$\boxed{\text{Statistik und Stochastik}}$

Mathematik für Informatiker IV

Dozent: Prof. Dr. R. Korn

Sommersemester 2001

Philipp Becker / Stephan Fudeus / Christoph Thelen

Version vom 23. Juli 2001

Wir geben keine Garantie auf Richtigkeit des Skripts. Hinweise auf Fehler sowie Verbesserungsvorschläge bitte an mathe@3ph.de

LaTeX 2_ε-Version by 3ph

Inhaltsverzeichnis

Kapitel 1

Stochastik

1.1 Einführung

- Stochastik (Wahrscheinlichkeitsrechnung)
  - Modellierung (scheinbar) zufälliger Ereignisse (Glücksspiele, Aktienkurse, Stabilität von Betriebssystemen)
  - Ableiten von sich aus den Modellen ergebenden Gesetzmäßigkeiten
- Statistik
  - schließende Statistik (Schätzen von Parametern stochastischer Modelle, und testen, ob die stochastischen Modelle die Realität adäquat beschreiben)
  - deskriptive Statistik (beschreibende Statistik)
    (Übersichtliche Darstellung von Daten durch Angabe von Maßzahlen wie z.B. Mittelwert, Minimum, Maximum & graphische Aufarbeitung (z.B. Balkendiagramme, Kreisdiagramme, Gesichterdarstellung))

Typische Beispiele:

i) Münzwurf (Zahl oder Kopf)
$\longrightarrow$ stochastisches Modell
Zahl erscheint mit Wahrscheinlichkeit p
Kopf erscheint mit Wahrscheinlichkeit $1-p,\ p \in (0,1)$
$\longrightarrow$ schließende Statistik

(a) Welchen Wert hat p?
$\Longrightarrow$ Werfe die Münze „hinreichend oft“ und schätze p aus den beobachteten Ergebnissen heraus.

(b) Gilt die Annahme $p = \frac{1}{2}$ („faire Münze“) oder gilt sie nicht?
$\Longrightarrow$ Ziehe eine Stichprobe (d.h. werfe die Münze n-mal) und entscheide dann über Annahme oder Ablehnung der Hypothese $p = \frac{1}{2}$ („Test“).

ii) Volkszählung
$\longrightarrow$ deskriptive Statistik; sehr große Datenmenge
Übersichtliche Darstellung der Daten, z.B. als Alterspyramide, Häufigkeitsdiagramme für Religion, Geschlecht, Berufsgruppen, Angabe von mittlerem Einkommen.

Wichtiger Unterschied:

Modell $\longrightarrow$ Wahrscheinlichkeiten für das Eintreten bestimmter Ereignisse	*Realität* $\longrightarrow$ Beobachtung von relativen Häufigkeiten des Eintretens einzelner Ereignisse bei wiederholt durchgeführten Experimenten

1.2 Theoretische Grundbegriffe der Stochastik

Ziel:
Beschreibung eines zufälligen Experiments (mögliche Ergebnisse, sinnvolle Zuordnung von Wahrscheinlichkeiten für das Auftreten einzelner Ergebnisse).

Definition 1:
Die Menge Ω der möglichen Ausgänge eines zufälligen Experiments heißt Ereignisraum . Die Elemente $\omega \in \Omega$ heißen Elementarereignisse (Ergebnisse, Ausgänge).

Beispiel:
Münzwurf $\Omega = \{\text{Kopf}, \text{Zahl}\}$
$1 \times$ würfeln $\Omega = \{1, 2, 3, 4, 5, 6\}$

Oft interessieren uns auch zusammengefasste Ereignisse (z.B. „Gerade Zahl gewürfelt“), sogenannte Ereignisse. Aus mathematischen Gründen ist die Definition „*Jede Teilmenge E von Ω heißt Ereignis*“ nicht sinnvoll, da sie zu Widersprüchen führen kann.

Sinnvolle Forderungen an das System f der Ereignisse:

(1) $f \neq \varnothing$

(2) $\Omega \in f$ „sicheres Ereignis liegt in f“

(3) $A \in f \Longrightarrow A^{\complement} = \Omega \backslash A \in f$ „das Gegenereigniss liegt auch in f“

(4) $A_1, A_2, \ldots \in f \Longrightarrow A_1 \cup A_2 \cup \ldots \in f$ „Vereinigung von Ereignissen ist ein Ereignis“

(5) $A_1, A_2, \ldots \in f \Longrightarrow A_1 \cap A_2 \cap \ldots \in f$ „Schnitt von Ereignissen ist ein Ereignis“

Definition 2:
Ein System f von Teilmengen E von Ω, das die Eigenschaften (1)-(4) besitzt, heißt eine σ-Algebra über Ω. Die Elemente von f heißen Ereignisse.

Beispiel:

i) Münzwurf: $\Omega = \{K, Z\} \longrightarrow$ minimale σ-Algebra: $f_{\text{min}} = \{\Omega, \varnothing\}$
maximale σ-Algebra: $f_{\text{max}} = \{\Omega, \varnothing, \{K\}, \{Z\}\}$

ii) $1\times$ Würfeln:
maximale σ-Algebra $f_{\text{max}} = \{A | A \subseteq \Omega\} =$ „Potenzmenge von Ω“ $= \Re(\Omega)$
„Menge aller möglichen Teilmengen von Ω“

iii) Ist z.B. $\Omega = [0, \infty)$, so kann auf $\Re([0, \infty))$ keine Wahrscheinlichkeit definiert werden ($\rightarrow$ schwierig!)

Nächster Schritt: Wie definiert man Wahrscheinlichkeiten?

Erfahrungstatsache:
Wird ein Experiment n-mal unter gleichen Bedingungen wiederholt, so scheinen sich die relativen Häufigkeiten des Eintretens eines Ereignisses A, $rh(A)$, mit

$$rh(A) = \frac{\sharp \text{ Eintreten von } A}{\sharp \text{ Versuche}}$$

um einen Wert zwischen 0 und 1 zu stabilisieren. Sieht man $rh(A)$ als „Messwert" für die Wahrscheinlichkeit von A, $P(A)$, an, so sollen die Wahrscheinlichkeiten dieselben Eigenschaften wie die relativen Häufigkeiten haben.

Es gilt (s. Übung)

$$0 \leq rh(A) \leq 1 \qquad \forall A \in f \qquad \text{„positiv"}$$
$$rh(\Omega) = 1 \qquad \text{„normiert"}$$

A_1, A_2 mit $A_1 \cap A_2 = \varnothing$
$\Longrightarrow rh(A_1 \cup A_2) = rh(A_1) + rh(A_2)$ „(σ)-additiv"

Sogar $A_1, A_2, \ldots$ mit $A_i \cap A_j = \varnothing$ für $i \neq j$
$\Longrightarrow rh(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} rh(A_i)$

Definition 3:
Ein Wahrscheinlichkeitsraum (Ω, f, P) besteht aus einer nicht-leeren Menge Ω (Ereignisraum), einer σ-Algebra f über Ω (System der Ereignisse) und einer Wahrscheinlichkeit P auf f, d.h. einer Funktion $P : f \to [0,1]$ mit

(6) $P(A) \geq 0 \quad \forall A \in f$

(7) $P(\Omega) = 1$

(8) $A_1, A_2, \ldots \in f$ mit $A_i \cap A_j = \varnothing$ für $i \neq j \Longrightarrow P(\dot{\bigcup}_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$
($\dot{\bigcup} \equiv$ paarweise disjunkt)

Beispiel: Münzwurf
$\Omega = \{K, Z\},\ f = \{\varnothing, \Omega, \{K\}, \{Z\}\}$
$P(\varnothing) = 0,\ P(\Omega) = 1,\ P(K) = p,\ P(Z) = 1 - p$ für ein $p \in [0,1]$

Folgerung 4:
Es gelten (für $A, B, A_i \in f$)

(9) $P(\varnothing) = 0$

(10) $A \cap B = \varnothing \Longrightarrow P(A \dot{\cup} B) = P(A) + P(B)$

(11) $P\left(A^{\complement}\right) = 1 - P(A)$

(12) $A \subseteq B \Longrightarrow P(A) \leq P(B)$

(13) $A, B \in f \Longrightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

bzw. allgemeiner:

(14) $P(\bigcup_{j=1}^n A_j) = \sum_{j=1}^n P(A_j) - \sum_{1\le j<i\le n} P(A_i \cap A_j)$
$\qquad + \sum_{1\le j<i<k\le n} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_n)$

(15) $P(\bigcup_{i=1}^\infty A_i) \le \sum_{i=1}^\infty P(A_i)$

Beweis:

(B9) $\Omega = \Omega \,\dot\cup\, \varnothing \,\dot\cup\, \varnothing \,\dot\cup\, \ldots \xrightarrow{(7)} 1 = P(\Omega) = P(\Omega \,\dot\cup\, \varnothing \,\dot\cup\, \varnothing \,\dot\cup\, \ldots)$
$= P(\dot\bigcup_{i=1}^\infty A_i) \overset{(8)}{=} \sum_{i=1}^\infty P(A_i) = \underbrace{P(\Omega)}_{=1} + \sum_{i=2}^\infty P(\varnothing)$
$(A_1 = \Omega, A_2 = A_3 = \ldots = \varnothing)$
$\Longrightarrow \sum_{i=2}^\infty \underbrace{P(\varnothing)}_{\ge 0} = 0 \Longrightarrow P(\varnothing) = 0$

(B10) Wähle in (8) $A_1 := A$, $A_2 := B$, $\varnothing =: A_3 = A_4 = \ldots$
$$\Longrightarrow P(A \,\dot\cup\, B) = P(\dot\bigcup_{i=1}^\infty A_i) = P(A) + P(B) + \underbrace{\sum_{i=3}^\infty P(\varnothing)}_{=0}$$

(B11) $\Omega = A \,\dot\cup\, A^\complement \overset{(10)+(7)}{\longrightarrow} 1 = P(\Omega) = P\left(A \,\dot\cup\, A^\complement\right) = P(A) + P\left(A^\complement\right) \Longrightarrow (11)$

(B12) - (B15): Übung □

1.3 Diskrete Wahrscheinlichkeitsräume

Ist die Ereignismenge Ω endlich oder höchstens abzählbar (also $\Omega = \{\omega_1, \ldots, \omega_n\}$ oder $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots\}$), so heißt (Ω, f, P) ein diskreter Wahrscheinlichkeitsraum. Da hier

$$A = \dot{\bigcup_{\omega \in A}} \{\omega\} \qquad \text{und} \qquad P(A) = \sum_{\omega \in A} P\{\omega\}$$

gelten, läßt sich die Wahrscheinlichkeit P durch einen Wahrscheinlichkeitsvektor p angeben mit

$$p = (p_1, p_2, \ldots),\ p_i \geq 0,\ \sum_{i=1}^{\infty} p_i = 1,\ p_i = P(\{\omega_i\}),\ i \in \mathbb{N}$$

Beispiel:
Eine faire Münze wird so lange geworfen, bis zum ersten mal „Kopf" auftritt. Es sei $\omega_j =$ „Erster Kopf im j-ten Wurf"
$\Longrightarrow$ Modell:

$$\Omega = \{1, 2, 3, \ldots\} = \mathbb{N},\ f = \Re(\Omega),\ p_j = \left(\frac{1}{2}\right)^j \geq 0 \ \left(\sum_{j=1}^{\infty} \left(\frac{1}{2}\right)^j = \frac{1}{1 - \frac{1}{2}} - 1 = 1\right)$$

Definition 5:
Ist $\Omega = \{\omega_1, \ldots, \omega_n\}$ endlich und gilt für alle $\omega \in \Omega$

$$(16) \quad p_\omega = P(\{\omega\}) = \frac{1}{|\Omega|} = \frac{1}{n}$$

so heißt (Ω, f, P) ein Laplace-Raum ($f = \Re(\Omega) =$ Potenzmenge von Ω).

Folgerung 6:
In einem Laplace-Raum sind alle Elementarereignisse ω gleich wahrscheinlich, und es gilt für alle $A \subseteq \Omega$:

$$(17) \quad P(A) = \frac{|A|}{|\Omega|} = \frac{\text{Anzahl der } \omega \in A}{\text{Anzahl der } \omega \in \Omega} = \frac{\text{Anzahl der für } A \text{ günstigen Fälle}}{\text{Anzahl der möglichen Fälle}}$$

Beweis:

$$A = \dot{\bigcup_{\omega \in A}} \{\omega\} \Longrightarrow P(A) = \sum_{\omega \in A} P(\{\omega\}) \stackrel{(16)}{=} \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{\sum_{\omega \in A} 1}{|\Omega|} = \frac{|A|}{|\Omega|}$$

Konsequenz:
Will man in einem Laplace-Raum die Wahrscheinlichkeit $P(A)$ berechnen, so muss man die Anzahl der Elemente von A und Ω bestimmen $\Longrightarrow$ Kombinatorik

Abzählregeln: „Urnenmodelle" (auch auf andere Situationen anwendbar)
Gegeben sei eine Urne mit N unterscheidbaren Kugeln. Hieraus werden $n \leq N$ Stück gezogen

(a) als geordnete (bzgl. der Reihenfolge) Stichprobe <u>mit</u> Wiederholung („<u>Ziehen mit Zurücklegen</u> unter Beachtung der Reihenfolge")

$\Longrightarrow$ <u>Ergebnis</u>: $(x_1, \ldots, x_n)$ mit $x_i \in \{1, \ldots, N\}$
$\longrightarrow N \cdot N \cdot \ldots \cdot N = \boxed{N^n \text{ Möglichkeiten}}$
(z.B. Befragung von n Personen, die ein Produkt mit den Noten $(1, \ldots, N)$ bewerten können)

(b) als geordnete Stichprobe <u>ohne</u> Wiederholung („<u>Ziehen ohne Zurücklegen</u> unter Beachtung der Reihenfolge")

$\Longrightarrow$ <u>Ergebnis</u>: $(x_1, \ldots, x_n)$ mit $x_i \neq x_j;\ x_i \in \{1, \ldots, N\}$

$$\longrightarrow N \cdot (N-1) \cdot (N-2) \cdot \ldots \cdot (N-n+1) = \boxed{\frac{N!}{(N-n)!} \text{ Möglichkeiten}}$$

(z.B. Direktes Beobachten der Ziehung der Lottozahlen)

(c) <u>Ziehen ohne Zurücklegen</u> (ohne Beachtung der Reihenfolge; ungeordnete Stichprobe)
$\Longrightarrow$ <u>Ergebnis</u>: $(x_1, \ldots, x_n)$ mit $1 \leq x_1 < x_2 < \ldots < x_n \leq N$
$\longrightarrow$ Teile die Anzahl aus (b), also $\frac{N!}{(N-n)!}$, durch die Anzahl der Möglichkeiten, n Zahlen verschieden anzuordnen, also durch $n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$

$$\longrightarrow \boxed{\frac{N!}{(N-n)! \cdot n!} =: \binom{N}{n} \text{ Möglichkeiten}}$$

Anwendungen:

1) <u>Lotto</u>: Ziehe 6 aus 49 Kugeln, wobei (ohne Zusatzzahl) 6 günstige Kugeln („rot") und 43 ungünstig („schwarz") sind
$\longrightarrow P(\text{„}k \text{ Richtige"})=?$ für $k \in \{0, 1, 2, 3, 4, 5, 6\}$
$\Longrightarrow$ Ziehen ohne Zurücklegen (ohne Reihenfolge):
$\Omega = \{(x_1, \ldots, x_6) | x_i \in \{1, \ldots, 49\}, 1 \leq x_1 < \ldots < x_6 \leq 49\}$
$\xrightarrow{(c)} |\Omega| = \binom{49}{6} = 13.983.816$ mögliche Ausgänge (als Menge)
$A_k =$„k Richtige" $\Longrightarrow |A_k| = \binom{6}{k} \cdot \binom{43}{6-k}$

$$\Longrightarrow P(A_k) = \frac{\binom{6}{k} \cdot \binom{43}{6-k}}{\binom{49}{6}} = \begin{cases} 0,436 & k = 0 \\ 0,413 & k = 1 \\ 0,132 & k = 2 \\ 0,018 & k = 3 \\ 0,00097 & k = 4 \\ 0,0002 & k = 5 \\ 0,00000007 & k = 6 \end{cases}$$

Allgemein:

Definition 7:
N, R, n seien natürliche Zahlen mit $N \geq R,\ N \geq n$ (N: alle Kugeln; R: rote Kugeln; n: gezogene Kugeln)
Dann heißt die Wahrscheinlichkeit, die durch

$$(18) \quad p_k = \frac{\binom{R}{k} \cdot \binom{N-R}{n-k}}{\binom{N}{n}},\ k = 0, 1, \ldots, n$$

gegeben ist, die hypergeometrische Verteilung $Hg(N, R, n)$.

2) Ziehen mit Zurücklegen (mit Beachtung der Reihenfolge)
Gleiches Spiel wie in 1), jetzt aber mit Zurücklegen der gezogenen Kugeln
$\xrightarrow{(a)}$ insgesamt 49^6 Möglichkeiten

- in jedem Zug gibt es 6 Möglichkeiten für die Richtigen und 43 für die Falschen
$\hookrightarrow$ also: $6^k \cdot 43^{6-k}$ günstige Zahlenmengen
- bei k Richtigen gibt es $\binom{6}{k}$ Möglichkeiten der verschiedenen Reihenfolge

$$\Longrightarrow p_k = \frac{\binom{6}{k} \cdot 6^k \cdot 43^{6-k}}{49^6} = \binom{6}{k} \cdot \left(\frac{6}{49}\right)^k \cdot \left(\frac{43}{49}\right)^{6-k},\ k = 0, 1, \ldots, 6$$

Allgemein:

Definition 8:
Es sei $n \in \mathbb{N},\ 0 \leq p \leq 1$ (n: Anzahl der Versuche; p: Erfolgswahrscheinlichkeit pro Versuch). Dann heißt die durch

$$(19) \quad p_k = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k},\ k = 0, 1, \ldots, n$$

gegebene Wahrscheinlichkeit Binomialverteilung $B(n, p)$.

1.4 Zweistufige Experimente, bedingte Wahrscheinlichkeit, Unabhängigkeit

Kann man vor Beendigung eines Experiments bereits Zusatzinformationen erhalten, so ändert sich oft die Wahrscheinlichkeit für das Eintreten eines bestimmten Ereignisses. Führt z.B. der FCK zur Halbzeit mit 3:0 gegen Bayern, so hat sich die Wahrscheinlichkeit eines Sieges des FCK gegenüber der Einschätzung vor dem Anpfiff deutlich erhöht.
Erhält man dagegen statt dem Halbzeitstand eine irrelevante Information, z.B. ob in China gerade ein Sack Reis umgefallen ist, so hat dies sicher keinen Einfluss auf die Einschätzung der Siegeschancen des FCK. Dies wird in den Begriffen der bedingten Wahrscheinlichkeit und der Unabhängigkeit formalisiert:

Definition 9:
Es sei (Ω, f, P) ein Wahrscheinlichkeitsraum. Ist $B \in f$ mit $P(B) > 0$, so heißt

$$(20) \quad P(A|B) := \frac{P(A \cap B)}{P(B)} \text{ für } A \in f$$

die bedingte Wahrscheinlichkeit von A gegeben B.

Beispiel zur Definition: „Zweimaliges Würfeln“
$\Omega = \{(x,y)|x,y \in \{1,2,\ldots,6\}\},\ f = \Re(\Omega),\ P((x,y)) = \frac{1}{36}\ \forall (x,y) \in \Omega$
Sei $A =$ „Augensumme mindestens 8“
$B =$ „3 im ersten Wurf“
$\Longrightarrow P(A) = \frac{15}{36} = \frac{5}{12}$

Nachdem B eingetreten ist, können nur noch Paare der Gestalt $(3,y)$, $y \in \{1,\ldots,6\}$ auftreten. Aller anderen Elemente von Ω sind unmöglich. Das heißt, „unter der Bedingung B“ ist die Menge der möglichen Ereignisse gerade gleich

$$\Omega \cap B = B$$

Die Menge der günstigen Ereignisse „gegeben B“ ist dann $A \cap B$, alle andere Elemente von A können nun nicht mehr auftreten. Im Beispiel:

$$P(A|B) = \frac{\frac{2}{36}}{\frac{6}{36}} = \frac{2}{6} = \frac{\sharp \text{ günstig, gegeben } B}{\sharp \text{ möglich, gegeben } B} = \frac{1}{3}$$

d.h. die Wahrscheinlichkeit für A gegeben B ist gegenüber der unbedingten Wahrscheinlichkeit, $P(A)$, gesunken.

In der Praxis erhält man oft bedingte Wahrscheinlichkeiten. Zur Berechnung der unbedingten („totalen“) Wahrscheinlichkeit dient der folgende Satz:

Satz 10: „Totale Wahrscheinlichkeit“
Es sei (Ω, f, P) ein Wahrscheinlichkeitsraum, $B_1, B_2, \ldots \in f$ bilden eine Zerlegung von Ω (d.h. $\bigcup_{i=1}^{\infty} B_i = \Omega,\ B_i \cap B_j = \varnothing$ für $i \neq j$) mit $P(B_i) > 0$.

$$\Longrightarrow (21) \quad P(A) = \sum_{i=1}^{\infty} P(A|B_i) \cdot P(B_i) \quad \forall A \in f$$

Beweis:

$$A = \dot{\bigcup_{i=1}^{\infty}}(A \cap B_i) \Longrightarrow P(A) = P\left(\dot{\bigcup_{i=1}^{\infty}}(A \cap B_i)\right) = \sum_{i=1}^{\infty} P(A \cap B_i)$$
$$\overset{(20)}{=} \sum_{i=1}^{\infty} P(A|B_i) \cdot P(B_i)$$

$\square$

Satz 11: „Bayesche Regel"
Unter den Bedingungen von Satz 10 gilt (wenn zusätzlich $P(A) > 0$)

$$(22) \quad P(B_i|A) = \frac{P(A|B_i) \cdot P(B_i)}{P(A)}, \qquad i = 1, 2, \ldots$$

Beweis:

$$P(A|B_i) \cdot P(B_i) \overset{(20)}{=} P(A \cap B_i) = P(B_i \cap A) \overset{(20)}{=} P(B_i|A) \cdot P(A) \Longrightarrow (22) \quad \square$$

Anwendung: „Zweistufige Experimente"
Eine Person unterzieht sich einem HIV-Test. Der Test führt zu den Ergebnissen:

$B =$ „Test sagt HIV-positiv"
$B^{\complement} =$„Test sagt HIV-negativ" ($\Omega = B \cup B^{\complement}$)

Mit welcher Wahrscheinlichkeit ist die Person tatsächlich HIV-positiv, wenn das Testergebnis „HIV-positiv" vorliegt?

Sei $A =$ „Person ist HIV-positiv"
Bereits bekannt:

(i) Qualität des Tests (Daten: Henze ('97): „ELISA"-Test S.108)

$$P(B|A) = 0.998 \qquad \text{„kranke Person wird als krank erkannt"}$$
$$P(B^{\complement}|A^{\complement}) = 0.998 \qquad \text{„gesunde Person wird als gesund erkannt"}$$

$\overset{?}{\Longrightarrow}$ Test irrt sich fast nie??

(ii) Wir nehmen an, daß die (unbedingte) Wahrscheinlichkeit einer HIV-Infizierung für die Bevölkerungsschicht, der die Person angehört, $a = 0.01\%$ ist, also $P(A) = 0.0001 \quad (= 0.01)$.

Gesucht: $P(A|B) =?$
Also: $\overset{\text{Bayes}}{\longrightarrow} P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

$$\text{totale W-keit: } P(B) = P(B|A) \cdot P(A) + P(B|A^{\complement}) \cdot P(A^{\complement})$$
$$= 0.998 \cdot 0.0001 + 0.002 \cdot 0.9999 \approx 0.0021 \quad (\approx 0.0120)$$
$$\Longrightarrow P(A|B) = \frac{0.998 \cdot 0.0001}{0.0021} = 0.048 \quad (= 0.832)$$

d.h. ist die Krankheit in der Bevölkerung sehr selten, so dominiert die Seltenheit die Trennschärfe des Tests („nur eine Erhöhung des Risikos auf 4.8%").

Ist die Krankheit in der Bevölkerungsgruppe dagegen häufig, so führt auch ein positives Testergebnis fast schon zu der Gewissheit, daß man tatsächlich krank ist.

Abschließend:
Satz 12: "Eigenschaften der bedingten Wahrscheinlichkeit"
Sei (Ω, f, P) ein Wahrscheinlichkeitsraum, $B \in f$ mit $P(B) > 0$. Dann gelten:

(a) $(\Omega, f, P(.|B))$ ist ein Wahrscheinlichkeitsraum

(b) Mit $f_B := \{A \subseteq B | A \in f\}$ ist auch $(B, f_B, P(.|B))$ ein Wahrscheinlichkeitsraum

(c) $A \cap B = \varnothing \Longrightarrow P(A|B) = 0$

(d) $A_1, \ldots, A_n \in f;\ P(A_1 \cap A_2 \cap \ldots \cap A_n) > 0$
$\Rightarrow P(A_1 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_n|A_1 \cap \ldots \cap A_{n-1})$

Beweis: Übung

Definition 13:

(a) $A, B \in f$ heißen (paarweise) unabhängig, falls gilt

$$P(A \cap B) = P(A) \cdot P(B) \tag{23}$$

(b) $A_1, A_2, \ldots$ (endlich oder höchstens abzählbar viele) $\in f$ heißen unabhängig, falls für die Auswahl endlich vieler $A_{j_1}, \ldots, A_{j_n}$ $(1 \leq j_1 < \ldots < j_n)$ gilt

$$P(A_{j_1} \cap \ldots \cap A_{j_n}) = P(A_{j_1}) \cdot \ldots \cdot P(A_{j_n}) \tag{24}$$

Bemerkung 14:

(a) A, B unabhängig mit $P(B) > 0$

$$\Longrightarrow \underline{P(A|B)} = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B)}{P(B)} = \underline{P(A)}$$

analog:
$P(A) > 0 \Longrightarrow P(B|A) = P(B)$,
d.h. die Wahrscheinlichkeit für das Eintreten von A ist unabhängig von der Information, das B bereits eingetreten ist.

(b) Gilt „A, B unabhängig", „B, C unabhängig" und „A, C unabhängig", so folgt daraus nicht notwendig auch, daß A, B, C unabhängig sind.

Beispiel: „Zweimaliges Würfeln"

Es seien A = "Im ersten Wurf gerade Zahl"
B = „Im zweiten Wurf gerade Zahl"
C = „Augensumme ist ungerade"

$$
\begin{aligned}
\Longrightarrow P(A\cap B) &= \frac{9}{36} = \frac{3}{6}\cdot\frac{3}{6} = \frac{1}{2}\cdot\frac{1}{2} = P(A)\cdot P(B) \\
&\overset{(23)}{\Longrightarrow} A,\ B \text{ unabhängig} \\
P(B\cap C) &= \frac{9}{36} = \frac{3}{6}\cdot\frac{3}{6} = \frac{1}{2}\cdot\frac{1}{2} = P(B)\cdot P(C) \\
&\Longrightarrow B,\ C \text{ unabhängig} \\
P(A\cap C) &= \frac{9}{36} = \frac{3}{6}\cdot\frac{3}{6} = \frac{1}{2}\cdot\frac{1}{2} = P(A)\cdot P(C) \\
&\Longrightarrow A,\ C \text{ unabhängig} \\
\textbf{aber: } P(A\cap B\cap C) &= \frac{0}{36} \neq \frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = P(A)\cdot P(B)\cdot (P(C)) \\
&\Longrightarrow A,\ B,\ C \text{ nicht unabhängig}
\end{aligned}
$$

(c) Aus A, B paarweise fremd (d.h. $A \cap B = \varnothing$) folgt i.a. nicht, daß A und B unabhängig sind, denn es gilt:

$$P(A\cap B) = 0 \neq P(A)\cdot P(B),\ \text{falls } P(A) > 0,\ P(B) > 0$$

1.5 Zufallsvariablen und Verteilungsfunktion

Oft interessiert uns bei einem Zufallsexperiment nicht das tatsächliche Ergebnis, sondern lediglich ein aus ihm abgeleiteter Wert (z.B. Klausur $\rightarrow$ Note, Würfeln mit mehreren Würfel $\rightarrow$ Augensumme) X, der eine (reellwertige) Funktion $X : \Omega \longrightarrow \mathbb{R}$ $(\mathbb{R}^d, \Omega', ..)$ ist.
Erfüllt X noch weitere technische Bedingungen, so nennen wir X eine Zufallsvariable (oder im Fall $\mathbb{R}^d$ einen Zufallsvektor).
Wir interessieren uns nun dafür, welche Werte die Zufallsvariable (kurz: „ZV") mit welchen Wahrscheinlichkeiten annimmt, d.h. wie die Wahrscheinlichkeit, die gemäß P auf den einzelnen $\omega \in \Omega$ liegt, auf ihre Werte verteilt werden.

$$
\begin{array}{lll}
\Omega & X(\omega) \leftarrow \text{ZV} & \\
\omega_1, \omega_2, \omega_3, \omega_4, \omega_5 & x_1 \text{ für } \omega \in \{\omega_1, \omega_2, \omega_3\} & p_1^x = p_1 + p_2 + p_3 \\
p_1, p_2, p_3, p_4, p_5 & x_2 \text{ für } \omega \in \{\omega_4, \omega_5\} & p_2^x = p_4 + p_5
\end{array}
$$

Definition 15:
Es sei (Ω, f, P) ein diskreter Wahrscheinlichkeitsraum.

(a) Eine Abbildung $X : \Omega \longrightarrow \mathbb{R}$ heißt eine reellwertige Zufallsvariable

(b) Eine Abbildung $X : \Omega \longrightarrow \mathbb{R}^n$ heißt n-dimensionaler, $\mathbb{R}^n$-wertiger Zufallsvektor

Definition 16:
Es sei $\Omega' = X(\Omega) := \{X(\omega) | \omega \in \Omega\}$ der Wertebereich einer reellwertigen Zufallsvariablen X auf einem diskreten Wahrscheinlichkeitsraum (Ω, f, P). Dann heißt die Funktion $P^X : \Re(\Omega') \longrightarrow [0, 1]$, definiert durch

$$(25a) \quad P^X(A) := P(X^{-1}(A)) = P(\{\omega | X(\omega) \in A\}) \quad \forall A \in \Re(\Omega')$$

die Wahrscheinlichkeitsverteilung der Zufallsvariablen X.

Beispiel 17:

(a) Urnenexperiment mit N Kugeln (R rote, $N - R$ schwarze) und Stichprobenumfang n. Es sei
X = Anzahl der roten Kugeln in der Stichprobe $(n \leq R,\ n \leq (N - R))$
Da X nur endlich viele Werte $\{0, 1, \ldots, n\}$ annehmen kann, ist die Wahrscheinlichkeitsverteilung von X durch den Wahrscheinlichkeitsvektor

$$p = (p_0, p_1, ..., p_n) \text{ mit } p_j = P^X(\{j\}) = P(\{\omega | X(\omega) = j\})$$

gegeben. Je nach „Ziehverfahren" erhält man:

i) Ziehen mit Zurücklegen („geordnet")

$$p_j = P^X(\{j\}) = \binom{n}{j} \left(\frac{R}{N}\right)^j \left(\frac{N-R}{N}\right)^{n-j}, \ j = 0, 1, ..., n$$

also $X \sim B\left(n, \frac{R}{N}\right)$, („$X$ ist binomial verteilt mit Parametern n und $\frac{R}{N}$")

ii) Ziehen ohne Zurücklegen („ungeordnet“)

$$p_j = P^X(\{j\}) = \frac{\binom{R}{j}\binom{N-R}{n-j}}{\binom{N}{n}}, \; j = 0, 1, ..., n$$

also $X \sim Hg(N, R, n)$ (Hypergeometrisch verteilt)

(b) Indikatorvariable („Null-Eins-Variable“)
Es sei $A \in f$, $X(\omega)$ definiert durch

$$X(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases} \quad \text{„Indikator von } A\text{“}$$

$$\Longrightarrow X : \Omega \longrightarrow \{0, 1\}$$

$$p_{\underline{1}} = P^X(\{1\}) = P(\{\omega | X(\omega) = 1\}) = \underline{P(A)}$$

$$p_{\underline{0}} = P^X(\{0\}) = \underline{1 - P(A)}$$

insbesondere: $X \sim B(1, P(A))$,
(„X ist Bernoulli-verteilt mit Parameter $P(A)$“)

(c) Weitere Verteilungen: s. Übungen

Problem: Was ist mit überabzählbarem Wahrscheinlichkeitsraum?
$\Rightarrow$ tiefer liegende technische Probleme!
Im diskreten Fall war jede Abbildung $X : \Omega \longrightarrow \mathbb{R}$ eine Zufallsvariable. Für allgemeines Ω ist dies unzweckmäßig. Will man z.B. die Wahrscheinlichkeit wissen, dass „$X \leq 3$“ gilt, so muss $\{\omega | X(\omega) \leq 3\}$ ein Ereignis sein, d.h. es muss gelten $\{\omega | X(\omega) \leq 3\} \in f$. Dies wird deshalb in der Definition einer Zufallsvariablen gefordert.

Definition 18:
(Ω, f, P) sei ein Wahrscheinlichkeitsraum, (Ω', f') ein Paar bestehend aus einer nichtleeren Menge Ω' und einer σ-Algebra f' über Ω'. Eine Abbildung $X : \Omega \longrightarrow \Omega'$ heißt dann eine (Ω'-wertige) Zufallsvariable, wenn für alle $A' \in f'$ gilt:

$$(25b) \quad X^{-1}(A') = \{\omega \in \Omega | X(\omega) \in A'\} \in f$$

(Man sagt auch: X ist $f - f'$-messbare Abbildung.)

Beispiel 19
Besonders wichtiger Fall: $\Omega' = \mathbb{R}$ (oder $\Omega' = \mathbb{R}^d$).
Man wählt dann meist

$$f' = \wp(= \wp^d) \quad \text{„Borel-}\sigma\text{-Algebra“}$$
$$= \text{„kleinste } \sigma\text{-Algebra, die alle halboffenen Intervalle }]a, b], \; a, b \in \mathbb{R} \text{ enthält“}$$

(Im Fall $\wp^d$ gilt: $]a, b] = \{x \in \mathbb{R}^d | a_i < x_i \leq b_i, \; i = 1, \ldots, d\}$

Es kann gezeigt werden:

$$\wp \subset \Re(\mathbb{R})$$

Gilt $\Omega' = I$ (wobei I Intervall), so wählen wir

$$f' = \wp(I) = \{A \subseteq I | \exists \widetilde{A} \in \wp : A = \widetilde{A} \cap I\}$$

Definition 20:

(a) Ist X eine reellwertige Zufallsvariable, so heißt die Funktion $P^X : \wp \longrightarrow [0;1]$ mit

$$(26) \quad P^X(B) = P(X^{-1}(B)) = P(\{\omega | X(\omega) \in B\}) \quad \forall B \in \wp$$

die <u>Wahrscheinlichkeitsverteilung</u> von X.

(b) Sei X wie in (a). Dann heißt die Funktion $F : \mathbb{R} \longrightarrow [0;1]$ mit

$$(27) \quad F(x) = P^X(]-\infty, x]) = P(X \leq x), \; x \in \mathbb{R}$$

die <u>Verteilungsfunktion</u> von X.

Existiert eine Funktion $f : \mathbb{R} \longrightarrow [0, \infty)$ mit

$$(28) \quad F(x) = \int_{-\infty}^{x} f(y)dy$$

so heißt f <u>Wahrscheinlichkeitsdichte</u> von X

Ziele: Verstehen der Begriffe und Beispiele

Ab jetzt: Wir betrachten nur noch entweder Zufallsvariablen, die nur höchstens abzählbar viele Werte annehmen („diskrete Zufallsvariablen“) oder solche, die eine Dichte besitzen („stetige Zufallsvariablen“).

Satz 21:

(a) Für jede Verteilungsfunktion F (einer Zufallsvariablen X) gilt:

$$(29) \quad \begin{array}{ll} i) \; F(x) \geq F(y) \text{ für } x \geq y & \text{„Monotonie“} \\ ii) \; \lim\limits_{h \downarrow 0} F(x+h) = F(x) & \text{„rechtsstetig“} \\ iii) \; \lim\limits_{x \to -\infty} F(x) = 0, \; \lim\limits_{x \to \infty} F(x) = 1 & \end{array}$$

(b) Jede Wahrscheinlichkeitsdichte f erfüllt

$$(30) \quad f(x) \geq 0, \quad \int_{-\infty}^{\infty} f(x)dx = 1$$

(c) Besitzt die Zufallsvariable X eine Wahrscheinlichkeitsdichte f, so gilt

$$P(X \in]a,b]) = \int_a^b f(x)dx \quad \text{für } a \leq b$$

Beweis:

(a) elementar, aber länglich (als Übung: Fall mit Dichte)

(b)

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)dx &= \lim_{k\to\infty} \int_{-k}^{k} f(x)dx = \lim_{k\to\infty} \left[\int_{-\infty}^{k} f(x)dx - \int_{-\infty}^{-k} f(x)dx\right] \\ &\overset{(28)}{=} \lim_{k\to\infty} [F(k) - F(-k)] \overset{(a)}{=} 1 - 0 = 1 \end{aligned}$$

(c)

$$\begin{aligned} P(X \in]a,b]) &= P(X \in]-\infty, b]) - P(X \in]-\infty, a]) \\ &= \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_a^b f(x)dx, \\ &\text{da }]-\infty, b] =]-\infty, a] \,\dot\cup\,]a,b] \end{aligned}$$

□

Es gilt weiter:
Satz 22: (ohne Beweis)
Jede nichtnegative Funktion $f(x)$ mit

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

definiert über $F(x) = \int_{-\infty}^{x} f(z)dz$ eine Verteilungsfunktion auf $\mathbb{R}$ mit zugehöriger Zufallsvariablen $X : \mathbb{R} \longrightarrow \mathbb{R}$ auf $(\mathbb{R}, \wp, P)$ mit

$$\begin{aligned} &P(]a,b]) = F(b) - F(a) \\ &X(\omega) = \omega \\ &P^X(A) = P(A) \quad \forall A \in \wp \end{aligned}$$

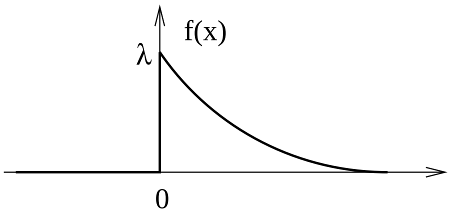
Beispiele 23: „Spezielle Verteilungen“

i) <u>Exponentialverteilung</u>:
Die Exponentialverteilung wird typischerweise zur Modellierung von Lebensdauern (z.B. von Glühbirnen) oder Dauer von Telefongesprächen, Bedienzeiten verwendet.
Sie ist durch die Dichte

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

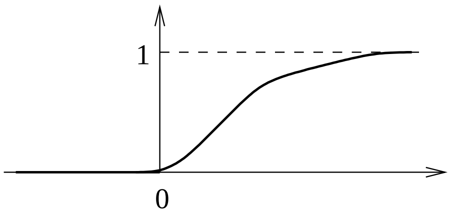
mit einem festen positiven Parameter λ gegeben.

Dichte:



$\Longrightarrow$ Verteilungsfunktion:

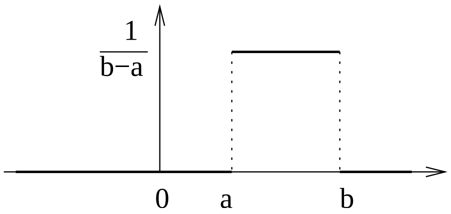
$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 0, & \text{für } x \leq 0 \\ \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x}, & \text{für } x \geq 0 \end{cases}$$



Besitzt X eine solche Verteilung, so schreibt man „$X \sim Exp(\lambda)$" (X ist exponential verteilt mit Parameter λ)

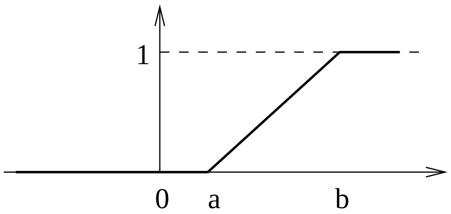
ii) Gleichverteilung
Dichte:

$$f(x) = \begin{cases} 0, & x \notin [a,b] \\ \frac{1}{b-a}, & x \in [a,b] \end{cases}$$



Verteilungsfunktion:

$$F(x) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & x \in [a,b] \\ 1, & x \geq b \end{cases}$$



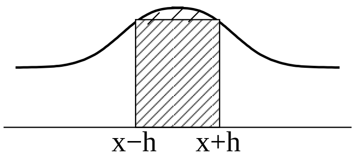
$\Longrightarrow$ „$X \sim U([a,b])$" (X ist auf $[a,b]$ gleichmäßig verteilt)

Für die Wahrscheinlichkeit, genau einen Punkt $x \in \mathbb{R}$ zu treffen, wenn X eine Zufallsvariable mit Dichte ist, gilt

$$P(X = x) = \int_x^x f(y)dy = 0.$$

Es gilt aber auch

$$P(X \in]x-h, x+h]) = \int_{x-h}^{x+h} f(y)dy \overset{h \text{ klein}}{\approx} f(x) \cdot 2h.$$



Also ist die Wahrscheinlichkeit für X, Werte „um x herum“ anzunehmen, (ungefähr) proportional zu $f(x)$, d.h. je größer $f(x)$, desto wahrscheinlicher wird X einen Wert „nahe bei x“ annehmen.

iii) Normalverteilung („Gauß-Verteilung“)
Die Normalverteilung ist die wichtigste Verteilung in Stochastik und Statistik ($\rightarrow$ später). Sie ist durch die Dichte

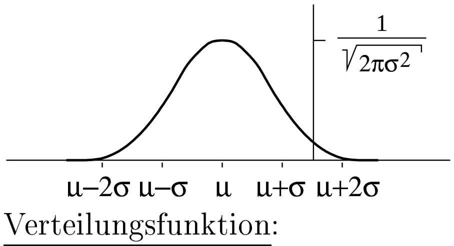
$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right),\ x \in \mathbb{R}$$

und gegebene Parameter $\mu \in \mathbb{R}$, $\sigma^2 \neq 0$ bestimmt („$X \sim \mathcal{N}(\mu, \sigma^2)$“). Ihre Verteilungsfunktion

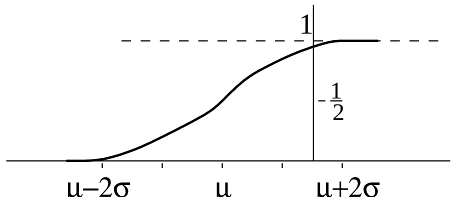
$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma^2}(y)dy$$

lässt sich nicht explizit berechnen.

Dichte:



Verteilungsfunktion:



Es gelten:

$$P(X \leq \mu) = \frac{1}{2} = P(X \geq \mu)$$

$$P(X \leq \mu - 2\sigma) = P(X \geq \mu + 2\sigma) \approx 0.03$$

$$P(X \leq \mu - 3\sigma) = P(X \geq \mu + 3\sigma) \approx 0.001$$

Im Fall $\mu = 0,\ \sigma^2 = 1$ heißt $\mathcal{N}(0,1)$ Standardnormalverteilung, und man schreibt $\varphi(x) = \varphi_{0,1}(x),\ \Phi(x) = \Phi_{0,1}(x)$
Wichtige Eigenschaften der Normalverteilung werden in der Übung behandelt.

Satz 24: „Transformationssatz"
Es sei h eine streng monotone, stetige Funktion mit Umkehrfunktion h^{-1}. Weiter sei X eine reellwertige Zufallsvariable mit Verteilungsfunktion F_X.

(a) Ist h wachsend, so ist die Verteilungsfunktion F_Y von $Y = h(X)$ durch

$$(31) \quad F_Y(y) = F_X(h^{-1}(y)) \text{ für alle } y \in h(\mathbb{R})$$

gegeben.

(b) Ist h fallend, so gilt

$$(32) \quad F_Y(y) = 1 - F_X(h^{-1}(y_+)) \text{ für } y \in h(\mathbb{R})$$

(c) Ist h stetig differenzierbar und besitzt F_X die Dichte f, so besitzt $Y = h(X)$ die Dichte g mit

$$(33) \quad g(y) = f(h^{-1}(y)) \cdot |(h^{-1})'(y)| \text{ für } y \in h(\mathbb{R}),\ g(y) = 0 \text{ für } y \notin h(\mathbb{R})$$

Beweis:

(a)

$$\begin{aligned}\underline{F_Y(y)} = &\underbrace{P(Y \leq y)}_{=P(\{\omega | Y(\omega) \leq y\})} = P(h(X) \leq y) \\ &\overset{y \in h(\mathbb{R})}{=} P(X \leq h^{-1}(y)) = \underline{F_X(h^{-1}(y))}, \quad y \in h(\mathbb{R})\end{aligned}$$

$$\begin{aligned}(\text{Gilt } &y \leq h(\mathbb{R}) \Longrightarrow F_Y(y) = 0 \\ &y \geq h(\mathbb{R}) \Longrightarrow F_Y(y) = 1)\end{aligned}$$

(b) Sei $y \in h(\mathbb{R})$

$$\begin{aligned}\Longrightarrow \underline{F_Y(y)} &= P(Y \leq y) = P(h(X) \leq y) \overset{y \in h(\mathbb{R})}{=} P(X \geq h^{-1}(y)) \\ &= 1 - P(X < h^{-1}(y)) = 1 - \lim_{\varepsilon \downarrow 0} P(X \leq h^{-1}(y + \varepsilon)) \\ &= 1 - \lim_{\varepsilon \downarrow 0} F_X(h^{-1}(y + \varepsilon)) = \underline{1 - F_X(h^{-1}(y_+))}\end{aligned}$$

(c) <u>nur Fall:</u> „h wachsend“:

$$(*)\quad \int_{-\infty}^{y} g(z)dz = F_Y(y) \overset{(a)}{=} F_X(h^{-1}(y) \overset{Vor.}{=} \int_{-\infty}^{h^{-1}(y)} f(x)dx$$

Leite beide Seiten von $(*)$ nach y ab

$$\Longrightarrow g(y) = f(h^{-1}(y))(h^{-1})'(y) \overset{h^{-1}\text{ wachsend}}{=} f(h^{-1}(y))\cdot|(h^{-1})'(y)| \text{ für } y \in h(\mathbb{R})$$

Fall „h fallend“: analog. □

Anwendung: Messen in anderen Skalen
(z.B.: kg → g, m → km, min → h, C → F)

$$Y = aX,\ a > 0 \overset{(31)}{\Longrightarrow} F_Y(y) = F_X(h^{-1}(y)) = F_X\left(\frac{y}{a}\right), g(y) \overset{(33)}{=} f\left(\frac{y}{a}\right)\cdot\frac{1}{a}, y \in \mathbb{R}$$

$$Y = aX + b,\ a > 0 \overset{(31)}{\Longrightarrow} F_Y(y) = F_X\left(\frac{y-b}{a}\right),\ g(y) \overset{(33)}{=} f\left(\frac{y-b}{a}\right)\cdot\frac{1}{a},\ y \in \mathbb{R}$$

1.6 Erwartungswert und Varianz

Wiederholt man ein Experiment sehr häufig, so wird man beobachten, dass sich das arithmetische Mittel der beobachteten Ergebnisse, $\frac{1}{N} \cdot \sum_{i=1}^{N} x_i$, um einen Wert μ herum mit wachsendem N stabilisiert. Diesem Wert, den man bei theoretisch unendlicher Durchführung des Experimentes erhalten würde, entspricht in der Wahrscheinlichkeitsrechnung der Begriff des Erwartungswertes.

Definition 25:
Es sei X eine reellwertige Zufallsvariable.

(a) Nimmt X nur höchstens abzählbar viele Werte $\{x_1, x_2, \ldots\}$ (X ist eine diskrete Zufallsvariable) mit den Wahrscheinlichkeiten

$$p_i := P(X = x_i), i \in \mathbb{N}$$

an, so heißt

$$(34) \quad E(X) = \sum_{i=1}^{\infty} (x_i \cdot p_i)$$

der Erwartungswert von X, falls dieser Wert eindeutig definiert ist (also z.B. falls $\sum_{i=1}^{\infty} |x_i| \cdot p_i < \infty$ gilt).

(b) Besitzt die Verteilungsfunktion F_X von X eine Dichte $f(x)$, so heißt

$$(35) \quad E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

der Erwartungswert von X, falls dieser Wert eindeutig definiert ist (also z.B. falls $\int_{-\infty}^{\infty} |x| \cdot f(x)\, dx < \infty$ gilt).

Beispiele:

(a) „Würfeln“ $\quad X$=„Augenzahl“
$X \in \{1, \ldots, 6\}$
$p_i = \begin{cases} \frac{1}{6} & i = 1, 2, \ldots, 6 \\ 0 & \text{sonst} \end{cases}$
$\overset{(34)}{\Longrightarrow} E(X) = \sum_{i=1}^{6} \left(i \cdot \frac{1}{6}\right) = \frac{1}{6} \cdot \sum_{i=1}^{6} i = \frac{7}{2}$

- Da hier alle auftretenden Werte gleich wahrscheinlich sind, stimmt der Erwartungswert mit dem Mittelwert überein.
- Der Erwartungswert $\frac{7}{2}$ tritt hier nie als Ergebnis auf. Er ist somit kein erwarteter Wert.

(b) Poisson-Verteilung (Verteilung der seltenen Ereignisse)
X heißt Poisson-verteilt mit Parameter $\lambda > 0$, („$X \sim Pn(\lambda)$“) falls

$$p_k = P(X = k) = e^{-\lambda} \cdot \frac{\lambda^k}{k!} \text{ für } k = 0, 1, 2, \dots$$

Da $X \geq 0$ gilt und X diskrete Zufallsvariable ist, ist $E(X)$ gemäß (34) eindeutig definiert.

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \cdot p_k = \sum_{k=1}^{\infty} k \cdot \frac{\lambda^k}{k!} \cdot e^{-\lambda} = e^{-\lambda} \cdot \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} \\ &= e^{-\lambda} \cdot \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \cdot e^{-\lambda} \cdot \underbrace{\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}}_{=e^{\lambda}} \\ &= \lambda \cdot e^{-\lambda} \cdot e^{+\lambda} = \lambda \end{aligned}$$

(c) Exponentialverteilung
$X \sim Exp(\lambda)$ ($\Longrightarrow P(X \geq 0) = 1$)
$\Longrightarrow E(X)$ ist somit (da $X \geq 0$) eindeutig durch (35) gegeben

$$\begin{aligned} \Longrightarrow E(X) &= \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_0^{\infty} x \cdot \lambda \cdot e^{-\lambda x}\, dx \\ &= -x \cdot e^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx = \underbrace{-x \cdot e^{-\lambda x}\Big|_0^{\infty}}_{=0} - \underbrace{\frac{1}{\lambda} \cdot e^{-\lambda x}\Big|_0^{\infty}}_{0-\frac{1}{\lambda}} \\ &= \frac{1}{\lambda} \end{aligned}$$

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

Satz 26: Eigenschaften des Erwartungswertes

(a) Falls $E(X), E(Y)$ existieren, so existiert auch $E(aX + bY)$ für $a, b \in \mathbb{R}$ mit

(36) $\quad E(aX + bY) = aE(X) + bE(Y) \qquad$ „Linearität“

$$\left(\text{allgemeiner: } E\left(\sum_{j=1}^{n} a_j X_j\right) = \sum_{j=1}^{n} a_j E(X_j)\right)$$

(b) Falls $E(X), E(Y)$ existieren mit $P(X \geq Y) = 1$, so gilt

(37) $\quad E(X) \geq E(Y)$

(c) Falls $E(X) < \infty$, so gilt

(38) $\quad |E(X)| \leq E(|X|) \qquad$ „Monotonie“

(d) Ist X eine diskrete, reellwertige Zufallsvariable, die nur die Werte $\{x_1, x_2, \ldots\}$ annimmt, so gilt für jede reellwertige Abbildung h:

$$(39) \quad E(h(X)) = \sum_{i=1}^{\infty} h(x_i) \cdot p_i, \text{ falls } \sum_{i=1}^{\infty} |h(x_i)| \cdot p_i < \infty$$

(e) Ist X eine Zufallsvariable mit Dichte $f(x)$, die nur höchstens endlich viele Sprungstellen besitzt und sei h eine stetige Funktion auf $Bild(X) = \{x \in \mathbb{R} | f(x) > 0\}$. Dann gilt

$$(40) \quad E(h(X)) = \int_{-\infty}^{\infty} h(x) \cdot f(x)dx, \text{ falls } \int_{-\infty}^{\infty} |h(x)| \cdot f(x)dx < \infty$$

Beweis:
Nur für den Fall diskreter Zufallsvariablen auf einem diskreten Wahrscheinlichkeitsraum (Ω, f, P) (allgemeiner Fall: Krengel (1998) mit Approximation).

Vorbemerkung:

$$\begin{aligned}
E(X) &= \sum_{j=1}^{\infty} x_j \underbrace{p_j}_{=P(\{\omega | X(\omega)=x_j\})} = \sum_{j=1}^{\infty} x_j \sum_{\omega \in X^{-1}(x_j)} P(\{\omega\}) \\
&= \underbrace{\sum_{j=1}^{\infty} \sum_{\omega \in X^{-1}(x_j)}}_{\sum_{\omega \in \Omega}} \underbrace{X(\omega)}_{=x_j} P(\{\omega\}) \\
&= \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}) = \sum_{k=1}^{\infty} X(\omega_k) P(\{\omega_k\})
\end{aligned}$$

zu a) $\underline{E(aX)} \overset{Vorb.}{=} \sum_{k=1}^{\infty} aX(\omega_k) \cdot P(\{\omega_k\}) = a \cdot \sum_{k=1}^{\infty} X(\omega_k) P(\{\omega_k\}) = \underline{aE(X)}$

$$\begin{aligned}
E(aX + bY) = E(Z) &\overset{Vorb.}{=} \sum_{k=1}^{\infty} Z(\omega_k) \cdot P(\{\omega_k\}) \\
&= \sum_{k=1}^{\infty} (aX + bY)(\omega_k) \cdot P(\{\omega_k\}) \\
&= \sum_{k=1}^{\infty} (aX(\omega_k) + bY(\omega_k)) \cdot P(\{\omega_k\}) \\
&= \sum_{k=1}^{\infty} aX(\omega_k) \cdot P(\{\omega_k\}) + \sum_{k=1}^{\infty} (bY(\omega_k)) \cdot P(\{\omega_k\}) \\
&= aE(X) + bE(Y)
\end{aligned}$$

Existenz der Reihen folgt, indem man analog zu vorangegangen Umformungen $E(|aX + bY|) \leq |a|E(|X|) + |b|E(|Y|) < \infty$ zeigt

zu b) $E(X) \overset{Vorb.}{=} \sum_{k=1}^{\infty} X(\omega_k) P(\{\omega_k\}) \geq \sum_{k=1}^{\infty} Y(\omega_k) P(\{\omega_k\}) = E(Y)$

zu c,d) Übung

zu e) aus allgemeinem Transformationssatz (siehe Krengel(1998)) □

Zur Messung der mittleren (quadratischen) Abweichung einer Zufallsvariablen von ihrem Erwartungswert definiert man die Varianz:

Definition 27:
Es sei X eine reellwertige Zufallsvariable mit Erwartungswert $E(|X|) < \infty$.
Die Größe

$$(41) \quad Var(X) := E((X - E(X))^2)$$

(falls sie endlich ist) heißt die Varianz von X.

$$(42) \quad \sigma(X) := \sqrt{Var(X)}$$

heißt dann die Standardabweichung von X.

Bemerkung 28:
$Var(X)$ misst die mittleren quadratischen Abweichungen von X von $E(X)$. Dies sichert, dass die Abweichungen alle ein positives Vorzeichen besitzen. Die mittleren Abweichungen mit Original-Vorzeichen sind nämlich 0, denn es gilt:

$$E(X - E(X)) \overset{(36)}{=} E(X) - E(E(X)) = E(X) - E(X) = 0$$

Satz 29: „Eigenschaften der Varianz"
Es sei X eine reellwertige Zufallsvariable mit existierender Varianz.
Dann gelten:

(a) $Var(X) = E(X^2) - (E(X))^2$

(b) $Var(X) \geq 0$.
Außerdem $Var(X) = 0 \Longleftrightarrow P(X = c) = 1$ für eine Konstante $c \in \mathbb{R}$

(c) $Var(aX + b) = a^2 \cdot Var(X)$ für alle $a, b \in \mathbb{R}$

Beweis: Übung

Beispiel 30:

(a) Würfeln

$$E(X) = \frac{7}{2}$$

$$E(X^2) = \sum_{j=1}^{6} j^2 \cdot \underbrace{\frac{1}{6}}_{p_j} = \frac{1}{6} \cdot \sum_{j=1}^{6} j^2 = \frac{91}{6}$$

$$\overset{\text{Satz 29 (a)}}{\Longrightarrow} Var(X) = E(X^2) - (E(X))^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}$$

(b) Gleichverteilung $U([0,1])$
$X \sim U([0,1])$
$\Longrightarrow E(X) = \frac{1}{2}, E(X^2) = \int_{-\infty}^{\infty} h(x)f(x)dx = \int_0^1 x^2 \cdot 1dx = \frac{x^3}{3}\Big|_1^0 = \frac{1}{3}$
$\Longrightarrow Var(X) = E(X^2) - (E(X))^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$

Satz 31: „Tschebyscheffsche Ungleichung“
Es sei X einer reellwertige Zufallsvariable mit endlicher Varianz.
Dann gilt für alle $\varepsilon > 0$:

$$(43)\ P(|X - E(X)| \geq \varepsilon) \leq \frac{Var(X)}{\varepsilon^2}$$

Beweis:

(i) Sei $\widetilde{\varepsilon} > 0$ und $A = (X^2)^{-1}([\widetilde{\varepsilon}, \infty)) = \{\omega | X^2(\omega) \geq \widetilde{\varepsilon}\} = \{X^2 \geq \widetilde{\varepsilon}\}$
$\Longrightarrow E(X^2) \geq E(\widetilde{\varepsilon} \cdot 1_A) = \widetilde{\varepsilon} \cdot E(1_A) = \widetilde{\varepsilon} \cdot (0 \cdot P(1_A = 0) + 1 \cdot P(1_A = 1)) = \widetilde{\varepsilon} P(A)$
$\Longrightarrow P(|X| \geq \varepsilon) = \underbrace{P(X^2 \geq \varepsilon)}_{P(A)} \leq \frac{E(X^2)}{\varepsilon^2}$ für $\varepsilon^2 = \widetilde{\varepsilon}$

(ii) Sei nun $Y = X - E(X) \overset{29c}{\Longrightarrow} Var(Y) = Var(X),\ E(Y) = 0,\ Var(Y) = E(Y^2)$
$\Longrightarrow P(|X - E(X)| \geq \varepsilon) = P(|Y| \geq \varepsilon) \overset{(i)}{\leq} \frac{E(Y^2)}{\varepsilon^2} = \frac{Var(X)}{\varepsilon^2}$ □

Eine Zufallsvariable $X = 1_A$ heißt <u>Indikatorfunktion</u> (<u>Indikatorvariable</u>) von $A \in f$, falls

$$X(\omega) := 1_A(\omega) := \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{falls } \omega \notin A \end{cases}$$

1.7 Gemeinsame Verteilung und Unabhängigkeit

Es gibt viele Größen, bei denen man aus Kenntnis der einen Größe bereits auf die Größenordnung der Anderen schließen kann (z.B. Gewicht und Größe einer zufällig ausgewählten Person). Um diese Gemeinsamkeiten und Abhängigkeiten modellieren zu können, benötigen wir den Begriff der gemeinsamen Verteilung und Korrelation.

i) Diskreter Fall (Diskrete Zufallsvariablen)
   (Ω, f, P) Wahrscheinlichkeitsraum. Hierauf seien n diskrete Zufallsvariablen $X_1, \ldots, X_n$ definiert mit Werten $\Omega'_1, \ldots, \Omega'_n$. Dann heißt die Abbildung

$$\begin{aligned} X = (X_1, \ldots, X_n) : \Omega &\longrightarrow (\Omega'_1, \Omega'_2, \ldots, \Omega'_n) = \Omega'_1 \times \ldots \times \Omega'_n = \Omega' \\ \omega &\longmapsto (X_1(\omega), \ldots, X_n(\omega)) \end{aligned}$$

   ein (diskreter) Ω'-wertiger Zufallsvektor. Da X eine diskrete Zufallsvariable ist, kann man auch Ω' als diskret annehmen. Wir können deshalb als Werteraum (Bildraum) $(\Omega', \Re(\Omega'))$ betrachten.

Definition 32:
X sei ein diskreter Ω'-wertiger Zufallsvektor. Dann heißt P^X mit

$$P^X : \Re(\Omega') \longrightarrow [0,1] \text{ und}$$

$$(44) \quad P^X(A) = P(X^{-1}(A)) = P(\{\omega | X(\omega) \in A\}) \text{ für alle } A \in \Re(\Omega')$$

die gemeinsame Verteilung von $X = (X_1, \ldots, X_n)$. Die Randverteilungen der X_j sind gegeben durch

$$\begin{aligned} (45) \quad P^{X_j}(A_j) &= P^X(\Omega'_1 \times \ldots \times \Omega'_{j-1} \times A_j \times \Omega'_{j+1} \times \ldots \times \Omega'_n) \\ &= P(\{\omega | X_j(\omega) \in A_j\}) \text{ für alle } A_j \in \Re(\Omega'_j) \end{aligned}$$

($\rightarrow$ Stimmt mit der bereits bekannten Definition der Verteilung einer eindimensionalen Zufallsvariablen überein)

Bemerkung 33:
Wie im eindimensionalen Fall ist die gemeinsame Verteilung von X durch die Wahrscheinlichkeit der einzelnen möglichen Werte $x = (x_1, \ldots, x_n)$, die X annimmt, gegeben, d.h. durch:

$$\begin{aligned} (46) \quad p(x) &= P(X = x) = P(X_1 = x_1, \ldots, X_n = x_n) \\ &= P(\{\omega \in \Omega | X_1(\omega) = x_1 \ \underline{\text{und}} \ \ldots \ \underline{\text{und}} \ X_n(\omega) = x_n\}) \end{aligned}$$

für alle $x \in \Omega'$. Es gilt dann nämlich

$$(47) \quad P^X(A) = \sum_{\omega \in A} p(x)$$

(wobei die Summe höchstens abzählbar viele (positive) Summanden besitzt.)
Die Randverteilung P^{X_j} erhält man über

$$(48) \quad P^{X_j}(A_j) = \sum_{x \in \Omega' \text{ mit } x_j \in A_j} p(x)$$

Beispiele 34: „Zweimaliges Würfeln“
$\Longrightarrow \Omega = \{(x,y)|x,y \in \{1,\dots,6\}\} = \Omega', f = \Re(\Omega),\ P(\{(x,y)\}) = \frac{1}{36} \forall (x,y) \in \Omega$

(a) $X = (X_1, X_2)$ mit X_i = Ergebnis des i-ten Wurfs $\Longrightarrow \Omega' = \Omega$
$\Longrightarrow p(x) = P(X_1 = x_1, X_2 = x_2) = P(\{(x_1,x_2)\}) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$
$= P(X_1 = x_1) \cdot P(X_2 = x_2)$ für alle $(x_1, x_2) \in \Omega = \Omega'$

(b) $X = (X_1, X_1{+}X_2) \Longrightarrow \Omega' = \{(x,y)|x \in \{1,\dots,6\}, y \in \{x{+}1,\dots,x{+}6\}\}$
$\Longrightarrow p(x,y) = P(X_1 = x, X_1 + X_2 = y)$

$$= \begin{cases} 0 & \text{für } x \geq y \\ P(X_1 = x, X_2 = y - x) = P(\{\overbrace{(x,y)}^{\in\Omega}\}) & \text{für } x \in \{1,\dots,6\}, y \in \{x+1,\dots,x+6\} \end{cases}$$

Während sich also im Fall (a) die gemeinsame Wahrscheinlichkeit für (X_1, X_2) als Produkt der Einzelwahrscheinlichkeiten für X_1 und X_2 ergibt, ist dies im Fall (b) nicht so, denn es gilt z.B. $P(X_1 = 4, X_1 + X_2 = 6) = P(X_1 = 4, X_2 = 2) = \frac{1}{36} \neq \frac{1}{6} \cdot \frac{5}{36} = P(X_1 = 4) \cdot P(X_1 + X_2 = 6)$
Der Grund hierfür liegt darin, dass zwar X_1 und X_2, nicht aber X_1 und $X_1{+}X_2$ unabhängig sind.

Definition 35:
$X_1, \dots, X_n$ seien diskrete Zufallsvariablen mit Werten in den höchstens abzählbaren Mengen $\Omega'_1, \dots, \Omega'_n$. $X_1, \dots, X_n$ heißen unabhängig, wenn für die gemeinsame Verteilung P^X des Wahrscheinlichkeitsvektors $X = (X_1, \dots, X_n)$ gilt:

$$(49) \quad P^X(A_1 \times \ldots \times A_n) = P^{X_1}(A_1) \cdot \ldots \cdot P^{X_n}(A_n) \text{ für alle } A_i \in \Re(\Omega'_i)$$

d.h. wenn sich die gemeinsame Verteilung als Produkt der Randverteilungen ergibt.

Satz 36:
$X_1, \dots, X_n$ seien diskrete Zufallsvariablen mit Werten in $\Omega'_1, \dots, \Omega'_n$. Dann gilt:

$X_1, \dots, X_n$ sind unabhängig $\Longleftrightarrow$
(50) $P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n)$
$\forall x = (x_1, \dots, x_n)$

Beweis:
„$\Longrightarrow$“ $X_1, \dots, X_n$ seien unabhängig; sei $x = (x_1, \dots, x_n) \in \Omega'$
Wähle dann $A_j = \{x_j\} \Longrightarrow A_j \in \Re(\Omega'_j)$
$\overset{(49)}{\longrightarrow} P(X_1 = x_1, \dots, X_n = x_n) = P^X(A_1 \times \ldots \times A_n) = P^{X_1}(A_1) \cdot \ldots \cdot P^{X_n}(A_n) = \underline{P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n)}$

„$\Longleftarrow$“ Es gelte (50), es seien $A_j \in \Re(\Omega'_j)$

$$\begin{aligned}\Longrightarrow A_1 \times \ldots \times A_n &= \{x | x_1 \in A_1, \ldots, x_n \in A_n\} \\ &= \bigcup_{x \in A_1 \times \ldots \times A_n =: A} \{(x_1, \ldots, x_n)\} \\ &= \left(\bigcup_{x_1 \in A_1} \{x_1\}\right) \times \ldots \times \left(\bigcup_{x_n \in A_n} \{x_n\}\right)\end{aligned}$$

$$\begin{aligned}\Longrightarrow & \underline{P^X(A_1 \times \ldots \times A_n)} \\ &= P^X\left(\dot{\bigcup_{x \in A}} \{x_1, \ldots, x_n\}\right) \\ &= \sum_{x \in A} P^X(\{(x_1, \ldots, x_n)\}) \\ &= \sum_{x_1 \in A_1} \sum_{x_2 \in A_2} \ldots \sum_{x_n \in A_n} P(X_1 = x_1, \ldots, X_n = x_n) \\ &\overset{(50)}{=} \sum_{x_1 \in A_1} \ldots \sum_{x_n \in A_n} P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n) \\ &= \sum_{x_1 \in A_1} P(X_1 = x_1) \cdot \sum_{x_2 \in A_2} P(X_2 = x_2) \cdot \ldots \cdot \sum_{x_n \in A_n} P(X_n = x_n) \\ &= \sum_{x_1 \in A_1} P^{X_1}(\{x_1\}) \cdot \ldots \cdot \sum_{x_n \in A_n} P^{X_n}(\{x_n\}) \\ &= P^{X_1}\left(\bigcup_{x_1 \in A_1} \{x_1\}\right) \cdot \ldots \cdot P^{X_n}\left(\bigcup_{x_n \in A_n} \{x_n\}\right) \\ &= \underline{P^{X_1}(A_1) \cdot \ldots \cdot P^{X_n}(A_n)}\end{aligned}$$

□

ii) Zufallsvariablen mit Dichten

Bemerkung (+Definition) 37:
Im Fall stetiger Zufallsvariablen (also solchen mit Dichten) $X_1, \ldots, X_n$ lassen sich die Definitionen 32 und 35 der gemeinsamen Verteilung, der Randverteilung und der Unabhängigkeit exakt übernehmen, wenn man die Ω'_i jeweils durch $\mathbb{R}$ und $\Re(\Omega'_i)$ durch die Borel-σ-Algebra $\wp$ ersetzt.

Definition 38:
$X_1, \ldots, X_n$ reellwertige Zufallsvariablen.
Dann heißt die Funktion $F : \mathbb{R}^n \longrightarrow [0, 1]$ mit

$$(51) \quad F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) \quad \forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

die (gemeinsame) Verteilungsfunktion der Zufallsvariablen $X = (X_1, \ldots, X_n)$.

Existiert zu F aus (51) eine Dichte, d.h. eine Funktion $f : \mathbb{R}^n \longrightarrow [0, \infty)$ mit

$$(52) \quad F(x_1, \ldots, x_n) = \int_{-\infty}^{x_n} \int_{-\infty}^{x_{n-1}} \ldots \int_{-\infty}^{x_1} f(y_1, \ldots, y_n) dy_1 dy_2 \ldots dy_n$$

so heißt f die gemeinsame Dichte von $X = (X_1, \ldots, X_n)$

Exkurs: Mehrdimensionale Integrale

$$\begin{aligned}
\int_a^b \int_c^d (x + x^2 y) dx dy &= \int_a^b \frac{x^2}{2} + y \frac{x^3}{3} \Big|_c^d dy \\
&= \int_a^b \left[\frac{d^2}{2} - \frac{c^2}{2} + y\left(\frac{d^3}{3} - \frac{c^3}{3}\right) \right] dy \\
&= \left[\left(\frac{d^2}{2} - \frac{c^2}{2}\right) y + \left(\frac{d^3}{3} - \frac{c^3}{3}\right) \frac{y^2}{2} \right] \Big|_a^b \\
&= \left(\frac{d^2}{2} - \frac{c^2}{2} \right) (b - a) + \left(\frac{d^3}{3} - \frac{c^3}{3} \right) \frac{1}{2} (b^2 - a^2)
\end{aligned}$$

Satz 39:
$X_1, \ldots, X_n$ reellwertige Zufallsvariablen mit Verteilungsfunktionen $F_{X_1}, \ldots, F_{X_n}$ und gemeinsamer Verteilungsfunktion F. Dann gelten:

(a) $F_{X_j}(x_j) = (+\infty, \ldots, +\infty, \underset{j.\text{ Stelle}}{x_j}, +\infty, \ldots, +\infty) (= \lim\limits_{z \to \infty} F(z, \ldots, z, x_j, z, \ldots, z))$

(b) Hat F auch eine gemeinsame Dichte f, so besitzt x_j eine Dichte f_j mit

$$f_j(x_j) = \underbrace{\int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty}}_{n-1 \text{ Stück}} f(x_1, \ldots, x_n) dx_1 \cdot \ldots \cdot dx_{j-1} \cdot dx_{j+1} \cdot \ldots \cdot dx_n$$

(„Ausintegration der restlichen Variablen")

(c) $X_1, \ldots, X_n$ unabhängig $\iff F(x) = \prod_{j=1}^n F_{X_j}(x_j) \forall x \in \mathbb{R}^n$

(d) Hat $(X_1, \ldots, X_n)$ eine gemeinsame Dichte f, so gilt:
$X_1, \ldots, X_n$ unabhängig $\iff f(x) = \prod_{j=1}^n f_j(x_j)$, wobei f_j die Dichte von X_j ist.

Beweis:

(a)

$$\begin{aligned}
\underline{F_{X_j}(x_j)} = &P(X_j \leq x_j) \overset{\Omega = \{X_{\underline{k}} \leq \infty\}}{=} \\
&P(X_1 \leq \infty, \ldots, X_{j-1} \leq \infty, X_j \leq x_j, X_{j+1} \leq \infty, \ldots X_n \leq \infty) \\
= &\underline{F(\infty, \ldots, \infty, x_j, \infty, \ldots, \infty)}
\end{aligned}$$

(b)

$$\begin{aligned} F_{X_j}(x_j) &\overset{(a)}{=} F(\infty, \dots, \infty, x_j, \infty, \dots, \infty) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \int_{-\infty}^{x_j} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(y_1, \dots, y_n) dy_1, \dots, dy_n \end{aligned}$$

Vertauschung der Integrationsreihenfolge:

$$\begin{aligned} &\int_{-\infty}^{x_j} \left[\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(y_1, \dots, y_n) dy_1 \dots dy_{j-1} dy_{j+1} \dots dy_n \right] dy_j \\ &=: \int_{-\infty}^{x_j} f_j(y_j) dy_j \end{aligned}$$

($f_j \geq 0$, da $f \geq 0$, $\int_{-\infty}^{\infty} f_j(y_j) dy_j = 1$, da $F_{X_j}(\infty) = 1$)

(c) „$\Longrightarrow$“ aus Definition der Unabhängigkeit
„$\Longleftarrow$“ ähnlich wie im diskreten Fall

(d) aus (c)

Bemerkung 40:

(a) Teil (b) des Satzes ist nicht umkehrbar. Es gibt Fälle, in denen $X_1, \dots, X_n$ jeweils eine Dichte besitzen, aber $X = (X_1, \dots, X_n)$ keine gemeinsame Dichte besitzt.

(b) Analog zu Satz 26 (d)+(e) gelten: $h : \mathbb{R}^n \longrightarrow \mathbb{R}, h$ stetig, $Y = h(X_1, \dots, X_n) \Longrightarrow$ Falls $E(h(X_1, \dots, X_n))$ existiert (im Sinn von Definition 25), so gelten:

$$E(h(X_1, \dots, X_n)) = \sum_{j=1}^{\infty} h(x_1^{(j)}, \dots, x_n^{(j)}) P(X = \underbrace{x^{(j)}}_{\in \mathbb{R}^n (*)}) \text{ „diskreter Fall“} \tag{53}$$

$(*)$: mögliche Werte $x^{(1)}, x^{(2)}, \dots$ von X

$$E(h(X_1, \dots, X_n)) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h(x_1, \dots, x_n) \cdot f(x_1, \dots, x_n) dx_1 \dots dx_n \tag{54}$$

Mit dieser Bemerkung und den vorangegangenen Sätzen und Definitionen sind wir nun in der Lage, die angekündigten Maße für den Zusammenhang zweier Zufallsvariablen, die Kovarianz und die Korrelation zu definieren:

Definition 41:
X, Y seien reellwertige Zufallsvariablen mit $Var(x) < \infty, Var(y) < \infty$

(a)

$$\begin{aligned} Cov(X, Y) &= E((X - E(X)) \cdot (Y - E(Y))) \\ &= E(XY) - E(X) \cdot E(Y) \text{ heißt die \underline{Kovarianz} von } X \text{ und } Y \end{aligned}$$

(b)

$$Kor(X,Y) := \rho_{X,Y} := \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

(falls $Var(X) > 0, Var(Y) > 0$)
heißt Korrelation von XxY oder X und Y.

Satz 42:„Eigenschaften von Kovarianz und Korrelation"
X, Y reellwertige Zufallsvariablen mit $Var(X) < \infty, Var(Y) < \infty \Longrightarrow$

(a)

$$(55) \quad Var(X+Y) = Var(X) + Var(Y) + 2 \cdot Cov(X,Y), \text{ allgemeiner}$$

$$(56) \quad Var\left(\sum_{j=1}^{n} X_j\right) = \sum_{j=1}^{n} Var(X_j) + \sum_{i,j=1}^{n} Cov(X_i, X_j)$$

(b) X, Y unabhängig $\Longrightarrow$

$$(57) \quad Var(X+Y) = Var(X) + Var(Y)$$

allgemein: $X_1, \ldots, X_n$ unabhängig $\Longrightarrow$

$$(58) \quad Var(\sum_{j=1}^{n} X_j) = \sum_{j=1}^{n} Var(X_j)$$

(c) $Cov(X,Y) = Cov(Y,X), \quad Cov(aX+b, cY+d) = a \cdot c \cdot Cov(X,Y)$

(d) $|\rho_{X\ Y}| \leq 1$

(e) $\rho_{X\ Y} = +1 \Longleftrightarrow Y = aX + b$ für ein $a > 0$
$\rho_{X\ Y} = -1 \Longleftrightarrow Y = aX + b$ für ein $a < 0$

(f) X, Y unabhängig $\Longrightarrow Cov(X,Y) = 0 = \rho_{X\ Y}$ („X, Y sind unkorreliert")
Die Umkehrung gilt im allgemeinen nicht.

Bemerkung 43:
Da der Wertebereich der Korrelation auf $[-1,1]$ normiert ist, ist sie das „bessere" Maß für die Abhängigkeit zwischen X und Y als die Kovarianz, die nicht normiert und nicht invariant unter einem Maßstabswechsel ist. Ein Wert von $\rho_{X,Y}$ nahe bei 1 besagt, dass man bei einem hohen X-Wert in der Regel auch einen hohen Y-Wert beobachten wird, während ein niedriger X-Wert in der Regel gemeinsam mit einem niedrigen Y-Wert zusammen auftritt.

Beweis: (Satz 42)

(c)

$$\begin{aligned}Cov(X,Y) &= E(X \cdot Y) - E(X) \cdot E(Y)\\ &= E(Y \cdot X) - E(Y) \cdot E(X) = Cov(Y,X)\end{aligned}$$

$$\begin{aligned}Cov(aX+c, cY+d) =& E((aX+b-E(aX+c))(cY+d-E(cY+d)))\\ & E(a(X-E(X))c(Y+E(Y)))\\ & ac \cdot E((X-E(X))(Y+E(Y))) = ac \cdot Cov(X,Y)\end{aligned}$$

(a) nur für $n = 2$

$$\begin{aligned}Var(X+Y) &= E((X-Y)^2) - (E(X+Y))^2\\ &= E(X^2 - 2XY + Y^2) - ((E(X))^2 + 2E(X)E(Y) + (E(Y))^2)\\ &= E(X^2) - (E(X))^2 + E(Y^2) - (E(Y))^2 + 2(E(XY) - E(X)E(Y))\\ &= Var(X) + Var(Y) + 2Cov(X,Y)\end{aligned}$$

(f) nur für Fall mit Dichten:

$$\begin{aligned}E(XY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f(x,y)dxdy\\ &\overset{X,Y \text{ unabhängig}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \cdot f_X(x)f_Y(y)dxdy\\ &= \int_{-\infty}^{\infty} xF_X(x)dx \cdot \int_{-\infty}^{\infty} yf_Y(y)dy = E(X)E(Y)\end{aligned}$$

$$\underline{\underline{Cov(X,Y)}} = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = \underline{\underline{0}}$$

(b) aus (a) + (f)

(d) + (e)

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \overset{(*)}{=} \frac{E(XY)}{\sqrt{E(X^2)}\sqrt{E(Y^2)}}$$

($*$) o.B.d.A. (wegen (c): E(X)=E(Y)=0)

Cauchy'sche Ungleichung:

$$E(XY) \leq \sqrt{E(X^2)} \cdot \sqrt{E(Y^2)}$$

wobei Gleichheit nur im Fall $Y = aX + b$ gilt.

$$\Longrightarrow \rho^2_{X,Y} = \frac{(E(XY))^2}{E(X^2)E(Y^2)} \leq \frac{E(X^2)E(Y^2)}{E(X^2)E(Y^2)} = 1$$

1.8 Die Gesetze der großen Zahl und MC-Simulation

Die Gesetze der großen Zahl besagen, dass die empirische Beobachtung, dass sich das arithmetische Mittel $\frac{1}{N} \cdot \sum_{i=1}^{N} x_i$ der beobachteten Ergebnisse bei häufiger unabhängiger Wiederholung eines Experiments mit wachsendem N um einen Wert zu stabilisieren scheint, ihr Analogon im Modell besitzt, denn es gilt:

Satz 44: „Schwaches Gesetz der großen Zahl"
$X_1, X_2, \ldots$ sei eine Folge reellwertiger Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, f, P), die alle die selbe Verteilung besitzen mit $\mu = E(X_j)$, $\sigma^2 = Var(X_j) < \infty$ für alle $j \in \mathbb{N}$. Dann gilt:

$$(59) \quad \lim_{N\to\infty} P\left(\left|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\right| \geq \varepsilon\right) = 0 \quad \text{für alle } \varepsilon > 0.$$

Man sagt auch:
$\frac{1}{N}\sum_{i=1}^{N} X_i$ konvergiert stochastisch (oder in Wahrscheinlichkeit) gegen μ.

Beweis:

$$\begin{aligned} \text{Sei } Y_N &= \sum_{i=1}^{N} X_i \\ \Rightarrow E(Y_N) &= \sum_{i=1}^{N} E(X_i) = N \cdot \mu \\ Var(Y_N) &= \sum_{i=1}^{N} Var(X_i) = N \cdot \sigma^2 \end{aligned}$$

$$\begin{aligned} P(|\frac{1}{N}\sum_{i=1}^{N} X_i - \mu| \geq \varepsilon) = P(|\frac{1}{N}Y_N - \mu| \geq \varepsilon) &= P(|Y_N - N\mu| \geq N \cdot \varepsilon) \\ &\overset{31}{\leq} \frac{Var(Y_N)}{N^2\varepsilon^2} = \frac{\sigma^2}{N\varepsilon^2} \overset{N\to\infty}{\Longrightarrow} 0 \end{aligned}$$

Wichtiger Spezialfall:
Satz 45: „Bernoullis Gesetz der großen Zahl"
$A_1, A_2, \ldots$ seien unabhängige Ereignisse mit gleicher Wahrscheinlichkeit $p = P(A_i)$. Dann gilt:

$$(60) \quad \lim_{N\to\infty} P\left(\left|\frac{1}{N}\sum_{i=1}^{N} 1_{A_i} - p\right| \geq \varepsilon\right) = 0 \quad \forall \varepsilon > 0$$

Beweis:
Es gilt (Übung) $A_1, A_2, \ldots$ unabhängig $\Longleftrightarrow 1_{A_1}, 1_{A_2}, \ldots$ unabhängig.
Damit folgt Behauptung aus Satz 44, wenn man dort $X_i = 1_{A_i}$ wählt

Diskussion:

- Kann man also ein Experiment oft unter den gleichen Bedingungen durchführen, so besagt Satz 45, dass die relative Häufigkeit des Eintretens eines Ereignisses gegen seine Eintrittswahrscheinlichkeit p konvergiert. Wir haben also auch im Modell ein Analogon zur „Stabilisierungsbeobachtung“ in der Realität.

- Genau genommen werden wir für die folgende Anwendung der Monte-Carlo-Simulation noch eine schärfere Variante der Konvergenzaussage (59) benötigen, nämlich:

Satz 46: „Kolmogoroffsches Gesetz der großen Zahl“
$X_1, X_2, \dots$ Folge unabhängiger reellwertiger Zufallsvariablen auf (Ω, f, P), die alle dieselbe Verteilung mit $\mu = E(X_i)$ besitzen.

$$(61) \quad P\left(\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} X_i = \mu\right) = 1$$

Man sagt dann auch: $\frac{1}{N}\sum_{i=1}^{N} X_i$ konvergiert fast sicher gegen μ.

Bemerkung 47:

(a) (61) besagt, dass für fast alle $\omega \in \Omega$ gilt:

$$(*) \quad \frac{1}{N}\sum_{i=1}^{N} X_i(\omega) \overset{N\Rightarrow\infty}{\longrightarrow} \mu$$

(genauer: Die Wahrscheinlichkeit der Menge aller dieser ω ist 1)
Hieraus folgt insbesondere auch (59), während umgekehrt aus (59) noch nicht mal für ein $\omega \in \Omega$ die Beziehung (*) folgen muss! Da wir allerdings als Ausgänge von Experimenten jeweils die Zahlen $X_i(\omega)$ beobachten (und zwar nur die!), ist (*) die für uns relevante Aussage.

(b) Da (61) (59) impliziert, bezeichnet man Satz 46 auch als starkes Gesetz der großen Zahl. Sein Beweis übersteigt den Rahmen dieser Vorlesung.

Anwendung: „Monte-Carlo-Simulation“

1. Experimentelle Bestimmung des Erwartungswertes
   Der Ausgang eines Experiments sei durch eine Zufallsvariable X beschrieben, von der wir annehmen, dass sie einen endlichen Erwartungswert μ besitzt.

$$\text{z.B. } X = \begin{cases} 1 \text{ falls Zahl} \\ 0 \text{ falls Kopf} \end{cases} \quad \text{beim Münzwurf}$$
$$\Longrightarrow \mu = E(X) = P(X=1) = P(\text{„Zahl“})$$

   Allerdings haben wir sonst keine Information über die Verteilungsfunktion von X.

Ziel: Bestimmung von μ. Methode: Simuliere X hinreichend oft, also:

- Führe das zugehörige Experiment N mal durch (mit N groß)
- Notiere die Ergebnisse $x_i = X_i(\omega)$ der einzelnen Ergebnisse
- Bilde $\frac{1}{N}\sum_{i=1}^N x_i \left(= \frac{1}{N}\sum_{i=1}^N X_i(\omega)\right)$
  (z.B. Werfe Münze hinreichend oft und bestimme die relative Häufigkeit von „Zahl")
  mit Satz 46: $\mu = E(X) \approx \frac{1}{N}\sum_{i=1}^N X_i$

2. Integration durch Monte-Carlo-Simulation

Problem: Bestimme $\int_0^1 f(x)\, dx$ für eine gegebene (aber eventuell komplizierte) stetige Funktion $f(x)$.

Hilfsmittel: Computer, der auf $[0,1]$ gleichmäßig verteilte Zufallsvariablen erzeugen kann.

Idee: Für $X \sim U([0,1])$ gilt nach Satz 26e):
$E(X) = \int_0^1 f(x) \cdot 1\, dx$
$\Longrightarrow$ „Monte-Carlo-Methode der Integration"

Simuliere N gleichmäßig verteilte Zufallszahlen $x_1, \ldots, x_n$ und berechne $f(x_1), \ldots, f(x_n)$

$$\overset{46}{\Longrightarrow} E(f(x)) \approx \frac{1}{N}\sum_{i=1}^N f(x_i)$$

$$\int_0^1 f(x)\, dx \approx \frac{1}{N}\sum_{i=1}^N f(x_i)$$

Es kann gezeigt werden (eventuell später), dass für den oben gemachten Approximationsfehler gilt:

$$E\left(\left|\int_0^1 f(x)\, dx - \frac{1}{N}\sum_{i=1}^N X_i\right|\right) \le const \cdot \frac{1}{\sqrt{N}}$$

wobei *const* eine von f abhängige Konstante ist.
Das heißt insbesondere, dass für jede weitere Stelle (mittlerer) Genauigkeit eine Erhöhung der Anzahl der Zufallszahlen um den Faktor 100 nötig ist!
Obwohl also die MC-Methode der Integration langsam konvergiert, ist sie in Problemen hoher Dimension oft die beste numerische Methode.
Sie ist auch auf Integrale der Form $\int_a^b f(x)\, dx$ anwendbar. Mann muss lediglich im 1. Schritt die Zufallszahl $x_1, \ldots, x_n$ durch $y_1 = a + (b-a)x_1, \ldots, y_n = a + (b-a)x_n$ ersetzen ($Y = a + (b-a)X \sim U([a,b])$) und dann mit $f(y_1), \ldots, f(y_n)$ weiterrechnen.

1.9 Der Zentrale Grenzwertsatz

Der Zentrale Grenzwertsatz (ZGS) ist ein sehr nützliches Hilfsmittel, um eine Verteilung von Summen unabhängiger, identisch verteilter Zufallsvariablen („Wie wahrscheinlich ist es z.B. beim 500-fachen Würfeln eine Augensumme von höchstens 1600 zu erzielen?“) durch eine geeignete Normalverteilung zu approximieren. Er kann auch verwendet werden, um die Genauigkeit der Approximation eines Erwartungswertes durch das arithmetische Mitteln von Ergebnissen wiederholter Versuche zu bestimmen (z.B. bei der MC-Methode).

Exkurs: Eigenschaften der Normalverteilung
Die durch die Dichte

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in \mathbb{R}$$

gegebene Verteilung $\mathcal{N}(\mu, \sigma^2)$ heißt Normalverteilung mit den Parametern $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}\backslash\{0\}$. Ihre Verteilungsfunktion wird mit

$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma^2}(t)\, dt,\ x \in \mathbb{R}$$

bezeichnet. Im Fall $(\mu, \sigma^2) = (0, 1)$ spricht man von der Standardnormalverteilung und verwendet die Bezeichnungen $\varphi(\cdot)$, $\Phi(\cdot)$.

Im Folgenden sei X $\mathcal{N}(\mu, \sigma^2)$-verteilt. Dann gelten:

(N1) $E(X) = \mu,\ Var(X) = \sigma^2$

(N2) $\Phi_{\mu,\sigma^2}(x) = \Phi\left(\frac{\mu-x}{\sigma}\right)$ für $x \in \mathbb{R}$ (wobei $\sigma > 0$)

(N3) $\Phi_{\mu,\sigma^2}(\mu) = \frac{1}{2}$ „Symmetrie“

(N4) $\Phi_{\mu,\sigma^2}(x + \mu) = 1 - \Phi_{\mu,\sigma^2}(\mu - x)$ für $x > 0$

(N5) $\Phi(x) = 1 - \Phi(-x)$ für $x > 0$

(N6) $P(X \in [\mu - \sigma, \mu + \sigma]) \approx 0.68$
$P(X \in [\mu - 2\sigma, \mu + 2\sigma]) \approx 0.95$
$P(X \in [\mu - 3\sigma, \mu + 3\sigma]) \approx 0.996$

(N7) $X \sim \mathcal{N}(\mu_1, \sigma_1^2),\ Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ unabhängig
$\Longrightarrow X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Satz 48: „Zentraler Grenzwertsatz für Summen unabhängiger Zufallsvariablen“
Es seien $X_1, X_2, \ldots$ unabhängige, identisch verteilte Zufallsvariablen mit $E(X_i) = \mu < \infty,\ 0 < Var(X_i) = \sigma^2 < \infty$. Dann gilt mit der Bezeichnung $S_n = \sum_{i=1}^{n} X_i$ für die Verteilungsfunktion $F_{S_n^*}$ von $S_n^* = \frac{S_n - E(S_n)}{\sqrt{Var(S_n)}}$

$$(62)\quad \lim_{n\to\infty} \sup_{X\in\mathbb{R}} |F_{S_n^*}(X) - \Phi(X)| = 0$$

Anwendung des ZGS:
Für die Verteilung P^{S_n} von S_n gilt wegen (62) für <u>große</u> n:

(63) P^{S_n} ist ungefähr gleich $\mathcal{N}(E(S_n),\ Var(S_n))$

So gilt z.B. im Problem „Wie wahrscheinlich ist bei 500-fachem Würfeln eine Augensumme kleiner gleich 1600?“

$$\begin{aligned} X_i &= \text{„Augensumme des } i\text{. Wurfs“} \\ E(X_i) &= \frac{7}{2} \\ Var(X_i) &= \frac{35}{12} \end{aligned}$$

$$\begin{aligned} \Longrightarrow \underset{(n=500)}{S_n} &= \sum_{i=1}^{500} X_i \\ E(S_n) &= 500 \cdot \frac{7}{2} = 1750 \\ Var(S_n) &\approx 1458.33 \end{aligned}$$

$$\begin{aligned} \Longrightarrow P(S_n \leq 1600) &\approx \Phi_{1750,1458.33}(1600) \\ &\overset{(N2)}{=} \Phi\left(\frac{1600-1750}{\sqrt{1458.33}}\right) \\ &= \Phi(-3.928) \approx 0 \\ P(S_n \leq 1700) &\overset{(63),(N2)}{\approx} \Phi(-1.31) \approx 0.095 \\ P(S_n \leq 1725) &\approx 0.26 \end{aligned}$$

Auf einen Beweis des ZGS wird im Rahmen dieser Vorlesung verzichtet. Ein mit den in dieser Vorlesung bereitgestellten Mitteln durchführbarer, aber länglicher Beweis befindet sich z.B. im Krengl (1998).

Kapitel 2

Statistik

2.1 Beschreibende Statistik

Ziele:

- Stelle vorliegende Daten übersichtlich in tabellarischer und graphischer Form dar
- Berechne einfache statistische Maßzahlen, die die Daten charakterisieren

Ausgangspunkt einer jeden statistischen Untersuchung ist die sogenannte Urliste, in der die gesammelten Daten vorhanden sind.
Bei den beobachteten Daten unterscheidet man zwischen qualitativen („artmäßigen") Merkmalen und quantitativen („mengenmäßigen") Merkmalen. Auch zwischen diskreten und stetigen Merkmalen wird unterschieden.

Wir betrachten im folgenden nur ein Merkmal.

2.1.1 Häufigkeitsverteilung, Histogramm und Kontingenztafel

- Eine Häufigkeitsverteilung gibt die absoluten und/oder relativen Häufigkeiten der einzelnen Ausprägungen eines Merkmals in den vorhandenen Daten an.
  $\Longrightarrow$ nur sinnvoll bei einem Merkmal mit wenigen Ausprägungen.

Bei einem diskreten Merkmal mit vielen Ausprägungen (d.h. mit vielen möglichen Werten) oder einem stetigen Merkmal ist eine Klassenbildung (d.h. Einteilung des Wertebereichs des Merkmals in „gleich lange" Teilintervalle) sinnvoll.

Beachte:		
wenige Klassen	$\rightarrow$	einfache Struktur der Häufigkeitsverteilung, aber eventuell wenig Information
viele Klassen	$\rightarrow$	viel Informationen, aber eventuell viele schwach besetzte Klassen

Praxis:
Wähle 10-20 Klassen bei Merkmalen mit vielen Ausprägungen (DIN-Empfehlung: Ist N die Anzahl der möglichen Ausprägungen, so soll die Anzahl der Klassen nicht größer als $\sqrt{N}$ sein).

- Eine Möglichkeit zur graphischen Darstellung quantitativer Daten ist das Histogramm, bei dem auf der x-Achse der Wertebereich der Merkmals und auf der y-Achse die (relativen) Häufigkeiten der einzelnen Klassen abgetragen werden.

- Will man mehr als ein Merkmal darstellen, so ist eine zweidimensionale Verallgemeinerung des Histogramms möglich, aber im Allgemeinen unübersichtlich (dreidimensionales Gebilde). Eine mögliche Darstellung zweier Merkmale in einem Graphen ist z.B. die Punktwolke, in der alle auftretenden Daten (x_i, y_i) dargestellt sind.
  Haben die Merkmale A, B die Ausprägungen $a_1, \ldots, a_m$ und $b_1, \ldots, b_k$ bzw. sind in diese Klassen eingeteilt, so hat die zugehörige Kontingenztafel folgendes Aussehen:

	b_1	$\ldots$	b_k	Summe
a_1	h_{11}	$\ldots$	h_{1k}	$h_{1.}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
a_m	h_{m1}	$\ldots$	h_{mk}	h_m
Summe	$h_{.1}$	$\ldots$	$h_{.k}$	N (oder 1)

$$\begin{aligned} \text{wobei } h_{ij} &= \text{Anzahl der Beobachtungen von } (a_i, b_j) \\ \text{oder } h_{ij} &= \frac{\text{Anzahl der Beobachtungen}}{N} \\ \text{mit } h_{i.} &= \sum_{j=1}^{k} h_{ij} \text{ und } h_{.j} = \sum_{i=1}^{m} h_{ij} \end{aligned}$$

2.1.2 Einfache Maßzahlen

(a) Lagemaße (geben an, wo die Daten auf der x-Achse liegen)

i) Mittelwert: $\overline{x}_n = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$

ii) Modalwert: $x_n^{(M)}$ = der/ein am häufigsten vorkommende(r) Wert in den Daten

iii) Median: $\dot{x}_n$ = mittlerer Wert der Daten,
d.h. sind $x_1, \ldots, x_n$ in geordneter Reihenfolge als $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ aufgeschrieben, so gilt

$$\dot{x}_n = \begin{cases} x_{(\frac{n+1}{2})} & \text{falls } n \text{ ungerade} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}) & \text{falls } n \text{ gerade} \end{cases}$$

Bemerkung:

1. Der Mittelwert ist sehr empfindlich gegenüber Ausreißern, (d.h. außergewöhnlich großen oder kleinen Werten), der Median nicht.
2. Bei einer Wahrscheinlichkeitsverteilung entsprechen der Erwartungswert dem Mittelwert, der „wahrscheinlichste" Wert (bzw. ein/das Maximum der Dichte) dem Modalwert sowie ein Wert $\dot{x}$ mit $P(X \geq \dot{x}) \geq \frac{1}{2}$ und $P(X \leq \dot{x}) \geq \frac{1}{2}$ dem Median.

(b) Streuungsmaße
(beschreiben die Streuung der Daten unabhängig von ihrer Lage)

i) Stichprobenvarianz $\tilde{s}_n^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x}_n)^2$
Empirische Varianz $s_n^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x}_n)^2$
Wir werden später sehen, daß s_n^2 bessere Eigenschaften als $\tilde{s}_n^2$ hat.

ii) Spannweite $R = x_{max} - x_{min}$ = größter Wert - kleinster Wert

iii) Quartilsabstand $\tilde{x}_{0.75} - \tilde{x}_{0.25}$ = Abstand zwischen dem 75%-Quantil der Daten (mindestens 75% der Daten sind kleiner/gleich $\tilde{x}_{0.75}$ und mindestens 25% der Daten sind größer/gleich $\tilde{x}_{0.25}$) und dem 25%-Quantil der Daten.

Bemerkung:
In $[\tilde{x}_{0.25}, \tilde{x}_{0.75}]$ liegen mindestens 50% der Daten.

(c) Schiefe

$$S = \frac{1}{n} \cdot \frac{\sum_{i=1}^n (x_i - \overline{x}_n)^2}{s_n^{\frac{3}{2}}}$$

misst, ob die Daten symmetrisch ($S = 0$), linksssschief ($S < 0$) oder rechtsschief ($S > 0$) um den Mittelwert liegen.

(d) Korrelationskoeffizient
ist ein Maß für den (linearen) Zusammenhang zwischen zwei reellwertigen Merkmalen.

$$\rho_{x,y} = \frac{\sum_{i=1}^n (x_i - \overline{x}_n)(y_i - \overline{y}_n)}{\sqrt{\sum_{i=1}^n (x_i - \overline{x}_n)^2}\sqrt{\sum_{i=1}^n (y_i - \overline{y}_n)^2}} \in [-1, 1]$$

Gilt $|\rho_{x,y}| = 1$, so liegen alle Beobachtungspaare (x_i, y_i) auf einer Geraden. Ein betragsmäßig hoher Korrelationskoeffizient spricht für (fast) lineare Abhängigkeit von x und y.

Zur Erinnerung:
X, Y unabhängig $\Longrightarrow X, Y$ unkorreliert.
Sind X, Y gemeinsam normalverteilt, so gilt "$\Longleftrightarrow$"

2.2 Schätzen von Verteilungsparametern

2.2.1 Punktschätzer

Ziel: Gewinne mit Hilfe der Beobachtungen („Daten") Informationen über die Parameter, die das Modell ($\hat{=}$ Wahrscheinlichkeitsverteilung) bestimmen.

Beispiel 2:
Bei einer Abfüllmaschine für Getränke werden die Abfüllmengen $x_1, \ldots, x_N$ pro Flasche gemessen. Wir modellieren die Abfüllmenge durch eine Zufallsvariable X und wollen

(a) $E(X)$ schätzen: naheliegend, wähle $\overline{X}_N = \frac{1}{N}\sum_{i=1}^N X_i$ („Schätze Erwartungswert durch Mittelwert").

(b) $Var(X)$ schätzen: $Var(X) = E(X - E(X))^2$

  (i) bei bekanntem $E(X)$: wähle $\hat{s}_N^2 = \frac{1}{N}\sum_{i=1}^N (X_i - E(X))^2$

  (ii) bei unbekanntem $E(X)$: wähle $s_N^2 = \frac{1}{N-1}\sum_{i=1}^N (X_i - \overline{X_N})^2$

Definition 3:
Z sei eine reellwertige Zufallsvariable, deren Verteilung von einem unbekannten Parameter $\theta \in \Theta \subseteq \mathbb{R}^d$ abhängt (aber sonst bekannt ist).

(a) Sind $Z_1, \ldots, Z_N$ unabhängig und identisch verteilt („u.i.v.") wie Z, so heißt $Z_1, \ldots, Z_N$ eine Stichprobe. Zusammen mit obiger Verteilungsannahme spricht man von einem statistischen Modell.

(b) Eine Funktion $T : \mathbb{R}^N \longrightarrow \Theta$ heißt eine Schätzfunktion oder ein (Punkt-)Schätzer von θ. Hat man bereits Werte („Realisierungen") $z_1, \ldots, z_N$ beobachtet, so heißt die Zahl $T(z_1, \ldots, z_N)$ ein Schätzwert für θ.

Im Beispiel 2(a):

$$\theta = E(X),\ \Theta = [0, \infty),\ T(X_1, \ldots, X_N) = \frac{1}{N}\sum_{i=1}^N X_i$$

Definition 4:

(a) Ein Schätzer $T(X_1, \ldots, X_N)$ für θ heißt erwartungstreu, falls
(1) $E_\theta(T(X_1, \ldots, X_N)) = \theta$ „ im Mittel richtig"
gilt, wobei $E_\theta(\ldots)$ den Erwartungswert bezüglich der Verteilung P_θ bezeichnet.

(b) Ein Schätzer $T(X_1, \ldots, X_N)$ für θ heißt (stark) konsistent, falls
(2) $P_\theta(\lim_{N\to\infty} T(X_1, \ldots, X_N) = \theta) = 1$
gilt, d.h. bei unendlichem Stichprobenumfang ergibt der Schätzer fast sicher den wahren Wert von θ.

Satz 5:

(a) $\overline{X}_N$ ist ein erwartungstreuer Schätzer für $\theta = E(X)$

(b) $\hat{s}{}_N^2$ ist ein erwartungstreuer Schätzer für $\theta = Var(X)$, falls $E(X)$ bekannt ist.
s_N^2 ist immer ein erwartungstreuer Schätzer für $\theta = Var(X)$.
Insbesondere ist die Stichprobenvarianz
$\tilde{s}{}_N^2 = \frac{1}{N} \cdot \sum_{i=1}^N (X_i - \overline{X}_N)^2$ <u>nicht</u> erwartungstreu.

Beweis (nur für (a))

$$\begin{aligned} E_\theta(T(X_1, \ldots, X_N)) &= E_\theta\left(\frac{1}{N}\sum_{i=1}^N X_i\right) \\ &= \frac{1}{N}\sum_{i=1}^N E_\theta(X_i) \\ &= \frac{1}{N}\sum_{i=1}^N \theta \\ &= \theta \end{aligned}$$

Satz 6:

(a) $\overline{X}_N$ ist ein (stark) konsistenter Schätzer für $\theta = E(X)$

(b) Sowohl $\hat{s}{}_N^2$ als auch s_N^2 und $\tilde{s}{}_N^2$ sind in den jeweiligen Situationen (stark) konsistente Schätzer für $\theta = Var(X)$

(c) Ist f eine stetige Funktion, so ist

$$\frac{1}{N} \cdot \sum_{i=1}^N f(X_i)$$

ein (stark) konsistenter Schätzer für $\theta = E(f(X))$ und
$f(\overline{X}_N)$ ein (stark) konsistenter Schätzer für $\theta = f(E(X))$ □

Beweis:

(a) folgt aus dem starken Gesetz der großen Zahl

(c) folgt aus dem starken Gesetz der großen Zahl und der Tatsache, dass folgendes gilt:
$Y_N \underset{N\to\infty}{\longrightarrow} Y$ fast sicher
$\Longrightarrow f(Y_N) \underset{N\to\infty}{\Longrightarrow} f(Y)$ fast sicher.

(b) folgt aus (c) z.B. für $\hat{s}{}_N^2$ wähle $f(X) = (x - E(X))^2$

Beispiel 7:
Die Zeiten bis zum ersten Ausfall einer Computerfestplatte seien $Exp(\theta)$-verteilt

Ziel: Schätze θ
Wir wissen: $E(X) = \frac{1}{\theta}$ (wobei $X \sim Exp(\theta)$) $\quad (X$: Ausfallzeit$)$
Wir beobachten deshalb $X_1, \ldots, X_N$ u.i.v Zufallsvariablen („Festplattenausfallzeiten") und schätzen wegen Satz 6(c) mit $f(x) = \frac{1}{x}$, θ durch

$$\hat{\theta} = \frac{1}{\overline{X}_N} \text{ („Schätze } f(E(X)) \text{ durch } f(\overline{X}_N)\text{")}$$

Nach Satz 6(c) ist $\hat{\theta}$ ein (stark) konsistenter Schätzer für θ (aber nicht erwartungstreu!)

Bemerkung 8:
Unter Regularitätsannahmen kann auch gezeigt werden, dass die Quantile der Daten für $N \longrightarrow \infty$ auch fast sicher gegen die Quantile der dem statistischen Modell zugrunde liegenden Verteilungsfunktionen konvergieren. Dies gilt insbesondere für Median und Quartilsabstand. Sie sind deshalb (stark) konsistente Schätzer.

Liste (stark) konsistenter Schätzer:
$X_1, \ldots X_N$ u.i.v. mit Erwartungswert μ, Varianz σ^2, Median m

zu schätzender Parameter	Schätzer
$\mu = E(X)$	$\overline{X}_N$
$\theta = f(\mu)$	$f(\overline{X}_N)$
$\theta = E(f(X))$	$\frac{1}{N} \cdot \sum_{i=1}^{N} f(X_i)$
$\sigma^2 = Var(X)$	s_N^2 (auch $\tilde{s}_N^2$)
$m = med(X)$	$\dot{X}'_N$
$q_{0.25}$ 25%-Quantil	$\tilde{x}_{0.25}$
$q_{0.75}$ 75%-Quantil	$\tilde{x}_{0.75}$

Problem:
Wie konstruiert man allgemein „gute" Schätzer?
hier nur: Maximum-Likelihood-Methode.

2.2.2 Die Maximum-Likelihood-Methode

Mit Hilfe der Ml-Methode lassen sich systematisch „gute" Schätzer konstruieren (in der Regel...). Sie basiert darauf, den Schätzer $\hat{\theta} = T(X_1, \ldots, X_N)$ für θ zu wählen, so dass die von uns beobachteten Daten $x_1, \ldots, x_N$ maximale Wahrscheinlichkeit für ihr Auftreten unter allen möglichen Werten $\theta \in \Theta$ besitzen. Diesem liegt das Prinzip zugrunde, dass wir viel eher „sehr wahrscheinliche Daten" als „extrem seltenen Daten" beobachten werden.

Annahme:
Die den Daten zugrunde liegende Verteilung ist entweder diskret oder besitzt eine Dichte.

Definition 9:
Ist die Verteilung P_θ im statistischen Modell eine diskrete Verteilung und werden die Daten $x_1, \ldots, x_N$ beobachtet, so heißt die Funktion

$$(3) \quad L(\theta) := L(\theta, x_1, \ldots, x_N) = P_\theta(X_1 = x_1) \cdot \ldots \cdot P_\theta(X_N = x_N) = \Pi_{j=1}^{N} P_\theta(X_j = x_j)$$

die zur Beobachtung $x_1, \ldots, x_N$ gehörende Likelihood-Funktion. Jeder Parameter $\hat{\theta} \in \Theta$, für den $L(\theta)$ maximal wird, also

$$(4) \quad L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$$

erfüllt, heißt ein Maximum-Likelihood-Schätzwert (kurz Mls) für θ. Jeder Schätzer $\hat{\theta}(X_1, \ldots, X_N) = T(X_1, \ldots, X_N)$, bei dem für alle möglichen Beobachtungen $x_1, \ldots, x_N$ $T(x_1, \ldots, x_N)$ ein Mls ist, heißt ein Maximum-Likelihood-Schätzer (kurz MlS) für θ.

Beispiel:
In der Mensa wurden an N verschiedenen Tagen $x_1, \ldots, x_N (\in \mathbb{N})$ ankommende Kunden an Essensausgabe 1 pro Stunde (zw. 12 und 13 Uhr) gezählt. Wir gehen davon aus, dass die Kundenzahl X pro Stunde $Pn(\theta)$-verteilt ist mit $\theta > 0$.

Ziel: Bestimme MlS für θ
$\Longrightarrow L(\theta) = L(\theta, x_1, \ldots, x_N) = \prod_{j=1}^{N} e^{-\theta} \frac{\theta^{x_j}}{x_j!} = e^{-N\theta} \prod_{j=1}^{N} \frac{\theta^{x_j}}{x_j!}$ ist in θ zu maximieren!
Um dies durchzuführen ist es oft sinnvoll, die log-Likelihood-Funktion

$$\tilde{L}(\theta) := ln(L(\theta)) = -N \cdot \theta + \sum_{j=1}^{N} (x_j \cdot ln(\theta) - ln(x_j!))$$

$$\Longrightarrow \tilde{L}(\theta) = -N \cdot \theta + ln(\theta) \cdot \sum_{j=1}^{N} x_j - \sum_{j=1}^{N} ln(x_j!)$$

zu verwenden ($\tilde{L}(\theta)$ wird genau dort Maximal, wo auch $L(\theta)$ maximal wird).
$\Longrightarrow$ Maximiere $\tilde{L}(\theta)$.

$$\Longrightarrow 0 \stackrel{!}{=} \tilde{L}'(\theta) = -N + \frac{1}{\theta} \sum_{j=1}^{N} x_j \Longleftrightarrow \theta = \frac{1}{N} \sum_{j=1}^{N} x_j = \overline{x}_N$$

Also ist $\hat{\theta} = \overline{x}_N$ Kandidat für Mls.
Wegen $\tilde{L}(0) = -\infty$, $\tilde{L}(+\infty) = -\infty$, $\tilde{L}(\theta) > -\infty$ und der Eindeutigkeit des Extremums ist $\hat{\theta}$ Mls = MlS für θ :

$$\hat{\theta} = \overline{X}_N$$

Definition 10:
Besitzt die Verteilung P_θ eine Dichte $f_\theta(x)$ und werden die Daten $x_1, \ldots, x_N$ beobachtet, so heißt

$$(5) \quad l(\theta) = l(\theta, x_1, \ldots, x_N) = \prod_{j=1}^{N} f_\theta(x_j)$$

die zu $x_1, \ldots, x_N$ gehörende Likelihood-Funktion. Jeder Parameter $\overset{\wedge}{\theta} \in \Theta$, der $l(\theta)$ maximiert, heißt ein Maximum-Likelihood-Schätzwert für θ. Der zugehörige Schätzer $T(X_1, \ldots, X_N)$ (siehe Def.9) heißt ein Maximum-Likelihood-Schätzer für θ.

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

$$\Longrightarrow \hat{l}((\mu, \sigma^2), x_1, \ldots, x_N) = \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x_j-\mu)^2}{2\sigma^2}} = \left(\frac{1}{\sqrt{2\pi}}\right)^N \cdot \left(\frac{1}{\sigma^2}\right)^{\frac{N}{2}} \cdot e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{N}(x_j-\mu)^2}$$

(a) $\sigma^2 = \sigma_0^2$ bekannt, also $\theta = \mu$ („Schätze Erwartungswert bei bekanntem σ^2")
$\Longrightarrow l(\theta) = \hat{l}((\theta, \sigma_0^2), x_1, \ldots, x_N)$ wird maximal (als Funktion von θ), wenn $\sum_{j=1}^{N}(x_j - \theta)^2$ minimal wird. Da dies eine nach oben geöffnete Parabel ist, genügt es, die Nullstelle ihrer Ableitung zu bestimmen, also

$$0 \overset{!}{=} \left(\sum_{j=1}^{N}(x_j - \theta)^2\right)' = -2\sum_{j=1}^{N}(x_j - \theta) = 2N \cdot \theta - 2\sum_{j=1}^{N} x_j$$

$$\Longrightarrow \theta = \frac{1}{N}\sum_{j=1}^{N} x_j \text{ also } \overset{\wedge}{\theta} = \overline{x_N} \text{ ist Mls, } \overset{\wedge}{\theta}(X_1, \ldots, X_N) = \overline{X_N} \text{ ist MlS.}$$

(b) $\mu = \mu_0$ bekannt, also $\theta = \sigma^2$ („Schätze Varianz bei bekanntem μ")
$\Longrightarrow l(\theta) = \hat{l}((\mu_0, \theta), x_1, \ldots, x_N)$ ist als Funktion von θ kompliziert. $\Longrightarrow$ Wähle deshalb die log-Likelihood-Funktion.

$$\tilde{l}(\theta) = ln(l(\theta)) = -N \cdot ln(\sqrt{2\pi}) - \frac{N}{2} ln(\theta) - \frac{1}{2\theta}\sum_{j=1}^{N}(x_j - \mu_0)^2$$

Ableiten von $\tilde{l}(0) = -\infty = \tilde{l}(+\infty)$ (Übung) liefert den Mls als

$$0 \overset{!}{=} -\frac{N}{2}\frac{1}{\theta} + \frac{1}{2\theta^2}\sum_{j=1}^{N}(x_j - \mu_0)^2$$

$$\Longrightarrow \ldots \Longrightarrow \overset{\wedge}{\theta}(X_1, \ldots, X_N) = \frac{1}{N}\sum_{j=1}^{N}(x_j - \mu_0)^2 \text{ ist Ml-Schätzer für } \theta$$

(c) $\underline{\mu, \sigma \text{ unbekannt, also } \theta = (\mu, \sigma^2)}$
Man kann zeigen:

$$\overset{\wedge}{\theta}(X_1, \ldots, X_N) = \left(\overline{X}_N, \frac{1}{N}\sum_{j=1}^{N}(X_j - \overline{X}_N)^2\right) \text{ ist MlS für } \theta = (\mu, \sigma^2).$$

Bemerkung:
Es lassen sich die folgenden Eigenschaften zeigen:

1. Ist $g : \Theta \longrightarrow \Theta'$ eine bijektive Abbildung und $\overset{\wedge}{\theta}$ ein MlS für θ, so ist $g(\overset{\wedge}{\theta})$ ein MlS für $g(\theta)$ („Funktionale Invarianz").
2. Unter Regularitätsannahme kann man zeigen, dass Ml-Schätzer (stark) konsistent sind.

2.2.3 Intervallschätzer („Konfidenzintervall")

Ziel: Beschreibe zusätzlich zu einem (Punkt-)Schätzer noch seine Genauigkeit, d.h. gib ein <u>zufälliges Intervall</u> an, das den wahren Wert θ mit hoher Wahrscheinlichkeit überdeckt.

Modell: $X_1, \ldots, X_N$ u.i.v. mit Verteilungsfunktion $F_\theta, \theta \in \Theta \subseteq \mathbb{R}^d$

Definition 11:
Ein <u>Konfidenzintervall</u> (oder <u>Vertrauensbereich</u>) zum Sicherheitsniveau $1 - \alpha$ (typisch $0 < \alpha \approx 0$) für θ ist ein zufälliges Intervall $[T_1, T_2]$ mit Grenzen $T_i = g_i(X_1, \ldots, X_N), i = 1, 2$, so dass gilt

(1) $\quad P_\theta(\theta \in [T_1, T_2]) \geq 1 - \alpha \quad \forall \theta \in \Theta$

Problem: Wie konstruiert man ein Konfidenzintervall?

(i) <u>Konfidenzintervall für den Erwartungswert einer Normalverteilung</u>
Situation: $X_1, \ldots, X_N$ u.i.v gemäß $\mathcal{N}(\theta, \sigma^2)$ mit $\theta \in \mathbb{R}$, σ^2 sei bekannt.

Hilfssatz 1:
$X_1, \ldots, X_N$ u.i.v $\mathcal{N}(\theta, \sigma^2) \Longrightarrow \overline{X}_N \sim \mathcal{N}(\theta, \frac{\sigma^2}{N})$ (Beweis: ✓)
$\overset{\text{Hilfssatz 1}}{\Longrightarrow} Z := \frac{\overline{X}_N - \theta}{\sqrt{\frac{\sigma^2}{N}}} \overset{\sigma \geq 0}{=} \sqrt{N} \cdot \frac{\overline{X}_N - \theta}{\sigma} \sim \mathcal{N}(0, 1)$
$\Longrightarrow$ Ist $q = q_{1-\frac{\alpha}{2}}$ (und somit $-q = q_{\frac{\alpha}{2}}$) das $1 - \frac{\alpha}{2}$- Quantil von $\mathcal{N}(0, 1)$, so gilt:

$$\begin{aligned}
\underline{1 - \alpha} &= \left(1 - \frac{\alpha}{2}\right) - \frac{\alpha}{2} = P_\theta(Z \leq q) - P_\theta(Z \leq -q) \\
&= P_\theta(-q \leq Z \leq q) \overset{Z = \sqrt{N} \cdot \frac{\overline{X}_N - \theta}{\sigma}}{=} P_\theta\left(-\frac{\sigma}{\sqrt{N}} \cdot q \leq \overline{X}_N - \theta \leq \frac{\sigma}{\sqrt{N}} \cdot q\right) \\
&= \underline{P_\theta\left(\overline{X}_N - \frac{\sigma}{\sqrt{N}} q \leq \theta \leq \overline{X}_N + \frac{\sigma}{\sqrt{N}} \cdot q\right)}
\end{aligned}$$

also: $[T_1, T_2] = [\overline{X}_N - \frac{\sigma}{\sqrt{N}} \cdot q, \overline{X}_N + \frac{\sigma}{\sqrt{N}} \cdot q]$ mit $q = q_{1-\frac{\alpha}{2}}$ ist ein $(1-\alpha)$ Konfidenzintervall für θ bei bekanntem $\sigma > 0$.

Beispiel:
Die Länge eines Gegenstandes werde $N = 10$ mal gemessen
$\Longrightarrow \overline{X}_{10} = 23.7\mu m$, Messgerät hat laut Hersteller $\sigma = 0.6\mu m$
$\overset{\text{oben}}{\Longrightarrow}$ 95%-Konfidenzintervall für $\mu : \overline{X}_{10} \pm \frac{0.6}{\sqrt{10}} \cdot 1.96 = [23.33; 24.07]$

(ii) Konfidenzintervall für θ bei unbekanntem σ^2
Situation: wie in (i), jetzt aber σ^2 unbekannt.

Idee: Schätze σ^2 durch $s_N^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (X_i - \overline{X}_N)^2$ und verwende
Hilfssatz 2:
$X_1, \ldots, X_N$ u.i.v. $\mathcal{N}(\theta, \sigma^2)$, $s_N^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (X_i - \overline{X}_N)^2$
$\Longrightarrow Z = \sqrt{N} \cdot \frac{\overline{X}_N - \theta}{s_N} \sim t_{N-1}$ („t-verteilt mit N-1 Freiheitsgeraden")

Beachte: Die t-Verteilung (genauer: ihre Dichte) ist symmetrisch um Null und besitzt eine ähnliche Form wie $\mathcal{N}(0,1)$. Sie ist tabelliert (für verschiedene N, siehe Blatt) und für große N gilt $t_N \approx \mathcal{N}(0,1)$

Mit gleichem Vorgehen wie in (i) erhält man nun:
$[T_1, T_2] = [\overline{X}_N - \frac{s_N}{\sqrt{N}} \cdot t_{N-1,1-\frac{\alpha}{2}}, \overline{X}_N + \frac{s_N}{\sqrt{N}} \cdot t_{N-1,1-\frac{\alpha}{2}}]$ ist ein $(1-\alpha)$ Konfidenzintervall für θ (bei unbekanntem σ^2), wobei $t_{N,\alpha}$ das α-Quantil von t_N ist.

Beispiel: Selbe Daten wie in (i), nur ergab sich jetzt aus den Daten
$s_N = 0.6\mu m \Longrightarrow$ 95%-Konfidenzintervall für $\theta = \overline{X}_{10} \pm \frac{0.6}{\sqrt{10}} \underbrace{t_{9,0.975}}_{=2.262} = [23.28; 24.12]$

(iii) Konfidenzintervall für die Varianz einer Normalverteilung
Situation: $X_1, \ldots, X_N \sim \mathcal{N}(\mu, \theta)$, θ zu schätzen, μ unbekannt.

Hilfssatz 3:
$X_1, \ldots, X_N$ u.i.v. $\mathcal{N}(\mu, \theta) \Longrightarrow \frac{N-1}{\sigma^2} \cdot s_N^2 \sim \chi_{N-1}^2$ („Chi-Quadrat verteilt mit $N-1$ Freiheitsgeraden")

Bemerkung: Wieder sind die Werte von χ_N^2 tabelliert und man erhält aus Hilfssatz 3 analog zu (i):

$$[T_1, T_2] = \left[\frac{(N-1)s_N^2}{\chi^2_{N-1,1-\frac{\alpha}{2}}}, \frac{(N-1)s_N^2}{\chi^2_{N-1,\frac{\alpha}{2}}} \right]$$

ist ein $(1-\alpha)$ Konfidenzintervall für θ, wobei $\chi_{N,\alpha}^2$ das α-Quantil von χ_N^2 ist.

(iv) Approximative Konfidenzintervalle für den Erwartungswert einer beliebigen Verteilung
Situation: $X_1, \ldots, X_N$ u.i.v. mit $E(X_i) = \theta, Var(X_i) = \sigma^2$
$\overset{ZGS}{\Longrightarrow} \sqrt{N} \cdot \frac{\overline{X}_N - \theta}{\sigma} = \frac{\sum_{i=1}^{N}(X_i - E(X-i))}{\sqrt{Var(\sum_{i=1}^{N} X_i)}} \underset{\text{für } N\text{groß}}{\overset{\text{approximativ}}{\sim}} \mathcal{N}(0,1)$

Analog zu (i) und (ii) erhält man

(a) Approx. $(1-\alpha)$ Konfidenzintervall für θ bei bekanntem σ^2:
$[T_1, T_2] = \overline{X}_N \pm \frac{\sigma}{\sqrt{N}} \cdot q_{1-\frac{\alpha}{2}}$ (q_α α- Quantil von $\mathcal{N}(0,1)$)

(b) Approx. $(1-\alpha)$ Konfidenzintervall für θ bei unbekanntem σ^2:
$[T_1, T_2] = \overline{X}_N \pm \frac{s_N}{\sqrt{N}} \cdot t_{N-1,1-\frac{\alpha}{2}}$ ($t_{N,\alpha}$ α-Quantil von t_N)

2.2.4 Kovarianz- und Korrelationsschätzer

Modell: $(X_1, Y_1), \ldots, (X_N, Y_N)$ u.i.v. (als Zufallspaare) mit
$\mu_X = E(X_i),\ \sigma_X^2 = Var(X_i),\ \mu_Y = E(Y_i),\ \sigma_Y^2 = Var(Y_i)$
$c = Cov(X_i, Y_i),\ \rho = \rho_{X_i,Y_i} = Corr(X_i, Y_i)$

Schätzer für Kovarianz und Korrelation
$Cov(X_i, Y_i) = E((X_i - E(X_i))(Y_i - E(Y_i)))$
$\hookrightarrow$ plausibel: $\hat{c}_N = \frac{1}{N-1} \cdot \sum_{i=1}^N (X_i - \overline{X}_N)(Y_i - \overline{Y}_N)$ (durch $\frac{1}{N-1}$ wird $\hat{c}_N$ erwartungstreu)
$\rho_{X_i,Y_i} = Corr(X_i, Y_i) = \frac{Cov(X_i,Y_i)}{\sigma_X \cdot \sigma_Y}$
$\hookrightarrow$ plausibel: $\hat{\rho}_N = \frac{\hat{c}_N}{s_{N_X} \cdot s_{N_Y}}$ mit $s_{N_X}^2 = \frac{1}{N-1} \cdot \sum_{i=1}^N (X_i - \overline{X}_N)^2, s_{N_Y}$ analog.

2.3 Statistische Testverfahren

2.3.1 Testen von Hypothesen, Fehler, Gaußtest

Bei vielen Experimenten oder Aufgaben ist man weniger an genauen Schätzwerten für gewisse Messgrößen interessiert, sonder man will qualitative Aussagen der Form „Jungen sind bei Geburt schwerer als Mädchen“ oder „Betriebssystem X ist stabiler als System Y“ überprüfen. Man will also entscheiden, ob eine gewisse Hypothese gültig ist oder nicht.
Ein statistisches Verfahren, das mit Hilfe einer relevanten Stichprobe $X_1, \ldots, X_N$ die Entscheidung über Annahme oder Verwerfung (Ablehnung) der Hypothese fällt, heißt ein Test (der Hypothese).

Beispiel:
Um zu überprüfen, ob ein neuer Lack tatsächlich die vom Hersteller angegebene mittlere Trockenzeit von $\mu_0 = 10\text{min}$ hat, werden 50 Proben untersucht. Wir wollen hierbei die Nullhypothese

$$H_0 : \ \mu \overset{(\le)}{=} \mu_0 = 10\text{min}$$

gegen die Alternativhypothese

$$H_1 : \ \mu > \mu_0 = 10\text{min}$$

testen, wobei wir für die Trockenzeit X annehmen, daß $X \sim \mathcal{N}(\mu, \sigma^2)$ gilt und die vom Hersteller angegebene Standardabweichung von $\sigma = 1.54\text{min}$ als bekannt voraussetzen.

Unsere Messungen ergeben: $\overline{X}_{50} = 10.89\text{min}$
$\Longrightarrow$ Ist dies eine systematische Abweichung von $\mu_0 = 10\text{min}$ oder ist sie im Rahmen von Zufallsschwankungen tolerierbar?

Hierzu: Nehme an, das H_0 richtig ist und berechne die Wahrscheinlickeit für die Beobachtung eines mindestens so großen Wertes („Werte, die der Nullhypothese noch stärker wiedersprechen“) wie den obigen:

$$\Longrightarrow P_0(\overline{X}_{50} \ge 10.89) \overset{\overline{X}_N \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{N})}{=} 1 - \Phi\left(\frac{10.89 - 10}{1.54/\sqrt{50}}\right)$$
$$\overset{1-\Phi(x)=\Phi(-x)}{=} \Phi(-4.087) \le 0.00003$$

Konsequenz:
Entweder haben wir also einen sehr unwahrscheinlich großen Wert für $\overline{X}_{50}$ beobachtet oder die Nullhypothese ist falsch! Das Kriterium, für welche der beiden Konsequenzen wir uns entscheiden sollen, heißt ein statistischer Test. Er ist durch seinen Annahmebereich (d.h. Werte von $\overline{X}_N$, für die H_0 nicht verworfen wird) und seinen Ablehnungs- oder Verwerfungsbereich (d.h. Werte von $\overline{X}_N$, für die H_0 zugunsten von H_1 verworfen wird) bestimmt.

Fehler:
Da die Entscheidung eines Tests immer nur Informationen über einen Teil aller möglichen Ereignisse verwenden kann, kann diese Entscheidung auch falsch sein. Man hat also nach Durchführung eine Tests entweder

- die richtige Entscheidung getroffen

oder

- H_0 abgelehnt, obwohl H_0 richtig ist (Fehler 1. Art)
- H_0 beibehalten, obwohl H_0 falsch ist (Fehler 2. Art)

Ideal:
Halte Wahrscheinlichkeiten für Fehler 1. Art und Fehler 2. Art unter einer vorgegebenen Schranke.
mathemat. Gründe: Dies ist nicht gleichzeitig möglich!

Traditioneller Ansatz:
Halte den Fehler 1. Art unter einer gegebenen Wahrscheinlichkeit α („Test zum (Signifikanz-)Niveau α und versuche, den Fehler 2. Art möglichst klein zu halten ($\longrightarrow$ theoretisches Problem).
Typische Werte für α: 0.05, 0.01, 0.001

Im Beispiel:
Wähle $\alpha = 0.05$ (d.h. lehne H_0 ab, wenn die Wahrscheinlichkeit für einen Fehler 1. Art (d.h. die Wahrscheinlichkeit unter H_0 noch größere Werte als den beobachteten zu erhalten) höchstens 5% beträgt!)
$\Longrightarrow$ Bestimme 95%-Quantil q für $\overline{X}_N$:

$$\sqrt{N} \cdot \frac{\overline{X}_N - \mu_0}{\sigma} \sim \mathcal{N}(0,1) \text{ unter } H_0 \Longrightarrow q = q_{0.95} = 1.65$$

Wir lehnen H_0 daher ab, wenn $\sqrt{N} \cdot \frac{\overline{X}_N - \mu_0}{\sigma} > 1.65$ gilt, also im Beispiel

$$\sqrt{50} \cdot \frac{10.89 - 10}{1.54} = 4.087 > 1.65$$

$\Longrightarrow$ H_0 kann auf dem 5%-Niveau abgelehnt werden. Wir können deshalb annehmen, daß H_1 richtig ist, d.h. die tatsächliche mittlere Trockenzeit μ ist größer als $\mu_0 = 10$.

Bemerkung:
Da oben H_1 nur in eine Richtung ging („$\mu > \mu_0$“), spricht man auch von einem einseitigen Test . Will man lediglich testen, ob $H_1 : \mu \neq \mu_0$ gilt, so muß die Fehlerwahrscheinlichkeit auf „beide Seiten“ verteilt werden, d.h. man lehnt ab, wenn man entweder einen sehr großen oder einen sehr kleinen Wert von $\overline{X}_N$ beobachtet. Die obige Rolle von $q_{0.95}$ wird dann von $q_{0.025} = -q_{0.975}$ und $q_{0.975}$ übernommen.

Allgemein heißt obiger Test <u>Gauß-Test</u> und sieht wie folgt aus:

<u>Gauß-Test ($X_1, \ldots, X_N$ u.i.v. $\mathcal{N}(\mu, \sigma^2)$, σ^2 bekannt)</u>

1. Wähle die zu testende Hypothese
   (a) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu > \mu_0$
   (b) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu < \mu_0$
   (c) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$
2. Lege das Signifikanzniveau α fest
3. Berechne: $Z = \sqrt{N} \cdot \frac{\overline{X}_N - \mu_0}{\sigma}$
4. Lehne H_0 ab, falls $Z \begin{cases} > q_{1-\alpha} & \text{im Fall i)} \\ < q_\alpha & \text{im Fall ii)} \\ < q_{\frac{\alpha}{2}} \ \underline{\text{oder}} \ > q_{1-\frac{\alpha}{2}} & \text{im Fall iii)} \end{cases}$
   Ansonsten wird H_0 beibehalten

Wichtig!
Nur bei Ablehnung von H_0, also Annahme von H_1, kann man etwas über die Fehlerwahrscheinlichkeiten sagen (hat dann also H_1 „bewiesen“ (bei gegebenem Restrisiko α)). Wird H_0 nicht abgelehnt, so heißt dies nicht, daß H_0 „bewiesesen“ wurde.
Will man folglich etwas „beweisen“, so formuliert man dies als Alternativhypothese H_1!

Hilfsmittel:

$$X_1, \ldots, X_N \text{ u.i.v. } \mathcal{N}(\mu_0, \sigma^2),\ S_N = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X}_N)^2} \Longrightarrow t := \sqrt{N} \cdot \frac{\overline{X}_N - \mu_0}{S_N} \sim t_{N-1}$$

2.3.2 Der 1-Stichproben-t-Test

Situation:
$X_1, \ldots, X_N$ u.i.v. $\mathcal{N}(\mu, \sigma^2)$, wobei μ, σ^2 unbekannt sind. Wir wollen eine der drei Hypothesen testen

i) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu > \mu_0$

ii) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu < \mu_0$

iii) $H_0 : \mu = \mu_0$ gegen $H_1 : \mu \neq \mu_0$

Wir wissen bereits (Hilfsmittel):

$$t = \sqrt{N} \cdot \frac{\overline{X}_N - \mu_0}{s_N} \sim t_{N-1} \text{ (unter } H_0)$$

und man erhält in Analogie zum Gauß-Test den

1-Stichproben-t-Test: ($X_1, \dots, X_N$ u.i.v. $\mathcal{N}(\mu, \sigma^2)$, μ, σ^2 unbekannt)

1. Wähle die zu testende Hypothese

   i) $H_0 : \mu = \mu_0 \longleftrightarrow H_1 : \mu > \mu_0$

   ii) $H_0 : \mu = \mu_0 \longleftrightarrow H_1 : \mu < \mu_0$

   iii) $H_0 : \mu = \mu_0 \longleftrightarrow H_1 : \mu \neq \mu_0$

2. Lege das Signifikanzniveau α fest

3. Berechne $t = \sqrt{N} \cdot \frac{\overline{X}_N - \mu_0}{s_N}$

4. Lehne H_0 ab, falls $t \begin{cases} > t_{N-1,1-\alpha} & \text{im Fall i)} \\ < t_{N-1,\alpha} & \text{im Fall ii)} \\ > t_{N-1,1-\frac{\alpha}{2}} \ \underline{\text{oder}} \ < t_{N-1,\frac{\alpha}{2}} & \text{im Fall iii)} \end{cases}$
   Ansonsten wird H_0 beibehalten

Bemerkung:

1. Die Wahrscheinlichkeit für den bei diesem Vorgehen gemachten Fehler 2. Art („Behalte H_0 bei, obwohl H_0 falsch ist") hängt vom wahren aber unbekannten Wert μ ab. Ist z.B. im Fall i) $\mu > \mu_0$, aber doch fast gleich μ_0, so kann die Wahrscheinlichkeit für den Fehler 2. Art fast gleich $1 - \alpha$ werden (also maximal!)
   Ist hingegen $\mu \gg \mu_0$, so wird die Wahrscheinlichkeit sehr klein und verschwindet asymptotisch (d.h. für die Nummer Anzahl der Beobachtungen für $N \to \infty$).

2. Man verwendet den 1-Stichproben-t-Test oft, um die Mittelwerte zweier verbundener (d.h. abhängiger) Stichproben zu vergleichen. Er heißt dann auch Differenztest: $X_1, \dots, X_N$ u.i.v., $Y_1, \dots, Y_N$ u.i.v mit $(X_1, Y_1), \dots, (X_N, Y_N)$ u.i.v. und $D_i := X_i - Y_i$ u.i.v. $\mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$.

   Mit den D_i führt man dann den 1-Stichproben-t-Test mit $\mu_0 = 0$,
   $s_N = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^N (D_i - \overline{D}_N)^2}$ durch.

Beispiel: Senkt ein Medikament den Blutdruck?
$N = 20$ Personen, X_i =Blutdruck der i. Person vor Medikamenteneinnahme,
Y_i =Blutdruck der i. Person nach Medikamenteneinnahme.
Teste $H_0 : \mu = 0$ „kein Einfluss" gegen $H_1 : \mu > 0$
Wähle $\alpha = 0.01$
Es wurde gemessen: $\overline{D}_{20} = 10.8$, $s_{20} = 20.5$

$$\Longrightarrow t = \sqrt{20} \cdot \frac{10.8 - 0}{20.5} = 2.356 < t_{19.099} = 2.539$$
$$t_{19.095} = 1.729$$

$\Longrightarrow$ H_0 kann auf dem 1%-Niveau nicht abgelehnt werden!

2.3.3 Vergleich der Mittelwerte zweier normalverteilter Stichproben: Der 2-Stichproben-t-Test

Situation:
$X_1, \dots, X_N$ u.i.v. $\mathcal{N}(\mu_1, \sigma^2)$, $Y_1, \dots, Y_N$ u.i.v. $\mathcal{N}(\mu_2, \sigma^2)$, σ^2 unbekannt.
Wir wollen eine der folgenden Nullhypothesen testen:

i) $H_0 : \mu_1 \leq \mu_2 \longleftrightarrow H_1 : \mu_1 > \mu_2$

ii) $H_0 : \mu_1 \geq \mu_2 \longleftrightarrow H_1 : \mu_1 < \mu_2$

iii) $H_0 : \mu_1 = \mu_2 \longleftrightarrow H_1 : \mu_1 \neq \mu_2$

Beispiel: Elektrischer Widerstand von Kupferdraht
Es wird vermutet, daß sich der el. Widerstand von Kupfer senkt, wenn es mit einem geringen Anteil anderer Metalle legiert wird.
$\Longrightarrow$ Experiment: 60 Messungen $X_1, \dots, X_{60}$ bei Standardkupferdraht, 40 Messungen $Y_1, \dots, Y_{40}$ bei legiertem Kupferdraht

Hypothese:
$H_0 : \mu_1 \overset{(\leq)}{=} \mu_2$ „ mittlerer Widerstand der Legierung ist (mind.) so hoch wie bei Standardkupferdraht“
$H_1 : \mu_1 > \mu_2$ „Legierung hat geringeren Widerstand“

$$\longrightarrow \underline{\text{Daten}}: \quad \begin{array}{ll} \overline{X}_{60} = 0.125 \text{ Ohm} & \qquad \overline{Y}_{40} = 0.114 \text{ Ohm} \\ s_{60,X} = 0.005 \text{ Ohm} & \qquad s_{40,Y} = 0.006 \text{ Ohm} \end{array}$$

$\overset{\text{Hilfsmittel}}{\longrightarrow} \overline{X}_N - \overline{Y}_M \sim \mathcal{N}(\mu_1 - \mu_2, \sigma^2(\frac{1}{N} + \frac{1}{M}))$, also

$$\frac{(\overline{X}_N - \overline{Y}_M) - (\mu_1 - \mu_2)}{\sigma \cdot \sqrt{\frac{1}{N} + \frac{1}{M}}} \sim \mathcal{N}(0, 1),$$

aber σ ist unbekannt.

<u>Also:</u>
Schätze σ^2 durch das gewichtete Mittel $s^2_{N,M}$ auch bei empirischen Varianzen $s^2_{N,X}$, $s^2_{M,Y}$:

$$s^2_{N,M} = \frac{s^2_{N,X} \cdot (N-1) + s^2_{M,Y} \cdot (M-1)}{N + M - 2}$$

Dann lässt sich zeigen: Unter Annahme von $H_0 : \mu_1 = \mu_2$ gilt

$$t(X_1, \dots, X_N; Y_1, \dots, Y_M) = \frac{\overline{X}_N - \overline{Y}_M}{s_{N,M} \cdot \sqrt{\frac{1}{N} + \frac{1}{M}}} \sim t_{N+M-2}$$

Man erhält also den folgenden Test:

<u>2-Stichproben-t-Test ($X_1, \ldots, X_N$ u.i.v. $\mathcal{N}(\mu_1, \sigma^2)$; $Y_1, \ldots, Y_M$ u.i.v $\mathcal{N}(\mu_2, \sigma^2)$), σ^2 unbekannt</u>

1. Wähle die zu testende Hypothese

   i) $H_0 : \mu_1 = \mu_2 \longleftrightarrow H_1 : \mu_1 > \mu_2$
   ii) $H_0 : \mu_1 = \mu_2 \longleftrightarrow H_1 : \mu_1 < \mu_2$
   iiii) $H_0 : \mu_1 = \mu_2 \longleftrightarrow H_1 : \mu_1 \neq \mu_2$

2. Lege das Signifikanzniveau α fest

3. Berechne $t = \frac{\overline{X}_N - \overline{Y}_M}{s_{N,M}\sqrt{\frac{1}{N}+\frac{1}{M}}}$ mit $s_{N,M} = \sqrt{\frac{s^2_{N,x}(N-1)+s^2_{M,y}(M-1)}{N+M-2}}$

4. Lehne H_0 ab, falls $t \begin{cases} > t_{N+M-2,1-\alpha} & \text{im Fall i)} \\ < t_{N+M-2,\alpha} & \text{im Fall ii)} \\ > t_{N+M-2,1-\frac{\alpha}{2}} \ \underline{\text{oder}} \ < t_{N+M-2,\frac{\alpha}{2}} & \text{im Fall iii)} \end{cases}$

   Ansonsten wird H_0 beibehalten

Im Beispiel:
$\alpha = 0.05$

$$s^2_{60,40} = \frac{59 s^2_{60,X} + 39 s^2_{40,Y}}{60+40-2} = 29.38 \cdot 10^{-6}$$

$$\Longrightarrow t = \frac{0.125 - 0.114}{\sqrt{\frac{1}{60}+\frac{1}{40}} \cdot s_{60,40}} = 8.18 > 1.645 = t_{98,0.95}$$

$\Longrightarrow$ Lehne H_0 ab, d.h. der legierte Kupferdraht besitzt einen geringeren Widerstand.

Bemerkung:
Gleichheit der Varianzen innerhalb der beiden Stichproben kann mit einem sogenannten F-Test überprüft werden.

2.3.4 Kontingenztafeln und der Chi-Quadrat-Unabhängigkeitstest

Häufig will man wissen, ob zwei Eigenschaften/Merkmale eines Objekts sich beeinflussen oder unabhängig sind (z.B. Größe/Gewicht, Grundstücklage/Grundstückpreis, ...). Sind die Merkmale qualitativ oder quantitativ mit wenigen Ausprägungen, so kann man den folgenden χ^2-Unabhängigkeitstest durchführen. Hierfür ordnet man den Einzelausprägungen der Merkmale jeweils Zahlen von 1,..., n bzw 1, ..., m zu.

Beispiel:
X= Hosenfarbe ($\rightarrow$ 1=„blau“, 2=„schwarz“, 3=„grün“, 4=„sonstige“)
Y= Pulloverfarbe ($\rightarrow$ 1=„grau“, 2=„weiß“, 3=„sonstige“))
und kann sich dann hierauf (d.h. X, Y nehmen jeweils nur endlich viele natürliche Zahlen an) beschränken.

Modell: $(X_1,Y_1),\ldots,(X_N,Y_N)$ u.i.v. mit Werten in $\{(k,l)|k = 1,\ldots,m;\ l = 1,\ldots,n\}$ und $p_{k,l} = P(X_j, = k,\ Y_j = l),\ k = 1,\ldots,m;\ l = 1,\ldots,n;\ j = 1,\ldots,N$

Beispiel: Zusammenhang Berufsleistung „X" (1=gut, 2=mittel, 3=schlecht) $\longleftrightarrow$ Diplomnote „Y" (1, 2, 3)
Es sei nun $Z_{k,l}$=Anzahl der Daten (X_j,Y_j)mit $X_j = k, Y_j = l$
Wir stellen dies in einer Kontingenztafel dar.

		Y 1	…	n	Summe
	1	Z_{11}	…	Z_{1n}	$Z_{1\cdot}$
X	⋮	⋮		⋮	⋮
	m	Z_{m1}	…	Z_{mn}	$Z_{m\cdot}$
	Summe	$Z_{\cdot 1}$	…	$Z_{\cdot n}$	N

$Z_{k\cdot} = Z_{k1} + \ldots + Z_{kn}$
$Z_{\cdot l} = Z_{1l} + \ldots + Z_{ml}$

Wie entscheidet man über Unabhängigkeit?
Es gelten:

$$P(X_j = k) = p_{k1} + p_{k2} + \ldots + p_{km} =: p_{k\cdot}$$

$$P(Y_j = l) = p_{1l} + p_{2l} + \ldots + p_{nl} =: p_{\cdot l}$$

Unter $H_0 : X_j, Y_j$ sind unabhängig gilt dann:

$$(\star)\quad p_{kl} = P(X_j = k, Y_j = l) = P(X_j = k)\cdot P(Y_j = l) = p_{k\cdot}\cdot p_{\cdot l}$$

Da alle Größen $p_{kl}, p_{k\cdot}, p_{\cdot l}$ in $(\star)$ unbekannt sind, schätzt man sie aus den Daten Z_{kl} und betrachtet die dann entstehenden Abweichungen von $(\star)$ genauer.
Schätze

$$p_{k\cdot} \text{ durch } \hat{p}_{k\cdot} = \frac{1}{N} Z_{k\cdot} \quad k = 1,\ldots,m$$
$$p_{\cdot l} \text{ durch } \hat{p}_{\cdot l} = \frac{1}{N} Z_{\cdot l} \quad l = 1,\ldots,n$$

und dann wegen $(\star)$

$$p_{kl} \text{ durch } \hat{p}_{kl} = \hat{p}_{k\cdot}\cdot\hat{p}_{\cdot l}$$

Weichen nun die Werte $\frac{1}{N}Z_{kl}$ und $\hat{p}_{kl}$ nur wenig von einander ab, so spricht dies für die Unabhängigkeit von X_j und Y_j. Große Abweichungen sprechen hingegen für Abhängigkeit von X_j und Y_j.
$\Longrightarrow$ summiere Abweichung

$$D = \sum_{k=1}^{m}\sum_{l=1}^{n} \frac{(Z_{kl} - N\cdot\hat{p}_{kl})^2}{N\cdot\hat{p}_{kl}} = \sum_{k=1}^{m}\sum_{l=1}^{n} \frac{(Z_{kl} - \frac{1}{N}Z_{k\cdot}\cdot Z_{\cdot l})^2}{\frac{1}{N}Z_{k\cdot}\cdot Z_{\cdot l}} \qquad „\chi^2 - \text{Statistik}"$$

Hilfssatz (ohne Beweis)
Für $N \cdot \hat{p}_{kl} \geq 5$ für alle (k,l) gilt: $D \overset{\approx}{\sim} \chi^2_{(m-1)(n-1)}$

$\Longrightarrow$ χ^2-Unabhängigkeitstest

1. Teste $H_0 : X_j, Y_j$ sind unabhängig $(j = 1, ..., N)$
   gegen $H_1 : X_j, Y_j$ sind abhängig.
2. Lege das Signifikanzniveau α fest.
3. Berechne $D = \sum_{k=1}^{m} \sum_{l=1}^{n} \frac{(Z_{kl} - \frac{1}{N} Z_{k\cdot} \cdot Z_{\cdot l})^2}{\frac{1}{N} Z_{k\cdot} \cdot Z_{\cdot l}}$
4. Lehne H_0 ab, falls $D > \chi^2_{(m-1)(n-1),1-\alpha}$

Beispiel (Fortsetzung): $N = 400, \alpha = 0,01$

		Diplomnote			
		1	2	3	
	1	63	49	9	121
Leistung	2	60	79	28	167
	3	29	60	23	112
		152	188	60	400

$$\Longrightarrow D = \frac{(63 - \frac{121 \cdot 152}{400})^2}{\frac{121 \cdot 152}{400}} + \ldots + \frac{(23 - \frac{112 \cdot 60}{400})^2}{\frac{112 \cdot 60}{400}} = 20.34 > 13.28 = \underline{\chi^2_{4;0.99}}$$

$\Longrightarrow$ H_0 kann abgelehnt werden, d.h. die Leistung im Beruf und die Diplomnote hängen zusammen.

2.4 Lineare Regression

Ziel: - Beschreibung der Abhängigkeit zweier MerkmaleX, Y
- Vorhersage von Y aus X (z.B. Gewicht und Größe)

Situation: Xunabhängige (von uns kontrollierbare) Variable
Y (von X) abhängige Variable
und (als Paare) unabhängige Datenpaare$(X_1, Y_1), \ldots, (X_N, Y_N)$
$\Longrightarrow$ Regressionsmodell $Y_j = g(X_j) + e_j$
mit: g - Regressionsfunktion
e_j Messfehler u.i.v mit $E(e_j) = 0$ und $Var(e_j) = \sigma_e^2$

Ziel: Schätze die Regressionsfunktion $g(.)$ so gut wie möglich!
Idee: Lege eine Klasse $g^{(\theta)}(.)$ von Fuktionen fest mit $\theta \in \Theta$ und wähle die Funktion $f^{(\theta)}(.)$ aus, die die Summe der quadratischen Abweichungen

$$\sum_{i=1}^{N}(Y_i - g^{(\theta)}(X_i))^2 \quad \text{minimiert}$$

„Methode der kleinsten Quadrate“

Modell:

$Y_i = b_1 + b_2 \cdot X_i + e_i$
„lineare Regression“ $(\theta = (b_1, b_2), \Theta = \mathbb{R}^2)$

$\Longrightarrow$ Aufgabe: Bestimme $(\hat{b}_1, \hat{b}_2)$ als Lösung von

$$\min_{b_1, b_2} \sum_{i=1}^{N}(Y_i - b_1 - b_2 \cdot X_i)^2 \quad (\text{“kleinste Quadrate-Schätzer für } (b_1, b_2)\text{“})$$

$$\text{Lösung :} \quad \hat{b}_2 = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y}_N)(X_i - \overline{X}_N)}{\sum_{i=1}^{N}(X_i - \overline{X}_N)^2}, \; \hat{b}_1 = \overline{Y}_N - \hat{b}_2 \cdot \overline{X}_N$$

(aus Bestimmung der Nullstellen der jeweiligen Ableitungen)

Beispiel: Werbeausgaben/Umsatzänderung einer FastFood-Kette

Anstieg der Werbeausgaben in %	0	1	3	5	6	6	7	9	12
Veränderung des Umsatzes in %	2,1	3,0	6,1	6,9	3,5	8,4	7,9	7,4	8,0

$$\Longrightarrow \overline{X}_N = 5,44 \qquad \hat{c}_N = 6,86$$

$$\overline{Y}_N = 5,92 \qquad \hat{b}_2 = 0,48$$

$$s_{N,X}^2 = 14,28 \qquad \hat{b}_1 = 3,3$$

Anhang A

Verschiedenes

A.1 Übungen zur Klausur 1

(1) **Aufgabe:**

(a) Beim Skat erhält jeder der drei Spieler zehn Karten aus einem Pack mit 32 Karten. Es gibt vier Asse. Wie groß ist die Wahrscheinlichkeit, daß irgendein Spieler 3 Asse erhält? Was ist an der letzetn Fragestellung für den Fall von 2 Assen komplizierter? Wie hoch ist die Wahrscheinlichkeit, daß kein Spieler 4 Asse erhält?

(b) Eine Computerlieferung von 30 Computern enthält 2 beschädigte Exemplare. Der Abnehmer der Lieferung untersucht in einer Stichprobe zunächst 10 Stück. Ist keiner der untersuchten PC's beschädigt, so akzeptiert er die Lieferung. Berechnen Sie die Wahrscheinlichkeit, daß er die Lieferung ablehnen wird.

Lösung:

(a) Für Spiele 1: Ziehen ohne Zurücklegen aus einem Topf mit 32 Karten. Er erhält 10 Karten, davon $k\,(\leq 4)$ aus dem "Untertopf" der Asse und $10-k$ aus dem "Resttopf"

$$\begin{aligned}\Longrightarrow P(\text{“Spieler 1 erhält 3 Asse”}) &= \frac{\sharp \text{ günstige Fälle}}{\sharp \text{ mögliche Fälle}}\\ &= \frac{\binom{4}{3}\cdot\binom{28}{7}}{\binom{32}{10}} \text{ (reicht als Ergebnis)}\end{aligned}$$

Lösung bezieht sich nur auf "<u>Spieler 1</u> erhält 3 Asse"

(b) analog

(2) **Aufgabe:**
In einer Stadt werden Tuberkulose-Untersuchungen mittels Röntgenaufnahmen durchgeführt. Hierbei werden 90% der Kranken als infiziert erkannt. 1% der Gesunden wird fälschlicherweise als "Tbc-verdächtig" registriert. Wenn man weiter weiß, daß 0.1% der gesamten Bevölkerung an Tbc erkrankt ist, wie hoch ist dann die Wahrscheinlichkeit dafür, daß ein als "Tbc-verdächtig"

registrierter Einwohner wirklich krank ist?

Lösung:
TE = Tuberkulose erkannt
$\overline{TE}$ = Tuberkulose-frei erkannt
TK = Tuberkulosekrank
$\overline{TK}$ = gesund
$\Omega = TK \,\dot{\cup}\, \overline{TK}$

<u>bekannt:</u>
$P(TE|TK) = 0.9, \qquad P(\overline{TE}|TK) = 0.1$
$P(TE|\overline{TK}) = 0.01, \qquad P(\overline{TE}|\overline{TK}) = 0.99$
$P(TK) = 0.001, \qquad P(\overline{TK}) = 0.999$
$\Longrightarrow P(TK|TE) \stackrel{\text{Bayes}}{\Longrightarrow} \frac{P(TE|TK)\cdot P(TK)}{P(TE)}$

<u>1. Schritt:</u>
$P(TE) \stackrel{\text{tot. W-keit}}{\Longrightarrow} P(TE|TK)\cdot P(TK) + P(TE|\overline{TK})\cdot P(\overline{TK})$
$= 0.9\cdot 0.001 + 0.01\cdot 0.999 = 0.01089$
$\Longrightarrow P(TK|TE) = \frac{0.9\cdot 0.001}{0.01089} = \underline{0.0826446}$

(3) **Aufgabe:**
Es seien X, Y die erste bzw. zweite Augenzahl beim zweimaligen Würfeln. Wir nehmen an, daß alle möglichen Paare (x, y) dieselbe Wahrscheinlichkeit haben. Untersuchen Sie die folgenden Paare von Zufallsvariablen auf Unabhängigkeit:

(a) X, Y
(b) X und $Z = X + Y$
(c) $Z = X + Y$ und $V = X - Y$

Lösung:

(a) $P(X = i,\, Y = j) = \frac{1}{36} \qquad i, j \in \{1, \ldots, 6\}$
$P(X = i)\cdot P(Y = j) = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36}$
$\Longrightarrow X, Y$ unabhängig

(b) $P(X = 1,\, Z = 6) = P(X = 1,\, Y = 5) \stackrel{(a)}{=} P(X = 1)\cdot P(Y = 6) = \frac{1}{36}$
$P(X = 1)\cdot P(Z = 6) = \frac{1}{6}\cdot\frac{5}{36} \neq \frac{1}{36}$
$\Longrightarrow X, Z$ sind nicht unabhängig.

(4) **Aufgabe:**
Zeigen Sie, daß die Funktion

$$f(x) = \begin{cases} 1 - x, & x \in [0, 1] \\ 1 + x, & x \in [-1, 0] \\ 0, & \text{sonst} \end{cases}$$

eine Wahrscheinlichkeitsdichte ist. Die zugehörige Verteilung heißt Dreiecksverteilung.

Lösung:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \int_{-1}^{0}(1+x)dx + \int_{0}^{1}(1-x)dx \\ &= \left(x+\frac{1}{2}x^2\right)\Big|_{-1}^{0} + \left(-\frac{1}{2}x^2\right)\Big|_{0}^{1} \\ &= \frac{1}{2}+\frac{1}{2} = 1\end{aligned}$$

$$E(X) = \int_{-\infty}^{infty} x\, f(x) = \int_{-1}^{0} x\,(1+x)dx + \int_{0}^{1} x\,(1-x)dx = \ldots = 0$$

(5) **Aufgabe:**
Der Zufallsvektor (X,Y) habe die Dichte

$$f(x,y) = \begin{cases} \frac{4}{5}(x^2+xy-y^2+1) & \text{für } (x,y)\in[0,1]\times[0,1] \\ 0 & \text{sonst}\end{cases}$$

(a) Berechnen Sie die Verteilungsfunktion $F(x,y)$ sowie $E(X)$, $E(Y)$ und $Cov(X,Y)$.

(b) Sind X und Y unabhängig?

Lösung aus der Vorlesung:
Bestimme die Dichte von X:

$$\begin{aligned}f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 \frac{4}{5}(x^2+xy-y^2+1)dy \\ &= \frac{4}{5}\left(x^2y+\frac{1}{2}xy^2-\frac{1}{3}y^3+y^3+y\right)\Big|_0^1 \\ &= \frac{4}{5}\left(x^2+\frac{1}{2}x-\frac{1}{3}+1\right)\end{aligned}$$

$f(x,y) = f_X(x)\cdot f_Y(y) \iff X,Y$ unabhängig (in diesem Fall nicht!)

Lösung aus der Übung:

(a) (i) $(x,y)\in[0,1]\times[0,1]$

$$\begin{aligned}F(x,y) &= \int_0^x \int_0^y \frac{4}{5}(u^2+ut-t^2+1)\,dt\,du \\ (\star)\quad &= \frac{4}{5}\left(\frac{1}{3}x^3y+\frac{1}{4}x^2y^2+\frac{1}{3}xy^3+xy\right)\end{aligned}$$

(ii) $x<0$ oder $y<0$
$\Longrightarrow F(x,y)=0$

(iii) $x\geq 1$ und $y\geq 1$

$$\begin{aligned}1 \geq F(x,y) = P(X\leq x, Y\leq y) \geq P(X\leq 1, Y\leq 1) &= F(1,1) \\ &\overset{(\star)}{=} 1\end{aligned}$$

$\Longrightarrow F(x,y)=1$

(iv) $x \in [0,1], y > 1$

$$\begin{aligned} F(x,y) &= P(X \leq x, Y \leq y) \\ &= P(X \leq x, Y \leq 1) \\ &= F(x,y) \overset{(\star)}{=} \frac{4}{5}\left(\frac{x^3}{3} + \frac{x^2}{4} + \frac{2x}{3}\right) \end{aligned}$$

(v) $x > 1,\ y \in [0,1]$
$F(x,y) = F(1,y) \overset{(\star)}{=} \frac{4}{5}\left(\frac{4}{3}y + \frac{1}{4}y^2 - \frac{1}{3}y^3\right)$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} f_X(x)x\,dx = \int_0^1 f_X(x)x\,dx = \frac{6}{10} \\ f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\ &= \int_0^1 \frac{4}{5}(x^2 + xy - y^2 + 1)dy = \frac{4}{5}\left(x^2 + \frac{1}{2}x + \frac{2}{3}\right) \\ E(Y) &= \int_0^1 \frac{4}{5}\left(\frac{4}{3}y + \frac{1}{2}y^2 - y^3\right)dy = \frac{7}{15} \\ E(XY) &= \int_0^1 \int_0^1 xy \cdot \frac{4}{5}(x^2 + xy - y^2 + 1)dx\,dy \\ &= \frac{4}{5}\int_0^1 \int_0^1 (x^3y + x^2y^2 - xy^3 + xy)dx\,dy \\ &= \frac{4}{5}\int_0^1 \left[\frac{1}{4}y + \frac{1}{3}y^2 - \frac{1}{2}y^3 + \frac{1}{2}y\right]dy \\ &= \frac{4}{5}\left(\frac{1}{8} + \frac{1}{9} - \frac{1}{8} + \frac{1}{4}\right) = \frac{13}{45} \\ Cov(x,y) &= E(XY) - E(X)\cdot E(Y) \\ &= \frac{13}{45} - \frac{6}{10}\cdot\frac{7}{15} = \frac{2}{225} \neq 0 \end{aligned}$$

(b) $\xrightarrow{(a)} Cov(X,Y) \neq 0$
$\Longrightarrow X,\ Y$ sind <u>nicht</u> unabhängig

A.2 Griechische Buchstaben

A	α	Alpha
B	β	Beta
Γ	γ	Gamma
Δ	δ	Delta
E	ϵ, ε	Epsilon
Z	ζ	Zeta
H	η	Eta
Θ	θ, ϑ	Theta
I	ι	Jota
K	κ	Kappa
Λ	λ	Lambda
M	μ	My
N	ν	Ny
Ξ	ξ	Xi
O	o	Omikron
Π	π, ϖ	Pi
P	ρ, ϱ	Rho
Σ	σ, ς	Sigma
T	τ	Tau
Υ	υ	Ypsilon
Φ	φ	Phi
X	χ	Chi
Ψ	ψ	Psi
Ω	ω	Omega

Index