mit 40 Mitarbeitern beträgt 8%. Wie viele Arbeiter sind an Tag **anwesend**?  
 $X$  = Anwesende Mitarbeiter                       $n = 40, p = 1 - 0.08$                        $X \sim \text{Bin}(40, 0.92)$
- Im Flugzeug sind 30 Plätze besetzt und generell sitzen 70% Männer im Flugzeug. Wie viele Frauen sind dabei?  
 $X$  = Anzahl Frauen in der Businessklasse                       $n = 30, p = 1 - 0.7$                        $X \sim \text{Bin}(30, 0.3)$

### Eigenschaften

- Der Ergebnisraum einer binomialverteilten Zufallsvariable ist endlich:  $\Omega = \{0, 1, 2, 3, 4, \dots, n\}$
- Eine binomialverteilte Zufallsgrösse kann auch als Summe  $S$  von  $n$  unabhängigen Wiederholungen einer bernouilliverteilte Zufallsgrösse  $X_i$  mit Parameter  $p$  beschrieben werden, d.h.  $S = \sum_{i=1}^n X_i$
- **Erwartungswert**  $E[S] = n \cdot p$
- **Varianz**  $Var[S] = n \cdot p \cdot (1 - p)$



**Beispiel: Bestehen von Wahr**

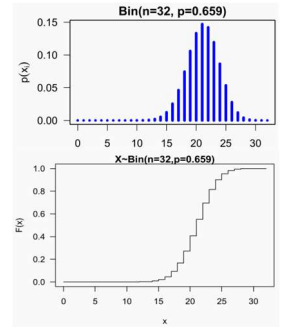
X = Anzahl Studenten aus einer Klasse mit 32 Studenten, welche das Modul Wahr bestehen.

Jeder Student stellt für sich ein Bernoulli-Experiment dar. Grundsätzlich können alle Studenten das Modul bestehen oder nicht. Der Wertebereich ist  $X \in \{0, 1, 2, 3, 4, \dots, 32\}$ . Somit ist  $X \sim \text{Bin}(32, 0.659)$

Wie gross ist die Wahrscheinlichkeit, dass 28 Studenten bestehen? `dbinom(x = 28, size = 32, prob = 0.659)`

Wahrscheinlichkeitsverteilung `pk <- dbinom(x = 0:k, size = n, prob = p)`

Wahrscheinlichkeiten für alle möglichen Werte, die von der Zufallsvariable X angenommen werden können: Von 0 bis 32 im oben genannten Beispiel.  $X = k$  mit Vektor von 0:32 übergeben. `plot(0:k, pk, type = "h")`



**Aussehen der Verteilung:**

- Symmetrie für  $p = 0.5$
- Bei wachsendem n auch für  $p \neq 0.5$  immer symmetrischer
- Faustregel: Falls  $np(1 - p) > 10$ , gilt  $\text{Bin}(n,p)$  als symmetrisch.

Kumulative Verteilungsfunktion `pbinom(x, size = n, prob = p)`

Wahrscheinlichkeit, dass 3 oder weniger Studenten

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

Wahr bestehen: `pbinom(3, size = 32, prob = 0.659)`

$$= \binom{32}{0} \cdot (0.341)^{32} + \binom{32}{1} \cdot 0.659 \cdot (0.341)^{31} + \binom{32}{2} \cdot (0.659)^2 \cdot (0.341)^{30} + \binom{32}{3} \cdot (0.659)^3 \cdot (0.341)^{29} = 4.213564 \cdot 10^{-11} \text{ oder mit R:}$$

**Checkliste Binomialverteilung**

Alle folgende Punkte müssen zutreffen, um von einer Binomialverteilung ausgehen zu können:

- Gibt es eine vorher festgelegte Anzahl an durchzuführenden Versuchen n?
- Hat jeder dieser Versuche nur zwei mögliche Ausgänge (Kopf oder Zahl / Erfolg oder Misserfolg / 0 oder 1)?
- Ändern sich die Wahrscheinlichkeiten p in den einzelnen Versuchen nicht?
- Sind alle Versuche unabhängig?

Quantile `qbinom(alpha, size = n, prob = p)`

Als  $\alpha$ -Quantil bezeichnet man denjenigen Wert x, sodass für die Wahrscheinlichkeit der Zufallsgrösse X gerade gilt:  $P(X \leq x) \geq \alpha$ . → Umkehrfunktion der kumulativen Verteilungsfunktion

In 90% der Fälle bestehen höchstens wie viele Studenten die Wahr Prüfung? `qbinom(0.9, size = 32, prob = 0.659)`

**Geometrische Verteilung**

Beispiel: Eile mit Weile → Um das Haus zu verlassen, muss eine 6 gewürfelt werden. X = "Anzahl Fehlwürfe, bis man das Haus verlassen kann." Wir möchten die Wahrscheinlichkeiten für  $X = 0, 1, 2, \dots$  Fehlwürfe berechnen.

$$P(X = 0) = \frac{1}{6} \qquad P(X = 2) = \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} \qquad P(X = k) = \left(\frac{5}{6}\right)^k \cdot \frac{1}{6}$$

$$P(X = 1) = \frac{5}{6} \cdot \frac{1}{6} \qquad P(X = 3) = \left(\frac{5}{6}\right)^3 \cdot \frac{1}{6}$$

Führen wir ein **Bernoulli-Experiment** mit Erfolgswahrscheinlichkeit p so lange durch, bis wir den **ersten Erfolg erzielen**, erhalten wir die geometrische Verteilung. Die Anzahl Fehlversuche vor dem ersten Erfolg X hat dann die folgende Verteilung:  $P(X = k) = p \cdot (1 - p)^k$

Hierbei ist **p fix als Wahrscheinlichkeit für einen Erfolg** gegeben. **Notation:  $X \sim \text{Geom}(p)$**

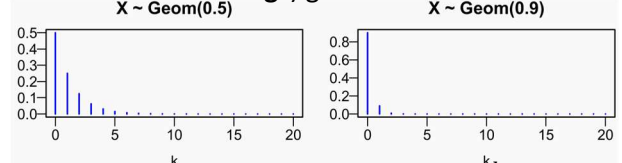
**Eigenschaften**

- Der Ergebnisraum einer geometrisch verteilten Zufallsvariable ist abzählbar unendlich:  $\Omega = \{0, 1, 2, 3, 4, \dots\}$
- Erwartungswert:  $E[X] = (1 - p)/p$  **Varianz:  $\text{Var}(X) = (1 - p)/p^2$**

Achtung: In vielen Büchern findet man eine etwas andere Definition, bei der nicht die **Anzahl der Misserfolge** bis zum ersten Erfolg, sondern die **Anzahl der Versuche (einschliesslich des ersten Erfolgs)** gezählt werden.

**Aussehen der Verteilung:**

- Die Wahrscheinlichkeitsfunktion ist monoton abfallend.
- Je grösser p, desto stärker abfallend →



Wahrscheinlichkeitsberechnung - `dgeom(k, prob = p)`

Zur Bestimmung der Wahrscheinlichkeiten verwenden wir die folgende Formel  $P(X = k) = p \cdot (1 - p)^k$

**Beispiel:** Eile mit Weile Wie gross ist die Wahrscheinlichkeit, dass man mit 3 Fehlwürfen das Haus verlassen kann? → `dgeom(3, prob = 1/6)`

Kumulative Verteilungsfunktion - `pgeom(x, prob = p)`

$$F(x) = P(X \leq x) = \sum_{k=0}^x p \cdot (1 - p)^k = 1 - (1 - p)^{x+1}$$

**Beispiel: Eile mit Weile** Wie gross ist die Wahrscheinlichkeit, weniger oder gleich 3 Fehlwürfe zu haben?

$$P(X \leq 3) = 1 - \left(1 - \frac{1}{6}\right)^{3+1} = 0.5177469 \rightarrow \text{pgeom}(3, \text{prob} = 1/6)$$

Wahrscheinlichkeiten und Quantile

Wie gross ist die Wahrscheinlichkeit, dass man zwischen 1 und 4 Fehlwürfe hat?  $\rightarrow \text{sum}(\text{dgeom}(1:4, \text{prob} = 1/6))$

In 90% der Fälle gibt es höchstens wie viele Fehlwürfe?  $\rightarrow \text{qgeom}(0.9, \text{prob} = 1/6)$

Erzeugen einer Zufallszahl -  $\text{rgeom}(1, \text{prob} = 1/6)$

Bemerkenswerte Eigenschaft

Die geometrische Verteilung besitzt eine bemerkenswerte Eigenschaft:  $P(X > k + j \mid X \geq k) = P(X > j)$

- **Gedächtnislosigkeit**, d.h. Wahrscheinlichkeit dafür, dass der erste Erfolg noch mindestens j weitere Fehlversuche lang auf sich warten lässt, ist unabhängig davon, wie viele Misserfolge man schon beobachtet hat.
- Die geometrische Verteilung wird u.a. auch dazu verwendet, um zu modellieren, wie lange ein technisches Gerät funktioniert (z.B. gemessen Anzahl Einsätzen). Gedächtnislosigkeit bedeutet, dass das Bauteil nicht altert!

Negative Binomialverteilung

Eine Verallgemeinerung der geometrischen Verteilung ist die negative Binomialverteilung. Hier zählt man die Anzahl der Misserfolge vor dem r-ten Erfolg. Die **Erfolgswahrscheinlichkeit** p ist konstant. **Notation:  $X \sim \text{NBin}(r, p)$**

- Serie der Experimente wird abgebrochen, sobald der r-te Erfolg auftritt.
- Gab es vorher k Fehlversuche wurden also insgesamt k+r Versuche durchgeführt.
- Binomialkoeffizient gibt die Anzahl Möglichkeiten die ersten r-1 Erfolge und die k Fehlversuche anzuordnen.
- Jede Anordnung hat die Wahrscheinlichkeit  $p^{r-1}(1-p)^k$  die wir dann mit p multiplizieren müssen, da im k+r Versuch der r-te Erfolg auftreten soll.

$$P(X = k) = \binom{k+r-1}{k} p^r (1-p)^k$$

Eigenschaften

- Der Ergebnisraum einer negativ binomial verteilten Zufallsvariable ist abzählbar unendlich:  $\Omega = \{0, 1, 2, 3, 4, \dots\}$
- **Erwartungswert:  $E(X) = r \cdot (1-p)/p$**       **Varianz:  $\text{Var}(X) = r \cdot (1-p)/p^2$**

Beispiel

Wir kaufen so lange Rubel-Lose, bis wir drei Gewinne haben. Ein Gewinn habe die Wahrscheinlichkeit 0.05. Die Anzahl der Nieten X bis zum dritten Gewinn ist  $\text{NBin}(3, 0.05)$ .

```
dnbinom(x = , size = r, prob = p) # Wahrscheinlichkeitsfunktion
pnbinom(q = , size = r, prob = p) # Kum. Verteilungsfunktion
qnbinom(p = , size = r, prob = p) # Quantilfunktion
rnbino(n = , size = r, prob = p) # Zufallszahlen
```

Beispiel: Rubbellose

Die Anzahl der Nieten X bis zum dritten Gewinn ist  $\text{NBin}(3, 0.05)$

In 99% der Fälle muss man höchstens wie viele Lose kaufen?  $\text{qnbinom}(0.99, \text{size} = 3, \text{prob} = 0.05) + 3$

Hypergeometrische Verteilung

Ziehen wir k Kugeln aus einer Urne mit m weissen und n schwarzen Kugeln, und **legen** nach jedem Zug die gezogene Kugel **wieder in die Urne**, gilt für die Anzahl X der weissen gezogenen Kugeln:  $X \sim \text{Bin}\left(k, \frac{m}{n+m}\right)$

Wenn wir die Kugeln nicht zurücklegen, ist es eine Hypergeometrische Verteilung

Zufallsvariable: X = "Anzahl Erfolge in k Versuchen": **Notation:  $X \sim \text{Hyp}(m, n, k)$**

Eigenschaften

- Der Ergebnisraum einer hypergeometrischen Zufallsvariable ist abzählbar endlich:  $\Omega = \{0, 1, \dots, \min(k, m)\}$
- **Erwartungswert:  $E(X) = k \cdot \frac{m}{m+n}$**       **Varianz:  $\text{Var}(X) = k \cdot \frac{m}{m+n} \cdot \left(1 - \frac{m}{m+n}\right) \cdot \frac{n+m-k}{m+n-1}$**

In einer Lieferung von 1000 Schrauben sind 27 defekt. Der Empfänger überprüft 30 zuf. ausgewählte Schrauben.

m = Anzahl defekter Schrauben = 27

k = Stichprobengrösse = 30

n = Anzahl intakter Schrauben = 973

$X \sim \text{Hyp}(27, 973, 30)$

Mit welcher Wahrscheinlichkeit sind alle Schrauben in der Stichprobe in Ordnung?  $\text{dhyper}(0, m=27, n=973, k=30)$

$\text{dhyper}(x = , m = m, n = n, k = k)$  # Wahrscheinlichkeitsfunktion

$\text{phyper}(q = , m = m, n = n, k = k)$  # Kum. Verteilungsfunktion

$\text{qhyper}(p = , m = m, n = n, k = k)$  # Quantilfunktion

$\text{rhyper}(nn = , m = m, n = n, k = k)$  # Zufallszahlen

## Poisson-Verteilung

Die Poisson-Verteilung eignet sich für Vorfälle, die im Laufe der Zeit eintreten oder sich an einem bestimmten Ort ereignen. Man interessiert sich für die Anzahl Vorkommnisse in einer bestimmten Zeitspanne oder einem festgelegten Gebiet.  $X =$  "Anzahl Ereignisse pro Zeiteinheit/Gebiet". Beispiele:

- Unfälle in einer Fabrik, auf Strassen, oder anderswo
- Defekte in Geräten, an Fahr- oder Flugzeugen
- Das Eintreffen von Kunden an einem Schalter

Die Poisson-Verteilung besitzt einen Parameter. Dieser beschreibt, mit welcher Rate die einzelnen Ereignisse, welche voneinander unabhängig sind, auftreten. Die Wahrscheinlichkeitsverteilung für die Anzahl Ereignisse  $X$  ist:

$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$  Hierbei ist  $\lambda$  fix gegeben:  $\lambda$  Rate, mit welcher die Ereignisse, in einer vorgegebenen Zeiteinheit oder Gebiet eintreffen. **Notation:  $X \sim \text{Pois}(\lambda)$**

### Eigenschaften

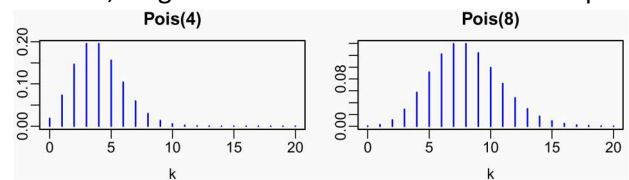
- Der Ergebnisraum einer Poisson-verteilten Zufallsvariable ist abzählbar unendlich:  $\Omega = \{0, 1, 2, 3, 4, \dots\}$ . Es gibt keine klar definierte Obergrenze.
- **Der Erwartungswert ist  $E(X) = \lambda$ . Wir "erwarten" also innerhalb einer Zeit/ Flächen-Einheit  $\lambda$  Ereignisse.**
- **Die Varianz ist  $\text{Var}(X) = \lambda$ .**

### Beispiele:

- Wenn  $X$  die Anzahl tödliche Verkehrsunfälle pro Jahr in der Schweiz beschreibt, so gilt bei durchschnittlich 350 tödlichen Unfällen  $X \sim \text{Pois}(350)$ .
- Wenn  $X$  die Anzahl Kunden am Postschalter pro Stunde beschreibt, so gilt bei durchschnittlich 32 Kunden pro Stunde  $X \sim \text{Pois}(32)$ .

### Aussehen der Verteilung

- Für kleine  $\lambda$  stark rechtsschief.
- Je grösser  $\lambda$ , desto symmetrischer. Generell gute Symmetrie ab  $\lambda > 10$



### Wahrscheinlichkeitsberechnung $\text{dpois}(x = k, \text{lambda} = \text{lambda})$

Wie gross ist die Wahrscheinlichkeit, dass in einer Stunde 30 Kunden am Postschalter eintreffen?

$P(X = 30) = ((32)^{30} \cdot e^{-32})/30! = 0.06814215 \rightarrow \text{dpois}(x= 30, \text{lambda} = 32)$

### Kumulative Verteilungsfunktion $\text{ppois}(q, \text{lambda} = \text{lambda})$

Kunden am Postschalter pro Stunde: Wie gross ist die Wahrscheinlichkeit, dass in einer Stunde weniger oder gleich 30 Kunden am Postschalter eintreffen?  $\rightarrow \text{ppois}(q = 30, \text{lambda} = 32)$

## Binomial vs. Poisson

Binomial	Poisson
bekanntes Anzahl von $n$ Einzelversuchen, d.h. Wertebereich von $X: \{0, \dots, n\}$	Grösse der Population ist unbekannt, d.h. Wertebereich von $X$ hat keine klar definierte Obergrenze.
Erfolgswahrscheinlichkeit $p$ für den Einzelversuch	Rate $\lambda$ , mit welcher das Ereignis auftritt

Anzahl Wahlberechtigter, die bei einer Umfrage unter 5000 Personen eine Gesetzesvorlage befürworten.  $\rightarrow B$

Anzahl Schadensfälle einer Versicherung.  $\rightarrow P$

Anzahl Kunden, die einen Kredit beantragen.  $\rightarrow P$

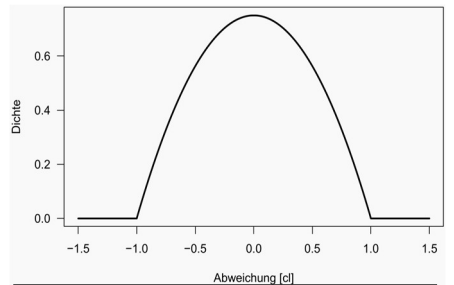
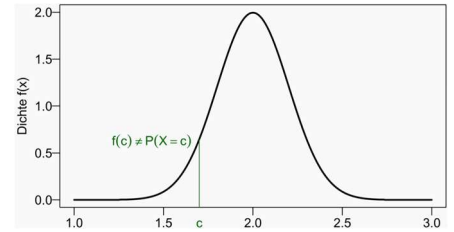
Anzahl Fahrzeuge, die bei einer Kontrolle von 100 angehaltenen Fahrzeugen beanstandet werden müssen.  $\rightarrow B$

## Rechnen mit diskreten Wahrscheinlichkeitsverteilungen in R - ?distributions

$P(X = 3)$	<code>dbinom(x=3, size=13, prob=1/3)</code>
$P(X \leq 3)$	<code>pbinom(q=3, size=13, prob=1/3)</code>
$P(X \geq 3) = 1 - P(X \leq 2)$	<code>1-pbinom(q=2, size=13, prob=1/3)</code>
$P(2 \leq X \leq 9) = P(X \leq 9) - P(X \leq 1)$	<code>pbinom(q=9, size=13, prob=1/3) - pbinom(q=1, size=13, prob=1/3)</code>

# Stetige Wahrscheinlichkeitsverteilung

- X kann jeden beliebigen Wert eines Intervalls annehmen.
- Überall dort, wo gemessen wird → Länge, Gewicht, Temperatur, Zeit
- Nicht mehr möglich, einzelnen Ergebnissen eine Wahrscheinlichkeit zuzuordnen, sondern nur einem Intervall.
- Die Rolle der Wahrscheinlichkeitsfunktion p(x) im diskreten Fall übernimmt die Dichtefunktion f<sub>x</sub>(x).



```
f <- function(x) ifelse(x >= -1 & x <= 1, -3/4*x^2 + 3/4, 0)
```

Stetige Wahrscheinlichkeitsdichten f<sub>x</sub>(x) beschreiben die Verteilung unendlich vieler Ergebnisse einer der Zufallsvariable X. Die Werte von f<sub>x</sub>(x) sind keine Wahrscheinlichkeiten! f<sub>x</sub>(x) beschreibt NUR die Form der Verteilung.

## Wahrscheinlichkeitsdichte

- f<sub>x</sub>(x) kann Werte > 1 annehmen.
- Wenn X stetig ist gilt: P(X = c) = 0 für ein beliebiges c ∈ [min(X), max(X)]  
Begründung: Falls P(X = c) > 0, wäre ∑<sub>c</sub> P(X = c) = ∞ da Summe über unendlich viele Werte.

### Eigenschaften von f<sub>x</sub>(x):

- f<sub>x</sub>(x) ≥ 0, nimmt keine negativen Werte an.
- ∫<sub>-∞</sub><sup>∞</sup> f<sub>x</sub>(x) dx = 1 das Integral über den ganzen Bereich ist gleich 1 (entspricht der Fläche unter der Kurve).

Die Funktion f<sub>x</sub>(x) ist stückweise stetig.

Sei X die Abweichung in cl beim Abfüllen einer 33cl Flasche. Wir nehmen an, dass X die folgende Dichte hat:

$$f_X(x) = \begin{cases} -\frac{3}{4}x^2 + \frac{3}{4} & -1 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

$$\int_{-\infty}^{\infty} f_X(x) dx = \underbrace{\int_{-\infty}^{-1} 0 dx}_{=0} + \int_{-1}^1 -\frac{3}{4}x^2 + \frac{3}{4} dx + \underbrace{\int_1^{\infty} 0 dx}_{=0}$$

$$= \left[ -\frac{1}{4}x^3 + \frac{3}{4}x \right]_{-1}^1 = \left( -\frac{1}{4} + \frac{3}{4} \right) - \left( \frac{1}{4} - \frac{3}{4} \right) = \frac{1}{2} + \frac{1}{2} = 1$$

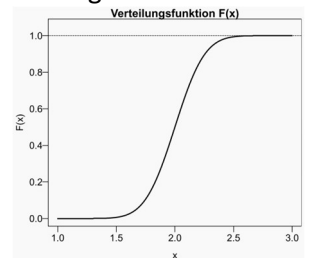
- Die Funktion ist für alle x grösser oder gleich 0.
- Die Funktion ist stückweise stetig.

Für **stetige** Zufallsvariablen gilt: P(a ≤ X ≤ b) = P(a < X ≤ b) = P(a ≤ X < b) = P(a < X < b)

Wahrscheinlichkeiten werden als Integrale über bestimmte Bereiche berechnet: P(a ≤ X ≤ b) = ∫<sub>a</sub><sup>b</sup> f<sub>x</sub>(x) dx

## (Kumulative) Verteilungsfunktion

- Die kumulative Verteilungsfunktion F<sub>X</sub>(x) ist im stetigen Fall definiert: F<sub>X</sub>(x) = P(X ≤ x) = ∫<sub>-∞</sub><sup>x</sup> f<sub>X</sub>(z) dz
- Die Verteilungsfunktion gibt Wahrscheinlichkeit an, dass Zufallsvariable X einen Wert kleiner gleich x annimmt.
- Bestimmt durch Integral über Dichtefunktion vom linken Rand bis zum Wert x.
- Die Verteilungsfunktion ist für stetige Verteilungen stetig.
- Die Dichtefunktion f(x) ergibt sich als Ableitung der Verteilungsfunktion



$$F: f_X(x) = F'_X(x) = \frac{d}{dx} F_X(x)$$

- An den Ränder gilt:  $\lim_{x \rightarrow -\infty} F_X(x) = 0$  und  $\lim_{x \rightarrow \infty} F_X(x) = 1$

## Erwartungswert

Ist ein Lagemass der Zufallsvariable und entspricht dem Schwerpunkt (in x-Richtung) der Wahrscheinlichkeitsdichte.

**Formel:** Sei X eine stetige Zufallsvariable mit Dichtefunktion f(x) → Erwartungswert  $\mu = E(X) = \int_{-\infty}^{\infty} (x \cdot f(x)) dx$   
Im Vergleich zum diskreten Fall wurde die Summe durch ein Integral und die Wahrscheinlichkeitsfunktion p(x) = P(X = x) durch die Dichte f(x) ersetzt.

Beispiel: Abfüllen einer 33cl Flasche Bestimmen Sie den Erwartungswert für die Verteilung. (siehe Funktion oben)

$$E(X) = \int_{-\infty}^{\infty} (x \cdot f(x)) dx = \int_{-1}^1 -\frac{3}{4}x^3 + \frac{3}{4}x dx = \left[ -\frac{3}{16}x^4 + \frac{3}{8}x^2 \right]_{-1}^1 = \left( -\frac{3}{16} + \frac{3}{8} \right) - \left( -\frac{3}{16} + \frac{3}{8} \right) = 0$$

## Varianz und Standardabweichung

Die Varianz Var(X) ist ein Streuungsmass einer Zufallsvariablen. Sie gibt die erwartete quadratische Abweichung der Zufallsvariable X von ihrem Erwartungswert an, d.h. es wird durch die Breite der Verteilung bestimmt.

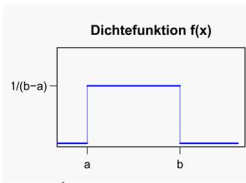
**Formel:** Sei X eine stetige Zufallsvariable mit Dichtefunktion f(x) und Erwartungswert E(X) = μ. Dann ist die Varianz definiert als:  $Var(X) = E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$  Die Standardabweichung σ ist die Wurzel aus

der Varianz. Alternativ:  $Var(X) = E(X^2) - (E(X))^2$  mit  $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$  (siehe Funktion oben)

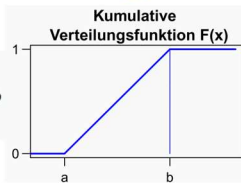
$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-1}^1 x^2 \left( -\frac{3}{4}x^2 + \frac{3}{4} \right) dx = \int_{-1}^1 -\frac{3}{4}x^4 + \frac{3}{4}x^2 dx = \left[ -\frac{3}{20}x^5 + \frac{1}{4}x^3 \right]_{-1}^1 = \left( -\frac{3}{20} + \frac{1}{4} \right) - \left( \frac{3}{20} - \frac{1}{4} \right) = \frac{1}{5}$$

Daraus ergibt sich  $Var(X) = \frac{1}{5} - 0^2 = 0.2$  und Standardabweichung  $\sigma = \sqrt{\frac{1}{5}}$

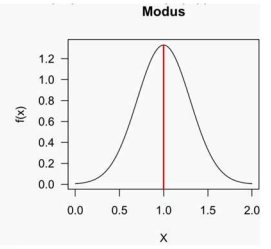
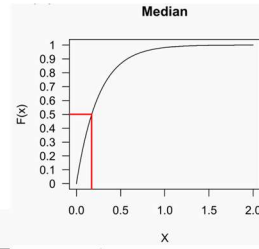
## Uniformverteilung (stetige Gleichverteilung)



$$f(x) = \begin{cases} 0 & \text{falls } x < a \\ 1/(b-a) & \text{falls } a \leq x \leq b \\ 0 & \text{falls } x > b \end{cases}$$



$$F(x) = \begin{cases} 0 & \text{falls } x < a \\ \frac{x-a}{b-a} & \text{falls } a \leq x \leq b \\ 1 & \text{falls } x > b \end{cases}$$



Bei der Uniformverteilung treten alle Werte eines Intervalls gleich oft ein.  $\Rightarrow$  **Notation:  $X \sim \text{Unif}([a,b])$ .**

- **Erwartungswert  $E(X) = (a + b)/2$** , d.h. entspricht der Mitte des Intervalls  $[a, b]$ .
- **Varianz  $\text{Var}(X) = (b - a)^2/12$** . Die Varianz erhöht sich, wenn die Intervalllänge  $(b - a)$  grösser wird.

$X$  = Wartezeit auf den Bus, welcher exakt alle 8 Minuten fährt. Jede Wartezeit tritt im Intervall  $[0,8]$  gleich häufig auf, wenn man zufällig an die Station kommt. Damit hat  $X$  eine Uniformverteilung:  $X \sim \text{Unif}([0,8])$

### Wahrscheinlichkeitsrechnung mit R

<code>dunif(x = x, min = a, max = b)</code>	Dichtefunktion
<code>punif(q = x, min = a, max = b)</code>	Kumulative Verteilungsfunktion
<code>qunif(p = alpha, min = a, max = b)</code>	Quantile
<code>runif(n = n, min = a, max = b)</code>	Zufallsvariablen

Erwartungswert = ist immer gleich (nicht abhängig vom Stichprobenumfang)

Mittelwert = bezieht sich auf die Stichprobe und ändert sich je nach Stichprobengröße

### Exponentialverteilung

**Beispiel:** An einem Schalter treffen im langfristigen Schnitt 3.6 Personen pro Minute ein.

$X$  = "Anzahl eintreffender Personen pro Minute" kann mit  $\text{Pois}(\lambda = 3.6)$  modelliert werden.

Gesucht:  $Y$  = "Zeitdauer bis nächste Person eintrifft"

Während die Anzahl der eintreffenden Personen nur ganzzahlige Werte annehmen kann, ist dies für die Wartezeit nicht der Fall. Diese ist von metrischem Charakter.  $\Rightarrow Y$  ist eine stetige Zufallsvariable mit stetiger Verteilung.

Für die Wartezeiten zwischen dem unabhängigen Eintreffen von Ereignissen ist die Exponentialverteilung ein geeignetes Modell. Deren Dichtefunktion ist wie folgt definiert:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & | \text{falls } x \geq 0 \\ 0 & | \text{falls } x < 0 \end{cases}$$

$\lambda$  ist dabei die mittlere Rate, mit welcher die Ereignisse passieren.

**Notation:  $X \sim \text{Exp}(\lambda)$**

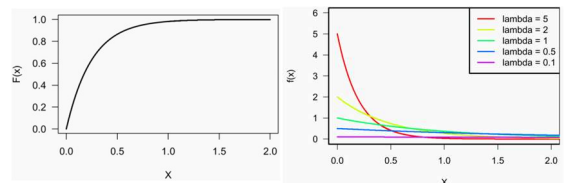
Aussehen der Wahrscheinlichkeitsdichte

Je grösser  $\lambda$ , desto grösser ist der exponentielle Abfall der Kurve.

**Beispiele:** Zeitintervall zwischen dem Eintreffen von zwei Kunden

Kumulative Verteilungsfunktion

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & | \text{falls } x \geq 0 \\ 0 & | \text{falls } x < 0 \end{cases}$$



Erwartungswert und Varianz

$$E[X] = \frac{1}{\lambda} = \text{erwartete Wartezeit zwischen zwei Ereignissen und } \text{Var}(X) = \frac{1}{\lambda^2}$$

Wahrscheinlichkeitsrechnung mit R

<code>dexp(x = x, rate = 1/lambda)</code>	Dichtefunktion
<code>pexp(q = x, rate = 1/lambda)</code>	Kumulative Verteilungsfunktion
<code>qexp(p = alpha, rate = 1/lambda)</code>	Quantile
<code>rexp(n = n, rate = 1/lambda)</code>	Zufallszahlen <b>Lambda ist hier die Rate siehe Beispiel</b>

**Beispiel: Wartezeit am Schalter: Im Mittel muss man am Schalter 1.9 Minute warten.**

$\Rightarrow X$  = Wartezeit  $\rightarrow X \sim \text{exp}(\lambda)$  mit  $\lambda = 1/1.9 = 0.526$  = Anzahl Kunden pro Minute

Wie gross ist die Wahrscheinlichkeit, dass man weniger als 2 Minuten warten muss? `pexp(2, rate = 0.526)`

### Gedächtnislosigkeit

Wie die geometrische Verteilung weist die Exponentialverteilung die Eigenschaft der Gedächtnislosigkeit auf, das Eintreffen des letzten Kunden hat keinen Einfluss auf das Eintreffen des nächsten und aller folgenden Kunden.

## Weibullverteilung

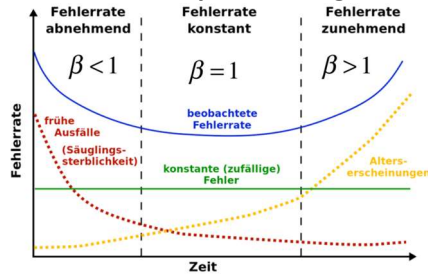
Bei der Modellierung der Lebensdauer und Ausfallhäufigkeit von Bauteilen spielt jedoch auch der Betrachtungszeitpunkt (= "Lebensphase" des Bauteils) eine Rolle. Dies wird in der Weibullverteilung (gedächtnisbehaftet), eine Erweiterung der Exponentialverteilung, berücksichtigt.

Dichtefunktion der Weibullverteilung ist wie folgt definiert:

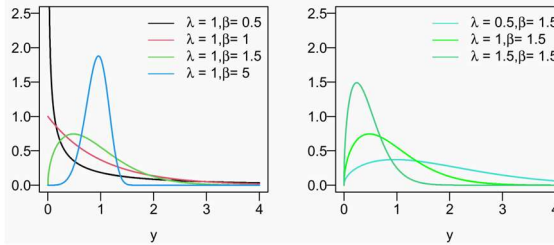
$$f(x) = \begin{cases} \lambda \cdot \beta \cdot (\lambda \cdot x)^{\beta-1} \cdot e^{-(\lambda x)^\beta} & \text{falls } x \geq 0 \\ 0 & \text{falls } x < 0 \end{cases}$$

**1/λ wird Skalenparameter genannt und β Formparameter.**

**Notation: X ~ Weibull(λ, β).**



Aussehen der Dichte



**Eigenschaften → dweibull(x, shape = beta, scale = 1/lambda)**

- Für β = 1 entspricht die Weibullverteilung der Exponentialverteilung.

- Die Kumulative Verteilungsfunktion ist  $F(x) = \begin{cases} 1 - e^{-(\lambda x)^\beta} & | \text{falls } x \geq 0 \\ 0 & | \text{falls } x < 0 \end{cases}$

- Erwartungswert  $E[X] = \frac{1}{\lambda} \Gamma(1 + 1/\beta)$

**Varianz  $Var[X] = \frac{1}{\lambda^2} [\Gamma(1 + 2/\beta) - \Gamma^2(1 + 1/\beta)]$**

## Gammafunktion

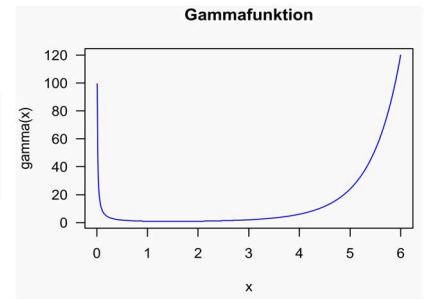
Die Gammafunktion ist eine Verallgemeinerung der Fakultät für reelle Zahlen.

Sie ist für α > 0 definiert als das Integral

**Eigenschaften:**

- $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- Für α = 1:  $\Gamma(1) = \int_0^\infty e^{-t} dt = 1$
- Für natürliche Zahlen n gilt:  $\Gamma(n) = (n - 1)!$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$$



**Wahrscheinlichkeitsrechnung mit R**

- dweibull(x, shape = beta, scale = 1/lambda)
- pweibull(q, shape = beta, scale = 1/lambda)
- qweibull(p, shape = beta, scale = 1/lambda)
- rweibull(n, shape = beta, scale = 1/lambda)

- Dichtefunktion f(x)
- Kumulative Verteilungsfunktion
- Quantile
- Zufallszahlen

## Gamma-Verteilung

**Beispiel: Höhlenexpedition:** Sie haben eine Taschenlampe und eine Ersatztaschenlampe für eine Höhlenexpedition eingepackt. Die Lebensdauer der einzelnen Batterien ist exponentialverteilt mit λ = 1/24 Stunden.

Wie gross ist die Wahrscheinlichkeit, dass Sie beide Taschenlampen zusammen für mindestens 1 Tag Höhlenexpedition reichen?

$$f(x) = e^{-\beta x} \cdot x^{\alpha-1} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \quad \text{für } x \geq 0$$

T<sub>1</sub> = Lebensdauer der Taschenlampe

T<sub>2</sub> = Lebensdauer der Ersatztaschenlampe

Die Lebensdauer der beiden Taschenlampen ist T = T<sub>1</sub> + T<sub>2</sub>. **Gesucht: P(T > 24)**

→ Die Summe unabhängiger, exponentialverteilter Zufallsvariablen ist Gamma verteilt.

Die Dichtefunktion einer Zufallsvariable Y mit Gamma-Verteilung ist:

$$f(y) = e^{-\lambda y} \cdot y^{\alpha-1} \cdot \frac{\lambda^\alpha}{\Gamma(\alpha)} \quad \text{mit } y > 0$$

**α > 0 ist der Formparameter. λ > 0 ist die Rate. bzw. 1/λ ist der Skalenparameter. Notation: X ~ G(α, λ).**

Ist α ∈ N spricht man auch von der Erlangverteilung.

**Aussehen der Dichte**

**Erwartungswert: E[Y] = α/λ**

**Varianz: Var(Y) = α/λ<sup>2</sup>**

Die Streuung steigt mit zunehmendem Erwartungswert.

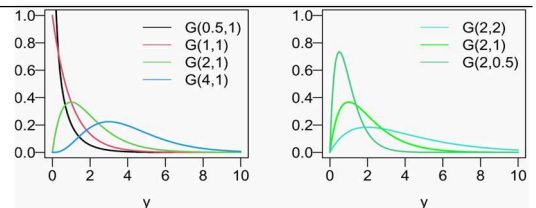
Die Gamma-Verteilung eignet sich zur Modellierung von positiven Größen wie Überlebenszeiten, Wartezeiten, Schadenshöhen.

**Beispiel Höhlenexpedition**

Wie gross ist die Wahrscheinlichkeit, dass Sie beide Taschenlampen zusammen für mindestens 1 Tag Höhlenexpedition reichen?  $P(T > 24) = 1 - P(T \leq 24) = 1 - \text{pgamma}(24, \text{shape} = 2, \text{rate} = 1/24)$

Wie gross ist die Wahrscheinlichkeit, dass Sie beide Taschenlampen genau in der spannenden Phase der Expedition (30h bis 40h nach Beginn) ihren Geist aufgeben?  $P(30 \leq T \leq 40) = P(T \leq 40) - P(T \leq 30)$

$\text{pgamma}(40, \text{shape} = 2, \text{rate} = 1/24) - \text{pgamma}(30, \text{shape} = 2, \text{rate} = 1/24)$



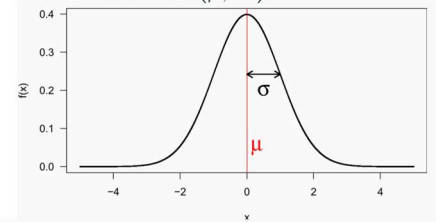
## Normalverteilung - Gaussverteilung

Die meistverwendete statistische Verteilung. Glockenform eine Art "Naturkonstante" in der Welt des Zufalls ist.

**Erkenntnis:** Die Verteilung der Mittelwerte von Zufallszahlen nimmt mit zunehmender Stichprobengröße - unabhängig von der ursprünglichen Verteilung - immer mehr Form einer Glockenkurve an.

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

Notation:  $X \sim \mathcal{N}(\mu, \sigma^2)$



### Dichtefunktion einer normalverteilten Zufallsvariable X

Die Dichtefunktion ist symmetrisch um  $x = \mu$ .

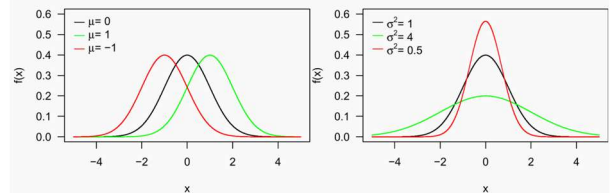
Auf beiden Seiten nähert sie sich asymptotisch der 0 an.

### Einfluss der Parameter

- $\mu$  verschiebt die Glockenkurve entlang der x-Achse. Die Glockenkurve hat stets bei  $\mu$  ihr Maximum.  $\mu$  entspricht dem Erwartungswert.
- Die Form der Dichtefunktion wird über  $\sigma$  bestimmt: Je kleiner  $\sigma$ , desto steiler ist der Gipfel; je grösser, desto flacher.  $\sigma^2$  entspricht der Varianz.
- $N(0, 1)$  heisst Standard-Normalverteilung ( $\mu = 0$  und  $\sigma^2 = 1$ ).

$E(X) = \mu$        $Var(X) = \sigma^2$

Einfluss der Parameter



Bei der Normalverteilung fallen fast alle Werte innerhalb drei Standardabweichungen vom Mittelwert

### 68–95–99.7–Prozent–Regel:

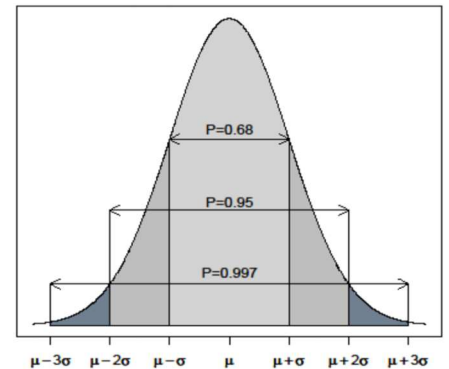
- Innerhalb  $[\mu - \sigma, \mu + \sigma]$  befinden sich ca. 68% der Wahrscheinlichkeitsmasse
- Innerhalb  $[\mu - 2\sigma, \mu + 2\sigma]$  ca. 95%
- Innerhalb  $[\mu - 3\sigma, \mu + 3\sigma]$  ca. 99.7%

Die Regel gilt für alle Normalverteilungen - unabhängig vom Erwartungswert  $\mu$  und der Standardabweichung  $\sigma$ .

**Notation  $X \sim N(\mu, \sigma^2)$**

Erwartungswert: 179 m und Standardabweichung 8  $\Rightarrow X \sim N(179, 8^2)$

- 68% aller Studierenden haben eine Körpergröße zwischen 171 m - 187 m.
- 95% aller Studierenden haben eine Körpergröße zwischen 163 m - 195 m.
- 99.7% aller Studierenden haben eine Körpergröße zwischen 155 m - 203



$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{1}{2} \frac{(z - \mu)^2}{\sigma^2}\right) dz$$

### Kumulative Verteilungsfunktion:

Das Integral über die Dichtefunktion existiert, kann aber nicht in geschlossener Form angegeben werden, d.h. es gibt keine "Formel" für die Verteilungsfunktion.

### Wahrscheinlichkeitsrechnung mit R

- dnorm(x = x, mean = mu, sd = sigma)
- pnorm(x, mean = mu, sd = sigma)
- qnorm(alpha, mean = mu, sd = sigma)
- rnorm(n, mean = mu, sd = sigma)

- Werte der Dichtefunktion an der Stelle x
- Kumulative Verteilungsfunktion
- Quantile
- Zufallszahl

## Lognormal-Verteilung

Die Verteilung ist rechtschief.

Es können nur positive Werte angenommen werden.

Bei Transformation mit dem Logarithmus erhalten wir eine Normalverteilung.  $\rightarrow$  Lognormal-Verteilung

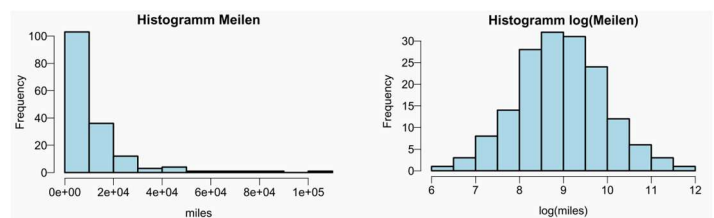
### Verwendung

Die Lognormal-Verteilung ist geeignet für stetige Variablen, welche nur positive Werte annehmen und die eine rechtsschiefe Verteilung aufweisen. Wenn Unterschiede besser durch Verhältnisse als durch Differenzen ausgedrückt werden (multiplikative Überlagerungen von vielen kleinen Einflüssen).

**Notation  $X \sim \text{logN}(\mu, \sigma^2)$**

- Erwartungswert:  $E(X) = e^{(\mu + \frac{\sigma^2}{2})}$
- Varianz:  $Var(X) = e^{(2\mu + \sigma^2)} (e^{\sigma^2} - 1)$
- Für die logarithmierte Zufallsvariable gilt:  $E(\log(X)) = \mu$  und  $Var(\log(X)) = \sigma^2$ .

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



### Beispiel: Meilen Verteilung: $X \sim \text{logN}(8.94, (0.977)^2)$

Mit welcher Wahrscheinlichkeit fliegt eine Person mehr als 20000 Meilen?  $1 - \text{pnorm}(20000, 8.94, 0.977)$

## Transformation von Zufallszahlen

Häufig ist die Verteilung einer Zufallsvariable  $X$  bekannt, man ist aber an der Verteilung einer anderen Zufallsvariable  $Y$  interessiert, die über Transformationsfunktion  $g$  mit  $X$  verknüpft ist.  $Y = g(X)$

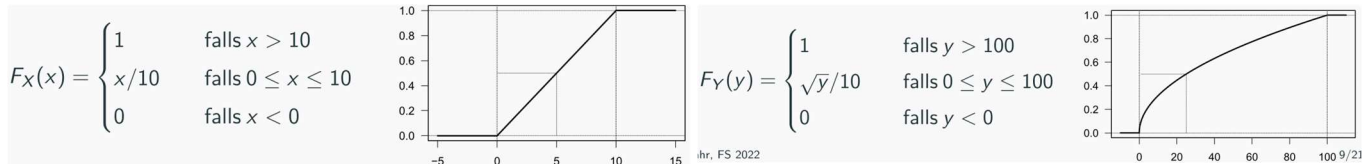
**Gegeben:**  $X$  Zufallsvariable Verteilungsfunktion von  $X$ :  $F_X(x) = P(X \leq x)$  Transformationsfunktion  $g$

**Ziel:** Aus  $X$  und  $g$  die Verteilung der Zufallsvariable  $Y = g(X)$  zu bestimmen.

⇒ Transformation kann die Verteilungsform ändern!

Die kumulative Verteilungsfunktion  $F_Y$  lässt sich aus  $F_X$  bestimmen: Wenn  $g$  **streng monoton zunehmend** (d.h.  $u > v \Rightarrow g(u) > g(v)$ ), dann ist  $g^{-1}$  die Umkehrfunktion von  $g$  (d.h.  $y = g(x)$  und  $x = g^{-1}(y)$ ) und  $F_Y(y) = F_X(g^{-1}(y))$

**Beispiel:** Quadratische Transformation Die Umkehrfunktion von  $y = g(x) = x^2$  ist  $g^{-1}(y) = \sqrt{y}$



Was passiert mit den Grenzen? Die **Grenzen müssen mittransformiert** werden.

Wenn  $g$  **streng monoton abnehmend** (d.h.  $u > v \Rightarrow g(u) < g(v)$ ), dann gilt:  $F_Y(y) = 1 - F_X(g^{-1}(y))$

**Beispiel** Monoton abnehmende Transformation  $y = g(x) = -x \rightarrow$  Umkehrfunktion  $g^{-1}(y) = -y$

### Transformation der Dichte

Transformation gilt egal, ob  $g$  monoton steigend oder fallend ist.

$$\Rightarrow f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|} = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right|$$

**Beispiel:** Quadratische Transformation  $y = g(x) = x^2$  ist  $g^{-1}(y) = \sqrt{y}$

$$f_X(x) = \begin{cases} \frac{1}{10} & \text{falls } 0 \leq x \leq 10 \\ 0 & \text{sonst} \end{cases} \quad \text{Für } 0 \leq y \leq 100: f_Y(y) = f_X(g^{-1}(y)) \cdot \left| \frac{d}{dy} g^{-1}(y) \right| = \frac{1}{10} \cdot \frac{1}{2\sqrt{y}} = \frac{1}{20\sqrt{y}}$$

Beispiel zur Transformation der stetigen Zufallsvariable  $X$  mit Dichtefunktion

Gegeben ist die Transformationsfunktion  $y = g(x)$

$$g(x) = 1 - \frac{\sqrt{4-x^2}}{2} \quad f_X(x) = \begin{cases} \frac{x}{2} & x \in [0, 2] \\ 0 & \text{sonst} \end{cases}$$

1. Umkehrfunktion von  $g(x)$  bestimmen (nach  $y$  umstellen)  $\rightarrow$

$$x = \sqrt{4y(2-y)}$$

2. Verteilungsfunktion von  $f(x)$  bestimmen (aufleiten/integrieren)  $\rightarrow F(x) = \frac{x^2}{4}$

3. Umkehrfunktion in Verteilungsfunktion  $F(x)$  einsetzen (überall wo  $x$  ist)  $\rightarrow \frac{(\sqrt{4y(2-y)})^2}{4} = \frac{4y(2-y)}{4} = y(2-y) = 2y - y^2$

4. Dichtefunktion erhalten durch Ableiten der erhaltenen Funktion  $\rightarrow (2y - y^2)' = 2 - 2y \quad y \in [0, 1]$

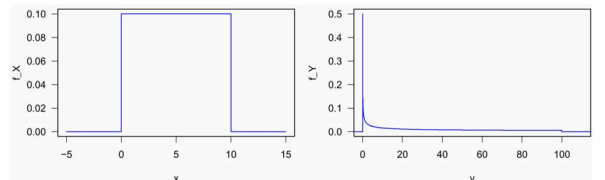
### Transformation von Kennzahlen

Oft möchten wir den **Erwartungswert** von  $Y = g(X)$  berechnen.

In diesem Fall gilt dann allgemein im

**diskreten** Fall  $E(Y) = E(g(X)) = \sum_i g(x_i) P(X = x_i)$

und im **stetigen** Fall  $(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$



Und für die Varianz:  $Var(Y) = E((Y - E(Y))^2) = E((g(X) - E(g(X)))^2) = E(g(X)^2) - E(g(X))^2$

Im Normalfall lassen sich diese Formeln nicht vereinfachen. Die Ausnahme bilden lineare Transformationen.

**Rechenregeln für Kenngrößen:** Ein Unternehmen stellt ein Produkt her und bietet es zum Preis von 6 CHF je Stück an. Die monatliche Absatzmenge sei eine Zufallsvariable  $X$  mit  $E(X) = 1000$  Stück und  $Var(X) = 500$  Stück<sup>2</sup>. Die Kosten je Stück sind durch folgende Gleichung bestimmt:  $Y = 250 + 3X$  Der monatliche Umsatz ergibt sich aufgrund folgender Gleichung:  $Z = 6X$ .

a) Bestimmen Sie den Erwartungswert und die Varianz des monatlichen Umsatzes.

$$E(Z) = E(6X) = 6 \cdot E(X) = 6000$$

$$Var(Z) = Var(6X) = 6^2 \cdot Var(X) = 36 \cdot 500 = 18000$$

b) Bestimmen Sie den Erwartungswert und die Varianz der monatlichen Kosten.

$$E(Y) = E(250 + 3X) = 250 + 3 \cdot E(X) = 3250$$

$$Var(Z) = Var(250 + 3X) = 3^2 \cdot Var(X) = 9 \cdot 500 = 4500$$

c) Bestimmen Sie den Erwartungswert und die Varianz des monatlichen Gewinns.

$$\text{Gewinn } G = \text{Umsatz} - \text{Kosten} = 6X - (250 + 3X) = 3X - 250$$

$$E(G) = E(3X - 250) = 3 \cdot E(X) - 250 = 2750$$

$$Var(Z) = Var(3X - 250) = 3^2 \cdot Var(X) = 9 \cdot 500 = 4500$$

Für den Fall einer **linearen Transformation**  $g(x) = a \cdot x + b$  gilt für den diskreten Fall:

$$E[Y] = a \cdot E[X] + b$$

$$Var[Y] = a^2 \cdot Var(X)$$

Ähnlich kommt man im stetigen Fall zum gleichen Resultat.



Standardisierung

Ziel: Transformierte Zufallsvariable Z mit  $E(Z) = 0$  und  $Var(Z) = 1$ .

**Zweck:** • Vergleich zweier Zufallsvariablen • Anwendung bei vielen statistischen Methoden

**Standardisierung von X:** Gegeben ist Zufallsvariable X mit Erwartungswert  $E(X) = \mu$  und Varianz  $Var(X) = \sigma^2$ .

Transformiert man X mit der linearen Transformationsfunktion  $g(x) = \frac{x-\mu}{\sigma}$  erhält man solch eine Zufallsvariable Z.

Durch die Transformationsfunktion  $g(x) = \frac{x-\mu}{\sigma}$  wird eine beliebige ZV so transformiert, dass ihr Erwartungswert 0 und die Varianz 1 ist.

**Beispiel: Standardnormalverteilung  $N(0, 1)$**

Die Verteilungsfunktion der Standardnormalverteilung wird meist mit  $\Phi$  bezeichnet.

- Sie ist punktsymmetrisch, d.h. es gilt  $\Phi(-x) = 1 - \Phi(x) \rightarrow \text{pnorm}(x)$  und  $1 - \text{pnorm}(x)$
- Eine beliebige Verteilungsfunktion  $N(\mu, \sigma^2)$  lässt sich schreiben als:  $F_Y(y) = \Phi\left(\frac{y-\mu}{\sigma}\right)$

Dichte der Standardnormalverteilung  $\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right)$

Transformationsfunktion  $g(x) = \frac{x-\mu}{\sigma} \Rightarrow g^{-1}(y) = \sigma \cdot y + \mu$

Inversionsmethode

Von Uniform zu Bernoulli

Wir wollen  $x_1, \dots, x_n$  Zufallszahlen einer Bernoulli(p) Verteilung aus  $u_1, \dots, u_n$  Zufallszahlen einer Gleichverteilung  $Unif([0, 1])$  erzeugen.

**Vorgehen:**

1. Ziehe  $u_1, \dots, u_n$  gleichverteilte Zufallsvariablen
2. Ist  $u_i \leq p$ , setzt man  $x_i = 0$ , ansonsten ist  $x_i = 1$  (Beispiel:  $p = 0.2$ ).

$$\text{Transformation: } x_i = g(u_i) = \begin{cases} 0 & u_i \leq 0.2 \\ 1 & u_i > 0.2 \end{cases}$$

Von Uniform zu Exponential

Gegeben: Zufallsvariable  $U \sim Unif([0, 1])$ .

Wahl einer Transformation: Umkehrfunktion der Verteilung

$$F_X(x) = 1 - e^{-\lambda x} \Rightarrow g(x) = F_X^{-1}(x) = (\log(1 - x)) / (-\lambda)$$

Verteilung von  $Y = g(U) \rightarrow F_Y(y) = F_U(g^{-1}(y)) = g^{-1}(y) = 1 - e^{-\lambda x}$

**Allgemein (Inversionsmethode):**

Wenn man auf eine Zufallsvariable  $U \sim Unif([0, 1])$  eine monotone Transformation  $g(x)$  anwendet, so ist die Umkehrfunktion  $g^{-1}$  die Verteilung der entstehenden Zufallsvariable.

**Umgekehrt gilt auch:**

Man kann jede stetige Zufallsvariable X so transformieren, dass eine uniform verteilte Zufallsvariable U entsteht.

Exponentialverteilung nach Uniform

Für  $X \sim Exp(\lambda)$  gilt  $F_X(x) = P(X \leq x) = 1 - e^{-\lambda x}$

Transformationsfunktion:  $F_Y(y) = F_X(g^{-1}(y)) = g(g^{-1}(y)) = y$  mit  $0 \leq y \leq 1$

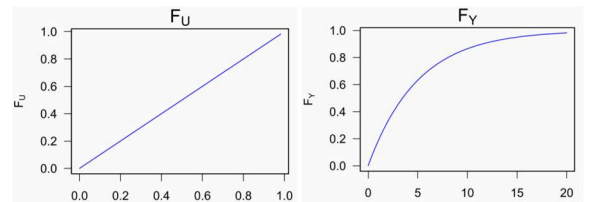
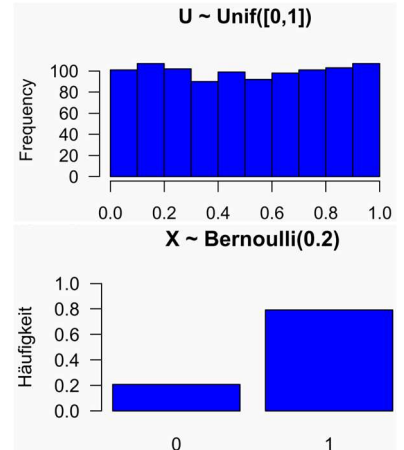
$\rightarrow$  Wir erhalten eine auf  $[0, 1]$  gleichverteilte Zufallsvariable.

Summe von Zufallszahlen

In manchen Fällen folgt die Summe zweier unabhängiger Zufallsvariablen X, Y mit bekannter Verteilung wieder einer Verteilung desselben Typs oder einer anderen bekannten Verteilung:

- X, Y sind normalverteilt  $\Rightarrow S \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$
- X, Y sind Poisson-verteilt  $\Rightarrow S \sim Pois(\lambda_x + \lambda_y)$
- X, Y sind binomialverteilt mit gleichem p  $\Rightarrow S \sim Bin(n_x + n_y, p)$
- X, Y sind exponentialverteilt mit gleichem  $\lambda \Rightarrow S \sim G(2, \lambda)$

Es wird vorausgesetzt, dass die Zufallsvariablen unabhängig sind!



Erwartungswerte und Varianz lassen sich für lineare Transformationen, wie Summen, einfach angeben. Es seien a und b skalare Grössen und S = X + Y Zufallsvariablen. Dann gelten die folgenden Rechenregeln:

Für den **Erwartungswert**

$$E(S) = E(X) + E(Y)$$

$$E(aX + b) = a \cdot E(X) + b$$

Für die **Varianz**

$$Var(S) = Var(X) + Var(Y) + 2 \cdot Cov(X, Y)$$

$$Var(aX + b) = a^2 \cdot Var(X)$$

Der Term  $Cov(X, Y)$  beschreibt den linearen Zusammenhang der beiden Zufallsvariablen. Er ist nur in Ausnahmefällen gleich null. → Falls X und Y unabhängig sind gilt:  $Cov(X, Y) = 0$ , d.h.  $Var(S) = Var(X) + Var(Y)$ .

**Beispiel: AB 12 – Aufgabe 1 Portfolio / für den Rest siehe R Skript**

In einem Portfolio befinden sich 2 Aktien. Wir kennen folgende Kennzahlen:

- Aktie A:  $E(X) = 7$  und  $\sigma(X) = \sqrt{Var(X)} = 25$  70% in Aktie A |  $Cor(X, Y) = -0.4$
- Aktie B:  $E(Y) = 15$  und  $\sigma(Y) = \sqrt{Var(Y)} = 45$  30% in Aktie B | Jahresrendite Portfolio:  $Z = a \cdot X + b \cdot Y$

Was ist die erwartete Rendite des Portfolios?  $E(Z) = a \cdot E(X) + b \cdot E(Y) = 0.7 \cdot 7 + 0.3 \cdot 15 = 9.4$

Was ist die Varianz des Portfolios?  $a^2 \cdot Var(X) + b^2 \cdot Var(Y) + 2 \cdot a \cdot b \cdot \sigma(X) \cdot \sigma(Y) \cdot Cor(X, Y)$

$$0.7^2 \cdot 25^2 + 0.3^2 \cdot 45^2 + 2 \cdot 0.7 \cdot 0.3 \cdot 25 \cdot 45 \cdot (-0.4) = 299.5 \quad \sigma(Z) = \sqrt{Var(Z)} = 17.31$$

### Gesetz der grossen Zahlen

Das Gesetz besagt, dass sich das arithmetische Mittel mit wachsendem n = Stichprobengrösse dem Erwartungswert  $\mu$  nähert. Die Varianz des arithmetischen Mittels ist für feste n proportional zu  $\sigma^2$  und wird mit wachsendem n kleiner. → Man erhält "im Durchschnitt" das richtige Resultat und die Varianz wird mit wachsendem n kleiner.

### Tschebyschow-Ungleichung

Es sei X eine Zufallsvariable mit Erwartungswert  $\mu$  und Varianz  $\sigma^2$ . Dann gilt die Tschebyschow-Ungleichung für jedes feste  $\epsilon > 0$ :  $P(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$

- Die Verteilung geht auf der rechten Seite nur durch die Varianz ein.
- Ungleichung ist auch dann anwendbar, wenn die Verteilung nicht bekannt ist.
- Die obere Schranke kann sehr grob sein, d.h. die exakten Werte für die Wahrscheinlichkeit auf der linken Seite sind oft viel kleiner als die obere Schranke auf der rechten Seite.

Arithmetisches Mittel:  $P(|\bar{X}_n - \mu| > \epsilon) \leq \lim_{n \rightarrow \infty} \frac{\sigma^2}{n \cdot \epsilon^2} = 0$

Wenn wir uns eine beliebig kleine Toleranz  $\epsilon > 0$  vorgeben, so geht die Wahrscheinlichkeit für die Abweichung zwischen  $\bar{X}_n$  und  $\mu$  um mehr als  $\epsilon$  gegen 0, wenn n gross genug ist. Wir sagen dann, dass  $\bar{X}_n$  **stochastisch** gegen  $\mu$  **konvergiert** und schreiben:  $\bar{X}_n \xrightarrow{P} \mu$  für  $n \rightarrow \infty$

### Zentraler Grenzwertsatz

Der Zentrale Grenzwertsatz macht eine Aussage über die Verteilung. Das arithmetische Mittel von  $X_1, X_2, \dots, X_n$  verhält sich mit zunehmenden n wie eine normalverteilte Zufallsvariable mit dem Erwartungswert gleich dem Erwartungswert der einzelnen  $X_i$  und einer gegen 0 gehenden Varianz:  $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$   $\mu = E(X_i)$  und  $\sigma^2 = Var(X_i)$  d.h. für grosses n ist  $\bar{X}_n$  approximativ normalverteilt. Der Zentrale Grenzwertsatz gilt im Wesentlichen unabhängig von der Verteilung der  $X_i$ !

Die Verteilungsfunktion nähert sich für wachsende n an die Verteilungsfunktion der Standardnormalverteilung.

### Satz von Moivre-Laplace

Spezialfall des zentralen Grenzwertsatzes für **binomialverteilte** Zufallsvariablen. Sei  $X \sim \text{Bin}(n, p)$ , dann gilt für grosses n:  $X \sim \mathcal{N}(np, np(1-p))$  **Faustregel:** Ab  $np(1-p) > 9$  ist die Approximation gerechtfertigt.

Es gilt dann  $E(X_i) = p$  und  $Var(X_i) = p(1-p)$

### Bivariate Normalverteilung

**Dichtefunktion  $M_1, M_2$  unabhängig:**  $f_M(m_1, m_2) = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2}} \exp\left(-\frac{(m_1-\mu_1)^2}{2\sigma_1^2} - \frac{(m_2-\mu_2)^2}{2\sigma_2^2}\right) = f_{(M_1)}(m_1)f_{(M_2)}(m_2)$

**Abhängigkeit: Linearer Zusammenhang → Kovarianz** von X und Y:  $Cov(X, Y) = E[(X - E[X])(Y - E[Y])]$

Mass für den Zusammenhang zweier Zufallsvariablen mit gemeinsamer Wahrscheinlichkeitsverteilung.

Standardisiert die Kovarianz, um besser interpretierbares Mass zu erhalten, das nicht von den Einheiten abhängt.

Korrelation von X und Y:  $\rho = Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$

- Es gilt:**
- $-1 \leq Cor(X, Y) \leq 1$ , sofern  $Var(X) > 0$  und  $Var(Y) > 0$ .
  - $|Cor(X, Y)| = 1$  genau dann, wenn  $a \neq 0, b \in \mathbb{R}$  existiert, so dass  $Y = a \cdot X + b$

Die **gemeinsame Dichte** von  $X = (X_1, X_2)$  ist:  $f_X(x_1, x_2) = \frac{1}{2\pi\sqrt{\det[\Sigma]}} \exp\left(-\frac{1}{2}(x_1 - \mu_1, x_2 - \mu_2)\Sigma^{-1}\begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{pmatrix}\right)$

**Erwartungswert**  $E[X] = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$       **Kovarianzmatrix**  $\Sigma = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

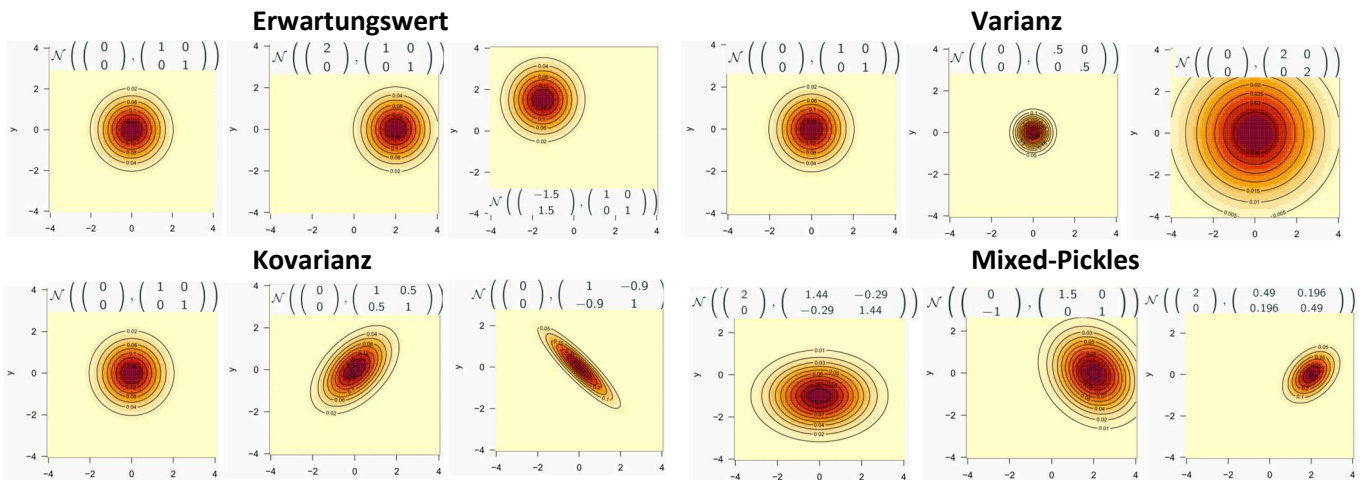
$$\Sigma^{-1} = \frac{1}{\det[\Sigma]} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \quad \det[\Sigma] = \begin{vmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{vmatrix} = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2$$

**Kurznotation:**  $(X_1, X_2) \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$

Spezialfall: Falls  $X_1, X_2$  unkorreliert sind, d.h.  $\rho = 0$  (unabhängig), dann vereinfacht sich die Kovarianzmatrix und

$$(X_1, X_2) \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right) \rightarrow \det[\Sigma] = \sigma_1^2\sigma_2^2 \text{ und } \Sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 \\ 0 & \frac{1}{\sigma_2^2} \end{pmatrix} \rightarrow f_X(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$$

Erwartungswerte und Kovarianzmatrix (Zusammenhang) mit Contour Plot mit Isolinien interpretieren:



### Bedingte Verteilung

Die bedingte Verteilung von  $X_1$  gegeben  $X_2 = x_2$  ist eine Normalverteilung:

$$(X_1 | X_2 = x_2) \sim \mathcal{N}\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), \sigma_1^2(1 - \rho^2)\right)$$

### Randdichte

Randverteilung von  $X_1$  kann man aus der gemeinsamen Dichte bestimmen, indem man über alle Werte von  $X_2$  integriert:  $f_{(X_1)}(x_1) = \int_{-\infty}^{\infty} f_X(x_1, x_2) dx_2$

Wenn  $X = (X_1, X_2) \sim \mathcal{N}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right)$ , dann gilt  $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$  und  $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$

Kovarianz: Mass für den Zusammenhang zweier Zufallsvariablen: Standardisierung  $\rightarrow$  Korrelation:

### Gemeinsame Verteilung

X und Y sind zwei Zufallsvariablen. Ihre gemeinsame Verteilungsfunktion  $F_{XY} : \mathbb{R} \rightarrow [0, 1]$  ist definiert als  $F_{XY}(x, y) = P(X \leq x, Y \leq y) = P(\{X \leq x\} \cap \{Y \leq y\})$  Diese Definitionen sind unabhängig davon, ob die einzelnen Zufallsvariablen alle stetig, alle diskret oder einige stetig und einige diskret sind.

### Diskrete gemeinsame Verteilung

Neben der Verteilungsfunktion ist auch eine gemeinsame Wahrscheinlichkeitsfunktion definiert als  $p(x_i, y_j) = P(X = x_i, Y = y_j)$  die für jedes Paar  $(x_i, y_j)$  die Wahrscheinlichkeit angibt, mit der diese beiden Werte gleichzeitig realisiert werden.

Aus der gemeinsamen Wahrscheinlichkeitsfunktion  $p_{XY}$ , können wir die Wahrscheinlichkeitsfunktion  $p_Y$  von Y (Randverteilung von Y) bestimmen

Die **Randverteilung** von Y erhält man durch das Summieren der Spalten der Tabelle  $p_Y(y) = \sum_{i=1}^{\infty} p_{XY}(x_i, y)$

Analog erhält man die Randverteilung von X durch Summieren der Zeilen der Tabelle:  $p_X(x) = \sum_{j=1}^{\infty} p_{XY}(x, y_j)$

Die bedingte Verteilung von X gegeben Y ist:

wobei wir dies im Fall  $p_Y(y_i) = 0$  als  $p_{X|Y}(x_i | y_i) = 0$  
$$p_{X|Y}(x_i | y_i) = P(X = x_i | Y = y_i) = \frac{P(X = x_i, Y = y_i)}{P(Y = y_i)} = \frac{p_{XY}(x_i, y_i)}{p_Y(y_i)}$$

Stetige gemeinsame Verteilung

Sind X und Y stetige Zufallsvariablen, so kann ihre Verteilung

$$P(X \in [a, b], Y \in [c, d]) = \int_a^b \int_c^d f_{XY}(x, y) dy dx = \int_c^d \int_a^b f_{XY}(x, y) dx dy$$

durch ihre gemeinsame Dichtefunktion beschrieben werden. Dies ist eine **nicht** negative Funktion  $f_{XY}$ , für die gilt Zusammenhang Dichtefunktion  $f_{XY}$  und  $F_{XY}$ :  $f_{XY} = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$

Analog zum diskreten Fall können wir die sogenannte stetige Randdichte von X aus der gemeinsamen Dichte von X und Y bestimmen, wobei über alle Werte von Y integriert, wird:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

Die bedingte Verteilung von X gegeben Y gibt die Verteilung der Zufallsvariablen X an,

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$$

wenn schon bekannt ist, welchen Wert Y angenommen hat:  $f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}$   $f_{X|Y}(x|y) = 0, f_Y(y) = 0$

Unabhängigkeit

Zufallsvariablen  $X_1, X_2, \dots, X_n$  heissen stochastisch unabhängig, wenn die gemeinsame Verteilungsfunktion das Produkt der Randverteilungsfunktionen ist:  $F_{X_1, \dots, X_n}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot F_{X_2}(x_2) \cdot \dots \cdot F_{X_n}(x_n)$

Diskreten Fall gilt die gemeinsame Wahrscheinlichkeitsfunktion:  $p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot \dots \cdot p_{X_n}(x_n)$

und im stetigen Fall für die Dichte  $f_{X_1, \dots, X_n}(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot f_{X_2}(x_2) \cdot \dots \cdot f_{X_n}(x_n)$

Falls keine Unabhängigkeit  $\rightarrow$  Kovarianz und Korrelation

Zwei Studierende gehen unabhängig voneinander zu einem zufällig gleichverteilten Zeitpunkt zwischen 12 und 13 Uhr zur Mensa. Sie haben ausgemacht, dass jeder nach der Ankunft 10 Minuten auf den Anderen wartet. Kommt dieser innerhalb der 10 Minuten, so essen sie gemeinsam zu Mittag. Wie gross ist die Wahrscheinlichkeit, dass die Studenten zusammen essen? Zufallsvariablen:  $X \sim \text{Unif}[0, 1]$  (Zeit ab 12 Uhr gerechnet)  $Y \sim \text{Unif}[0, 1]$

Die gemeinsame Verteilung ergibt sich  $F_{XY}(x, y) = P(X \leq x, Y \leq y) = P(X \leq x) \cdot P(Y \leq y) = F_X(x) \cdot F_Y(y)$

Für die Berechnung der Wahrscheinlichkeit gibt es mehrere Wege:

1. Simulation  $\rightarrow N < 100000$

```
x <- runif(N, min = 0, max = 1)
y <- runif(N, min = 0, max = 1)
sum(x <= y + 1/6 & x >= y - 1/6)/N
```

2. Geometrie

$$1 - \left(\frac{5}{6}\right)^2 = \frac{11}{36}$$

3. Integration der Dichte:

Varianz ( $s_x^2$ )

Mittlere quadratische Abweichung der Beobachtungen vom arithmetischen Mittel ( $\bar{x}$ )

$$s_x^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2 = \left[ \frac{1}{(n-1)} \sum_{i=1}^n x_i^2 \right] - \frac{n}{(n-1)} (\bar{x})^2$$

In R  $\rightarrow \text{var}(A)$  oder  $\text{sd}(A)^2$

Standardabweichung ( $s_x$ )

Quadratwurzel der Varianz. Gleiche Einheit wie Beobachtung (besser interpretierbar)

$$s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2}$$

In R  $\rightarrow \text{sd}(A)$  oder  $\text{sqrt}(\text{var}(A))$

Übungsblatt 7 – Aufgabe 1

$$f(x) = \begin{cases} cx & \text{falls } 2 \leq x \leq 4 \\ 0 & \text{sonst} \end{cases}$$

In einem Betrieb verdienen Aushilfskräfte zwischen 2000 und 4000 CHF brutto. Die Zufallsvariable X soll den Lohn einer Hilfskraft in Tausend CHF beschreiben. X kann als stetig angenommene Lohnstruktur der Aushilfskräfte mit folgender Dichtefunktion  $f(x)$  beschrieben werden: Bestimmen Sie die Konstante c in der Dichtefunktion:

$\rightarrow$  Das Integral über die Dichtefunktion muss stets 1 ergeben:  $\int_2^4 cx dx = \left[ \frac{c \cdot x^2}{2} \right]_2^4 = 8c - 2c = 6c \quad c = \frac{1}{6}$ .

Wie gross ist die Wahrscheinlichkeit, dass eine zufällig herausgegriffene Aushilfskraft zwischen 2.5-3 TCHF verdient?

$$\int_2^4 \frac{1}{6} x dx = \frac{1}{12} x^2 \Big|_{2.5}^3 = \frac{11}{48} = 0.229$$

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_{-\infty}^2 0 dx + \int_2^4 \frac{1}{6} x^2 dx + \int_4^{\infty} 0 dx = \int_2^4 \frac{1}{6} x^2 dx$$

Bestimmen Sie den Erwartungswert des Lohns X gemäss dieser Dichte?

$\rightarrow x \cdot f(x)$  über den ganzen Wertebereich der Zufallsvariablen integrieren  $E(X) = \int_2^4 \frac{1}{6} x^2 dx = \left[ \frac{1}{18} x^3 \right]_2^4 = \frac{1}{18} 4^3 - \frac{1}{18} 2^3 = \frac{56}{18} = 3.111$

Bestimmen Sie die Varianz des Lohns X gemäss dieser Dichte?

Für die Varianz müssen wir  $(x - \mu)^2 \cdot f(x)$  über den ganzen Wertebereich der Zufallsvariablen integrieren.

$$\text{Var}(X) = \int_2^4 \left(x - \frac{56}{18}\right)^2 \cdot \left(\frac{1}{6} x\right) dx = \int_2^4 \left(x^2 - \frac{112}{18} x + \left(\frac{56}{18}\right)^2\right) \cdot \left(\frac{1}{6} x\right) dx = \int_2^4 \left(\frac{1}{6} x^3 - \frac{112}{108} x^2 + \frac{3136}{1944} x\right) dx$$

$$\text{Var}(X) = \left(\frac{1}{24} x^4 - \frac{112}{324} x^3 + \frac{3136}{3888} x^2\right) \Big|_2^4 = \left(\frac{1}{24} 4^4 - \frac{112}{324} 4^3 + \frac{3136}{3888} 4^2\right) - \left(\frac{1}{24} 2^4 - \frac{112}{324} 2^3 + \frac{3136}{3888} 2^2\right) = 0.321$$