

Dipl.-Kfm. Ansgar Ostermann

# Grundwissen Stochastik für die Wertpapieranalyse



## **Vorwort**

Seit vielen Jahren bin ich im Handel mit High Yield Bonds aktiv, dem ohne Anwendung mathematischer Methoden ein wesentliches Fundament fehlen würde. Es entstand daher die Idee, ausgewählte elementare Grundlagen der Stochastik, also der Wahrscheinlichkeitsrechnung und Statistik, in einem Skript zusammenzufassen.

Das vorliegende Skript führt in jene Grundlagen der Stochastik ein, die erforderlich sind, um die wissenschaftliche Fachliteratur zur Wertpapieranalyse nachvollziehen zu können. Es steht dabei nicht der mathematische Formalismus im Vordergrund, sondern die Verdeutlichung von Rechentechniken.

Ich freue mich über Anregungen, Kritik, Fehlerhinweise und Verbesserungsvorschläge, die Sie an folgende E-Mail-Adresse richten können:

[ieff-professional@freenet.de](mailto:ieff-professional@freenet.de)

Berlin, im Mai 2026  
Ansgar Ostermann

# Gliederung

## 1. Mittelwerte

- 1.1. Grundlagen
- 1.2. Modalwert
- 1.3. Median
- 1.4. Arithmetisches Mittel
- 1.5. Geometrisches Mittel
- 1.6. Harmonisches Mittel
- 1.7. Durchschnittliche absolute und durchschnittliche relative Änderung von Zahlenreihen
- 1.8. Übungsaufgaben zu Mittelwerten

## 2. Grundlagen der Wahrscheinlichkeitsrechnung

- 2.1. Definition der Wahrscheinlichkeit nach Laplace
- 2.2. Additionsgesetze der Wahrscheinlichkeitsrechnung
- 2.3. Bedingte Wahrscheinlichkeit
- 2.4. Wahrscheinlichkeit des Komplementärereignisses
- 2.5. Unabhängige Ereignisse
- 2.6. Multiplikationsgesetze der Wahrscheinlichkeitsrechnung

## 3. Zufallsvariablen und Wahrscheinlichkeitsverteilungen

- 3.1. Zufallsvariablen
- 3.2. Wahrscheinlichkeitsverteilungen diskreter Zufallsvariablen
- 3.3. Wahrscheinlichkeitsverteilungen stetiger Zufallsvariablen
- 3.4. Parameter von Wahrscheinlichkeitsverteilungen
  - 3.4.1. Erwartungswert
  - 3.4.2. Varianz und Standardabweichung
- 3.5. Mehrdimensionale Wahrscheinlichkeiten
  - 3.5.1. Grundlagen
  - 3.5.2. Randverteilung
  - 3.5.3. Bedingte Wahrscheinlichkeitsverteilung
  - 3.5.4. Erwartungswert, Varianz, Kovarianz, Korrelationskoeffizient
  - 3.5.5. Unabhängigkeit
- 3.6. Gesetz der großen Zahlen
- 3.7. Übungsaufgaben zu Wahrscheinlichkeitsverteilungen

## 4. Diskrete Wahrscheinlichkeitsverteilungen

- 4.1. Binominalverteilung
- 4.2. Geometrische Verteilung
- 4.3. Hypergeometrische Verteilung
- 4.4. Poissonverteilung

## **5. Stetige Wahrscheinlichkeitsverteilungen**

- 5.1. Gleichverteilung
- 5.2. Exponentialverteilung
- 5.3. Normalverteilung
- 5.4. Logarithmische Normalverteilung
- 5.5. Chi-Quadrat-Verteilung
- 5.6. Studentverteilung

## **6. Schiefe**

## **7. Konfidenzintervalle**

## **8. Grundlagen der Integralrechnung**

# 1. Mittelwerte

## 1.1. Grundlagen

- Die Mittelwerte gehören zu einer wichtigen Klasse von Parametern, den Lageparametern. Sie geben Auskunft über die Lage einer Verteilung auf der verwendeten Skala (= x-Achse).
- Die wichtigsten **Mittelwerte** sind:
  - (i) Modalwert
  - (ii) Median
  - (iii) Arithmetisches Mittel
    - a) Ungewogenes arithmetisches Mittel
    - b) Gewogenes arithmetisches Mittel, Gewichtung mit relativen Häufigkeiten
    - c) Gewogenes arithmetisches Mittel, Gewichtung mit absoluten Häufigkeiten
  - (iv) Geometrisches Mittel
    - a) Ungewogenes geometrisches Mittel
    - b) Gewogenes geometrisches Mittel
  - (v) Harmonisches Mittel
    - a) Ungewogenes harmonische Mittel
    - b) Gewogenes harmonisches Mittel

## 1.2. Modalwert

- Der Modalwert  $\bar{x}_D$  (= Modus, häufigster Wert) ist der Wert in einem Datensatz, der am häufigsten vorkommt. Er ist ein Lageparameter der deskriptiven Statistik, der bei nominalskalierten, ordinalskalierten und metrisch skalierten Daten bestimmt werden kann. Es können mehrere Modalwerte existieren, wenn verschiedene Werte gleich häufig auftreten (bimodale Verteilung, multimodale Verteilung).
- Bestimmung des Modalwertes erfolgt durch einfaches Abzählen der Häufigkeiten.

### Beispiel:

In der Datenreihe

1, 2, 2, 3, 4, 4, 4, 5

ist der Modalwert  $\bar{x}_D = 4$ , da er am häufigsten vorkommt.

- **Nominalskalierte Daten**

Bei nominalskalierten Daten (= nominale Daten) handelt es sich um Daten, die in keinerlei Reihenfolge gebracht werden können – beispielsweise Geschlecht, Haarfarbe, Telefonnummer. Feststellbar ist hier lediglich, ob zwei statistische Einheiten im Hinblick auf ein nominalskaliertes Merkmal die gleichen Ausprägungen aufweisen – d.h. ob etwa beide befragten Personen blond sind oder ob sie über unterschiedliche Haarfarben verfügen. Die Analyse nominaler Daten erfolgt über Häufigkeiten (= Anzahl) und den Modalwert (= häufigster Wert).

### Beispiele:

|                 |                                |
|-----------------|--------------------------------|
| Farben:         | rot, blau, grün                |
| Nationalität:   | deutsch, französisch, spanisch |
| Beruf:          | Lehrer, Ingenieur, Arzt        |
| Familienstand:  | ledig, verheiratet, geschieden |
| Postleitzahlen: | 10117, 12435, 37671            |

- **Ordinalskalierte Daten**

Bei ordinalskalierten Daten (= ordinale Daten) handelt es sich um Daten, die zwar in eine Reihenfolge gebracht werden können ("*größer als*", "*besser als*", ...), da allerdings die Abstände zwischen den einzelnen Werten nicht quantifizierbar sind, kann mit ihnen nicht gerechnet werden, obwohl es sich auf den ersten Blick um "*normale Zahlen*" handelt. Typische Beispiele sind Ranglisten, Schulnoten, Zufriedenheits-Skalen, Dienstgrade.

### Beispiele:

|                 |                                            |
|-----------------|--------------------------------------------|
| Ranglisten:     | erster Platz, zweiter Platz, dritter Platz |
| Schulnoten:     | 1, 2, 3, 4, 5, 6                           |
| Zufriedenheit:  | sehr zufrieden, zufrieden, unzufrieden     |
| Schulabschluss: | Hauptschule, Realschule, Abitur            |

Gütekategorie: A, B, C

- **Metrisch skalierte Daten**

Bei metrischen Daten (auch Kardinaldaten) handelt es sich um Daten, die quantitativ messbare Abstände bzw. Merkmalsausprägungen haben. Sie erlauben präzise Berechnungen (Mittelwert, Standardabweichung, ...).

Beispiele:

Länge, Volumen, Gewicht, Temperatur, Preis, Zeitdauer, ...

- Beispiel:

368 Studenten beurteilen ihre Berufsaussichten.

Möglichkeiten: sehr gut, gut, mittelmäßig, schlecht, sehr schlecht, keine Meinung.

Die meisten Studenten (124) beurteilen ihre Berufsaussichten als sehr schlecht.

Der Modalwert lautet: sehr schlecht.

### 1.3. Median

- Der Median  $\bar{x}_Z$  (= Zentralwert) ist der mittlere Wert in einer Folge von nach aufsteigender Größe sortierten Werten. Ober- bzw. unterhalb des Median liegt jeweils die Hälfte der Werte. Gegenüber dem arithmetischen Mittelwert ist der Median robuster gegenüber Ausreißern.
- Jede Merkmalsausprägung eines wenigstens ordinal messbaren Merkmals, die die geordnete Reihe der Beobachtungswerte in zwei gleiche Teile zerlegt, heißt Median.
- Gibt es  $n$  geordnete Beobachtungswerte und ist  $n$  eine ungerade Zahl, so gibt es einen Median:

$$\bar{x}_Z = x_{\frac{n+1}{2}}$$

#### Beispiel:

Bei einer Klausur wurden von 7 Teilnehmern folgende (geordnete) Punktezahlen erzielt:

8; 8; 11; 15; 20; 26; 30

Berechne den Median.

$$\bar{x}_Z = 15$$

- Gibt es  $n$  geordnete Beobachtungswerte und ist  $n$  eine gerade Zahl, so kommen alle Werte zwischen  $x_{\frac{n}{2}}$  und  $x_{\frac{n}{2}+1}$  als Median infrage. Üblicherweise verwendet man allerdings den Mittelwert der beiden mittleren geordneten Beobachtungswerte als Median.

#### Beispiel:

8 Weitspringer erreichen folgende Sprungweiten [m]:

4,60; 4,80; 4,80; 4,90; 5,00; 5,00; 5,10; 5,30

Berechne den Median.

$$\bar{x}_Z = \frac{4,90 \text{ m} + 5,00 \text{ m}}{2} = 4,95 \text{ m}$$

- Unterscheide:

Modus: Der am häufigsten vorkommende Wert.

Median: Der Wert in der Mitte.

Mittelwert: Der Durchschnitt (Summe aller Werte geteilt durch Anzahl).

## 1.4. Arithmetisches Mittel

- Das arithmetische Mittel  $\bar{x}_A$  (= Durchschnitt) beschreibt den Durchschnittswert einer Datenmenge.
- Die Bestimmung des arithmetischen Mittels ist nur sinnvoll für Merkmale, die auf einer metrischen Skala gemessen werden.
- **Ungewogenes arithmetisches Mittel**

Das ungewogene arithmetische Mittel  $\bar{x}_{AU}$  erhält man, indem man die Summe aller Einzelwerte einer Datenreihe durch die Anzahl  $n$  der Einzelwerte dividiert. Die Einzelwerte werden nicht durch Gewichtungsfaktoren ergänzt.

Gegeben seien  $n$  Einzelwerte  $x_i$  ( $i = 1, 2, 3, \dots, n$ ) eines Merkmals  $X$ . Für das ungewogene arithmetische Mittel gilt dann:

$$\bar{x}_{AU} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$x_i$  = Einzelwert  $i$

$i = 1; 2; 3 \dots$

$n$  = Anzahl der Einzelwerte

Beispiel:

Herr Meyer hat an 7 Tagen einer Woche folgende Mengen Bier getrunken (in Litern):

0,7; 1,6; 2,5; 3,2; 1,6; 2,4; 2,8

Berechne den durchschnittlichen Bierkonsum als arithmetische Mittel.

$$\bar{x}_{AU} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

$$\bar{x}_{AU} = \frac{0,7 + 1,6 + 2,5 + 3,2 + 1,6 + 2,4 + 2,8}{7} = 2,114286 \frac{\text{Liter}}{\text{Tag}}$$

- **Gewogenes arithmetisches Mittel, Gewichtung mit relative Häufigkeiten**

Das einfache arithmetische Mittel muss erforderlichenfalls um relative Häufigkeiten erweitert werden, wenn die Einzelwerte unterschiedliche Relevanz haben und einzelnen Werten somit ein stärkerer Einfluss auf den Mittelwert zukommen muss. Das um relative Häufigkeiten erweiterte gewogene arithmetische Mittel  $\bar{x}_{AGR}$  erhält man, indem man die Einzelwerte einer Datenreihe mit einer relativen Häufigkeit gewichtet und anschließend addiert.

$$\bar{x}_{AGR} = x_1 * h_1 + x_2 * h_2 + \dots + x_n * h_n = \sum_{i=1}^n x_i * h_i$$

$x_i = \text{Einzelwert } i$

$h_i = \text{relative Häufigkeit des Einzelwertes } x_i \text{ mit } h_i \leq 1 \wedge \sum_{i=1}^n h_i = 1$

Die Berechnung eines um relative Häufigkeiten erweiterten arithmetischen Mittels  $\bar{x}_{AR}$  kann beispielsweise erforderlich sein, wenn das im Jahresdurchschnitt eingesetzte Kapital berechnet werden soll, welches im Laufe des Jahres in verschiedenen Assets unterschiedlich lange gebunden ist. Die Summe der Gewichtungsfaktoren ist in diesem Fall die Zahl der Tage eines Jahres.

#### Beispiel:

Herr Meyer hat während eines Jahres folgende Aktien besessen:

360 Tage lang Aktien für insgesamt 4.000 EUR der X AG mit Anschaffungskosten in Höhe von 4.000 EUR.

40 Tage lang Aktien für insgesamt 2.000 EUR der Y AG mit Anschaffungskosten in Höhe von 2.000 EUR.

200 Tage lang Aktien für insgesamt 3.000 EUR der Z AG mit Anschaffungskosten in Höhe von 3.000 EUR.

Berechne das im Jahresdurchschnitt eingesetzte Kapital.

$$\bar{x}_{AGR} = \sum_{i=1}^n x_i * h_i$$

$$\bar{x}_{AGR} = 4.000 \text{ EUR} * \frac{360}{360} + 2.000 \text{ EUR} * \frac{40}{360} + 3.000 \text{ EUR} * \frac{200}{360}$$

$$\bar{x}_{AGR} = \frac{4.000 \text{ EUR} * 360 + 2.000 \text{ EUR} * 40 + 3.000 \text{ EUR} * 200}{360}$$

$$\bar{x}_{AGR} = 5.888, \bar{8} \text{ EUR}$$

- **Gewogenes arithmetischen Mittel, Gewichtung mit absoluten Häufigkeiten**

Das einfache arithmetische Mittel muss erforderlichenfalls um absolute Häufigkeiten erweitert werden, wenn die Einzelwerte unterschiedliche Relevanz haben und einzelnen Werten somit ein stärkerer Einfluss auf den Mittelwert zukommen muss. Das um absolute Häufigkeiten erweiterte gewogene arithmetische Mittel  $\bar{x}_{AGA}$  erhält man, indem man die Einzelwerte einer Datenreihe mit einer absoluten Häufigkeit gewichtet, anschließend addiert und diese Summe durch die Summe der absoluten Häufigkeiten dividiert.

$$\bar{x}_{AGA} = \frac{\sum_{i=1}^n x_i * h_i}{\sum_{i=1}^n h_i}$$

$x_i = \text{Einzelwert } i$

$h_i = \text{absolute Häufigkeit des Einzelwertes } x_i$

### Beispiel 1:

Während eines halben Jahres (= 120 Arbeitstage) wird täglich im Rahmen einer Untersuchung über den Publikumsverkehr beim Sozialamt die Anzahl der persönlich vorsprechenden Antragsteller festgehalten. Folgende Häufigkeitsverteilung hat sich ergeben:

|                          |   |   |    |    |    |    |    |    |    |   |    |                |
|--------------------------|---|---|----|----|----|----|----|----|----|---|----|----------------|
| Anzahl der Antragsteller | 0 | 1 | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9 | 10 |                |
| Anzahl der Tage          | 5 | 4 | 10 | 12 | 20 | 18 | 18 | 12 | 15 | 2 | 4  | $\Sigma = 120$ |

Wie hoch ist die durchschnittliche Anzahl der persönlich vorsprechenden Antragsteller pro Tag?

$$\bar{x}_{AGA} = \frac{\sum_{i=1}^n x_i * h_i}{\sum_{i=1}^n h_i}$$

$$\bar{x}_{AGA} = \frac{0 * 5 + 1 * 4 + 2 * 10 + 3 * 12 + 4 * 20 + 5 * 18 + 6 * 18 + 7 * 12 + 8 * 15 + 9 * 2 + 10 * 4}{120}$$

$$\bar{x}_{AGA} = 5 \text{ Antragsteller}$$

### Beispiel 2:

Der Primärenergieverbrauch (PEV) pro Kopf (in t Steinkohle-Einheiten [SKE]) im Jahr 2000 ist für verschiedene Kontinente und Regionen der Welt in der folgenden Tabelle aufgelistet, ebenso wie die Anteile an der gesamten Weltbevölkerung von 6.058 Millionen Menschen und der Zahl der Einwohner in diesen Regionen:

| Region            | PEV/Kopf | Anteil | Einwohner [Mio.] |
|-------------------|----------|--------|------------------|
| Europa            | 4,5      | 9,51%  | 576              |
| Ehem. UDSSR       | 4,5      | 4,80%  | 291              |
| Nordamerika       | 11,4     | 5,18%  | 314              |
| Lateinamerika     | 1,4      | 8,57%  | 519              |
| Afrika            | 0,5      | 13,11% | 794              |
| Asien, Australien | 1,1      | 58,83% | 3.564            |

Berechne den weltweiten durchschnittlichen Primärenergieverbrauch pro Kopf im Jahr 2000.

$$\bar{x}_{AGA} = \frac{\sum_{i=1}^n x_i * h_i}{\sum_{i=1}^n h_i}$$

$$\bar{x}_{AGA} = \frac{4,5 * 576 + 4,5 * 291 + 11,4 * 314 + 1,4 * 519 + 0,5 * 794 + 1,1 * 3.564}{6.058}$$

$$\bar{x}_{AGA} = 2,067531$$

Alternativ:

$$\bar{x}_{AGR} = \sum_{i=1}^n x_i * h_i$$

$$\bar{x}_{AGR} = 4,5 * 0,0951 + 4,5 * 0,048 + 11,4 * 0,0518 + 1,4 * 0,0857 + 0,5 * 0,1311 + 1,1 * 0,5883$$

$$\bar{x}_{AGR} = 2,06713$$

- **Durchschnittlicher Jahreszinssatz**

Der durchschnittliche Jahreszinssatz ist der Durchschnitt aller einzelnen Jahreszinssätze – eine einfache Kennzahl für die Einträglichkeit (Rentabilität) einer mehrjährigen Geldanlage, deren Zinssatz sich jedes Jahr ändert.

Berechnet wird der durchschnittliche Jahreszinssatz als ungewogenes arithmetisches Mittel der einzelnen Jahreszinssätze: Die einzelnen Jahreszinssätze werden addiert und durch die Anzahl der Jahre geteilt. Dies liefert eine einfache Einschätzung, wie viele Zinsen eine Geldanlage im Durchschnitt pro Jahr erbringt.

Beispiel:

1.000 EUR werden für drei Jahre zinsbringend angelegt. Im ersten Jahr beträgt der Zinssatz 1,0 Prozent, im zweiten Jahr 2,0 Prozent und im dritten Jahr 6,0 Prozent.

Der durchschnittliche Jahreszinssatz ergibt sich aus der Summe der drei Zinssätze (9,0 Prozent), geteilt durch die Anzahl der Jahre (3), und beträgt damit glatte 3,0 Prozent.

Nicht im durchschnittlichen Jahreszins berücksichtigt wird Zinseszins, der entsteht, wenn Zinsen nicht ausgezahlt, sondern dem Anlagekapital gutgeschrieben und anschließend mitverzinst werden.

Auch Steuerabzüge gehen nicht in den durchschnittlichen Jahreszins ein; ebenso wenig, wann die einzelnen Zahlungsströme erfolgen. Es gibt in der Finanzmathematik aber weitere Durchschnittswerte, die diese Details mit berücksichtigen. Dazu gehört z.B. die Rendite (als interner Zinssatz).

## 1.5. Geometrisches Mittel

- Das geometrische Mittel  $\bar{x}_G$  wird angewendet, wenn Einzelwerte sinnvoll durch Multiplikation verknüpft werden können. Dies ist beispielsweise bei der Berechnung durchschnittlicher Wachstumsraten, durchschnittlicher prozentualer Veränderungen oder durchschnittlicher Zinsen der Fall.
- Die Bestimmung des geometrischen Mittels ist nur sinnvoll für Merkmale, die auf einer metrischen Skala gemessen werden. Die Einzelwerte müssen positiv sein, da bei negativen Werten das geometrische Mittel nicht sinnvoll definiert ist.
- **Ungewogenes geometrisches Mittel**

Das ungewogene geometrische Mittel  $\bar{x}_{GU}$  erhält man, wenn das geometrische Mittel aus den Einzelwerten einer Datenreihe ohne Berücksichtigung von Gewichtungsfaktoren berechnet wird.

Gegeben seien  $n$  Einzelwerte  $x_i$  ( $i = 1, 2, 3, \dots, n$ ) eines Merkmals  $X$ . Für das ungewogene geometrische Mittel gilt dann:

$$\bar{x}_{GU} = \sqrt[n]{x_1 * x_2 * \dots * x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

$x_i$  = Einzelwert  $i$

$i = 1; 2; 3 \dots$

$n$  = Anzahl der Einzelwerte

Bei den relevanten Einzelwerten  $x_i$  handelt es sich oftmals um jährliche Wachstumsfaktoren  $q_i$ , die sich wie folgt aus jährlichen Wachstumsraten  $p_i$  [%] ergeben:

$$q_i = 1 + p_i[\%]$$

Der mittlere bzw. durchschnittliche Wachstumsfaktor  $\bar{q}_{GU}$  ist dann gleich dem ungewogenen geometrischen Mittel  $\bar{x}_{GU}$  der jährlichen Wachstumsfaktoren  $q_i$ .

$$\bar{x}_{GU} = \bar{q}_{GU} = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{\prod_{i=1}^n q_i}$$

Für die durchschnittliche relative Wachstumsrate pro Jahr  $\bar{p}$  gilt dann:

$$\bar{x}_{GU} = 1 + \bar{p}$$

$$\bar{p} = \bar{x}_{GU} - 1$$

$$\bar{q}_{GU} = 1 + \bar{p}$$

$$\bar{p} = \bar{q}_{GU} - 1$$

$p_i$  = jährliche Wachstumsrate bzw. jährlicher Zinssatz (z. B. 3% bzw. 0,03)  
 $q_i$  = jährlicher Wachstumsfaktor bzw. jährlicher Zinsfaktor  
 $\bar{p}$  = mittlere bzw. durchschnittliche Wachstumsrate bzw. mittlerer bzw. durchschnittlicher Zinssatz  
 $\bar{q}_{GU}$  = mittlerer bzw. durchschnittlicher Wachstumsfaktor bzw. mittlerer bzw. durchschnittlicher Zinsfaktor

### Beispiel 1:

Ein Unternehmen erzielte in den Jahren 2020 – 2024 folgende Umsätze:

| Jahr | Umsatz<br>[Mio. EUR] |
|------|----------------------|
| 2020 | 2,0                  |
| 2021 | 2,4                  |
| 2022 | 2,9                  |
| 2023 | 2,7                  |
| 2024 | 3,1                  |

Berechne das durchschnittliche relative Umsatzwachstum pro Jahr.

Zur Lösung bestimmt man zunächst die jährlichen Wachstumsraten  $p_i$  [%] in Prozent des Vorjahresumsatzes und ermittelt daraus die jährlichen Wachstumsfaktoren  $q_i = 1 + p_i$ . Anschließend berechnet man aus den jährlichen Wachstumsfaktoren das ungewogene geometrische Mittel, d.h. den mittleren Wachstumsfaktor.

| Jahr | Umsatz<br>[Mio. EUR] | Jährliche Wachstumsrate in Prozent des Vorjahresumsatzes<br>( $p_i$ ) | Wachstumsfaktor<br>( $q_i$ ) |
|------|----------------------|-----------------------------------------------------------------------|------------------------------|
| 2020 | 2,0                  |                                                                       |                              |
| 2021 | 2,4                  | +20,000000%                                                           | 1,200000                     |
| 2022 | 2,9                  | +20,833333%                                                           | 1,208333                     |
| 2023 | 2,7                  | -6,896552%                                                            | 0,931034                     |
| 2024 | 3,1                  | +14,814815%                                                           | 1,148148                     |

$$\bar{x}_{GU} = \bar{q}_{GU} = \sqrt[n]{\prod_{i=1}^n x_i}$$

$$\bar{x}_{GU} = \bar{q}_{GU} = \sqrt[4]{1,2 * 1,208333 * 0,931034 * 1,148148} = 1,11579092$$

$$\bar{p} = \bar{q}_{GU} - 1$$

Durchschnittliches relatives Umsatzwachstum pro Jahr = 11,579092%

### Beispiel 2:

Herr Meyer trägt 1.000 EUR am 01.01.2021 zur Bank. In den ersten beiden Jahren bekommt er 4%, in den folgenden drei Jahren 5,5% und im 6. Jahr 5% Zinsen.

Auf wieviel EUR ist sein Kapital nach 6 Jahren angewachsen?

Bei welcher konstanten Durchschnittsverzinsung hätte Herr Meyer das gleiche Endkapital erhalten?

| Jahr | Jährliche Wachstumsrate in Prozent (p <sub>i</sub> ) | Wachstumsfaktor (q <sub>i</sub> ) |
|------|------------------------------------------------------|-----------------------------------|
| 2021 | +4%                                                  | 1,040                             |
| 2022 | +4%                                                  | 1,040                             |
| 2023 | +5,5%                                                | 1,055                             |
| 2024 | +5,5%                                                | 1,055                             |
| 2025 | +5,5%                                                | 1,055                             |
| 2026 | +5%                                                  | 1,050                             |

$$\bar{x}_{GU} = \sqrt[n]{\prod_{i=1}^n x_i}$$

$$\bar{x}_{GU} = \sqrt[6]{1,04 * 1,04 * 1,055 * 1,055 * 1,055 * 1,05} = 1,0491451068$$

$$\bar{p} = \bar{x}_{GU} - 1$$

$$\text{Konstante Durchschnittsverzinsung} = 4,91451068\%$$

Gegenprobe:

$$K_n = K_0 * (1 + i_{nom})^n$$

$K_n$  = Endkapital

$K_0$  = Anfangskapital, Nominalkapital

$n$  = Laufzeit in Jahren (inkl. gebrochene Jahreszahlen)

$i_{nom}$  = nomineller Jahreszinssatz

$$K_n = 1.000 \text{ EUR} * (1 + 0,0491451068)^6 = 1.333,56244476 \text{ EUR}$$

$$K_n = 1.000 \text{ EUR} * 1,04 * 1,04 * 1,055 * 1,055 * 1,055 * 1,05 = 1.333,56244476 \text{ EUR}$$

- **Gewogenes geometrisches Mittel**

Das gewogene geometrische Mittel  $\bar{x}_{GG}$  erhält man, wenn das geometrische Mittel aus den Einzelwerten einer Datenreihe unter Berücksichtigung von Gewichtungsfaktoren berechnet wird.

$$\bar{x}_{GG} = \sqrt[\sum h]{x_1^{h_1} * x_2^{h_2} * ... * x_n^{h_n}} = \sqrt[\sum h]{\prod_{i=1}^n x_i^{h_i}}$$

$x_i$  = Einzelwert  $i$

$i = 1; 2; 3 \dots$

$n$  = Anzahl der Einzelwerte

$h_i$  = relative oder absolute Häufigkeit des Einzelwertes  $x_i$

Durch die Exponenten kann den Einzelwerten  $x_i$  eine bestimmte Gewichtung zugewiesen werden, die die Berechnung des geometrischen Mittels beeinflusst.

Beispiel:

Berechne das durchschnittliche Wachstum eines Unternehmens über drei Jahre, wobei das Wachstum in jedem Jahr unterschiedlich stark gewichtet wird:

Jahr 1: Wachstum in Höhe von 10%, Gewichtung = 2

Jahr 2: Wachstum in Höhe von 5%, Gewichtung = 1

Jahr 3: Wachstum in Höhe von 15%, Gewichtung = 3

$$\bar{x}_{GG} = \sqrt[\sum h]{\prod_{i=1}^n x_i^{h_i}}$$

$$\bar{x}_{GG} = \sqrt[6]{1,1^2 * 1,05^1 * 1,15^3} = 1,11603555766$$

Durchschnittliche gewichtete Wachstumsrate = 11,603555766%

Alternativ:

$$\bar{x}_{GG} = \sqrt[1]{1,1^{\frac{2}{6}} * 1,05^{\frac{1}{6}} * 1,15^{\frac{3}{6}}} = 1,11603555766$$

Durchschnittliche gewichtete Wachstumsrate = 11,603555766%

## 3.4. Parameter von Wahrscheinlichkeitsverteilungen

### 3.4.1. Erwartungswert

- **Erwartungswert allgemein**

Zur Charakterisierung von Häufigkeitsverteilungen werden in der deskriptiven (= beschreibenden) Statistik Lageparameter (Modalwerte, Median, arithmetisches Mittel, geometrisches Mittel, ...) und Streuungsparameter (Varianz, Standardabweichung, Variationskoeffizient, Spannweite, ...) bestimmt. In ähnlicher Weise können auch für Wahrscheinlichkeitsverteilungen Parameter bestimmt werden.

Den Erwartungswert kann man sich – zumindest für diskrete Zufallsvariablen – am arithmetischen Mittel aus der deskriptiven Statistik veranschaulichen.

Der Erwartungswert  $E(X)$  bzw.  $\mu_X$  ist eine statistische Kennzahl, die den langfristigen, durchschnittlichen Wert einer Zufallsvariablen bei häufiger Wiederholung eines Experiments angibt. Er wird berechnet, indem man alle möglichen Ereignisse mit ihren Eintrittswahrscheinlichkeiten gewichtet und aufsummiert.

$$E(X) = \mu_X = \sum_{i=1}^n x_i * P(X = x_i)$$

Der Erwartungswert  $E(X)$  ergibt sich somit als Summe der Produkte aus Wert ( $x_i$ ) und Wahrscheinlichkeit  $P(X = x_i)$ .

Der Erwartungswert ist ein theoretischer Mittelwert und nicht zwingend ein Wert, der tatsächlich bei einem Versuch eintritt.

Ein Erwartungswert von 0 in einem Spiel bedeutet langfristig weder Gewinn noch Verlust (fair), positiv bedeutet Gewinn, negativ Verlust.

#### **Rechenregeln im Kontext Erwartungswert:**

$$E(A + B) = E(A) + E(B) \Rightarrow \text{Linearität des Erwartungswertes}$$

$$E(c * A) = c * E(A)$$

Beispiel:

Würfel

$$E(X) = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = 3,5$$

- **Erwartungswert diskreter Zufallsvariablen**

Bei einem endlichen Zufallsversuch seien  $x_i$  ( $i = 1, \dots, n$ ) mögliche Zufallsereignisse, denen eine bestimmte Wahrscheinlichkeit  $f_X(x_i)$  zuzuordnen ist. Wiederholt man das zugrundeliegende Zufallsexperiment sehr oft und stellt jedesmal den Wert  $x_i$  fest, so erwartet man als Mittelwert den Wert  $E(X)$ . Deshalb nennt man  $E(X)$  auch den Mittelwert der Zufallsgröße  $x_i$ . Für den Erwartungswert einer diskreten Zufallsvariablen gilt:

$$E(X) = \sum_{i=1}^n x_i * f_X(x_i)$$

Beispiel 1:

Es wird ein Würfel geworfen. Die Wahrscheinlichkeit für die Augenzahl  $i$  beträgt dann  $f_X(i) = \frac{1}{6}$  für  $i = 1, \dots, 6$ . Für den Erwartungswert ergibt sich:

$$E(X) = \sum_{i=1}^6 i * f_X(i)$$

$$E(X) = 1 * \frac{1}{6} + 2 * \frac{1}{6} + 3 * \frac{1}{6} + 4 * \frac{1}{6} + 5 * \frac{1}{6} + 6 * \frac{1}{6} = 3,5$$

Beispiel 2:

Eine diskrete Zufallsvariable  $X$  hat folgende Wahrscheinlichkeitsverteilung:

|            |     |     |     |     |     |
|------------|-----|-----|-----|-----|-----|
| $x_i$      | 2   | 3   | 5   | 8   | 9   |
| $f_X(x_i)$ | 0,1 | 0,4 | 0,2 | 0,1 | 0,2 |

$$E(X) = \sum_i^n x_i * f_X(x_i)$$

$$E(X) = 2 * 0,1 + 3 * 0,4 + 5 * 0,2 + 8 * 0,1 + 9 * 0,2 = 5$$

Beispiel 3:

Zwei Spieler A und B würfeln abwechseln mit je zwei Würfeln. Dabei ist folgende Spielregel vereinbart:

A bekommt von B 3 EUR, wenn B 10 und mehr Augen würfelt.

B bekommt von A 1 EUR, wenn A weniger als zehn Augen würfelt.

Untersuche, ob die Spielregel beiden Spielern gleiche Gewinnchancen gewährt.

$$P(\geq 10) = P((4; 6) \cup (6; 4) \cup (5; 5) \cup (5; 6) \cup (6; 5) \cup (6; 6))$$

$$P(\geq 10) = P(4) * P(6) + P(6) * P(4) + P(5) * P(5) + P(5) * P(6) + P(6) * P(5) + P(6) * P(6)$$

$$P(\geq 10) = \frac{1}{6} * \frac{1}{6} + \frac{1}{6} * \frac{1}{6} + \frac{1}{6} * \frac{1}{6} + \frac{1}{6} * \frac{1}{6} + \frac{1}{6} * \frac{1}{6} + \frac{1}{6} * \frac{1}{6} = \frac{6}{36} = \frac{1}{6}$$

$$P(< 10) = P((1; 1) \cup (1; 2) \cup (1; 3) \cup (1; 4) \cup (1; 5) \cup (1; 6) \cup \\ P(2; 1) \cup (2; 2) \cup (2; 3) \cup (2; 4) \cup (2; 5) \cup (2; 6) \cup \\ P(3; 1) \cup (3; 2) \cup (3; 3) \cup (3; 4) \cup (3; 5) \cup (3; 6) \cup \\ P(4; 1) \cup (4; 2) \cup (4; 3) \cup (4; 4) \cup (4; 5) \cup \\ P(5; 1) \cup (5; 2) \cup (5; 3) \cup (5; 4) \cup \\ P(6; 1) \cup (6; 2) \cup (6; 3))$$

$$P(< 10) = \left(\frac{1}{6} * \frac{1}{6}\right) * 30 = \frac{5}{6}$$

$$E\{\text{Gewinn des Spielers A}\} = \frac{1}{6} * 3 - \frac{5}{6} * 1 = -\frac{1}{3}$$

$$E\{\text{Gewinn des Spielers B}\} = \frac{5}{6} * 1 - \frac{1}{6} * 3 = +\frac{1}{3}$$

$$E\{\text{Gewinn des Spielers B}\} > E\{\text{Gewinn des Spielers A}\} \Rightarrow$$

*Es ist kein faires Spiel.*

- **Erwartungswert stetiger Zufallsvariablen**

Für den Erwartungswert einer stetigen Zufallsvariablen gilt:

$$E(X) = \int_{-\infty}^{+\infty} x * f_X(x) dx$$

Beispiel:

Gegen sei folgende Dichtefunktion:

$$f_X(x) = \begin{cases} 4x^3 & \text{für } 0 < x < 1 \\ 0 & \text{sonst} \end{cases}$$

Bestimme  $E(X)$ .

$$E(X) = \int_{-\infty}^{+\infty} x * f_X(x) dx$$

$$E(X) = \int_0^1 x * 4x^3 \, dx = \int_0^1 4x^4 \, dx = \left[ \frac{4}{5} x^5 \right]_0^1 = 0,8 - 0 = 0,8$$

### 3.4.2. Varianz und Standardabweichung

- **Varianz allgemein**

Grundsätzlich ist die Varianz  $VAR(X)$  bzw.  $\sigma_X^2$  die Summe der quadrierten Differenzen zwischen den jeweiligen Einzelwerten und dem Mittelwert dividiert durch die Anzahl der Einzelwerte.

**Unkorrigierte Varianz:**

$$VAR(X) = \sigma_X^2 = \frac{1}{n} * \sum_{i=1}^n (x_i - \bar{x})^2$$

Die Varianz ist ein Streuungsparameter. Eine hohe Varianz bedeutet starke Streuung um den Mittelwert, eine niedrige Varianz bedeutet enge Verteilung um den Mittelwert

Bei bekannter Grundgesamtheit kommt in vorgenannter Formel  $\frac{1}{n}$  zur Anwendung.

Bei Stichproben kommt in vorgenannter Formel grundsätzlich  $\frac{1}{n-1}$  zur Anwendung (Bessel-Korrektur).

**Korrigierte Varianz :**

$$VAR(X) = \sigma_X^2 = \frac{1}{n-1} * \sum_{i=1}^n (x_i - \bar{x})^2$$

Die unkorrigierte Varianz ist ein verzerrter Schätzer für die Varianz der Grundgesamtheit. Sie unterschätzt die wahre Varianz, insbesondere bei kleinen Stichproben, denn in der Regel ist die Streuung innerhalb einer Stichprobe etwas geringer als in der gesamten Population, da extreme Werte oft nicht in der Stichprobe enthalten sind.

Die Bessel-Korrektur ist eine statistische Anpassung, die angewendet wird, um eine verzerrte Schätzung der Stichprobenvarianz zu korrigieren. Sie wird verwendet, weil die unkorrigierte Stichprobenvarianz dazu neigt, die wahre Varianz der Grundgesamtheit zu unterschätzen. Das ist vor allem bei kleinen Stichproben der Fall. Die Bessel-Korrektur besteht darin, den Nenner von  $n$  auf  $n - 1$  zu ändern, wodurch die Varianz größer wird.

**Rechenregeln im Kontext Varianz:**

$$VAR(A + B) = VAR(A) + VAR(B) \Rightarrow \text{Linearität der Varianz}$$

- **Varianz diskreter Zufallsvariablen**

**Grundformel (= Abweichungsquadrate):**

$$VAR(X) = \sum_{i=1}^n (x_i - E(X))^2 * f_X(x_i)$$

**Verschiebungssatzformel für die Varianz (= Rechenformel):**

$$VAR(X) = E(X^2) - (E(X))^2 = \sum_{i=1}^n x_i^2 * f_X(x_i) - (E(X))^2$$

**Überleitung:**

$$VAR(X) = \sum_{i=1}^n (x_i - E(X))^2 * f_X(x_i)$$

$$VAR(X) = \sum_{i=1}^n (x_i^2 - 2 * x_i * E(X) + (E(X))^2) * f_X(x_i)$$

$$VAR(X) = \sum_{i=1}^n (x_i^2 * f_X(x_i) - 2 * x_i * E(X) * f_X(x_i) + (E(X))^2 * f_X(x_i))$$

$$VAR(X) = \sum_{i=1}^n x_i^2 * f_X(x_i) - \sum_{i=1}^n 2 * x_i * E(X) * f_X(x_i) + \sum_{i=1}^n (E(X))^2 * f_X(x_i)$$

$$+ \sum_{i=1}^n x_i^2 * f_X(x_i) = \text{Erwartungswert von } X^2 = +E(X^2)$$

$$- \sum_{i=1}^n 2 * x_i * E(X) * f_X(x_i) = -2 * E(X) * \sum_{i=1}^n x_i * f_X(x_i) = -2 * E(X) * E(X) = -2 * (E(X))^2$$

$$+ \sum_{i=1}^n (E(X))^2 * f_X(x_i) = (E(X))^2 * \sum_{i=1}^n f_X(x_i) = + (E(X))^2 * 1$$

$$VAR(X) = E(X^2) - 2 * (E(X))^2 + (E(X))^2 = E(X^2) - (E(X))^2$$

$$VAR(X) = E(X^2) - (E(X))^2 = \sum_{i=1}^n x_i^2 * f_X(x_i) - (E(X))^2$$

Der Vorteil der Rechenformel liegt darin, dass man nicht für jeden einzelnen Wert  $x_i$  die Abweichung vom Erwartungswert berechnen muss.

- **Varianz stetiger Zufallsvariablen**

**Abweichungsquadrate (Grundformel):**

$$VAR(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 * f_X(x) dx$$

**Verschiebungssatz (Rechenformel):**

$$VAR(X) = \int_{-\infty}^{+\infty} x^2 * f_X(x) dx - (E(X))^2$$

- **Standardabweichung**

Die Quadratwurzel aus der Varianz heißt Standardabweichung  $\sigma_X$ :

$$\sigma_X = \sqrt{VAR(X)} = \sqrt{\sigma_X^2}$$

Die Standardabweichung zeigt die durchschnittliche Entfernung der Datenpunkte vom Mittelwert.

Der Vorteil der Standardabweichung gegenüber der Varianz liegt in ihrer besseren Interpretierbarkeit, da sie in der gleichen Maßeinheit wie die ursprünglichen Daten angegeben wird. Während die Varianz quadrierte Werte liefert, ermöglicht die Standardabweichung (Wurzel der Varianz) direkte, durchschnittliche Aussagen über die Streuung um den Mittelwert.

Werden beispielsweise Daten in Metern gemessen, wird die Standardabweichung ebenfalls in Metern angegeben, während die Varianz in Quadratmetern angegeben wird.

Die Standardabweichung ist qua Definition stets  $\geq 0$ , da sie als die positive Quadratwurzel aus der Varianz definiert ist. Obwohl die Quadratwurzel aus einer positiven Zahl mathematisch gesehen zwei Ergebnisse haben kann (ein positives und ein negatives, z.B.  $\sqrt{4} = \pm 2$ ) wird in der Statistik nur der positive Wert für die Standardabweichung verwendet.

Bei der Berechnung der Varianz werden die Abweichungen der Messwerte vom Mittelwert quadriert:  $(x_i - \bar{x})^2$ . Da das Quadrat einer reellen Zahl niemals negativ ist – unabhängig davon, ob die ursprüngliche Abweichung positiv oder negativ war – ist auch die Summe dieser Quadrate – und damit die Varianz – immer positiv oder 0. Die Standardabweichung  $\sigma = \sqrt{VAR(X)}$  soll den durchschnittlichen Abstand der Werte vom Mittelwert beschreiben. Ein negativer Abstand macht daher keinen Sinn.

- Beispiel 1:

Berechne Varianz (über Grund- und Rechenformel) und Standardabweichung des Würfel experiments.

$$VAR(X) = \sum_{i=1}^n (x_i - E(X))^2 * f_X(x_i)$$

$$VAR(X) = \sum_{i=1}^6 (x_i - 3,5)^2 * \frac{1}{6}$$

$$VAR(X) = (1 - 3,5)^2 * \frac{1}{6} + (2 - 3,5)^2 * \frac{1}{6} + (3 - 3,5)^2 * \frac{1}{6} + (4 - 3,5)^2 * \frac{1}{6} + (5 - 3,5)^2 * \frac{1}{6} + (6 - 3,5)^2 * \frac{1}{6} = 2,91\bar{6}$$

$$VAR(X) = E(X^2) - (E(X))^2 = \sum_{i=1}^n x_i^2 * f_X(x_i) - (E(X))^2$$

$$VAR(X) = \left(1^2 * \frac{1}{6} + 2^2 * \frac{1}{6} + 3^2 * \frac{1}{6} + 4^2 * \frac{1}{6} + 5^2 * \frac{1}{6} + 6^2 * \frac{1}{6}\right) - 3,5^2 = 2,91\bar{6}$$

$$\sigma_X = \sqrt{VAR(X)}$$

$$\sigma_X = \sqrt{2,91\bar{6}} = 1,70782512766$$

- Beispiel 2:

Eine diskrete Zufallsvariable X hat folgende Wahrscheinlichkeitsverteilung:

|            |     |     |     |     |     |
|------------|-----|-----|-----|-----|-----|
| $x_i$      | 2   | 3   | 5   | 8   | 9   |
| $f_X(x_i)$ | 0,1 | 0,4 | 0,2 | 0,1 | 0,2 |

Berechne Varianz (über Grund- und Rechenformel) und Standardabweichung

$$E(X) = \sum_i^n x_i * f_X(x_i)$$

$$E(X) = 2 * 0,1 + 3 * 0,4 + 5 * 0,2 + 8 * 0,1 + 9 * 0,2 = 5$$

$$VAR(X) = \sum_{i=1}^n (x_i - E(X))^2 * f_X(x_i)$$

$$VAR(X) = \sum_{i=2}^9 (x_i - 5)^2 * f_X(x_i)$$

$$VAR(X) = (2 - 5)^2 * 0,1 + (3 - 5)^2 * 0,4 + (5 - 5)^2 * 0,2 + (8 - 5)^2 * 0,1 + (9 - 5)^2 * 0,2 = 6,6$$

$$VAR(X) = E(X^2) - (E(X))^2 = \sum_{i=1}^n x_i^2 * f_X(x_i) - (E(X))^2$$

$$VAR(X) = (2^2 * 0,1 + 3^2 * 0,4 + 5^2 * 0,2 + 8^2 * 0,1 + 9^2 * 0,2) - 5^2 = 6,6$$

$$\sigma_X = \sqrt{VAR(X)}$$

$$\sigma_X = \sqrt{6,6} = 2,56904651573$$

### Beispiel 3:

Gegeben sei folgende Dichtefunktion:

$$f_X(x) = \begin{cases} 4x^3 & \text{für } 0 < x < 1 \\ 0 & \text{sonst} \end{cases}$$

Berechne die Varianz der stetigen Zufallsvariable.

$$E(X) = \int_{-\infty}^{+\infty} x * f_X(x) dx$$

$$E(X) = \int_0^1 x * 4x^3 dx = \int_0^1 4x^4 dx = \left[ \frac{4}{5} x^5 \right]_0^1 = \frac{4}{5} - 0 = \frac{4}{5}$$

$$VAR(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 * f_X(x) dx$$

$$VAR(X) = \int_0^1 \left( x - \frac{4}{5} \right)^2 * 4x^3 dx = \int_0^1 \left( x^2 - \frac{8}{5}x + \frac{16}{25} \right) * 4x^3 dx$$

$$VAR(X) = \int_0^1 4x^5 - \frac{32}{5}x^4 + \frac{64}{25}x^3$$

$$VAR(X) = \left[ \frac{4}{6}x^6 - \frac{32}{25}x^5 + \frac{64}{100}x^4 \right]_0^1 = \frac{2}{75}$$

$$VAR(X) = \int_{-\infty}^{+\infty} x^2 * f_X(x) dx - (E(X))^2$$

$$VAR(X) = \int_0^1 x^2 * 4x^3 dx - \left(\frac{4}{5}\right)^2 = \int_0^1 4x^5 dx - \frac{16}{25}$$

$$VAR(X) = \left[ \frac{4}{6}x^6 \right]_0^1 - \frac{16}{25} = \left(\frac{4}{6} - 0\right) - \frac{16}{25} = \frac{2}{75}$$

Beispiel 4:

Gegeben sei folgende Dichtefunktion

$$f_X(x) = \begin{cases} 0 & \text{für } x < 0 \\ \frac{1}{4}x & \text{für } 0 \leq x < 2 \\ 1 - \frac{1}{4}x & \text{für } 2 \leq x \leq 4 \\ 0 & \text{für } x > 4 \end{cases}$$

Berechne die Varianz der stetigen Zufallsvariable.

$$E(X) = \int_{-\infty}^{+\infty} x * f_X(x) dx$$

$$E(X) = \int_0^2 x * \frac{1}{4}x dx + \int_2^4 x * \left(1 - \frac{1}{4}x\right) dx = \int_0^2 \frac{1}{4}x^2 dx + \int_2^4 \left(-\frac{1}{4}x^2 + x\right) dx$$

$$E(X) = \left[ \frac{1}{12}x^3 \right]_0^2 + \left[ -\frac{1}{12}x^3 + \frac{1}{2}x^2 \right]_2^4 = \left(\frac{2}{3} - 0\right) + \left(\frac{8}{3} - \frac{4}{3}\right) = 2$$

$$VAR(X) = \int_{-\infty}^{+\infty} x^2 * f_X(x) dx - (E(X))^2$$

$$VAR(X) = \left( \int_0^2 x^2 * \frac{1}{4} x dx + \int_2^4 x^2 * \left(1 - \frac{1}{4} x\right) dx \right) - 2^2$$

$$VAR(X) = \left( \int_0^2 \frac{1}{4} x^3 dx + \int_2^4 \left(-\frac{1}{4} x^3 + x^2\right) dx \right) - 4$$

$$VAR(X) = \left( \left[ \frac{1}{16} x^4 \right]_0^2 + \left[ -\frac{1}{16} x^4 + \frac{1}{3} x^3 \right]_2^4 \right) - 4$$

$$VAR(X) = \left( (1 - 0) + \left( \frac{16}{3} - \frac{5}{3} \right) \right) - 4 = \left( 1 + \frac{11}{3} \right) - 4 = \frac{2}{3}$$

#### Beispiel 5:

Gegeben sei folgende Verteilungsfunktion:

$$F_X(x) = \begin{cases} 0 & \text{für } x < 2 \\ -\frac{1}{4}x^2 + 2x - 3 & \text{für } 2 \leq x \leq 4 \\ 1 & \text{für } x > 4 \end{cases}$$

Bestimme die Dichtefunktion der Zufallsvariablen X, den Erwartungswert  $E(X)$  und die Varianz  $VAR(X)$ .

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

$$E(X) = \int_{-\infty}^{+\infty} x * f_X(x) dx$$

$$E(X) = \int_2^4 x * \left(-\frac{1}{2}x + 2\right) dx = \int_2^4 \left(-\frac{1}{2}x^2 + 2x\right) dx$$

$$E(X) = \left[ -\frac{1}{6}x^3 + x^2 \right]_2^4 = \frac{16}{3} - \frac{8}{3} = \frac{8}{3}$$

$$VAR(X) = \int_{-\infty}^{+\infty} x^2 * f_X(x) dx - (E(X))^2$$

$$VAR(X) = \int_2^4 x^2 * \left(-\frac{1}{2}x + 2\right) dx - \left(\frac{8}{3}\right)^2$$

$$VAR(X) = \int_2^4 \left(-\frac{1}{2}x^3 + 2x^2\right) dx - \left(\frac{8}{3}\right)^2$$

$$VAR(X) = \left[-\frac{1}{8}x^4 + \frac{2}{3}x^3\right]_2^4 - \left(\frac{8}{3}\right)^2$$

$$VAR(X) = \left(\frac{32}{3} - \frac{10}{3}\right) - \left(\frac{8}{3}\right)^2 = \frac{2}{9}$$

## 3.5. Mehrdimensionale Wahrscheinlichkeiten

### 3.5.1. Grundlagen

- Bisher wurden einzelne Zufallsvariablen und ihre Wahrscheinlichkeitsverteilungen sowie Parameter dieser Wahrscheinlichkeitsverteilungen betrachtet. Es handelt sich dabei um **eindimensionale Zufallsvariablen**.

Im Folgenden soll die gemeinsame Verteilung mehrerer Zufallsvariablen betrachtet werden. Es handelt sich dabei um **mehrdimensionale Zufallsvariablen**.

Im Folgenden werden aus Gründen der praktischen Relevanz nur Paare von diskreten Zufallsvariablen vertiefend betrachtet.

- **Gemeinsame Wahrscheinlichkeitsfunktion**

Gegeben seien zwei diskrete Zufallsvariablen  $X$  und  $Y$  mit den Ausprägungen  $x_1, x_2, \dots$  und  $y_1, y_2, \dots$ . Die gemeinsame Wahrscheinlichkeitsfunktion  $f(x_i, y_j)$  – genauer:  $f_{XY}(x_i, y_j)$ ; zweidimensionale bzw. gemeinsame Wahrscheinlichkeitsverteilung – der Zufallsvariablen  $X$  und  $Y$  gibt dann die Wahrscheinlichkeit dafür an, dass die Zufallsvariable  $X$  den Wert  $x_i$  und die Zufallsvariable  $Y$  gleichzeitig den Wert  $y_j$  annimmt.

Für konkrete  $x_i$  und  $y_j$  gibt  $f_{XY}(x_i, y_j)$  den gemeinsamen Wahrscheinlichkeitswert in der gemeinsamen Wahrscheinlichkeitsfunktion an.

Es werden also Paare von Zufallsvariablen  $(X; Y)$  betrachtet, die man als zweidimensionale Zufallsvariablen bezeichnet.

Die gemeinsame Wahrscheinlichkeitsverteilung einer diskreten zweidimensionalen Zufallsvariable kann beschrieben werden durch ihre gemeinsame Wahrscheinlichkeitsfunktion, die wie folgt definiert ist:

$$f_{XY}(x_i, y_j) = P_{XY}(X = x_i, Y = y_j) \quad i, j = 1, 2, \dots$$

Für diese gemeinsame Wahrscheinlichkeitsfunktion gilt:

$$f_{XY}(x_i, y_j) \geq 0$$

und

$$\sum_{i=1}^n \sum_{j=1}^m f_{XY}(x_i, y_j) = 1$$

Die gemeinsame Wahrscheinlichkeitsfunktion wird meistens tabellarisch in Kontingenztafeln dargestellt.

### 3.5.4. Erwartungswert, Varianz, Kovarianz, Korrelationskoeffizient

- Erwartungswert der Randverteilung

Erwartungswertbildung kann auch bei mehrdimensionalen Wahrscheinlichkeitsverteilungen erfolgen.

Im zweidimensionalen Fall mit zwei diskreten Zufallsvariablen  $X$  und  $Y$  und der gemeinsamen Wahrscheinlichkeitsfunktion  $f_{XY}(x_i, y_j)$  ist der Erwartungswert  $E(X)$  der Zufallsvariablen  $X$  wie folgt definiert:

$$E(X) = \sum_{i=1}^n \sum_{j=1}^m x_i * f_{XY}(x_i, y_j)$$

Im zweidimensionalen Fall mit zwei diskreten Zufallsvariablen  $X$  und  $Y$  und der gemeinsamen Wahrscheinlichkeitsfunktion  $f_{XY}(x_i, y_j)$  ist der Erwartungswert  $E(Y)$  der Zufallsvariablen  $Y$  wie folgt definiert:

$$E(Y) = \sum_{i=1}^n \sum_{j=1}^m y_j * f_{XY}(x_i, y_j)$$

Beispiel:

Gegeben sei eine Maschine, die zwei besonders störanfällige Bauteile  $B_1$  und  $B_2$  besitzt.  $X$  sei die Anzahl der Ausfälle pro Tag von  $B_1$ .  $Y$  sei die Anzahl der Ausfälle pro Tag von  $B_2$ . Nimmt man an, dass  $X$  nur die Werte  $x_1 = 0, x_2 = 1, x_3 = 2$  und  $Y$  nur die Werte  $y_1 = 0, y_2 = 1, y_3 = 2$  annehmen kann, dann lässt sich folgende gemeinsame Wahrscheinlichkeitsfunktion darstellen:

| $x_i \backslash y_j$ | 0    | 1    | 2    |
|----------------------|------|------|------|
| 0                    | 0,30 | 0,14 | 0,02 |
| 1                    | 0,18 | 0,10 | 0,02 |
| 2                    | 0,12 | 0,06 | 0,06 |

Berechne  $E(X)$  und  $E(Y)$ .

$$E(X) = \sum_{i=1}^n \sum_{j=1}^m x_i * f_{XY}(x_i, y_j)$$

$$\begin{aligned} E(X) &= 0 * 0,30 + 0 * 0,14 + 0 * 0,02 + \\ &\quad 1 * 0,18 + 1 * 0,10 + 1 * 0,02 + \\ &\quad 2 * 0,12 + 2 * 0,06 + 2 * 0,06 = 0,78 \end{aligned}$$

$$E(Y) = \sum_{i=1}^n \sum_{j=1}^m y_j * f_{XY}(x_i, y_j)$$

$$E(Y) = 0 * 0,30 + 0 * 0,18 + 0 * 0,12 + \\ 1 * 0,14 + 1 * 0,10 + 1 * 0,06 + \\ 2 * 0,02 + 2 * 0,02 + 2 * 0,06 = 0,5$$

Wenn  $f_{XY}(x_i, y_j)$  eine gemeinsame Wahrscheinlichkeitsfunktion  $P(X = x_i; Y = y_j)$  ist, dann ist die Summe über alle möglichen Werte von  $y_j$  die Randwahrscheinlichkeitsfunktion  $P(X = x_i)$ , welche kurz als  $f_X(x_i)$  bezeichnet wird. Die Wahrscheinlichkeit für ein bestimmtes  $x_i$  ist daher gleich der Wahrscheinlichkeit über alle  $j$  für dieses gegebene  $x_i$ :

$$f_X(x_i) = \sum_{j=1}^m f_{XY}(x_i, y_j)$$

Die Ausgangsgleichung

$$E(X) = \sum_{i=1}^n \sum_{j=1}^m x_i * f_{XY}(x_i, y_j)$$

kann umgeschrieben werden in

$$E(X) = \sum_{i=1}^n \sum_{j=1}^m x_i * f_{XY}(x_i, y_j) = \sum_{i=1}^n x_i * \sum_{j=1}^m f_{XY}(x_i, y_j)$$

weil  $x_i$  nicht von  $y_j$  abhängt. Deshalb kann  $\sum_{i=1}^n x_i$  aus der inneren Summe herausgezogen werden.

Regel:

$$\sum_{i=1}^n \sum_{j=1}^m a_i * b_j = \sum_{i=1}^n a_i * \sum_{j=1}^m b_j$$

Beispiel:

Ohne Herausziehen (linke Seite):

$$\sum_{i=1}^2 \sum_{j=1}^3 i * j = 1 * 1 + 1 * 2 + 1 * 3 + 2 * 1 + 2 * 2 + 2 * 3 = 18$$

bzw.

$$\sum_{i=1}^2 \left( \sum_{j=1}^3 i * j \right) = \sum_{i=1}^2 (i * (1 + 2 + 3)) = 1 * 6 + 2 * 6 = 18$$

Alternativ mit Herausziehen (rechte Seite):

$$\sum_{i=1}^2 \left( \sum_{j=1}^3 i * j \right) = \left( \sum_{i=1}^2 i \right) * \left( \sum_{j=1}^3 j \right) = (1 + 2) * (1 + 2 + 3) = 3 * 6 = 18$$

$$E(X) = \sum_{i=1}^n x_i * \sum_{j=1}^m f_{XY}(x_i, y_j)$$

$$E(X) = \sum_{i=1}^n x_i * f_X(x_i)$$

$$E(Y) = \sum_{j=1}^m y_j * f_Y(y_j)$$

Der Erwartungswert der Zufallsvariablen  $X$  bzw.  $Y$  kann also auch mit Hilfe der Randverteilung von  $X$  bzw. von  $Y$  berechnet werden.

Beispiel:

Gegeben sei eine Maschine, die zwei besonders störanfällige Bauteile  $B_1$  und  $B_2$  besitzt.  $X$  sei die Anzahl der Ausfälle pro Tag von  $B_1$ .  $Y$  sei die Anzahl der Ausfälle pro Tag von  $B_2$ . Nimmt man an, dass  $X$  nur die Werte  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 2$  und  $Y$  nur die Werte  $y_1 = 0$ ,  $y_2 = 1$ ,  $y_3 = 2$  annehmen kann, dann lässt sich folgende gemeinsame Wahrscheinlichkeitsfunktion darstellen:

| $x_i \backslash y_j$ | 0    | 1    | 2    |
|----------------------|------|------|------|
| 0                    | 0,30 | 0,14 | 0,02 |
| 1                    | 0,18 | 0,10 | 0,02 |
| 2                    | 0,12 | 0,06 | 0,06 |

Berechne  $E(X)$  und  $E(Y)$  mit Hilfe der Randverteilung von  $X$  bzw.  $Y$ .

$$E(X) = \sum_{i=1}^n x_i * f_X(x_i)$$

$$E(X) = 0 * 0,46 + 1 * 0,3 + 2 * 0,24 = 0,78$$

$$E(Y) = \sum_{j=1}^m y_j * f_Y(y_j)$$

$$E(Y) = 0 * 0,6 + 1 * 0,3 + 2 * 0,1 = 0,5$$

- **Varianz bzw. Standardabweichung der Randverteilung**

Für die Varianz der Randverteilung der Zufallsvariablen  $X$  bzw.  $Y$  gilt:

$$VAR(X) = \sum_{i=1}^n \sum_{j=1}^m (x_i - E(X))^2 * f_{XY}(x_i, y_j)$$

$$VAR(X) = \sum_{i=1}^n (x_i - E(X))^2 * \sum_{j=1}^m f_{XY}(x_i, y_j)$$

$$VAR(X) = \sum_{i=1}^n (x_i - E(X))^2 * f_X(x_i)$$

$$VAR(Y) = \sum_{j=1}^m (y_j - E(Y))^2 * f_Y(y_j)$$

Daraus folgt gem. Verschiebungssatz:

$$VAR(X) = E(X^2) - (E(X))^2 = \sum_{i=1}^n x_i^2 * f_X(x_i) - (E(X))^2$$

$$VAR(Y) = E(Y^2) - (E(Y))^2 = \sum_{j=1}^m y_j^2 * f_Y(y_j) - (E(Y))^2$$

Beispiel:

Gegeben sei eine Maschine, die zwei besonders störanfällige Bauteile  $B_1$  und  $B_2$  besitzt.  $X$  sei die Anzahl der Ausfälle pro Tag von  $B_1$ .  $Y$  sei die Anzahl der Ausfälle pro Tag von  $B_2$ . Nimmt man an, dass  $X$  nur die Werte  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 2$  und  $Y$  nur die Werte  $y_1 = 0$ ,  $y_2 = 1$ ,  $y_3 = 2$  annehmen kann, dann lässt sich folgende gemeinsame Wahrscheinlichkeitsfunktion darstellen:

| $x_i \backslash y_j$ | 0    | 1    | 2    |
|----------------------|------|------|------|
| 0                    | 0,30 | 0,14 | 0,02 |
| 1                    | 0,18 | 0,10 | 0,02 |
| 2                    | 0,12 | 0,06 | 0,06 |

Berechne die Varianzen der Randverteilungen.

Varianz der Randverteilung der Zufallsvariablen  $X$ :

$$VAR(X) = E(X^2) - (E(X))^2 = \sum_{i=1}^n x_i^2 * f_X(x_i) - (E(X))^2$$

$$VAR(X) = 0^2 * 0,46 + 1^2 * 0,3 + 2^2 * 0,24 - 0,78^2 = 0,6516$$

Varianz der Randverteilung der Zufallsvariablen  $Y$ :

$$VAR(Y) = E(Y^2) - (E(Y))^2 = \sum_{j=1}^m y_j^2 * f_Y(y_j) - (E(Y))^2$$

$$VAR(Y) = 0^2 * 0,6 + 1^2 * 0,3 + 2^2 * 0,1 - 0,5^2 = 0,45$$

- **Erwartungswert der bedingten Wahrscheinlichkeitsverteilung**

Mit Hilfe bedingter Wahrscheinlichkeitsverteilungen lassen sich bedingte Erwartungswerte bestimmen.

Erwartungswert der Zufallsvariablen  $X$ , gegeben  $Y = y_j$ :

$$E(X|y_j) = \sum_{i=1}^n x_i * f_X(x_i|y_j)$$

Erwartungswert der Zufallsvariablen  $Y$ , gegeben  $X = x_i$ :

$$E(Y|x_i) = \sum_{j=1}^m y_j * f_Y(y_j|x_i)$$

Beispiel:

Gegeben sei eine Maschine, die zwei besonders störanfällige Bauteile  $B_1$  und  $B_2$  besitzt.  $X$  sei die Anzahl der Ausfälle pro Tag von  $B_1$ .  $Y$  sei die Anzahl der Ausfälle pro Tag von  $B_2$ . Nimmt man an, dass  $X$  nur die Werte  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 2$  und  $Y$  nur die Werte  $y_1 = 0$ ,  $y_2 = 1$ ,  $y_3 = 2$  annehmen kann, dann lässt sich folgende gemeinsame Wahrscheinlichkeitsfunktion darstellen:

| $x_i \backslash y_j$ | 0    | 1    | 2    |
|----------------------|------|------|------|
| 0                    | 0,30 | 0,14 | 0,02 |
| 1                    | 0,18 | 0,10 | 0,02 |
| 2                    | 0,12 | 0,06 | 0,06 |

Berechne den Erwartungswerte der bedingten Wahrscheinlichkeitsverteilung von  $X$  unter der Bedingung  $y_j = 0$  und den Erwartungswert der bedingten Wahrscheinlichkeitsverteilung von  $Y$  unter der Bedingung  $x_i = 2$ .

$$E(X|y_j) = \sum_{i=1}^n x_i * f_X(x_i|y_j)$$

$$E(X|0) = \sum_{i=1}^3 x_i * f_X(x_i|0) = 0 * 0,5 + 1 * 0,3 + 2 * 0,2 = 0,7$$

$$E(Y|x_i) = \sum_{j=1}^m y_j * f_Y(y_j|x_i)$$

$$E(Y|2) = \sum_{j=1}^3 y_j * f_Y(y_j|2) = 0 * 0,5 + 1 * 0,25 + 2 * 0,25 = 0,75$$

- **Varianz bzw. Standardabweichung der bedingten Wahrscheinlichkeitsverteilung**

Mit Hilfe bedingter Wahrscheinlichkeitsverteilungen lassen sich bedingte Varianzen bzw. bedingte Standardabweichungen bestimmen.

Varianz der Zufallsvariablen  $X$ , gegeben  $Y = y_j$ :

$$VAR(X|y_j) = \sum_{i=1}^n (x_i - E(X|y_j))^2 * f_X(x_i|y_j)$$

$$VAR(X|y_j) = \sum_{i=1}^n x_i^2 * f_X(x_i|y_j) - (E(X|y_j))^2$$

Varianz der Zufallsvariablen  $Y$ , gegeben  $X = x_i$ :

$$VAR(Y|x_i) = \sum_{j=1}^m (y_j - E(Y|x_i))^2 * f_Y(y_j|x_i)$$

$$VAR(Y|x_i) = \sum_{j=1}^m y_j^2 * f_Y(y_j|x_i) - (E(Y|x_i))^2$$

Beispiel:

Gegeben sei eine Maschine, die zwei besonders störanfällige Bauteile  $B_1$  und  $B_2$  besitzt.  $X$  sei die Anzahl der Ausfälle pro Tag von  $B_1$ .  $Y$  sei die Anzahl der Ausfälle pro Tag von  $B_2$ . Nimmt man an, dass  $X$  nur die Werte  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 2$  und  $Y$  nur die Werte  $y_1 = 0$ ,  $y_2 = 1$ ,  $y_3 = 2$  annehmen kann, dann lässt sich folgende gemeinsame Wahrscheinlichkeitsfunktion darstellen:

| $x_i \backslash y_j$ | 0    | 1    | 2    |
|----------------------|------|------|------|
| 0                    | 0,30 | 0,14 | 0,02 |
| 1                    | 0,18 | 0,10 | 0,02 |
| 2                    | 0,12 | 0,06 | 0,06 |

Berechne die Varianz der bedingten Wahrscheinlichkeitsverteilung von  $X$  unter der Bedingung  $y_j = 0$  und die Varianz der bedingten Wahrscheinlichkeitsverteilung von  $Y$  unter der Bedingung  $x_i = 2$ .

$$VAR(X|y_j) = \sum_{i=1}^n x_i^2 * f_X(x_i|y_j) - (E(X|y_j))^2$$

$$VAR(X|0) = \sum_{i=1}^3 x_i^2 * f_X(x_i|0) - 0,7^2 = 0^2 * 0,5 + 1^2 * 0,3 + 2^2 * 0,2 - 0,7^2$$

$$VAR(X|0) = 0,61$$

$$VAR(Y|x_i) = \sum_{j=1}^m y_j^2 * f_Y(y_j|x_i) - (E(Y|x_i))^2$$

$$VAR(Y|2) = \sum_{j=1}^3 y_j^2 * f_Y(y_j|2) - 0,75^2 = 0^2 * 0,5 + 1^2 * 0,25 + 2^2 * 0,25 - 0,75^2$$

$$VAR(Y|2) = 0,6875$$

- **Erwartungswert des Produktes  $X * Y$**

Der Erwartungswert des Produktes  $X * Y$ , nämlich  $E(XY)$ , dient i.d.R. der Berechnung von Kovarianz und Korrelationskoeffizient. Oft – allerdings nicht immer – ist der Ausdruck  $E(XY)$  auch selbst interpretierbar.

Wäre beispielsweise die Zufallsvariable  $X$  die Zahl der Krankmeldungen der einzelnen Mitarbeiter eines Unternehmens pro Jahr und  $Y$  die Krankheitsdauer je Krankmeldung in Tagen, so wäre  $E(XY)$  die zu erwartende Anzahl der insgesamt wegen Krankheit ausfallenden Tage in einem Jahr.

Es gilt:

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^m x_i * y_j * f_{XY}(x_i, y_j)$$

### Beispiel:

Gegeben sei eine Maschine, die zwei besonders störanfällige Bauteile  $B_1$  und  $B_2$  besitzt.  $X$  sei die Anzahl der Ausfälle pro Tag von  $B_1$ .  $Y$  sei die Anzahl der Ausfälle pro Tag von  $B_2$ . Nimmt man an, dass  $X$  nur die Werte  $x_1 = 0, x_2 = 1, x_3 = 2$  und  $Y$  nur die Werte  $y_1 = 0, y_2 = 1, y_3 = 2$  annehmen kann, dann lässt sich folgende gemeinsame Wahrscheinlichkeitsfunktion darstellen:

| $x_i \backslash y_j$ | 0    | 1    | 2    |
|----------------------|------|------|------|
| 0                    | 0,30 | 0,14 | 0,02 |
| 1                    | 0,18 | 0,10 | 0,02 |
| 2                    | 0,12 | 0,06 | 0,06 |

Berechne und interpretiere den Erwartungswert  $E(XY)$

$$E(XY) = \sum_{i=1}^n \sum_{j=1}^m x_i * y_j * f_{XY}(x_i, y_j)$$

$$\begin{aligned} E(XY) = & 0 * 0 * 0,3 + 0 * 1 * 0,14 + 0 * 2 * 0,02 + \\ & 1 * 0 * 0,18 + 1 * 1 * 0,1 + 1 * 2 * 0,02 + \\ & 2 * 0 * 0,12 + 2 * 1 * 0,06 + 2 * 2 * 0,06 = 0,5 \end{aligned}$$

$E(XY)$  ist hier der Erwartungswert für synchrone Ausfälle der Bauteile  $B_1$  und  $B_2$  an einem Tag.

### • Kovarianz

Bei zweidimensionalen Verteilungen lässt sich oft beobachten, dass die beiden Zufallsvariablen mehr oder weniger korreliert sind, d.h., dass bei großen Werten von  $X$  die Werte von  $Y$  tendenziell auch groß sind (= positive Korrelation) bzw. dass bei großen Werten von  $X$  die Werte von  $Y$  tendenziell klein sind (= negative Korrelation).

Ein Parameter, der die Stärke dieses Zusammenhangs beschreibt, ist die Kovarianz  $COV(X, Y)$ .

Positiver Wert der Kovarianz,  $COV(X, Y) > 0$ : Wenn Variable  $X$  steigt, steigt tendenziell auch Variable  $Y$ .

Negativer Wert der Kovarianz,  $COV(X, Y) < 0$ : Wenn Variable  $X$  steigt, sinkt tendenziell Variable  $Y$ .

Wert nahe 0,  $COV(X, Y) \approx 0$ : Es besteht kein linearer Zusammenhang zwischen den Variablen.

Ein positiver Wert der Kovarianz zeigt somit an, dass beide Variablen tendenziell gemeinsam steigen, während ein negativer Wert eine gegenläufige Entwicklung

signalisiert. Die Kovarianz sagt also nur etwas über die Richtung des Zusammenhangs (positiv/negativ) aus.

**Grundformel:**

$$COV(X, Y) = \sum_{i=1}^n \sum_{j=1}^m (x_i - E(X)) * (y_j - E(Y)) * f_{XY}(x_i, y_j)$$

bzw.

$$COV(X, Y) = E((X - E(X)) * (Y - E(Y)))$$

**Verschiebungssatzformel für die Kovarianz (= Rechenformel):**

$$COV(X, Y) = E(XY) - E(X) * E(Y)$$

**Überleitung:**

$$COV(X, Y) = E((X - E(X)) * (Y - E(Y)))$$

$$(X - E(X)) * (Y - E(Y)) = X * Y - X * E(Y) - Y * E(X) + E(X) * E(Y)$$

$$COV(X, Y) = E(X * Y - X * E(Y) - Y * E(X) + E(X) * E(Y))$$

Anwendung der Linearität des Erwartungswerts:  $E(A + B) = E(A) + E(B)$

$$COV(X, Y) = E(X * Y) - E(X * E(Y)) - E(Y * E(X)) + E(E(X) * E(Y))$$

Da  $E(Y)$  im Term  $E(X * E(Y))$  eine Konstante ist, kann sie aus dem Erwartungswert herausgezogen werden:  $E(X * E(Y)) = E(Y) * E(X)$

Da  $E(X)$  im Term  $E(Y * E(X))$  eine Konstante ist, kann sie aus dem Erwartungswert herausgezogen werden:  $E(Y * E(X)) = E(X) * E(Y) = E(Y) * E(X)$

Da  $E(X)$  und  $E(Y)$  Konstanten sind, ist ihr Produkt ebenfalls eine Konstante, der Erwartungswert einer Konstante ist die Konstante selbst:  $E(E(X) * E(Y)) = E(X) * E(Y)$

$$COV(X, Y) = E(X * Y) - E(Y) * E(X) - E(Y) * E(X) + E(Y) * E(X)$$

$$COV(X, Y) = E(XY) - E(X) * E(Y)$$

**Beispiel:**

Gegeben sei eine Maschine, die zwei besonders störanfällige Bauteile  $B_1$  und  $B_2$  besitzt.  $X$  sei die Anzahl der Ausfälle pro Tag von  $B_1$ .  $Y$  sei die Anzahl der Ausfälle pro Tag von  $B_2$ . Nimmt man an, dass  $X$  nur die Werte  $x_1 = 0$ ,  $x_2 = 1$ ,  $x_3 = 2$

und  $Y$  nur die Werte  $y_1 = 0, y_2 = 1, y_3 = 2$  annehmen kann, dann lässt sich folgende gemeinsame Wahrscheinlichkeitsfunktion darstellen:

| $x_i \backslash y_j$ | 0    | 1    | 2    |
|----------------------|------|------|------|
| 0                    | 0,30 | 0,14 | 0,02 |
| 1                    | 0,18 | 0,10 | 0,02 |
| 2                    | 0,12 | 0,06 | 0,06 |

Berechne  $COV(XY)$ .

$$COV(X, Y) = E(XY) - E(X) * E(Y)$$

$$COV(X, Y) = 0,5 - 0,78 * 0,5 = 0,11$$

- Korrelationskoeffizient**

Da die Kovarianz nicht standardisiert ist und stark von der Skalierung abhängt, lässt sich ihre Stärke kaum interpretieren, weshalb der Korrelationskoeffizient bevorzugt wird. Um die Stärke des Zusammenhangs (zwischen  $-1$  und  $+1$ ) zu bestimmen, wird die Kovarianz standardisiert, indem sie durch das Produkt der Standardabweichungen geteilt wird (Korrelation).

Der Korrelationskoeffizient ist die standardisierte Kovarianz und so standardisiert, dass dieser Koeffizient immer zwischen  $-1$  und  $+1$  liegt.

$$K(X, Y) = \frac{COV(X, Y)}{\sqrt{VAR(X) * VAR(Y)}}$$

$$K(X, Y) = \frac{COV(X, Y)}{\sigma_X * \sigma_Y}$$

$$-1 \leq K(X, Y) \leq +1$$

$$\sigma_X = \text{Standardabweichung der Zufallsvariable } X = \sqrt{VAR(X)}$$

$$\sigma_Y = \text{Standardabweichung der Zufallsvariable } Y = \sqrt{VAR(Y)}$$

$$K(X, Y) = 0 \Rightarrow \text{Unabhängigkeit, die Variablen sind unkorreliert}$$

$$|K(X, Y)| = 1 \Rightarrow \text{deterministischer Zusammenhang zwischen } X \text{ und } Y \text{ im Sinne einer perfekten Korrelation}$$

$$K(X, Y) = -1 \Rightarrow \text{totaler negativ-linearer Zusammenhang}$$

$$K(X, Y) = +1 \Rightarrow \text{totaler positiv-linearer Zusammenhang}$$

Bei  $K(X, Y) = 0$  besteht kein linearer Zusammenhang, eventuell kann aber z.B. ein quadratischer Zusammenhang bestehen.

Beispiel:

Gegeben sei eine Maschine, die zwei besonders störanfällige Bauteile  $B_1$  und  $B_2$  besitzt.  $X$  sei die Anzahl der Ausfälle pro Tag von  $B_1$ .  $Y$  sei die Anzahl der Ausfälle pro Tag von  $B_2$ . Nimmt man an, dass  $X$  nur die Werte  $x_1 = 0, x_2 = 1, x_3 = 2$  und  $Y$  nur die Werte  $y_1 = 0, y_2 = 1, y_3 = 2$  annehmen kann, dann lässt sich folgende gemeinsame Wahrscheinlichkeitsfunktion darstellen:

| $x_i \backslash y_j$ | 0    | 1    | 2    |
|----------------------|------|------|------|
| 0                    | 0,30 | 0,14 | 0,02 |
| 1                    | 0,18 | 0,10 | 0,02 |
| 2                    | 0,12 | 0,06 | 0,06 |

Berechne und interpretiere den Korrelationskoeffizient  $K(X, Y)$

$$K(X, Y) = \frac{COV(X, Y)}{\sqrt{VAR(X) * VAR(Y)}}$$

$$K(X, Y) = \frac{0,11}{\sqrt{0,6516 * 0,45}} = 0,2031402069$$

$K(X, Y) = 0,2031402069$  deutet auf einen nur schwachen Zusammenhang der Ausfälle von Bauteil  $B_1$  und Bauteil  $B_2$  hin.

## 5.4. Logarithmische Normalverteilung

- Die logarithmische Normalverteilung (= Lognormalverteilung) ist eine stetige Wahrscheinlichkeitsverteilung für eine Variable, die nur positive Werte annehmen kann. Eine Zufallsvariable  $X$  ist dann lognormalverteilt, wenn die mit dem natürlichen Logarithmus transformierte Zufallsvariable  $G = \ln(X)$  normalverteilt ist. Die Zufallsvariable  $X$  ist also dann lognormalverteilt, wenn ihr natürlicher Logarithmus  $\ln(X)$  normalverteilt ist.

Ist die Zufallsvariable  $A$  lognormalverteilt, ist ihr natürlicher Logarithmus  $\ln(A)$  normalverteilt.

Ist die Zufallsvariable  $B$  normalverteilt, ist ihre e-Funktion  $e^B$  lognormalverteilt. Der Satz "*Die Zufallsvariable  $X$  ist dann lognormalverteilt, wenn ihr natürlicher Logarithmus  $\ln(X)$  normalverteilt ist*" gilt auch in diesem Fall:

$$X \equiv e^B | \ln$$

$$\ln(X) = \ln(e^B) = B * \ln(e) = B * 1 = B$$

$\Rightarrow$  Da  $B$  per Annahme normalverteilt ist, ist auch  $\ln(X)$  normalverteilt.

Die Lognormalverteilung ist somit eine stetige Wahrscheinlichkeitsverteilung für positive Werte ( $x > 0$ ), deren Logarithmus normalverteilt ist. Sie ist rechtsschief (= linkssteil) und hat einen langen Ausläufer nach rechts.

- Wenn die Ausgangszufallsvariable  $y = f(x)$  lognormalverteilt ist, ist die transformierte Zufallsvariable  $g = \ln(y) = \ln(f(x))$  normalverteilt.

Die Transformation einer lognormalverteilten Zufallsvariable  $y$  durch die Logarithmusfunktion  $\ln(y)$  führt zu einer Normalverteilung.

Zusammenhänge:

Ist

$y = f(x)$  lognormalverteilt, dann ist

$g = \ln(y) = \ln(f(x))$  normalverteilt.

Wenn eine Zufallsvariable  $y = f(x)$  lognormalverteilt ist, ist ihr natürlicher Logarithmus  $\ln(y)$  bzw.  $\ln(f(x))$  normalverteilt. Sei  $y = f(x)$  eine lognormalverteilte Zufallsvariable. Durch die Transformation  $g = \ln(y) = \ln(f(x))$  wird die Variable auf den Logarithmus abgebildet. Nach der Definition der Lognormalverteilung ist die transformierte Variable  $g$  normalverteilt.

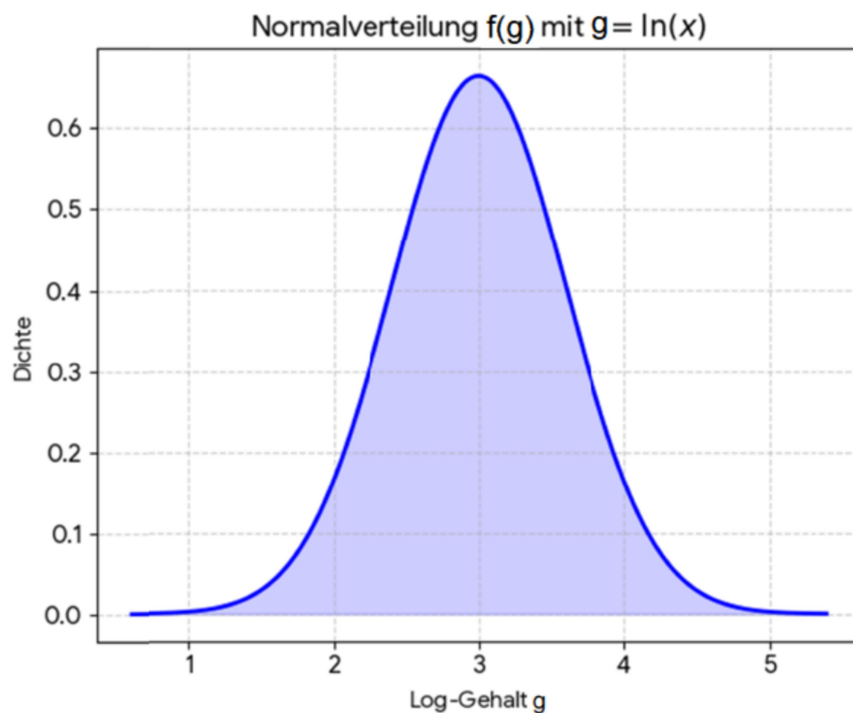
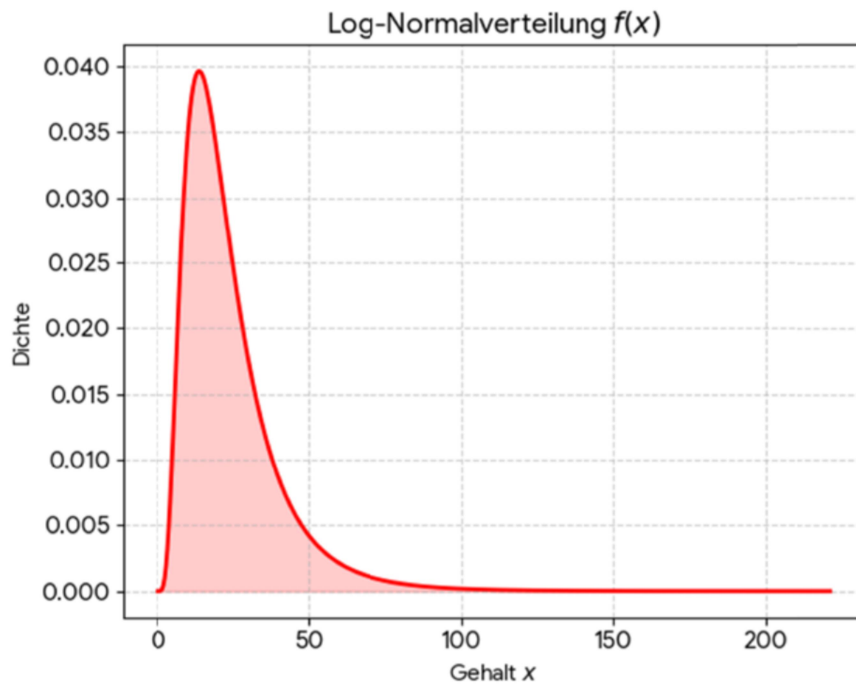
$$g = \ln(f(x)) \quad | e^{(\cdot)}$$

$$e^g = e^{\ln(f(x))}$$

$$g = \log_e(e^{\ln(f(x))}) = \ln(e^{\ln(f(x))}) = \ln(f(x))$$

Lognormalverteilung  $y = f(x)$ : Beschreibt Daten, die nur positive Werte annehmen können (*Wertebereich*:  $x > 0$ ) und eine Rechtsschiefe aufweisen.

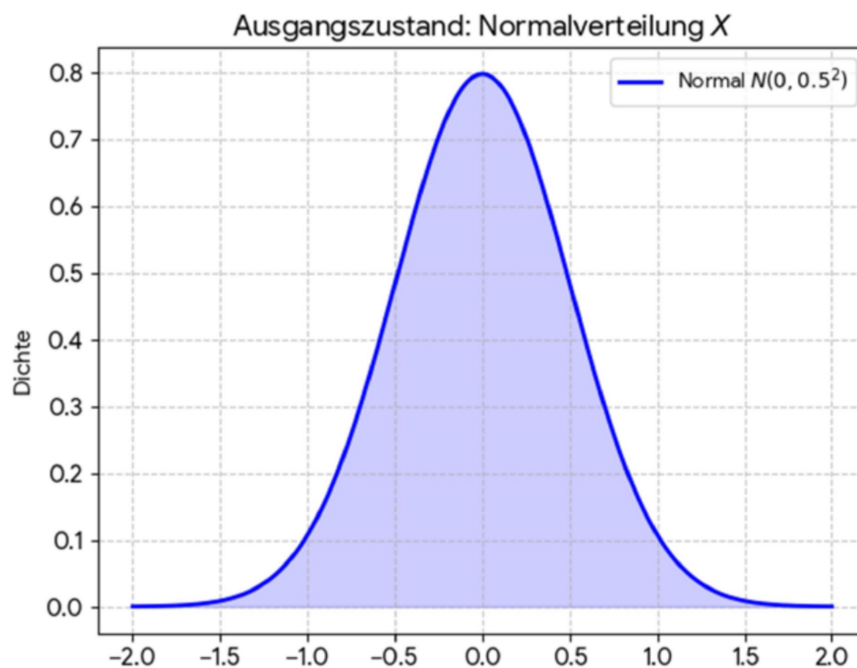
Normalverteilung  $g = \ln(y) = \ln(f(x))$ : Beschreibt Daten, die beliebige Werte annehmen können (*Wertebereich*:  $-\infty < x < +\infty$ ) und als Glockenkurve symmetrisch um den Mittelwert verteilt sind.

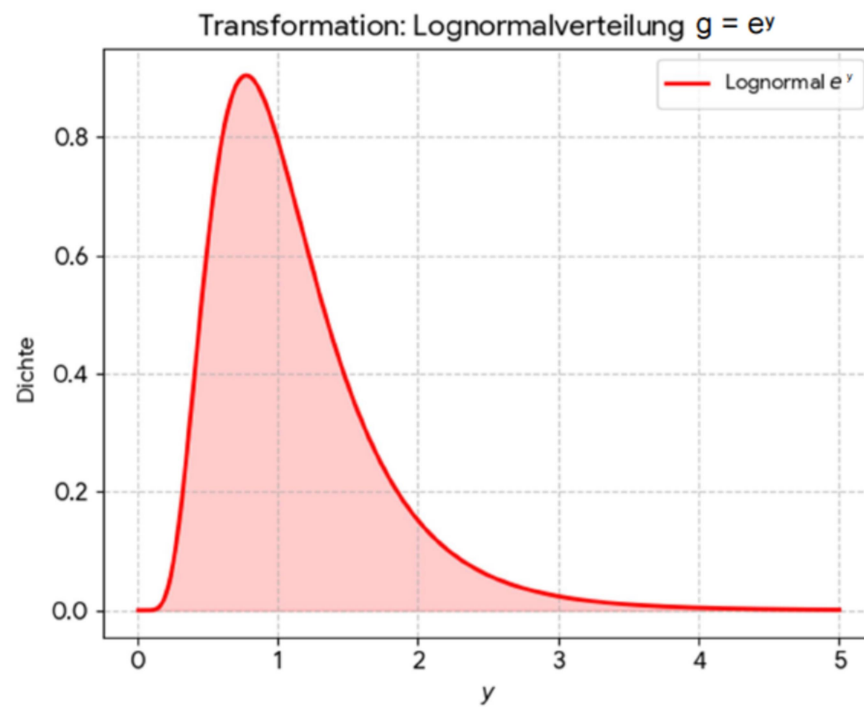


- Wenn die Ausgangszufallsvariable  $y = f(x)$  normalverteilt ist, ist die transformierte Zufallsvariable  $g = e^y = e^{f(x)}$  lognormalverteilt.

Die Transformation einer normalverteilten Zufallsvariable  $y$  durch die Exponentialfunktion  $g = e^y$  führt zu einer Lognormalverteilung.

Während die Normalverteilung theoretisch von  $-\infty$  bis  $+\infty$  definiert ist, bildet die Transformation  $e^y$  für die Dichtefunktion alle Werte im Intervall  $]0; +\infty[$  ab. Die Lognormalverteilung beginnt daher immer bei 0. Da die Exponentialfunktion  $e^y$  für alle reellen Zahlen  $y$  (inklusive sehr kleiner negativer Zahlen) stets einen Wert größer als 0 liefert, ist der Wertebereich von  $g$  immer auf das Intervall  $]0; +\infty[$  beschränkt. Durch  $g = e^y$  werden die negativen  $y$ -Werte "umgeklappt" und in der Dichtefunktion in den Bereich nahe 0 gedrängt. Die hohen positiven  $y$ -Werte werden hingegen durch das exponentielle Wachstum in der Dichtefunktion stark auseinandergezogen, was den flachen Auslauf nach rechts erzeugt.



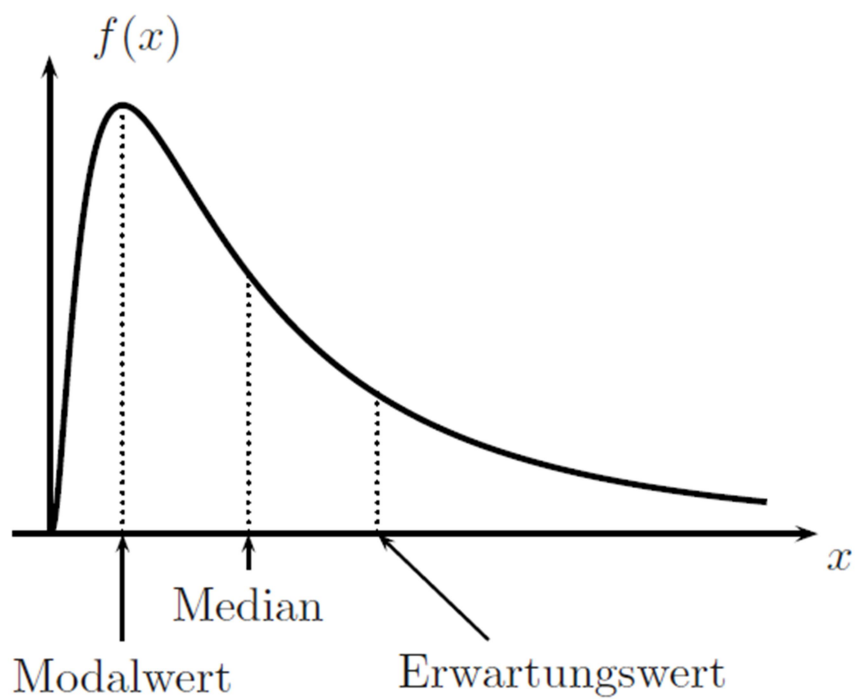


- **Dichtefunktion**

$$f_{LN}(x|\mu, \sigma) = \frac{1}{\sigma * x * \sqrt{2 * \pi}} * e^{-\frac{1}{2} * \left(\frac{\ln(x) - \mu}{\sigma}\right)^2}$$

$$x > 0$$

Dichtefunktion der Dichte der Lognormalverteilung:



Die Dichtefunktion der Lognormalverteilung ist rechtsschief, d.h. sie steigt schnell an und fällt langsam ab. Sie eignet sich zur Modellierung von Verteilungen mit vielen kleinen Werten auf der linken Seite und wenigen großen Werten auf der rechten Seite der x-Achse. Klassischer Anwendungsfall sind Einkommensverteilungen. So gibt es sehr wenige hochdotierte Positionen, wohingegen ein Großteil der Einkommen aus Tätigkeiten mit eher geringem Einkommen resultiert. Extrem niedrige Einkommen werden dann wieder schnell seltener.

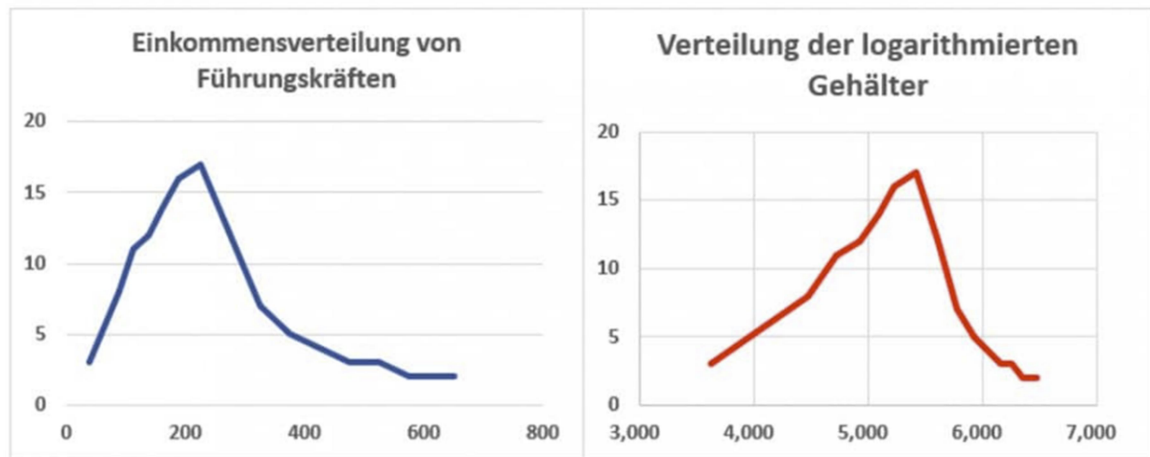
Beim Maximum der Dichte der Lognormalverteilung liegt der Modalwert  $e^{\mu-\sigma^2}$ . Rechts vom Modalwert liegt der Median  $e^{\mu}$ ; an dieser Stelle wird die Fläche unter der Dichte halbiert. Rechts vom Median liegt der Erwartungswert (= Schwerpunkt)  $e^{\mu+\frac{\sigma^2}{2}}$  der Verteilung. Die Funktionsparameter  $\mu$  und  $\sigma^2$  der Lognormalverteilung können anders als bei der Normalverteilung nicht als Erwartungswert und Varianz der Verteilung interpretiert werden, da sie Erwartungswert und Varianz der logarithmierten Variablen sind.

#### Beispiel:

Folgende Häufigkeiten für Gehaltsklassen von Führungskräfte im Bankensektor wurden erhoben:

| Gehaltsklasse | Klassenmitte<br>x | log. Klassenmitte<br>g=ln(x) | Häufigkeiten |
|---------------|-------------------|------------------------------|--------------|
| bis 75        | 37,5              | 3,624                        | 3            |
| 75-100        | 87,5              | 4,472                        | 8            |
| 100-125       | 112,5             | 4,723                        | 11           |
| 125-150       | 137,5             | 4,924                        | 12           |
| 150-175       | 162,5             | 5,091                        | 14           |
| 175-200       | 187,5             | 5,234                        | 16           |
| 200-250       | 225,5             | 5,418                        | 17           |
| 250-300       | 275,5             | 5,619                        | 12           |
| 300-350       | 325,5             | 5,785                        | 7            |
| 350-400       | 375,5             | 5,928                        | 5            |
| 400-450       | 425,5             | 6,053                        | 4            |
| 450-500       | 475,5             | 6,164                        | 3            |
| 500-550       | 525,5             | 6,264                        | 2            |
| 550-600       | 575,5             | 6,355                        | 1            |
| >600          | 650               | 6,477                        | 2            |

Die linke Grafik zeigt die Häufigkeitsverteilung der Gehälter als blaue Kurve an; deutlich ist die schiefe Verteilung erkennbar. Die Daten sind lognormalverteilt. Logarithmiert man die Klassenmittel ergibt sich die rote Kurve in der rechten Grafik. Die Daten sind normalverteilt (Glockenkurve).

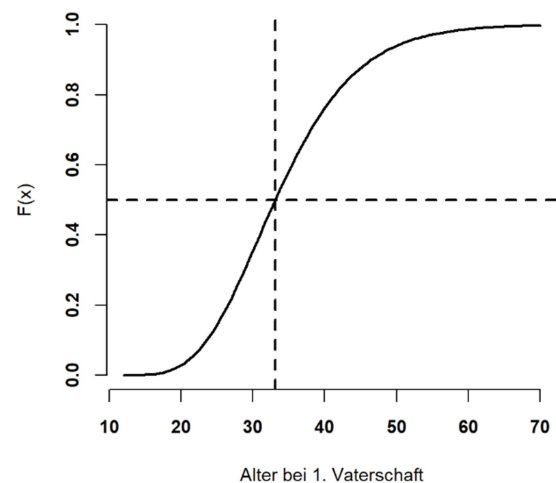
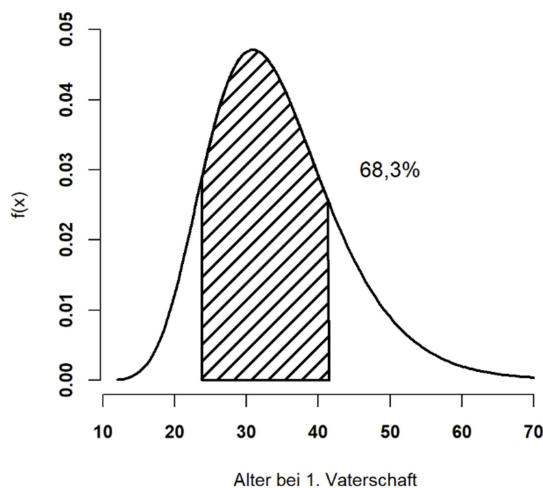


Wenn eine Zufallsvariable  $x$  (= Gehalt) lognormalverteilt ist, dann ist ihr natürlicher Logarithmus  $g = \ln(x)$  normalverteilt (Glockenkurve). Die folgende Abbildung verdeutlicht exemplarisch den Unterschied zwischen der rechtsschiefen Verteilung der tatsächlichen Gehälter  $x$  und der symmetrischen Glockenkurve der logarithmierten Gehälter  $g$ .

- **Verteilungsfunktion**

Die Verteilungsfunktion der Lognormalverteilung kann zurückgeführt werden auf die Verteilungsfunktion  $\Phi$  der Standardnormalverteilung. Daher gilt für die Verteilungsfunktion  $F$  der lognormalverteilten Zufallsvariablen  $X$ :

$$F_{LN}(x|\mu, \sigma) = \Phi * \frac{\ln(x) - \mu}{\sigma}$$



- **Erwartungswert**

$$E(X) = e^{\mu + \frac{\sigma^2}{2}}$$

- **Varianz**

$$VAR(X) = e^{2\mu + \sigma^2} * (e^{\sigma^2} - 1)$$

- Wenn  $X = \frac{a}{b}$  lognormalverteilt und  $b > 0$  eine Konstante ist, dann ist auch  $a$  lognormalverteilt und  $\ln(a)$  ist normalverteilt.

Durch Multiplikation einer Lognormalverteilung mit einer positiven Konstante bleibt die Form der Verteilung erhalten; lediglich die Parameter verschieben sich.

- **Aktienkurs-Modellierung**

Zunächst wird der Unterschied zwischen diskreten und stetigen Renditen erläutert.

Bei diskreter Verzinsung mit  $m$  Zinsperioden pro Jahr lautet die Formel für das Endkapital  $S_t$  nach  $n$  Jahren bei einem Anfangskapital  $S_0$  und einem Nominalzinsatz  $r$ :

$$S_t(\text{diskret}) = S_0 * \left(1 + \frac{r}{m}\right)^{m*n}$$

Um von der diskreten zur stetigen Verzinsung zu gelangen, lässt man die Häufigkeit der Zinsgutschriften  $m$  über alle Grenzen wachsen ( $m \rightarrow \infty$ ). Das Kapital wird also unendlich oft pro Jahr verzinst

$$S_t(\text{stetig}) = \lim_{m \rightarrow \infty} \left( S_0 * \left(1 + \frac{r}{m}\right)^{m*n} \right)$$

$$S_t(\text{stetig}) = S_0 * \lim_{m \rightarrow \infty} \left(1 + \frac{r}{m}\right)^{m*n}$$

Für die Eulersche Zahl gilt:

$$e = \lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x$$

Substitution von  $x = \frac{m}{r}$  führt zu:

$$e = \lim_{m \rightarrow \infty} \left(1 + \frac{r}{m}\right)^{\frac{m}{r}}$$

$$S_t(\text{stetig}) = S_0 * \lim_{m \rightarrow \infty} \left(1 + \frac{r}{m}\right)^{m*n}$$

$$S_t(\text{stetig}) = S_0 * \lim_{m \rightarrow \infty} \left(1 + \frac{r}{m}\right)^{\frac{m}{r} * n * r}$$

$$S_t(\text{stetig}) = S_0 * e^{n*r}$$

$$S_t(\text{stetig}) = S_t = S_0 * e^{r*n}$$

$$e^{r*n} = \left(\frac{S_t}{S_0}\right) | \ln$$

$$\ln(e^{r*n}) = \ln\left(\frac{S_t}{S_0}\right)$$

$$\ln((e^r)^n) = \ln\left(\frac{S_t}{S_0}\right)$$

$$n * \ln(e^r) = \ln\left(\frac{S_t}{S_0}\right)$$

$$n * r = \ln\left(\frac{S_t}{S_0}\right)$$

$$r = \frac{1}{n} * \ln\left(\frac{S_t}{S_0}\right) = \frac{1}{n} * (\ln(S_t) - \ln(S_0))$$

### **Kernaussage: Aktienkurse sind lognormalverteilt.**

Über identische Betrachtungsperioden im Zeitablauf (z.B. bei jährlicher Erhebung) sind bei einer Aktie die diskrete Renditen und die stetigen Renditen normalverteilt, während die Aktienkurse lognormalverteilt sind. Im Folgenden werden nur stetige Renditen  $r$  betrachtet.

Renditen bewegen sich symmetrisch um einen Mittelwert und können sowohl positive als auch negative Ausprägungen annehmen. Insoweit kann für Renditen eine Normalverteilung unterstellt werden.

Für die stetige Rendite einer Aktie vom Zeitpunkt  $t - 1$  zum Zeitpunkt  $t$  (Einperiodenfall) gilt:

$$r = \frac{1}{n} * \ln\left(\frac{S_t}{S_{t-1}}\right) = \frac{1}{1} * \ln\left(\frac{S_t}{S_{t-1}}\right)$$

$r$  = stetige Periodenrendite der Aktie

$n$  = Laufzeit in Jahren

$S_t$  = Aktienkurs am Periodenende

$S_{t-1}$  = Aktienkurs der Vorperiode; dieser Aktienkurs ist gegeben

$\frac{S_t}{S_{t-1}}$  = Wachstumsfaktor

$$r = \ln\left(\frac{S_t}{S_{t-1}}\right) \Leftrightarrow e^r = \frac{S_t}{S_{t-1}}$$

$$S_t = S_{t-1} * e^r$$

$$S_t(r) = S_{t-1} * e^r$$

Wird für die stetige Rendite  $r$  (= normalverteilte Zufallsvariable) eine Normalverteilung mit  $\mu$  und  $\sigma^2$  angenommen, so ist

$$e^{S_t(r)}$$

lognormalverteilt mit einem Aktienkurs-Erwartungswert

$$E(S_t) = S_{t-1} * e^{\mu + \frac{\sigma^2}{2}}$$

und einer Varianz

$$VAR(S_t) = e^{2\mu + \sigma^2} * (e^{\sigma^2} - 1)$$

Der Aktienkurs ist mithin lognormalverteilt und damit rechtsschief. Die Annahme, dass Renditen normalverteilt und Aktienkurse lognormalverteilt sind, ist zentraler Baustein des Black-Scholes-Optionspreismodells.

Hinweis: Folgender Rechenweg wäre falsch, weil nur jährlicher Zinszuschlag unterstellt werden würde.

$$r = \frac{S_t - S_{t-1}}{S_{t-1}} = \frac{S_t}{S_{t-1}} - 1$$

$$S_t = S_{t-1} * (r + 1)$$

Beispiel:

$$S_t = 6,54 \text{ GE}$$

$$S_{t-1} = 6,00 \text{ GE}$$

$$r = \frac{S_t}{S_{t-1}} - 1 = \frac{6,54}{6,00} - 1 = 9\%$$

Bei stetiger Verzinsung beträgt  $r$  jedoch 8,617769624%.

$$r = \frac{1}{n} * \ln\left(\frac{S_t}{S_0}\right) = \frac{1}{1} * \ln\left(\frac{6,54}{6,00}\right) = 8,617769624\%$$

Um den erwarteten Aktienkurs zu berechnen, wird der stochastische Prozess der geometrischen Brownschen Bewegung genutzt. Die geometrische Brownsche Bewegung ist Grundlage des Black-Scholes-Modells und modelliert Aktienkurse, indem angenommen wird, dass die prozentualen Renditen (und nicht die absoluten Preisänderungen) unabhängig voneinander sind und normalverteilten Schwankungen unterliegen. Über die Lognormalverteilung der Aktienkurse werden negative Aktienkurse ausgeschlossen.

$$S_t = S_0 * e^{\left(r - \frac{\sigma^2}{2}\right) * n + \sigma * \sqrt{n} * Z}$$

$S_t$  = Aktienkurs nach  $n$  Jahren

$S_0$  = Aktienkurs am Periodenbeginn; dieser Aktienkurs ist gegeben

$r$  = Erwartete jährliche Rendite (= Drift)

$\sigma$  = Jährliche Volatilität der Aktie (= Stärke der zufälligen Marktschwankungen)

$n$  = Laufzeit in Jahren

$Z$  = Z-Wert der Standardnormalverteilung

- Aufgabe 1:

Die Lebensdauer  $X$  [Stunden] eines elektronischen Bauteils ist lognormalverteilt. Der Logarithmus der Lebensdauer  $Y = \ln(X)$  ist normalverteilt mit dem Erwartungswert  $\mu = 4$  und der Standardabweichung  $\sigma = 0,5$ . Gesucht ist die Wahrscheinlichkeit dafür, dass die Lebensdauer des Bauteils zwischen  $e^4 \approx 54,6$  Stunden und  $e^4 * \sqrt{e} \approx 90$  Stunden liegt, also der Bereich  $P(55 \leq X \leq 90)$ .

$$e^4 = 54,6$$

$$\ln(e^4) = 4$$

$$e^4 * \sqrt{e} = 90$$

$$\ln(e^4 * \sqrt{e}) = \ln(e^{4,5}) = 4,5$$

$$z_1 = \frac{x - \mu}{\sigma}$$

$$z_1 = \frac{4 - 4}{0,5} = 0$$

$$z_2 = \frac{x - \mu}{\sigma}$$

$$z_2 = \frac{4,5 - 4}{0,5} = 1$$

$$P(0 \leq X \leq 1) = 0,8413 - 0,5 = 0,3413$$

Die Wahrscheinlichkeit, dass die Lebensdauer zwischen 55 und 90 Stunden liegt, beträgt 34,13%

- Beispiel 2:

- (i) Es soll der Kurs einer Aktie nach einem halben Jahr prognostiziert werden. Folgende Werte sind gegeben:

$$S_0 = 100 \text{ EUR}$$

$$r = 8\%$$

$$\sigma = 25\%$$

$$n = 0,5 \text{ Jahre}$$

- (ii) Welcher Kurs würde bei vorgenannten Parametern mit einer Wahrscheinlichkeit von 95% nicht unterschritten werden?

(i)

$$S_t = S_0 * e^{r*n}$$

$$S_t = 100 * e^{0,08*0,5}$$

$$S_t = 104,081077419 \text{ EUR}$$

(ii)

Wenn der Kurs mit 95 % Wahrscheinlichkeit nicht unterschritten werden soll, dürfen nur 5% der Werte darunter liegen. Gesucht ist der Kurs, der mit einer Wahrscheinlichkeit von 5% unterschritten wird. Der entsprechende Z-Wert ergibt sich aus einem einseitig linksseitigen Test in Höhe von  $-1,645$ .

$$S_t = S_0 * e^{\left(r - \frac{\sigma^2}{2}\right) * n + \sigma * \sqrt{n} * Z}$$

$$S_t = 100 * e^{\left(0,08 - \frac{0,25^2}{2}\right) * 0,5 + 0,25 * \sqrt{0,5} * -1,645} = 76,6115249906 \text{ EUR}$$

Mit einer Wahrscheinlichkeit von 95% liegt der Kurs nach einem halben Jahr über 76,60 €