VORLESUNGSSTRUKTUR: ST. U. WT

I EINLEITUNG
II STOCHASTISCHE GRUNDBEGRIFFE
III EINDIMENSIONALE VERTEILUNGEN
IV MEHRDIMENSIONALE VERTEILUNGEN
V FOLGEN STOCHASTISCHER GRÖSZEN
VI KLASSISCHE SCHLIESZENDE STATISTIK
VII BAYES'SCHE STATISTIK
VIII ERGÄNZUNGEN

I

EINLEITUNG

Statistik und Wahrscheinlichkeitstheorie

Offizielle Statistik
Angewandte Statistik
Theoretische Statistik } Stochastik
Wahrscheinlichkeitstheorie } Stochastik

Wozu Statistik u. WR ?
„Quantitative Erfassung von Massenphänomenen und Beschreibung von nichtdeterministischen Vorgängen"

1. HISTORISCHES UND GRUNDSÄTZLICHES

Stat. Erhebungen: seit ca. 4500 Jahren
Ägypten, China (Militär, Steuer)
seit 550 v. Chr. regelmäßiger „Zensus" im Röm. Reich
1754 erste Volkszählung in N.Ö. (inkl. Wien)

Universitätsstatistik: seit dem 17. Jhdt. Lehre von den Staatsmerkwürdigkeiten
Name Statistik: ital. statista = Staatsmann
Information zu Bevölkerung, Wirtschaft u.s.w.

Wahrscheinlichkeitsrechnung: Im 16. Jhdt.
Beschr. von Glücksspielen
Beschr. von Ungewissheit

verschiedene W-Begriffe: objektivistisch
subjektivistisch
axiomatisch

Stochastik: στοχαστικός
στοχάζεσθαι

Kausalität: deterministisch, stochastisch, fuzzy

Begleitende **B**eispiele: 6

B1: Stat. Qualitätskontrolle
N ... Losumfang, A ... Anzahl schlechter Stücke
n ... Stichprobenumfang, a —"— in der SP
Problem: Rückschluss auf A

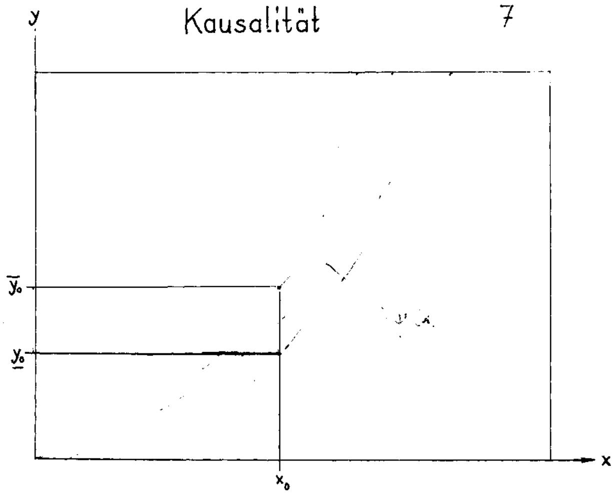
B2: Lebensdauer eines Produktes

B3: Wartezeit bei einer Bedienstelle

B4: Messung von Distanzen

B5: Bremsweg, Kovariable Geschwindigkeit

Kausalität 7



2. BESCHREIBENDE STATISTIK (S. 2)

Merkmale und Häufigkeiten

2.1 Diskrete Merkmale

höchstens abzählbar viele mögliche Werte, die sich nicht häufen

2.1.1 Artmerkmale

$A_1, A_2, \dots, A_m$ versch. mögl. Werte
n Beobachtungen
$H_n(A_j)$ = Anzahl der Beob. mit Wert A_j

$H_n(A_j)$ absolute Häufigkeit von A_j

$h_n(A_j) := H_n(A_j)/n$ relative –"–

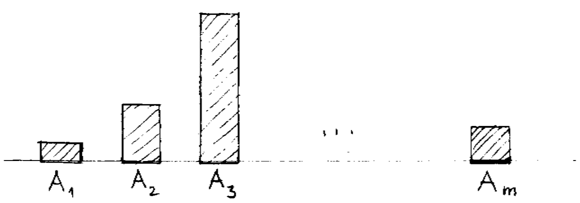
Darstellung von Häufigkeitsverteilungen:

- Strichlisten
- Balkendiagramme
- Kreisdiagramme

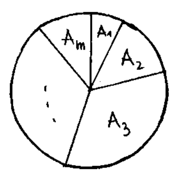
Häufigkeitstabelle

Merkmal-Ausprägung	Strichliste	Absolute Häufigkeit	Relative Häufigkeit
A_1	卌 卌 || ...	$H_n(A_1)$	$h_n(A_1)$
A_2	卌 卌 ...	$H_n(A_2)$	$h_n(A_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_j	卌 卌 卌 || ...	$H_n(A_j)$	$h_n(A_j)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_m	||| ...	$H_n(A_m)$	$h_n(A_m)$
	Summe	n	1

Balkendiagramm: Höhen der Balken proportional zu den Häufigkeiten



Kreisdiagramm: Sektorenflächen proportional zu den Häufigkeiten



2.1.2 Ordnungsmerkmale (größer - Beziehung)

oft durch Zahlen dargestellt

$z_1, z_2, \dots, z_m$ versch. mögliche Werte

Für n Beobachtungen

$h_n(z_j)$ für $j = 1(1)m$

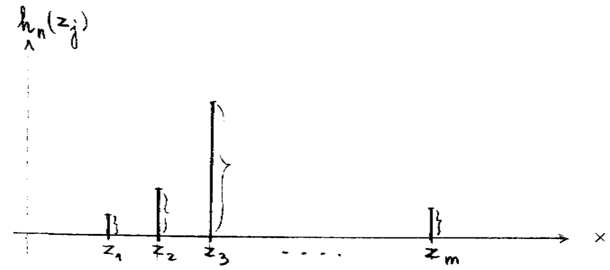
Darstellung der Häufigkeitsverteilung

Stabdiagramm

Summenkurve

2.6

Sind die Merkmalsausprägungen Zahlen, $z_1, \dots, z_m$, so stellt man die Häufigkeiten in Form eines <u>Stabdiagrammes</u> dar.



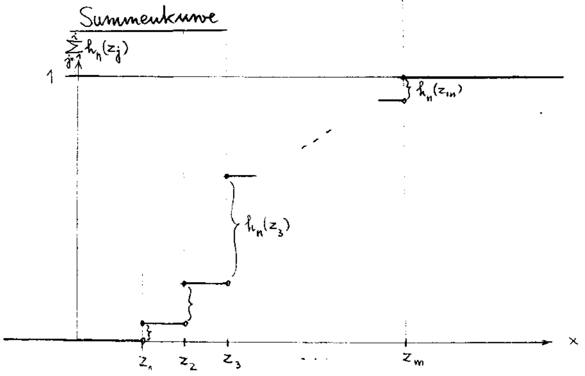
2.7

Manchmal werden Summenhäufigkeiten

$$\sum_{j=1}^{i} h_n(z_j), \quad i = 1(1)m$$

gewünscht und als reelle Funktion dargestellt:

<u>Summenkurve</u>



2.8

2.2 Kontinuierliche Merkmale

können alle Zahlen aus einem Intervall annehmen

Für n Beobachtungen $x_1, x_2, \dots, x_n$

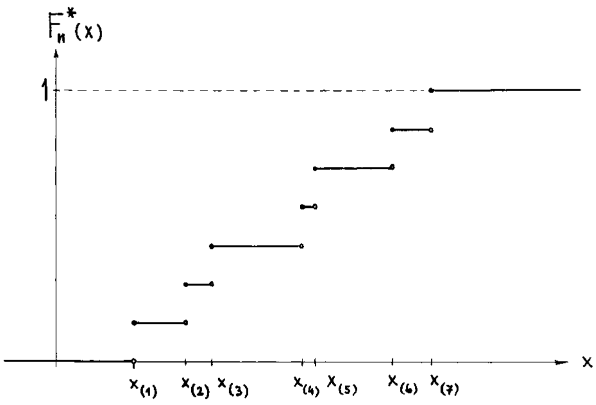
Darstellung der Häufigkeitsverteilung als

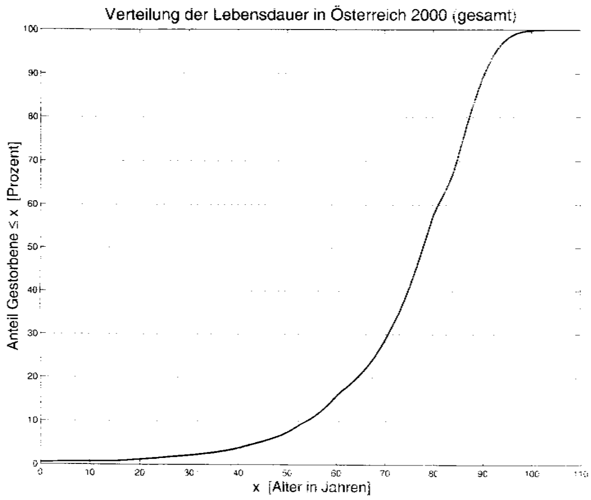
Empirische Verteilungsfunktion $F_n^*(\cdot)$

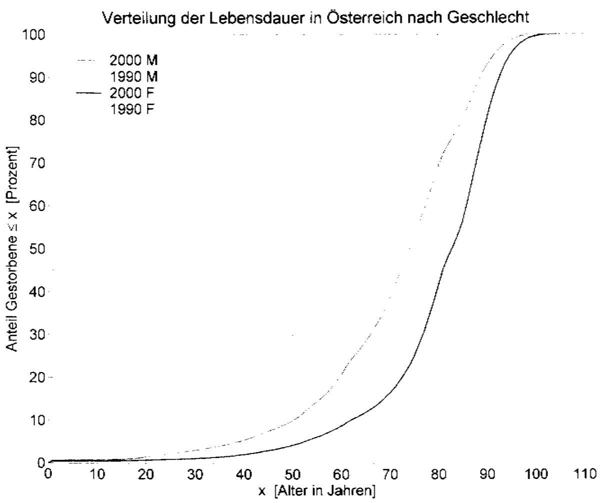
$F_n^*(x) := h_n\big((-\infty, x]\big)$ für jedes $x \in \mathbb{R}$

Ordnungsstatistik $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$

2.9







Für größere n bildet man Klassen

$[c_1, c_2], (c_2, c_3], (c_3, c_4], \ldots, (c_m, c_{m+1}]$

$K_1 \quad K_2 \quad K_3 \quad \ldots \quad K_m$

relative Häufigkeiten der Klassen

Darstellung: Histogramm

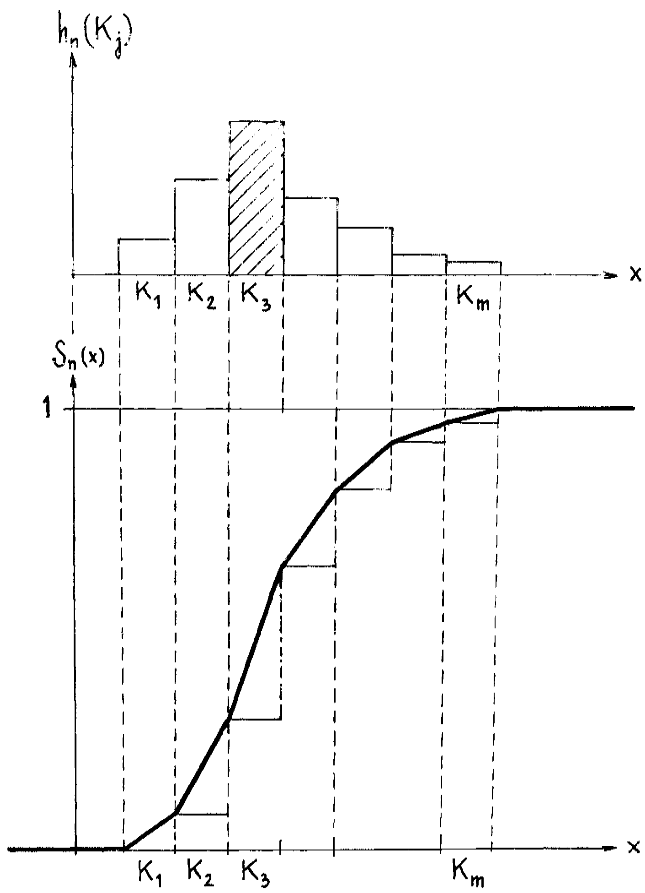
Summenpolygon $S_n(x)$

Theoretisches Modell: Wahrscheinlichkeitsdichte

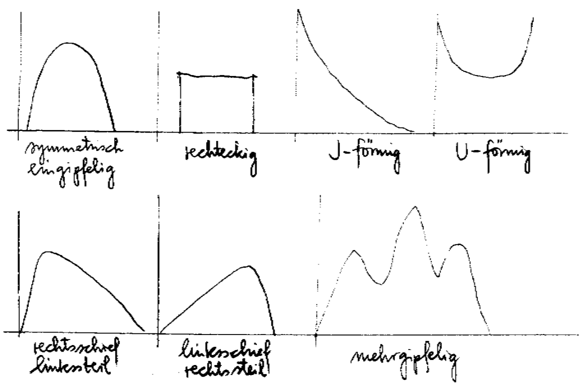
Verteilungsfunktion

Kontinuierliche Merkmale

Klassen K_j	Strichliste	absolute Häufigkeiten $H_n(K_i)$	relative Häufigkeiten $h_n(K_i)$	Summen-häufigkeiten $\sum_{j=1}^{i} h_n(K_j)$
$K_1=[c_1,c_2]$	||||| |||	$H_n(K_1)$	$h_n(K_1)$	$h_n(K_1)$
$K_2=(c_2,c_3]$	||||| ||||| ||||| ||	$H_n(K_2)$	$h_n(K_2)$	$h_n(K_1)+h_n(K_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$K_k=(c_k,c_{k+1}]$	||||| ||||| ||||| ||||| |||	$H_n(K_k)$	$h_n(K_k)$	$\sum_{j=1}^{k} h_n(K_j)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$K_m=(c_m,c_{m+1}]$	|||	$H_n(K_m)$	$h_n(K_m)$	1
		n	1	



Typen von Histogrammen (Verteilungen)



2.3 Lageparameter

numerische Daten $x_1, \dots, x_n$

a) Mittelwert

$$\overline{x_n} := \frac{1}{n} \sum_{i=1}^{n} x_i$$

Für klassierte Daten mit m Klassen und Klassenhäufigkeiten $H_n(K_j)$ sowie Klassenmitten z_j, $j = 1(1)m$:

$$\overline{x}_g := \frac{1}{n} \sum_{j=1}^{m} z_j H_n(K_j)$$

c) Empirische Fraktile

Für $p \in (0,1)$ jener Wert x_p, für den das Summenpolygon den Funktionswert p hat

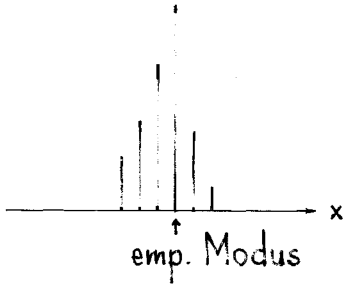
$$S_n(x_p) = p$$

<u>Bem.</u>: Für theoretische Verteilungen mittels der sog. Verteilungsfunktion

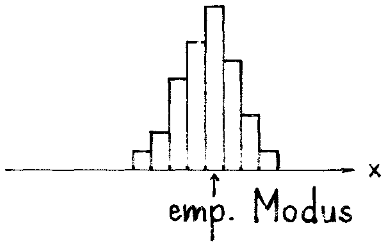
d) Modalwert (= empirischer Modus)

Jener Merkmalwert, der am häufigsten auftritt (Abb.)

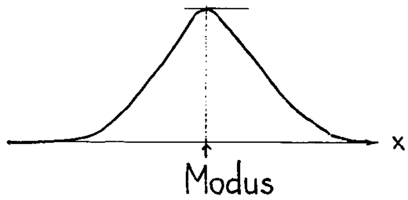
Diskrete Größen



Kontin. Größen



<u>Bem.</u>: Bei theoretischen kontinuierlichen Verteilungen



2.4 Streuungsparameter

Daten $x_1, \dots, x_n$

a) Spannweite $= x_{(n)} - x_{(1)} = \max_{i=1(1)n} x_i - \min_{i=1(1)n} x_i$

b) Quartilabstand $= x_{0,75} - x_{0,25}$

c) Mittlere absolute Abweichung

$\underset{\sim}{x}$ Median

$$MAD = \frac{1}{n} \sum_{i=1}^{n} |x_i - \underset{\sim}{x}|$$

mean absolute deviation

d) Mittlere quadratische Abweichung (empirische Varianz)

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

e) Empirische Streuung $s := \sqrt[+]{s^2}$

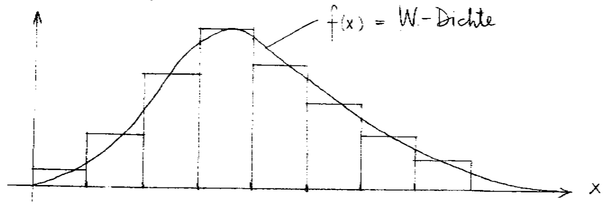
f) Empirischer Variationskoeffizient

$$VK := \frac{s}{\bar{x}}$$

Dimensionsloses Streumaß

<u>Bsp.</u>: Variabilität der Preise in versch. Ländern

Ziel der schließenden Statistik ist es, den empirisch gegebenen Verteilungen (Histogrammen) theoretische Verteilungen (Wahrscheinlichkeitsdichten) anzupassen, die erstere „gut" beschreiben



Bem. Die theoretischen Verteilungen braucht man für Vergleiche und Prognosen.

2.22

3. WAHRSCHEINLICHKEITSBEGRIFFE

Stat. Versuch: Wiederholbare Handlung mit mehr oder weniger ungewissem Ausgang

Wahrscheinlichkeiten: Zahlen, die den Grad des Vertrauens in Ereignisse angeben

Ereignis: Eine Menge von möglichen Versuchsausgängen

M ... Merkmalraum = Menge der möglichen Versuchsausgänge

3.1 Klassische Wahrscheinlichkeitsdefinition

m mögliche Versuchsausgänge

$M = \{a_1, a_2, \dots, a_m\}$

Ereignis $A = \{a_{i_1}, \dots, a_{i_g}\}$, $i_j \in \{1, \dots, m\}$

A enthält g Elemente

$W(A)$ ... Wahrscheinlichkeit von A

$$W(A) := \frac{g}{m}$$

Anwendung: B1 (SQK)

3.2 Geometrische Wahrscheinlichkeiten

Merkmalraum ∞, $M \subseteq \mathbb{R}^k$ mit $I(M) < \infty$

alle Teilräume von M mit gleichem Inhalt sind gleich wahrscheinlich

Für $A \subseteq M$ definiert man

$$W(A) := \frac{I(A)}{I(M)}$$

Bem.: $k=1$ Länge

$k=2$ Fläche

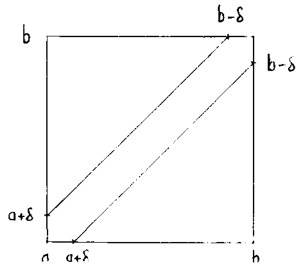
$k=3$ Volumen

Bsp.: Zufällige Ankunft in $[a,b]$

Teilintervall $[a_1,b_1] \subset [a,b]$

Bsp.: Rendezvous-Problem

2 Ankünfte in $[a,b]$, unabhängig



$$|t_1 - t_2| \le \delta$$

3.3 Grenzwert von Häufigkeiten

Konvergenzverhalten von $h_n(A)$ für $n \to \infty$

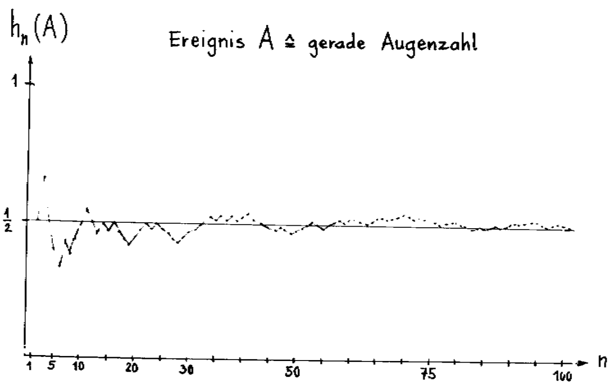
Empirisches Gesetz der großen Zahlen

$$h_n(A) \underset{n\to\infty}{\rightsquigarrow} p = W(A) \quad , \quad A \text{ fest}$$

Es gilt $\quad 0 \le h_n(A) \le 1 \quad \forall$ Ereignisse A

$A \wedge B$ unmöglich $\Rightarrow h_n(A \vee B) = h_n(A) + h_n(B)$

Bem.: Theorie der Kollektive



3.4 Axiomatische Wahrscheinlichkeiten

Ereignissystem $\mathcal{A}$ orthokomplementärer Verband
σ-vollständig

o unmögliches Ereignis

e sicheres Ereignis

$A^\perp$ das zu A komplementäre Ereignis

$A \wedge A^\perp = o$ und $A \vee A^\perp = e$

$$A_1, A_2, A_3, \ldots \text{ Folge mit } A_k \in \mathcal{A} \Rightarrow \begin{cases} \bigvee_{k=1}^{\infty} A_k \in \mathcal{A} \\ \bigwedge_{k=1}^{\infty} A_k \in \mathcal{A} \end{cases}$$

Wahrscheinlichkeitsverteilung W auf $\mathcal{A}$

$W: \mathcal{A} \rightarrow [0,1]$ mit

$W(\emptyset) = 0$ und $W(e) = 1$

Für jede Folge $A_1, A_2, A_3, \ldots$ von einander paarweise ausschließenden Ereignissen gilt

$$W\left(\bigvee_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} W(A_k) \qquad \ldots \ \sigma\text{- Add.}$$

<u>Bem.</u>: Endliche Additivität folgt daraus

$$W(A^{\perp}) = 1 - W(A)$$

3.8

3.5 Subjektive Wahrscheinlichkeiten

Bewertungen der Unsicherheit bedingt durch den Wissensstand H: $W(A|H)$

Ereignisse: A, B ; es muss gelten:

(1) $0 \leq W(A|H) \leq 1$

(2) Falls A und B einander ausschließen, gilt

$$W(A \vee B|H) = W(A|H) + W(B|H)$$

(3) Falls neue Information verfügbar: A eingetr.

$$W(A \wedge B|H) = W(B|A \wedge H) \cdot W(A|H)$$

3.9

3.6 Unscharfe Wahrscheinlichkeiten

Wahrscheinlichkeiten sind unscharfe Zahlen

S. 9 , die gewisse Bedingungen erfüllen

3.10

II

STOCHASTISCHE GRUNDBEGRIFFE

4. WAHRSCHEINLICHKEITSRÄUME (S. 24)

Spezialfall axiom. W. , M Merkmalraum
Spezialfall von $\mathcal{A}$: Ereignisfeld $\mathcal{E}$ ist ein System von Teilmengen von M

$A^{\perp} \mathrel{\hat{=}} A^c = M \setminus A$ zu A komplementäres Ereig.

$A \wedge B \mathrel{\hat{=}} A \cap B$ sowohl A als auch B Ereignis

$A \vee B \mathrel{\hat{=}} A \cup B$ entwer A oder B Ereignis

$o \mathrel{\hat{=}} \emptyset$ unmögliches Ereignis

$e \mathrel{\hat{=}} M$ sicheres Ereignis

Ein System $\mathcal{E}$ von Teilmengen von M heißt Ereignisfeld, wenn gilt:

$M \in \mathcal{E}$

$A \in \mathcal{E} \Rightarrow A^c \in \mathcal{E}$

$A_1, A_2, \ldots$ Folge mit $A_i \in \mathcal{E}$ $\forall i = 1(1)n$
$\Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{E}$

Bem.: Morgan'sche Regeln
$\bigcap_{i=1}^{\infty} A_i \in \mathcal{E}$

Wahrscheinlichkeitsverteilung W auf $\mathcal{E}$:

$W: \mathcal{E} \to [0,1]$

$W(M) = 1$

Für jede abzählbare Folge $A_1, A_2, A_3, \ldots$ von paarweise disjunkten Ereignissen gilt

$$W\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} W(A_k)$$

Def.: Die Struktur $(M, \mathcal{E}, W)$ heißt ein Wahrscheinlichkeitsraum.

4.1 Diskrete W-Räume

$M = \{a_1, a_2, \ldots\}$ höchstens abzählbar unendlich
kein Häufungspunkt

reelle Zahlen $p_1, p_2, \ldots$ mit $p_i \geq 0$, $p_i = W(a_i)$
und $\sum_{i=1}^{\infty} p_i = 1$

$\mathcal{E}$ = System aller Teilmengen von M

Für $A \subseteq M$ gilt $W(A) = \sum_{a \in A} W(a)$

Sonderfall: M endlich

4.2 Kontinuierliche W-Räume

$M = \mathbb{R}^k$, speziell $k = 1$

$\mathcal{E}$ = das am wenigsten umfassende Ereignisfeld auf $\mathbb{R}^k$, das alle Intervalle umfasst

W-Vtlg. auf $\mathcal{E}$ durch eine sog. Dichtefunktion bestimmt

$$f: \mathbb{R}^k \longrightarrow [0,\infty) \quad \text{mit} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k = 1$$

Für $B \in \mathcal{E}$ gilt $W(B) := \int \cdots \int_B f(x_1, \ldots, x_k)\, dx_1 \ldots dx_k$

4.3 Satz: In jedem W-Raum gilt:

(1) $A, B \in \mathcal{E}$ mit $A \subseteq B \Rightarrow W(A) \leqq W(B)$

(2) und $W(B \setminus A) = W(B) - W(A)$

Bew.: $B = A \cup (A^c \cap B) \Rightarrow W(B) = W(A) + W(A^c \cap B)$

$A^c \cap B = B \setminus A \Rightarrow W(B \setminus A) = W(B) - W(A)$ //

Folgerung: $W(A^c) = 1 - W(A)$

4.4 Additionstheorem: In jedem W-Raum gilt für beliebige Ereignisse A, B, A_k:

(1) $W(A \cup B) = W(A) + W(B) - W(A \cap B)$

(2) $W\left(\bigcup_{k=1}^{n} A_k\right) =$

$$= \sum_{\{i_1, \ldots, i_m\} \subseteq \{1, \ldots, n\}} (-1)^{m-1} W(A_{i_1} \cap \ldots \cap A_{i_m})$$

Bew.: (1) $A \cup B = A \cup (A^c \cap B)$
$B = (A \cap B) \cup (A^c \cap B)$

$$\begin{aligned} W(A \cup B) &= W(A) + W(A^c \cap B) \\ -W(B) &= -W(A \cap B) \pm W(A^c \cap B) \\ \hline W(A \cup B) - W(B) &= W(A) - W(A \cap B) \Rightarrow (1) \end{aligned}$$

(2) vollständige Induktion nach n

4.5 Bedingte Wahrscheinlichkeiten S. 30

Def.: Sind A und B Ereignisse eines W-Raumes mit $W(A) > 0$, so ist die durch A bedingte Wahrscheinlickeit von B

$$W(B|A) := \frac{W(A \cap B)}{W(A)}$$

Bem.: Ist A fest, so entsteht eine neue W-Vtlg., genannt Bedingte Verteilung W_A auf $A \cap \mathcal{E}$

$$W_A(B \cap A) = W(B|A)$$

4.6 Multiplikationstheorem

In jedem W-Raum gilt für Ereignisse A, B, A_k, falls die bed. W. $\exists$:

(1) $W(A \cap B) = W(B|A) \cdot W(A)$
$\qquad\qquad\qquad = W(A|B) \cdot W(B)$

(2) $W\left(\bigcap_{k=1}^{n} A_k\right) = W(A_1) \cdot W(A_2|A_1) \cdot W(A_3|A_1 \cap A_2) \cdots W(A_n | A_1 \cap \dots \cap A_{n-1})$

Bew.: (1) aus der Def. der bed. W.

(2) vollständige Induktion nach n.

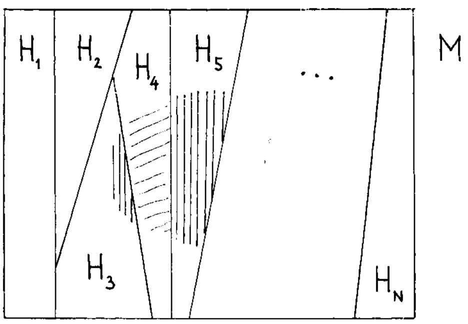
4.7 Satz v. d. vollständigen Wahrscheinlichkeit

S. 30 In jedem W-Raum $(M, \mathfrak{E}, W)$ gilt:

Ist $(H_k)_{k=1,\dots,N}$ eine endliche oder abzählbare Folge von paarweise disjunkten Ereignissen mit $W\left(\bigcup_{k=1}^{N} H_k\right) = 1$ und $W(H_k) > 0 \ \forall k = 1(1)N$, so gilt $\forall A \in \mathfrak{E}$

$$W(A) = \sum_{k=1}^{N} W(H_k) \cdot W(A|H_k)$$

Beweis: $A = \left[A \cap \left(\bigcup_{k=1}^{N} H_k\right)\right] \cup \left[A \cap \left(\bigcup_{k=1}^{N} H_k\right)^c\right]$



4.1(.

$W\left(\bigcup_{k=1}^{N} H_k\right) = 1 \Rightarrow W\left(\left(\bigcup_{k=1}^{N} H_k\right)^c\right) = 0$

$\Rightarrow W\left(A \cap \left(\bigcup_{k=1}^{N} H_k\right)^c\right) = 0$

$\Rightarrow W(A) = W\left(A \cap \left(\bigcup_{k=1}^{N} H_k\right)\right) = W\left(\bigcup_{k=1}^{N} (A \cap H_k)\right)$

$= \sum_{k=1}^{N} \underbrace{W(A \cap H_k)}_{W(A|H_k) \cdot W(H_k)} = \sum_{k=1}^{N} W(H_k) \cdot W(A|H_k)$ //

4.11

4.8 Bayes'sche Formel   S. 31

Unter den Vsn. des Satzes v. d. vollst. W. gilt:

$$W(H_i|A) = \frac{W(H_i)\cdot W(A|H_i)}{\sum_{k=1}^{N} W(H_k)\cdot W(A|H_k)} \quad \forall i = 1(1)N$$

Beweis: Zähler Multiplikationstheorem
Nenner Satz v. d. vollst. W

<u>Bem.</u>: $W(H_k)$ ... A-priori-Wahrsch. der H_k
$W(H_k|A)$ ... A-posteriori-W. " "

<u>Anwendung</u>: Qualitätssicherung

Produkte von verschiedenen Maschinen $H_1, \dots, H_m$

$W(H_k)$ ... W(Stück von Maschine H_k)

A ... Ausschussstück in Produktion

Frage 1: $W(A)$ ... S.v.v. W.

Frage 2: Wahrsch., dass ein Ausschussstück von der Maschine H_i stammt
Bayes'sche Formel $W(H_i|A)$

5. STOCHASTISCHE UNABHÄNGIGKEIT
S. 32

Unabhängigkeit von Ereignissen
Unabhängigkeit von Versuchen

<u>Definition</u>: Zwei Ereignisse A und B eines W-Raums heißen (stochastisch) unabhängig, i.Z. $A \perp\!\!\!\perp B$, wenn $W(A \cap B) = W(A)\cdot W(B)$.

5.1 Satz: Für Ereignisse A und B mit $W(A) > 0$, $W(B) > 0$, $W(A^c) > 0$ und $W(B^c) > 0$ gilt:

$$A \perp\!\!\!\perp B \Leftrightarrow \begin{cases} W(A|B) = W(A|B^c) = W(A) \\ W(B|A) = W(B|A^c) = W(B) \end{cases}$$

Anwendung: QS, Ziehungen <u>mit</u> Zurücklegen

<u>Definition</u>: Eine Familie $(A_i; i \in I)$ von Ereignissen eines W-Raumes heißen unabhängig, wenn für jede endliche Teilfamilie $A_{i_1}, \dots, A_{i_k}$ gilt:

$$W(A_{i_1} \cap \dots \cap A_{i_k}) = W(A_{i_1}) \cdots W(A_{i_k})$$

Anwendung: 1) mehrfache Ziehungen
2) zusammengesetzte Versuche

$$M = \{(x_1, \dots, x_n) : x_i \in M_i\} = M_1 \times \dots \times M_n$$

5.2 Produktwahrscheinlichkeitsräume

n W-Räume $(M_i, \mathcal{E}_i, W_i)$, $i = 1(1)n$
auf $M_1 \times \cdots \times M_n$ wird das am wenigsten umfassende Ereignisfeld betrachtet, das alle Mengen der Gestalt $A_1 \times \cdots \times A_n$ mit $A_i \in \mathcal{E}_i$ enthält.
Das so bestimmte Ereignisfeld wird mit $\bigotimes_{i=1}^{n} \mathcal{E}_i$ bezeichnet.
Auf $\bigotimes_{i=1}^{n} \mathcal{E}_i$ ist mittels der W_i folgendermaßen eine eindeutig bestimmte W-Vtlg. W bestimmt

$$W(A_1 \times \cdots \times A_n) = W_1(A_1) \cdots W_n(A_n) \quad \forall A_i \in \mathcal{E}_i$$

W heißt Produkt der W-Vtlgen $W_1, \dots, W_n$

<u>Anwendung</u>: n-maliges ziehen eines Probestückes mit Zurücklegen

Einzelversuch $M_i = \{0,1\}$, $W_i(\{1\}) = p = \frac{A}{N}$
$W_i(\{0\}) = 1-p$

Gesamtversuch $M = M_1 \times \cdots \times M_n = \{0,1\}^n$

Produktwahrscheinlichkeitsverteilung

$$W(\{(x_1, \dots, x_n)\}) = W_1(\{x_1\}) \cdots W_n(\{x_n\})$$

Da alle W_i gleich sind, gilt für n-Tupel, die genau k Einser enthalten:

$$W(\{(x_1, \dots, x_n)\}) = p^k (1-p)^{n-k}$$

6. STOCHASTISCHE GRÖSZEN S. 36

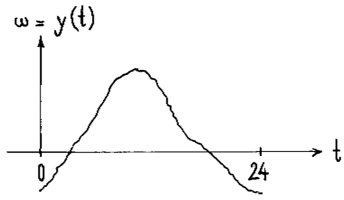
Größen die variieren bzw. nichtdeterministisch sind

<u>Beispiele</u>: Lebensdauern
Einkommen einer Firma nächstes Jahr
Wasserverbrauch einer Stadt im Jahr 2015
⋮

<u>Bemerkung</u>: Optimale Information
W-Vtlg.

Oft sind S.G. X Funktionen von Ausgängen von statistischen Experimenten

<u>Beispiel</u>: Temperaturverlauf in 24 Stunden (Experiment, Versuchsausgang ω)



SG $X(\omega)$ = maximale Temperatur (1-dim.)

2-dim. SG (Stoch. Vektor) $\underline{X}(\omega) = \begin{pmatrix} \max_{0 \le t \le 24} y(t) \\ \min_{0 \le t \le 24} y(t) \end{pmatrix}$

B1: X … Anzahl schlechter Stücke in der Stichprobe vom Umfang n

$\omega = (x_1, \dots, x_n) \in M_e^n$ mit $M_e = \{0,1\}$

$$X = X(\omega) = \sum_{i=1}^{n} x_i$$

Definition: Eine Abbildung $X: M \longrightarrow \mathbb{R}^m$ von einem W-Raum $(M, \mathcal{E}, W)$, für die das Urbild

$$X^{-1}(B) := \{\omega \in M : X(\omega) \in B\} \in \mathcal{E}$$

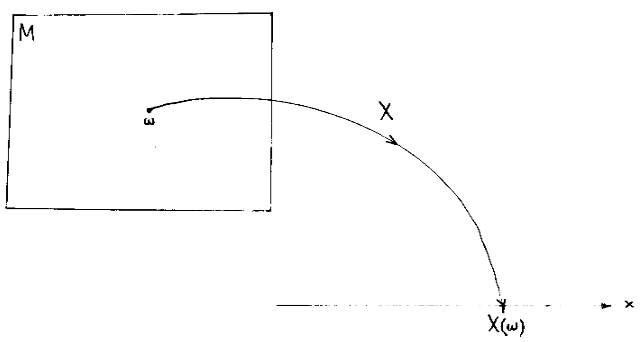
jedes Rechtecks $B = (a_1, b_1] \times \dots \times (a_m, b_m]$ in $\mathcal{E}$ liegt, nennt man m-dimensionale SG.

Sonderfall $m = 1$:

$$X: M \longrightarrow \mathbb{R} \quad \text{mit} \quad X^{-1}((a,b]) \in \mathcal{E} \quad \forall (a,b] \subseteq \mathbb{R}$$

Bemerkung: X messbare Funktion notwendig zur Berechnung der W-Verteilung von X

$$W\{X \in B\} := W\left(X^{-1}(B)\right)$$



III

EINDIMENSIONALE VERTEILUNGEN

7. VERTEILUNG STOCH. GRÖSZEN S. 39

Bestimmung der W-Verteilung einer 1-dim. SG X

B1: $X : M_e^n \longrightarrow \mathbb{N}_0$, $\quad M_X = \{0,1,2,\dots,n\}$

$$X((x_1,\dots,x_n)) = \sum_{i=1}^{n} x_i$$

Für $k \in M_X$ gilt $W\{X=k\} = W\left(X^{-1}(\{k\})\right)$

$$= W\left(\left\{(x_1,\dots,x_n) : \sum_{i=1}^{n} x_i = k\right\}\right)$$

$$= \sum_{\substack{(x_1,\dots,x_n): \\ \sum x_i = k}} W(\{(x_1,\dots,x_n)\}) = \sum_{\substack{(x_1,\dots,x_n): \\ \sum x_i = k}} p^k (1-p)^{n-k}$$

$$= \binom{n}{k} p^k (1-p)^{n-k} \quad \forall\, k = 0(1)n$$

B2: Lebensdauer eines Systems aus n Komponenten $\quad T = \min\{T_1, \dots, T_n\}$

7.1 Verteilungsfunktionen (1-dim.)

Definition: Die Verteilungsfunktion $F(\cdot)$ einer SG X ist durch ihre Funktionswerte

$$F(x) = W\{X \le x\} \quad \forall x \in \mathbb{R}$$

Bemerkung: $\{X \le x\} := X^{-1}((-\infty, x]) \in \mathcal{E}$

$\Rightarrow$ $W\{X \le x\}$ ist definiert

7.2 Satz: Für die VF $F(\cdot)$ einer SG X gilt:

(1) $\quad 0 \le F(x) \le 1 \quad \forall x \in \mathbb{R}$

(2) $\quad x_1 < x_2 \;\Rightarrow\; F(x_1) \le F(x_2)$

(3) $\quad \lim_{x \to -\infty} F(x) = 0$ und $\lim_{x \to +\infty} F(x) = 1$

(4) $\quad \lim_{h \downarrow 0} F(x+h) = F(x) \quad \forall x \in \mathbb{R}$ (rechtsstetig)

(5) $\quad W\{X=x\} = F(x) - \lim_{h \downarrow 0} F(x-h)$ (linkss. GW)

(6) $\quad a<b \;\Rightarrow\; W\{a < X \le b\} = F(b) - F(a)$

Zum Beweis: (1) Wahrsch., (2) $(-\infty, x_1] \subseteq (-\infty, x_2]$, (3)-(5) schwieriger, (6) $(-\infty, a] \subset (-\infty, b]$

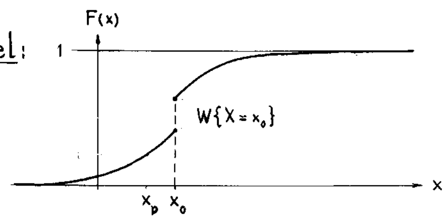
Definition: Ist $F(\cdot)$ eine 1-dim. VF, so heißt für $p \in [0,1]$ jede reelle Zahl x_p, für die gilt

$$F(x_p) = p, \quad \text{ein } p\text{-Fraktiles}.$$

Bemerkung: Für $X \sim F(\cdot)$ gilt $W\{X \le x_p\} = p$

Definition: $x_{0,5}$ heißt Median

Beispiel:



7.2 Typen von Verteilungen

a) diskrete Verteilungen
   VF treppenförmig

b) kontinuierliche Verteilungen
   $\exists$ Dichtefunktion $f: \mathbb{R} \to [0, \infty)$ integrierbar
   mit $F(x) = \int_{-\infty}^{x} f(t)\,dt \quad \forall x \in \mathbb{R}$

c) gemischte Verteilungen
   weder diskret noch kontinuierlich,
   eine Mischung von beidem $F(x) = \alpha F_1(x) + (1-\alpha) F_2(x$

8. DISKRETE VERTEILUNGEN

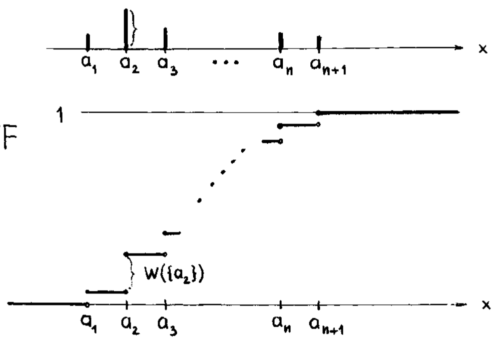
$\exists$ höchstens abzählbar viele $a_i \in \mathbb{R}$, die keinen Häufungspunkt haben, mit $W(\{a_i\}) = p_i > 0$, für die gilt: $\sum_{i=1}^{\infty} p_i = 1$

Die Wahrscheinlichkeit einer beliebigen Borelmenge B erhält man folgendermaßen:

$$W(B) = \sum_{a_i \in B} W(\{a_i\})$$

Bemerkungen: 1) meist $a_i \in \mathbb{N}_0$
2) $X \sim (M_X, p_i)$

Stabdiagramm



VF

8.1 Dirac-Verteilung δ_μ, $\mu \in \mathbb{R}$

$W(\{\mu\}) = 1$

$W(\{x\}) = 0 \quad \forall x \neq \mu$

Bemerkung: Determinismus

8.2 Diskete Gleichverteilung D_m, $m \in \mathbb{N}$

$X \sim D_m :\Leftrightarrow M_X = \{1, 2, \dots, m\}$

$W(\{1\}) = W(\{2\}) = \dots = W(\{m\}) = \frac{1}{m}$

Beispiele: 1) Würfeln, Roulette
2) SQK

8.3 Alternativverteilung A_p , $p \in (0,1)$

$X \sim A_p \;:\Leftrightarrow\; M_X = \{0,1\}$

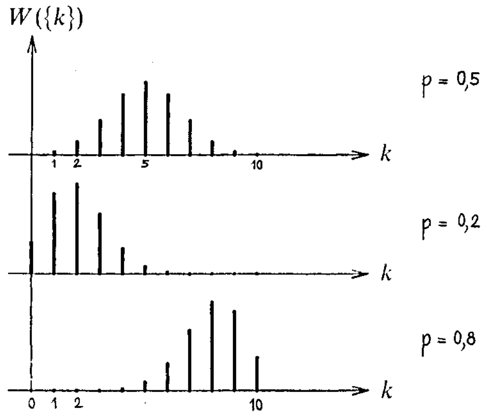
$$W(\{1\}) = p \quad \text{und} \quad W(\{0\}) = 1-p$$

8.4 Binomialverteilung $B_{n,p}$, $n \in \mathbb{N}$, $p \in (0,1)$

führt man n-mal unabhängig einen Alternativversuch durch, so ist die Anzahl der Einser $\sim B_{n,p} \Rightarrow M_X = \{0, 1, 2, \dots, n\}$

$$W(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k} \quad \forall k = 0(1)n$$

Stabdiagramme verschiedener Binomialverteilungen $B_{10,p}$



8.5 Hypergeometrische Verteilung $H_{N,A,n}$

Ziehungen ohne Zurücklegen von n Stück aus insgesamt N Stück, von denen A Ausschuss sind; $A \le N$, $n \le N$ (Stichprobenumfang n)

X = Anzahl der Ausschussstücke in der Stichprobe

$M_X = \{a_1, a_1+1, a_1+2, \dots, a_2\}$

mit $a_1 = \max\{0, n-(N-A)\}$, $a_2 = \min\{n, A\}$

$$W(\{a\}) = \frac{\binom{A}{a}\binom{N-A}{n-a}}{\binom{N}{n}} \quad \text{für } a = a_1(1)a_2$$

Begründung: klassische W-Definition für alle Auswahlen von n Elementen aus N vorhandenen

$\binom{N}{n}$ mögliche Auswahlen von n Elementen

$\binom{A}{a}$ mögliche Auswahlen von a Ausschussstücken aus den A Ausschussstücken

$\binom{N-A}{n-a}$ mögliche Auswahlen von n-a guten Stück aus den insgesamt N-A guten Stücken

Aus $\frac{g}{m}$ folgt die Formel //

Bemerkung: Approximation für $n < \frac{N}{10}$

$$H_{N,A,n} \approx B_{n,\frac{A}{N}}$$

8.6 Poisson-Verteilung P_μ, $\mu > 0$

$$X \sim P_\mu :\Leftrightarrow \begin{cases} M_X = \mathbb{N}_0 \\ W(\{k\}) = \frac{\mu^k e^{-\mu}}{k!} \quad \text{für } k = 0(1)\infty \end{cases}$$

<u>Bemerkungen</u>: $0! = 1$

Exponentialreihe $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

Anwendung für punktförmige Ereignisse:

- Versicherungsfälle
- Elementzerfallsvorgänge
- Astronomie
- u.a.

Poisson-Prozess, Voraussetzungen $\Rightarrow P_\mu$

Approximationen:

$$B_{n,p} \approx P_{n\cdot p} \quad \text{falls} \quad n > 50 \wedge p < \tfrac{1}{10}$$

$$H_{N,A,n} \approx P_{n\cdot\frac{A}{N}} \quad \text{falls} \quad n < \tfrac{N}{10} \wedge A < \tfrac{N}{10}$$

8.7 Geometrische Verteilung G_p, $p \in (0,1)$

Wird ein Alternativversuch A_p unabhängig durchgeführt, so ist die Anzahl X der Durchführungen, bis erstmals „1" auftritt, geometrisch verteilt, $X \sim G_p$, mit

$$X \sim G_p :\Leftrightarrow \begin{cases} M_X = \mathbb{N} \\ W\{X = n\} = p\cdot(1-p)^{n-1} \quad \forall\ n \in \mathbb{N} \end{cases}$$

<u>Bemerkung</u>: Geometrische Reihe

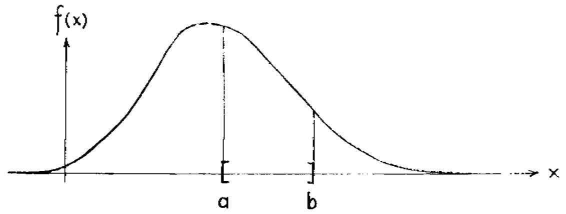
$$\sum_{n=1}^{\infty} p\cdot(1-p)^{n-1} = 1 \quad \text{falls} \quad 0 < p < 1$$

9. KONTINUIERLICHE EINDIMENSIONALE VERTEILUNGEN

Kontinuierliche SG X nehmen alle Werte eines Intervalls an (Kontinuum), ihre W-Vtlg. ist durch eine integrierbare Funktion $f(\cdot)$, genannt Dichtefunktion, bestimmt:

$$f: \mathbb{R} \longrightarrow [0,\infty) \quad \text{mit} \quad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

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



9.1 Satz: Für kontinuierliche Verteilungen bzw. kontinuierlich vt. SG X mit Dichte $f(\cdot)$ gilt:

1) VF $F(x) = \int_{-\infty}^{x} f(\xi)\,d\xi \quad \forall x \in \mathbb{R}$

2) $\exists \frac{d}{dx} F(x) \Rightarrow f(x) = F'(x)$

3) $W\{X = x\} = 0 \quad \forall x \in \mathbb{R}$

Beweis: 1) $F(x) = W((-\infty, x]) = \int_{-\infty}^{x} f(\xi)\,d\xi$

2) Differenziazion des Integrals

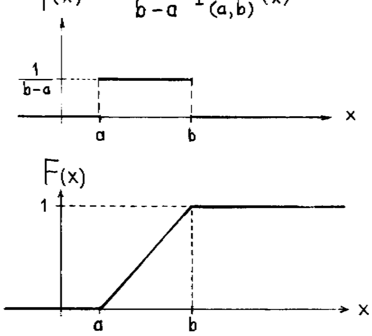
3) $W\{X = x\} = \int_{x}^{x} f(\xi)\,d\xi = 0$

<u>Definition</u>: Ist M eine beliebige Menge, so ist die Indikatorfunktion $I_A(\cdot)$ einer Teilmenge A von M folgendermaßen gegeben:

$$I_A(x) := \begin{Bmatrix} 1 & \text{für } x \in A \\ 0 & \text{für } x \notin A \end{Bmatrix} \quad \forall x \in M$$

9.2 Uniforme Verteilung $U_{a,b}$, $a < b \in \mathbb{R}$

$f(x) = \frac{1}{b-a} I_{(a,b)}(x)$



Rechteckverteilung

9.3 Exponentialverteilung Ex_τ, $\tau > 0$

$f(x) = \frac{1}{\tau} e^{-\frac{x}{\tau}} I_{(0,\infty)}(x)$

$F(x) = (1 - e^{-\frac{x}{\tau}}) I_{[0,\infty)}(x)$

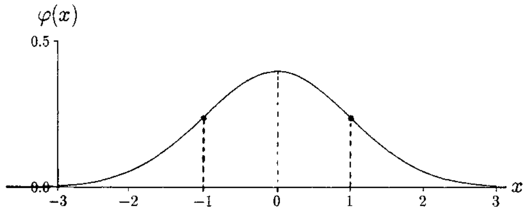
<u>Anwendung</u>: Wartezeiten
τ ... mittlere Wartezeit

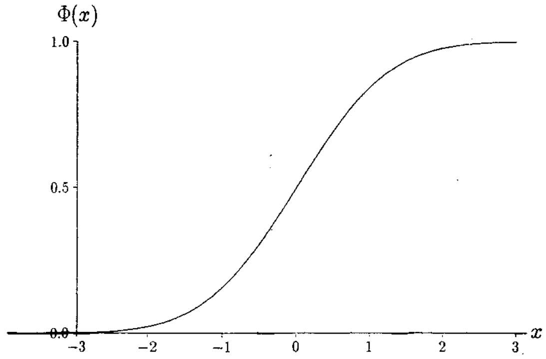
9.4 Standard-Normalverteilung $N(0,1)$

Dichte $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \forall x \in \mathbb{R}$

VF $\Phi(x) = \int_{-\infty}^{x} \varphi(x)\,dx$ tabelliert

N(0,1)-Verteilung





$$\Phi(-x) = 1 - \Phi(x) \qquad \forall x > 0$$

9.5 Allgemeine Normalverteilung $N(\mu, \sigma^2)$

Parameter $\mu \in \mathbb{R}$ und $\sigma^2 > 0$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \forall x \in \mathbb{R}$$

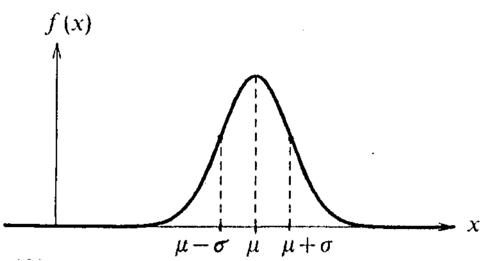
symmetrische, Gauß'sche Glockenkurve

Abbildung

Es gilt: Die Funktionswerte der VF $F(\cdot\,|\,\mu,\sigma^2)$ der $N(\mu,\sigma^2)$ erhält man durch:

$$F(x\,|\,\mu,\sigma^2) = \Phi\left(\frac{x-\mu}{\sigma}\right) \qquad \forall x \in \mathbb{R}$$

Normalverteilung



Beweis: $F(x\,|\,\mu,\sigma^2) = \int_{-\infty}^{x} f(t)\,dt$

Substitution $\frac{t-\mu}{\sigma} = u$, $dt = \sigma\, du$,

t	u
$-\infty$	$-\infty$
x	$\frac{x-\mu}{\sigma}$

$$\Rightarrow \int_{-\infty}^{x} f(t)\,dt = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}\,\cancel{\sigma}}\, e^{-\frac{u^2}{2}}\, \cancel{\sigma}\, du$$

$$= \Phi\left(\frac{x-\mu}{\sigma}\right) \quad //$$

Es gilt: Aus $X \sim N(\mu,\sigma^2)$ und $\alpha > 0$, $\beta \in \mathbb{R}$ folgt $\alpha \cdot X + \beta \sim N(\alpha\cdot\mu+\beta,\, \alpha^2\cdot\sigma^2)$

Beweis: Integraltransformation (VF)

9.6 Logarithmische Normalverteilung $LN(\mu,\sigma^2)$

53 Parameter $\mu \in \mathbb{R}$, $\sigma^2 > 0$

<u>Definition</u>: $X > 0$ ist vt $LN(\mu,\sigma^2)$, wenn $\ln X \sim N(\mu,\sigma^2)$

Es gilt: Die VF der $LN(\mu,\sigma^2)$ lautet

$$F_{\mu,\sigma^2}(x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \cdot I_{(0,\infty)}(x)$$

Beweis: $F_{\mu,\sigma^2}(x) = W\{X \leq x\} = W\{\ln X \leq \ln x\}$
$= F_N(\ln x \mid \mu,\sigma^2) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \quad \forall x > 0$ //

Es gilt: Die Dichtefunktion der $LN(\mu,\sigma^2)$ ist

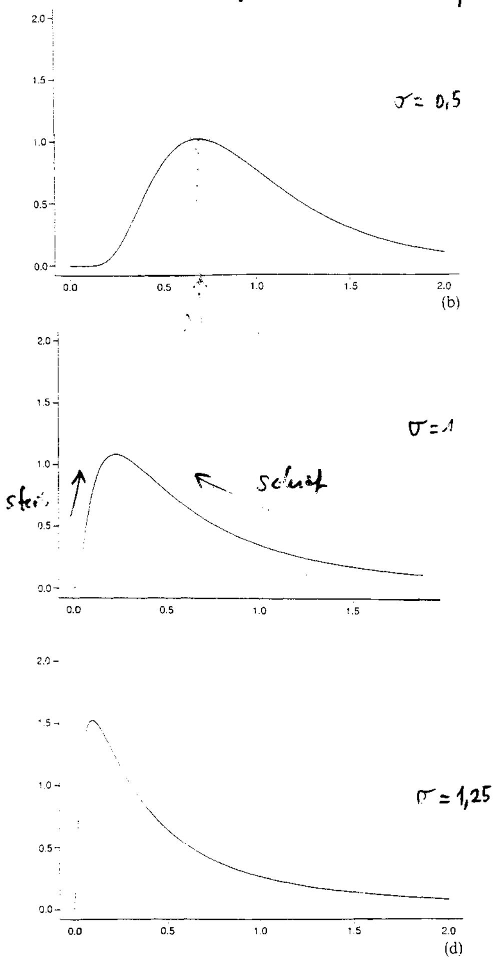
$$f(x\mid\mu,\sigma^2) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2} I_{(0,\infty)}(x)$$

Beweis: Ableitung der Verteilungsfunktion

<u>Bemerkung</u>: Dichte ist rechtsschief und linkssteil (Abbildung)

<u>Definition</u>: Falls eine Dichtefunktion genau ein Maximum hat, nennt man den Abszissenwert des Maximums den Modus der Dichte

$X \sim LN(\mu,\sigma^2)$, $\quad \mu + \frac{\sigma^2}{2} = 1$



9.7 t_n-Verteilung, $n \in \mathbb{N}$

Dichte $\quad f(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{n\pi}} \cdot \frac{1}{\left(1+\frac{x^2}{n}\right)^{\frac{n+1}{2}}} \quad \forall x \in \mathbb{R}$

<u>Es gilt</u>: $X \sim t_n \Rightarrow EX = 0 \quad \forall n > 1$

$\mathrm{Var} X = \frac{n}{n-2} \quad \forall n > 2$

<u>Bem.</u>: Für $n \to \infty$ konvergiert die Dichte gegen die Dichte der $N(0,1)$

9.8 χ^2_n - Verteilung, $\quad n \in \mathbb{N}$

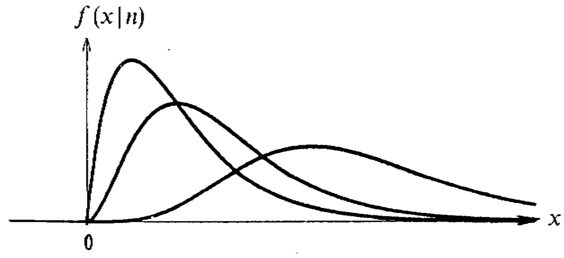
Dichte $\quad f(x) = \dfrac{x^{\frac{n}{2}-1}\, e^{-\frac{x}{2}}}{\Gamma(\frac{n}{2}) \cdot 2^{\frac{n}{2}}}\, I_{(0,\infty)}(x)$

Es gilt: $X \sim \chi^2_n \;\Rightarrow\; \mathbb{E}X = n$

$\qquad\qquad\qquad\qquad\qquad \mathrm{Var}\, X = 2n$

$$X \sim \chi^2_n \;\Rightarrow\; X^2 \sim \chi^2_1$$

Dichtefunktionen von χ^2_n-Verteilungen



10. GEMISCHTE VERTEILUNGEN

Definition: 1-dim. Verteilungen, die weder diskret noch kontinuierlich sind, heißen gemischte Verteilungen

10.1 Satz: Die VF $F(\cdot)$ einer gemischten Verteilung ist eine Konvexkombination einer diskreten VF $F_1(\cdot)$ und einer kontinuierlichen VF $F_2(\cdot)$,

$$F(x) = p \cdot F_1(x) + (1-p) \cdot F_2(x) \quad \forall x \in \mathbb{R}$$

mit $0 < p < 1$.

B2: Lebensdauer einer Glühlampe ... X
Zeit ist kontinuierlich, aber $W\{X=0\} > 0$, daher ist X nicht kontinuierlich verteilt.

Ist aber $X > 0$, so ist die Lebensdauer durch eine sog. modifizierte Dichtefunktion

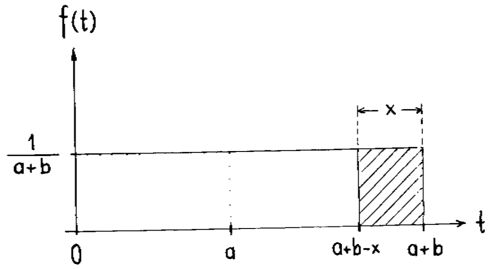
$f^*: [0,\infty) \to [0,\infty)$ mit

$\int_0^\infty f^*(x)\,dx = 1 - W\{X=0\}$ bestimmt.

Verteilung von X: Für $[0,b] \subseteq [0,\infty)$ gilt

$$W\{0 \le X \le b\} = W\{X=0\} + \int_0^b f^*(x)\,dx$$

<u>B3</u>: Wartezeit an einer Verkehrsampel
Grünphase a [sec.], Rotphase b [sec.]
Ankunft zufällig ... $T \sim U_{0,a+b}$



Wartezeit X:

$$X = \begin{cases} 0 & \text{für } T \in [0,a] \\ a+b-T & \text{für } T \in [a, a+b] \end{cases}$$

W-Vtlg. von X:

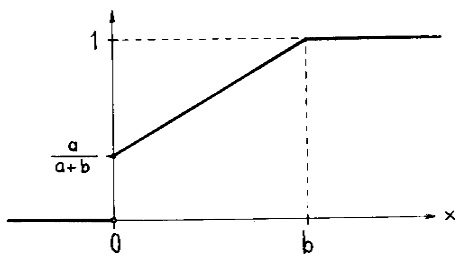
$$W\{X=0\} = W\{T \in [0,a)\} = \frac{a}{a+b} > 0$$

$$W\{0 < X \leq x\} = W\{T \in [a+b-x, a+b)\} = \frac{x}{a+b} \quad \text{für } 0 < x < b$$

$$W\{X > b\} = 0$$

Verteilungfunktion von X:

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



<u>Übung</u>: Bestimmung der modifizierten Dichte $f^*(\cdot)$

10.2 Mischverteilungen

Diese entstehen als Kombination von VF$_n$ $F_i(\cdot)$, $i=1(1)m$, Gewichte $\alpha_i > 0$ $\forall i=1(1)m$ mit $\sum_{i=1}^{m} \alpha_i = 1$:

$$F(x) = \sum_{i=1}^{m} \alpha_i F_i(x) \quad \forall x \in \mathbb{R}$$

<u>Beispiel</u>: Zusammengesetzte Warenlieferung

<u>Bemerkung</u>: Mischverteilungen können auch kontinuierliche Verteilungen sein.

11. ERWARTUNGSWERT

Mittlerer Wert einer Verteilung

Beispiel D_m Mittelwert

$$\frac{1}{m}\sum_{i=1}^{m} x_i = \sum_{i=1}^{m} x_i\, p(x_i)$$

11.1 Diskrete Verteilungen

$X \sim \left(M_X,\ p(x) = W\{X = x\},\ x \in M_X\right)$

Erwartungswert $\mathbb{E}X$ = Mittel μ der Verteilung von X

$$\mathbb{E}X := \sum_{x:\, p(x) > 0} x \cdot p(x)$$

11.2 Kontinuierliche Verteilungen

$X \sim f(\cdot)$

$$\mathbb{E}X := \int_{-\infty}^{\infty} x\, f(x)\, dx$$

Bem.: Massenschwerpunkt

Beispiele: $X \sim Ex_\tau \Rightarrow \mathbb{E}X = \tau$

$X \sim N(\mu, \sigma^2) \Rightarrow \mathbb{E}X = \mu$

$X \sim U_{a,b} \Rightarrow \mathbb{E}X = \frac{a+b}{2}$

11.3 Gemischte Verteilungen

$X \sim \left(M_X = \{x_1, \dots, x_m\} \cup \langle a, b\rangle,\ p(x_i),\ f^*(\cdot)\right)$

$0 < \sum_{i=1}^{m} p(x_i) < 1$ und $\int_a^b f^*(x)\, dx = 1 - \sum_{i=1}^{m} p(x_i)$

$$\mathbb{E}X := \sum_{i=1}^{m} x_i\, p(x_i) + \int_a^b x\, f^*(x)\, dx$$

Beispiele: Mittlere Wartezeit in Abschn. 10

Erwartungswerte von Lebensdauern

Bem.: Allgem. $\mathbb{E}X = \int_{-\infty}^{\infty} x\, dF(x)$

12. FUNKTIONEN VON STOCH. GRÖSZEN – ERWARTUNGSWERT

X Stoch. Gr., $\psi: \mathbb{R} \to \mathbb{R}$ (messbar)

$\Rightarrow$ $Y := \psi(X)$:

Wert von Y := ψ(Wert von X)

$\Rightarrow$ Y Stoch. Gr.

Frage: $\mathbb{E}Y$

i.A. $\mathbb{E}\,\psi(X) \neq \psi(\mathbb{E}X)$

Bem.: Für lineares $\psi(\cdot)$ Gleichheit

12.1 Satz vom unbewussten Statistiker

(1) Ist X disket vt. $X \sim p(\cdot)$, $\psi : \mathbb{R} \to \mathbb{R}$ messbar und $\exists\, \mathbb{E}[\psi(X)]$, so gilt:

$$\mathbb{E}[\psi(X)] = \sum_{x:\, p(x)>0} \psi(x)\, p(x)$$

(2) Ist X kontin. vt. $X \sim f(\cdot)$, $\psi : \mathbb{R} \to \mathbb{R}$ wie in (1), so gilt:

$$\mathbb{E}[\psi(X)] = \int_{-\infty}^{\infty} \psi(x) f(x)\, dx$$

(3) Allgem.: $\mathbb{E}[\psi(X)] = \int_{-\infty}^{\infty} \psi(x)\, dF(x)$

12.2 Varianz

Falls die folgende Erwartung existiert, heißt

$$\operatorname{Var} X := \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$$

Varianz von X bzw. Varianz der Verteilung

Berechnung mittels des SvuStat.

Für gemischte Verteilungen:

$$\operatorname{Var} X = \sum_{i=1}^{m} (x_i - \mathbb{E}X)^2 p(x_i) + \int_a^b (x - \mathbb{E}X)^2 f^*(x)\, dx$$

Bsp.: Wartezeit aus Abschnitt 10

$X \sim N(\mu, \sigma^2) \Rightarrow \operatorname{Var} X = \sigma^2$

$X \sim Ex_\tau \Rightarrow \operatorname{Var} X = \tau^2$

$X \sim A_p \Rightarrow \operatorname{Var} X = p(1-p)$

$X \sim B_{n,p} \Rightarrow \operatorname{Var} X = np(1-p)$

$X \sim U_{a,b} \Rightarrow \operatorname{Var} X = \frac{(b-a)^2}{12}$

12.3 Satz: Ist X eine St. Gr. mit $\exists \operatorname{Var} X$ und $c \in \mathbb{R}$, $d \in \mathbb{R}$, konstant, so gilt:

(1) $\operatorname{Var}(X+d) = \operatorname{Var} X$

(2) $\operatorname{Var}(cX) = c^2 \operatorname{Var} X$

(3) $\operatorname{Var}(cX+d) = c^2 \operatorname{Var} X$

Beweis: (3) $\operatorname{Var}(cX+d) = \mathbb{E}\left[\{cX+d-\mathbb{E}(cX+d)\}^2\right]$
$= \mathbb{E}\left[\{cX+d-c\mathbb{E}X-d\}^2\right] = \mathbb{E}\left[c^2(X-\mathbb{E}X)^2\right]$
$= c^2\, \mathbb{E}\left[(X-\mathbb{E}X)^2\right] = c^2 \operatorname{Var} X$ //

12.4 Verschiebungssatz: Ist X eine SG mit existierender Varianz, so gilt:

$$\operatorname{Var} X = \mathbb{E}[X^2] - (\mathbb{E}X)^2$$

Beweis: $\operatorname{Var} X = \mathbb{E}\left[(X-\mathbb{E}X)^2\right] = \mathbb{E}\left[X^2 - 2(\mathbb{E}X)X + (\mathbb{E}X)^2\right]$
$= \mathbb{E}(X^2) - 2(\mathbb{E}X)\mathbb{E}X + \mathbb{E}\left[(\mathbb{E}X)^2\right] = \mathbb{E}(X^2) - 2(\mathbb{E}X)^2 +$
$+ (\mathbb{E}X)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ //

12.5 Standardisierung

<u>Def.</u>: $\sqrt[+]{\operatorname{Var} X}$ heißt Streuung von X

<u>Def.</u>: Ist X SG mit $\exists \operatorname{Var} X = \sigma^2$ ($\Rightarrow \exists \mathbb{E}X = \mu$), so ist die zugeh. Standardisierte SG

$$\frac{X-\mu}{\sigma}$$

<u>Es gilt</u>: (1) $\mathbb{E}\left(\frac{X-\mu}{\sigma}\right) = 0$

(2) $\operatorname{Var}\left(\frac{X-\mu}{\sigma}\right) = 1$

Bew.: (1) $\frac{X-\mu}{\sigma} = \frac{1}{\sigma}(X-\mu)$ und $\mathbb{E}\left(\frac{1}{\sigma}(X-\mu)\right) = \ldots$

13. VERTEILUNG VON FUNKTIONEN STOCH. GRÖSZEN

X SG, $\psi: \mathbb{R} \to \mathbb{R}$ messbar $Y = \psi(X)$

$F(\cdot)$ VF von X

gesucht: VF $G(\cdot)$ von Y

13.1 Satz: $X \sim F(\cdot)$, $\psi: \mathbb{R} \to \mathbb{R}$ monoton und invertierbar, so hat $Y = \psi(X)$ die VF

$$G(x) = F(\psi^{-1}(x)) \quad \forall x \in \mathbb{R}$$

Bew.: $G(x) = W\{Y \le x\} = W\{\psi(X) \le x\}$
$= W\{X \le \psi^{-1}(x)\} = F(\psi^{-1}(x)) \quad \forall x \in \mathbb{R}$ //

13.2 Satz: Gilt zusätzlich zu den Vorauss. von Satz 13.1 dass X kontin. vt. ist mit Dichte $f(\cdot)$ und dass $\psi^{-1}(\cdot)$ differenzierbar ist, so ist $Y = \psi(X)$ auch kontin. vt. mit folgender Dichte $g(\cdot)$:

$$g(x) = f(\psi^{-1}(x)) \cdot \left|\frac{d}{dx}\psi^{-1}(x)\right|$$

Bew.: Kettenregel der Diff.-Rechnung

Beispiel: $X \sim N(0,1)$, gesucht: Vtlg. von X^2

VF $G(\cdot)$ von X^2: $G(x) = W\{X^2 \le x\}$ (Bem. zu Satz)

$= W\{-\sqrt{x} \le X \le \sqrt{x}\} = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) = 2\Phi(\sqrt{x}) - 1$

Differentiation von $G(\cdot)$ ergibt die Dichte $g(\cdot)$ von X^2

$$g(x) = \frac{d}{dx}G(x) = 2\Phi'(\sqrt{x}) \cdot \tfrac{1}{2} \cdot x^{-\frac{1}{2}} \quad \text{für } x > 0$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} x^{-\frac{1}{2}} I_{(0,\infty)}(x) \quad \hat{=} \quad \chi^2_1$$

(da $\Gamma(\frac{1}{2}) = \sqrt{\pi}$)

13.3 Satz: Gilt $X \sim N(\mu, \sigma^2)$ und sind $c, d \in \mathbb{R}$ mit $c \neq 0$, so gilt:

$$cX + d \sim N(c\mu + d, c^2\sigma^2)$$

Bew.: Satz 9.1 und Anwendung von Satz 13.2 //

13.4 Satz: (1) Ist $X \sim F(\cdot)$ und $\exists\, F^{-1}(\cdot)$

$$\Rightarrow \; F(X) \sim U_{0;1}$$

(2) Ist $F(\cdot)$ eine invertierbare VF und $Y \sim U_{0;1}$

$$\Rightarrow \; F^{-1}(Y) \text{ hat die VF } F(\cdot)$$

Bew.: (1) Ist $G(\cdot)$ die VF der SG $F(X)$, so gilt für $y \in [0,1]$: $G(y) = W\{F(X) \le y\} = W\{X \le F^{-1}(y)\}$

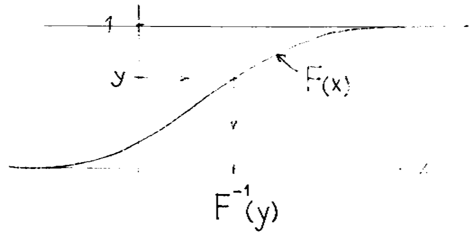
$= F(F^{-1}(y)) = y \;\Rightarrow\; G(\cdot)$ ist die VF der kontin. Gleichverteilung $U_{0;1}$

zu (2): Für die VF $H(\cdot)$ der SG $F^{-1}(Y)$ gilt:

$H(x) = W\{F^{-1}(Y) \le x\} = W\{Y \le F(x)\} = F(x)$

$\forall x \in \mathbb{R}$, da $Y \sim U_{0;1} \;\Rightarrow\; F^{-1}(Y) \sim F(\cdot)$ //

Bem.:



Anwendung: Computersimulation von Verteilungen mit Hilfe der Simulation der $U_{0;1}$

IV

MEHRDIMENSIONALE VERTEILUNGEN

14. STOCHASTISCHE VEKTOREN

Versuchsausgänge, die durch $m > 1$ reelle Zahlen $x_1, \dots, x_m$ beschrieben werden.

Stoch. Modell $(X_1, \dots, X_m)$, X_i SG $\forall\, i = 1(1)m$ genannt m-dim. Stochastischer Vektor

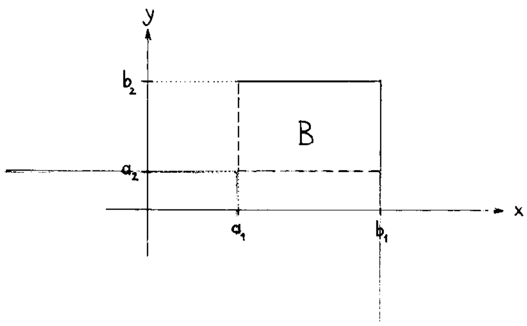
<u>Beispiel</u>: System aus m Komponenten
X_i Lebensdauer der i-ten Komponente

Spezialfall $m = 2$ $\quad \underline{X} = (X_1, X_2) = (X, Y)$

2-dim. VF $\quad F(x,y) := W\{X \le x, Y \le y\} \quad \forall (x,y) \in \mathbb{R}^2$

Mit Hilfe der 2-dim. VF erhält man die Wahrsch. von endlichen Rechtecken $B = (a_1, b_1] \times (a_2, b_2]$

$$W(B) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$



14.1 Randverteilungen

Ist $\underline{X} = (X_1, \dots, X_m)$ ein SV, so heißen die Verteilungen jedes Teilvektors $(X_{i_1}, \dots, X_{i_k})$, $k < m$ Randverteilungen.

<u>Spezialfall $k=1$</u>: Einzelverteilungen der X_i

Es gilt: $\underline{X} = (X, Y)$ 2-dim. SV mit VF $F(x,y)$

$\Rightarrow$ die 1-dim. VF$_n$ $F_1(\cdot)$ von X und $F_2(\cdot)$ von Y sind

$$F_1(x) = W\{X \leq x\} = \lim_{y \uparrow \infty} F(x,y) \quad \forall x \in \mathbb{R}$$
$$F_2(y) = W\{Y \leq y\} = \lim_{x \uparrow \infty} F(x,y) \quad \forall y \in \mathbb{R}$$

15. MEHRDIMENSIONALE DISKRETE VERT.

M abzählbare Menge von reellen m-Tupeln $(x_1, \dots, x_m)$ ohne Häufungspunkt mit zugehörigen Punktwahrscheinlichkeiten $p(x_1, \dots, x_m)$, sodass

$$\sum_{(x_1, \dots, x_m) \in M} p(x_1, \dots, x_m) = 1$$

Die W. einer Teilmenge A von M ist def. durch

$$W(A) := \sum_{(x_1, \dots, x_m) \in A} p(x_1, \dots, x_m)$$

<u>Beispiel</u>: Multinomialverteilun $M_{n; \theta_1, \dots, \theta_m}$

2-dim SV bzw. Verteilungen $(X, Y) \sim p(x,y)$

Darstellung in Tabellenform oder als Diagramm

X \ Y	b_1	b_2	$\cdots$	b_j	$\cdots$	b_l
a_1	$p(a_1,b_1)$	$p(a_1,b_2)$	$\cdots$	$p(a_1,b_j)$	$\cdots$	$p(a_1,b_l)$
a_2	$p(a_2,b_1)$	$p(a_2,b_2)$	$\cdots$	$p(a_2,b_j)$	$\cdots$	$p(a_2,b_l)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
a_i	$p(a_i,b_1)$	$p(a_i,b_2)$	$\cdots$	$p(a_i,b_j)$	$\cdots$	$p(a_i,b_l)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
a_k	$p(a_k,b_1)$	$p(a_k,b_2)$	$\cdots$	$p(a_k,b_j)$	$\cdots$	$p(a_k,b_l)$

15.1 Randverteilungen $(X,Y) \sim p(x,y)$

Punktwahrscheinlichkeiten der Einzelgrößen

$$p_1(x) = W\{X=x\} = \sum_{y \in M_Y} W\{(X,Y)=(x,y)\} = \sum_{y \in M_Y} p(x,y)$$

$$p_2(y) = W\{Y=y\} = \sum_{x \in M_X} W\{(X,Y)=(x,y)\} = \sum_{x \in M_X} p(x,y)$$

<u>Bem.</u>: Name Randverteilung siehe Tabelle

Für höherdimensionale SVn gibt es mehrdimensionale Randverteilungen

2-dim. Verteilungstabelle

X \ Y	b_1	b_2	$\cdots$	b_j	$\cdots$	b_ℓ	$p_1(x)$
a_1	$p(a_1,b_1)$	$p(a_1,b_2)$	$\cdots$	$p(a_1,b_j)$	$\cdots$	$p(a_1,b_\ell)$	$p_1(a_1)$
a_2	$p(a_2,b_1)$	$p(a_2,b_2)$	$\cdots$	$p(a_2,b_j)$	$\cdots$	$p(a_2,b_\ell)$	$p_1(a_2)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
a_i	$p(a_i,b_1)$	$p(a_i,b_2)$	$\cdots$	$p(a_i,b_j)$	$\cdots$	$p(a_i,b_\ell)$	$p_1(a_i)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
a_k	$p(a_k,b_1)$	$p(a_k,b_2)$	$\cdots$	$p(a_k,b_j)$	$\cdots$	$p(a_k,b_\ell)$	$p_1(a_k)$
$p_2(y)$	$p_2(b_1)$	$p_2(b_2)$	$\cdots$	$p_2(b_j)$	$\cdots$	$p_2(b_\ell)$	1

Randverteilung von X

Randverteilung von Y

16. MEHRDIMENSIONALE KONTIN. VTLGEN

$\underline{X} = (X_1, \dots, X_m) \sim f(x_1, \dots, x_m)$ m-dim. Dichtefunktion

$$f: \mathbb{R}^m \to [0,\infty) \quad \text{mit} \quad \int_{\mathbb{R}^m} \cdots \int f(x_1,\dots,x_m)\, dx_1 \cdots dx_m = 1$$

Als Ereignisfeld auf $\mathbb{R}^m$ das am wenigsten umfassende Ereignisfeld, das alle m-dim. Intervalle I_m enthält:

$$I_m := \mathop{\times}_{i=1}^{m} (a_i, b_i] \quad \text{mit} \quad a_i < b_i \in \mathbb{R}$$

Dieses Ereignisfeld nennt man das System der m-dim. Borelmengen $\mathcal{B}_m$.

Für $B \in \mathcal{B}_m$ ist die W. $W(B)$ def. durch

$$W(B) := \int_B \cdots \int f(x_1,\dots,x_m)\, dx_1 \cdots dx_m$$

Im Spezialfall $B = I_m = \mathop{\times}_{i=1}^{m} (a_i, b_i]$ gilt

$$W(I_m) = \int_{a_m}^{b_m} \cdots \int_{a_1}^{b_1} f(x_1,\dots,x_m)\, dx_1 \cdots dx_m$$

<u>Beispiel</u>: m-dim. Normalverteilung $N(\mu, \Sigma)$

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} \in \mathbb{R}^m, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1m} \\ \sigma_{21} & \cdots & \sigma_{2m} \\ \vdots & & \\ \sigma_{m1} & \cdots & \sigma_{mm} \end{pmatrix}$$

symmetrische, reelle, positiv definite quadratische Matrix m-dim. Dichtefunktion mittels der inversen Matrix

$$\Sigma^{-1} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \vdots & & \\ \alpha_{m1} & \cdots & \alpha_{mm} \end{pmatrix}$$

$$f(x_1, \dots, x_m) = \frac{1}{(2\pi)^{\frac{m}{2}} \sqrt{\operatorname{Det} \Sigma}} \, e^{-\frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_{ij} (x_i - \mu_i)(x_j - \mu_j)}$$

$$\forall (x_1, \dots, x_m) \in \mathbb{R}^m$$

16.1 2-dim. Normalverteilung $N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \varrho)$

$$\mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \quad \sigma_x^2 > 0, \ \sigma_y^2 > 0, \quad -1 < \varrho < 1, \quad \Sigma = \begin{pmatrix} \sigma_x^2 & \varrho\sigma_x\sigma_y \\ \varrho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}$$

Dichtefunktion $f(x,y) =$

$$= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\varrho^2}} \, e^{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\varrho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

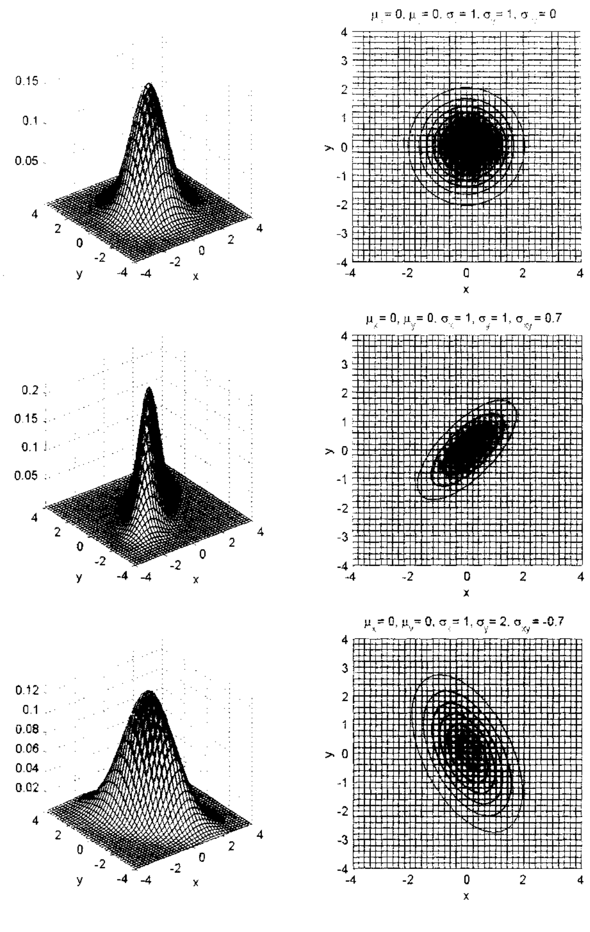
$$\forall (x,y) \in \mathbb{R}^2$$

Bem.: $\sigma_x^2 \mathrel{\hat{=}} \sigma_{11}$, $\sigma_y^2 \mathrel{\hat{=}} \sigma_{22}$, $\varrho\sigma_x\sigma_y \mathrel{\hat{=}} \sigma_{12} = \sigma_{21}$

ϱ ist ein Maß für den Zusammenhang von X und Y

Für reale Verteilungen müssen die Parameter $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \varrho$ aus Daten statistisch geschätzt werden.

$\varrho = 0$ stoch. unabhängig



16.2 Randdichten

$\underline{X} = (X,Y) \sim f(x,y)$, Randverteilung von X:

$$W\{X \in A\} = W\{X \in A, Y \text{ beliebig}\} = W\{X \in A, -\infty < Y < \infty\}$$

$$= \int_{-\infty}^{\infty} \left[\int_A f(x,y)\,dx \right] dy = \int_A \left[\int_{-\infty}^{\infty} f(x,y)\,dy \right] dx$$

$$= \int_A f_1(x)\,dx \qquad \forall A \in \mathcal{B}_1$$

$\Rightarrow f_1(\cdot)$ ist die Dichte der Randverteilung von X, genannt Randdichte von X

$$f_1(x) := \int_{-\infty}^{\infty} f(x,y)\,dy \quad \forall x \in \mathbb{R} \quad \text{Randdichte von } X$$

$$f_2(y) := \int_{-\infty}^{\infty} f(x,y)\,dx \quad \forall y \in \mathbb{R} \quad \text{Randdichte von } Y$$

16.3 Satz: $(X,Y) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \varrho)$

$$\Rightarrow X \sim N(\mu_x, \sigma_x^2)$$

$$Y \sim N(\mu_y, \sigma_y^2)$$

o. Bew.

17. ERWARTUNG VON FUNKTIONEN VON STOCH. VEKTOREN

$\underline{X} = (X_1, \dots, X_m)$ SV, $\psi: \mathbb{R}^m \to \mathbb{R}$ messbar

$\Rightarrow Z = \psi(X_1, \dots, X_m)$ SG

gesucht: $\mathbb{E}Z$

Beispiele:

1) $\psi(x_1, \dots, x_m) = \sum_{i=1}^{m} x_i$

2) $\psi(x_1, \dots, x_m) = \sum_{i=1}^{m} x_i^2$

3) $\psi(x_1, \dots, x_m) = \min_{i=1(1)m} x_i$

4) $\psi(x_1, \dots, x_m) = \max_{i=1(1)m} x_i$

17.1 Satz vom unbewussten Statistiker:

Ist $\underline{X} = (X_1, \dots, X_m)$ ein SV und $\psi: \mathbb{R}^m \to \mathbb{R}$ eine messbare Funktion, sodass $\exists\, \mathbb{E}\psi(\underline{X})$, so gilt:

(1) Im disketen Fall $\underline{X} \sim (M_{\underline{X}}, p(x_1, \dots, x_m))$

$$\mathbb{E}\psi(X_1, \dots, X_m) = \sum_{(x_1, \dots, x_m) \in M_{\underline{X}}} \psi(x_1, \dots, x_m)\, p(x_1, \dots, x_m)$$

(2) Im kontinuierlichen Fall $\underline{X} \sim f(x_1, \dots, x_m)$

$$\mathbb{E}\psi(X_1, \dots, X_m) = \int \cdots \int_{\mathbb{R}^m} \psi(x_1, \dots, x_m) f(x_1, \dots, x_m)\, dx_1 \cdots dx_m$$

(3) Im gemischten Fall eine Kombination

17.2 Linearität der Erwartungsbildung:

Sind $X_1, \dots, X_n$ 1-dim. SG$_n$ mit $\exists\, \mathbb{E}X_i \; \forall i = 1(1)n$ und $a_1, \dots, a_n$ reelle Konstanten, so gilt:

$$\mathbb{E}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i \mathbb{E}X_i$$

Bew.: Mittels des S.v.u.Stat.

Bem.: $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$

18. Kovarianz, Korrelation und Unabhängigkeit

Beschreibung von Zusammenhängen zwischen SGn

18.1 Kovarianz

Für zwei SGn X und Y mit $\exists$ Varianzen ist die Kovarianz $Cov(X,Y)$ definiert durch

$$Cov(X,Y) = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]$$

Bem.: Berechnung mittels S.v.u.Stat.

$Cov(X,Y)$ Maß für die Abhängigkeit von X und Y

Beispiel: $(X,Y) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \varrho) \Rightarrow Cov(X,Y) = \varrho\sigma_x\sigma_y$

$(X_1, \dots, X_m) \sim N(\mu, \Sigma) \Rightarrow Cov(X_k, X_l) = \sigma_{kl}$

18.2 Satz: Falls $\exists\, Cov(X,Y)$, so gilt:

$$Cov(X,Y) = \mathbb{E}(X\cdot Y) - (\mathbb{E}X)\cdot(\mathbb{E}Y)$$

Bew.: $Cov(X,Y) = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]$

$= \mathbb{E}[X\cdot Y - Y\cdot\mathbb{E}X - X\cdot\mathbb{E}Y + (\mathbb{E}X)(\mathbb{E}Y)]$

$= \mathbb{E}(X\cdot Y) - \mathbb{E}X\cdot\mathbb{E}Y - \mathbb{E}X\cdot\mathbb{E}Y + \mathbb{E}X\cdot\mathbb{E}Y$ //

18.3 Satz: Sind $X_1, \dots, X_n$ SGn mit $\exists\, Var X_i \ \forall i = 1(1)n$ und $c_1, \dots, c_n$ reelle Konstanten, so gilt:

$$Var\left(\sum_{i=1}^{n} c_i X_i\right) = \sum_{i=1}^{n} c_i^2 Var X_i + 2\sum_{i=1}^{n-1}\sum_{\substack{j=2\\ i<j}}^{n} c_i c_j Cov(X_i, X_j)$$

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

Bew.: $Var(X+Y) = \mathbb{E}\{[X+Y-\mathbb{E}(X+Y)]^2\}$

$= \mathbb{E}\{[X+Y-\mathbb{E}X-\mathbb{E}Y]^2\} = \mathbb{E}\{[X-\mathbb{E}X+Y-\mathbb{E}Y]^2\}$

$= \mathbb{E}\{(X-\mathbb{E}X)^2 + (Y-\mathbb{E}Y)^2 + 2(X-\mathbb{E}X)(Y-\mathbb{E}Y)\}$

$= \underbrace{\mathbb{E}[(X-\mathbb{E}X)^2]}_{Var X} + \underbrace{\mathbb{E}[(Y-\mathbb{E}Y)^2]}_{Var Y} + 2\underbrace{\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)]}_{Cov(X,Y)}$ //

18.4 Korrelationskoeffizient

Sind X und Y SGn mit $\exists$ Varianzen, so ist der Korrelationskoeffizient $\varrho_{X,Y}$ von X und Y def. als

$$\varrho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var X}\cdot\sqrt{Var Y}}$$

Es gilt: 1) $-1 \le \varrho_{X,Y} \le 1$

2) Falls $|\varrho_{X,Y}| = 1$, so gilt f.s. ein linearer Zusammenhang von X und Y, d.h. $W\{Y = kX + d\} = 1$

$\varrho_{X,Y} = 1$ positive lineare Kopplung

$\varrho_{X,Y} = -1$ negative —"—

Beispiel: $(X,Y) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \varrho) \Rightarrow \varrho_{X,Y} = \varrho$

Def.: Gilt $\varrho_{X,Y} = 0$, so heißen X und Y unkorreliert.

18.5 Satz: Für paarweise unkorrelierte SGn gilt:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var X_i \qquad \text{(vgl. Satz 18.3)}$$

18.6 Stochastische Unabhängigkeit SG$_n$

SG$_n$ X und Y sind unabhängig, wenn für ihre gemeinsame VF $G(\cdot,\cdot)$, wobei $X \sim F_1(\cdot)$ und $Y \sim F_2(\cdot)$, gilt:

$$G(x,y) = F_1(x)\cdot F_2(y) \quad \forall (x,y) \in \mathbb{R}^2 \qquad (X \perp\!\!\!\perp Y)$$

<u>Es gilt</u>: Die Unabhängigkeit ist gleichbedeutend mit

(1) $p(x,y) = p_1(x)\cdot p_2(y) \quad \forall (x,y)$ im diskreten Fall

(2) $f(x,y) = f_1(x)\cdot f_2(y) \quad \forall (x,y)$ im kontinuierl. Fall

<u>Beispiel</u>: $(X,Y) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \varrho)$, so sind X und Y genau dann unabhängig, wenn $\varrho = 0$.

18.7 Satz: Sind X und Y unabhängig, so gilt für alle messbaren Funktionen $\varphi(\cdot)$ und $\psi(\cdot)$, für die $\exists\, \mathbb{E}[\varphi(X)]$ und $\exists\, \mathbb{E}[\psi(Y)]$:

$$\mathbb{E}[\varphi(X)\cdot\psi(Y)] = \mathbb{E}[\varphi(X)]\cdot\mathbb{E}[\psi(Y)]$$

Bew.: für den kontin. Fall; $g(x,y) = f_1(x)\cdot f_2(y) \quad \forall (x,y) \in \mathbb{R}^2$; S.v.u.Stat. $\Rightarrow$

18.8 Satz: Für SG X und Y gilt: $X \perp\!\!\!\perp Y \Rightarrow \varrho_{X,Y} = 0$

Bew.: $Cov(X,Y) = \mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)] =$ Satz 18.7
$= \mathbb{E}[X-\mathbb{E}X]\cdot\mathbb{E}[Y-\mathbb{E}Y] = [\mathbb{E}X-\mathbb{E}X]\cdot[\mathbb{E}Y-\mathbb{E}Y] = 0$
$\Rightarrow \varrho_{X,Y} = 0$ //

19. BEDINGTE VERTEILUNGEN

(X,Y) 2-dim. SV

<u>gesucht</u>: Vtlg. von $Y|X=x$

19.1 Diskrete Vtlgen.

$(X,Y) \sim p(x,y) \quad \forall (x,y) \in M_{(X,Y)}$

$p_1(\cdot)$ Randverteilung von X

$p_2(\cdot)$ — " — " Y

Bedingte Verteilungen von $Y|X=x$

$X|Y=y$

$$p(y|x) := W\{Y=y \mid X=x\} = \frac{W\{X=x \wedge Y=y\}}{W\{X=x\}} = \frac{p(x,y)}{p_1(x)}$$

bei festem x, $\forall y \in M_Y$

$$p(x|y) := \frac{p(x,y)}{p_2(y)} \quad \forall x \in M_X, \; y \text{ fest}$$

19.2 Kontinuierliche Vtlgen..

$(X,Y) \sim f(x,y)$

Problem: $W\{X=x\} = 0$

Lösung: Bedingte Dichten von $Y|X=x$

$X|Y=y$

$f_1(\cdot)$ Randdichte von X

$f_2(\cdot)$ —"— " Y

Bedingte Dichte $f(y|x) := \dfrac{f(x,y)}{f_1(x)} \quad \forall x: f_1(x) > 0$

$$f(x|y) := \frac{f(x,y)}{f_2(y)} \quad \forall y: f_2(y) > 0$$

Beispiel: $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \varrho)$

$$f(y|x) = \frac{1}{\sqrt{2\pi}\,\sigma_2\sqrt{1-\varrho^2}} \exp\left\{-\frac{(y-[\mu_2+\varrho\frac{\sigma_2}{\sigma_1}(x-\mu_1)])^2}{2(1-\varrho^2)\sigma_2^2}\right\}$$

$$\triangleq N\left(\mu_2+\varrho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\ (1-\varrho^2)\sigma_2^2\right)$$

Definition: Für (X,Y) heißt die Funktion

$x \to \mathbb{E}(Y|X=x)$ bedingte Erwartung von $Y|X$,

auch genannt Regressionsfunktion von Y bezüglich X.

Analog $y \to \mathbb{E}(X|Y=y)$

Sonderfall: $(X,Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \varrho)$

$$\Rightarrow \mathbb{E}(Y|X=x) = \mu_2 + \varrho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$$

Regressionsgerade

20. FUNKTIONEN VON STOCH. VEKTOREN

$\underline{X} = (X_1, \cdots, X_m)$ SV, $\psi: \mathbb{R}^m \to \mathbb{R}^k$ messbar

$Z = \psi(\underline{X})$ k-dim. SV

gesucht: W-Vtlg. von Z

Beispiele: 1) X, Y SGn, $Z = X+Y$

2) $X_1, \cdots, X_m$ SGn, $Z = \min(X_1, \cdots, X_m)$

3) $X_1, \cdots, X_n$ SGn, $Z = \frac{1}{n}\sum_{i=1}^{n} X_i$

4) X, Y SGn, $Z = X \cdot Y$

20.1 Lineartransformation von Normalverteilungen

$\underline{X} = (X_1, \dots, X_m)^T \sim N(\mu, \Sigma)$

$A = (a_{ij})_{j=1(1)m}^{i=1(1)m}$ reguläre Matrix

$$\Rightarrow \underline{Z} = A\underline{X} \sim N(A\mu, A\Sigma A^T)$$

(ohne Beweis)

20.2 Satz: Sind $X_1, \dots, X_n$ unabhängige SGn, $X_i \sim F_i(\cdot)$, so gilt:

(a) $W\{\max(X_1, \dots, X_n) \le x\} = \prod_{i=1}^{n} F_i(x) \quad \forall x \in \mathbb{R}$

(b) $W\{\min(X_1, \dots, X_n) \le x\} = 1 - \prod_{i=1}^{n} [1 - F_i(x)] \quad \forall x \in \mathbb{R}$

Beweis: (a) $\{\max(X_1,\dots,X_n) \le x\} = \bigcap_{i=1}^{n} \{X_i \le x\}$

$$\Rightarrow W\{\max_{i=1(1)n} X_i \le x\} = W\Big(\bigcap_{i=1}^{n}\{X_i \le x\}\Big) = \prod_{i=1}^{n} \underbrace{W\{X_i \le x\}}_{=F_i(x)} \quad //$$

(b) analog, mittels $W\{\min_{i=1(1)n} X_i \le x\} = 1 - W\{\min_{i=1(1)n} X_i > x\}$

20.3 Ordnungsstatistiken

Sind $X_1,\dots,X_n$ SGn, so nennt man die der Größe nach geordneten SGn $X_{(1)} \le X_{(2)} \le \dots \le X_{(n-1)} \le X_{(n)}$ die zugehörigen Ordnungsstatistiken.

Bem.: $X_{(1)} = \min(X_1,\dots,X_n)$
$X_{(n)} = \max(X_1,\dots,X_n)$

20.4 Faltung von W.-Vtlgen.

X, Y unabhängige SGn mit VFn $X \sim F(\cdot)$, $Y \sim G(\cdot)$
gesucht: VF $H(\cdot)$ von $X+Y$

Es gilt: $H(x) = \int_{-\infty}^{\infty} F(x-t)\,dG(t) \quad \forall x \in \mathbb{R}$

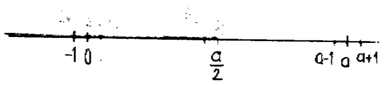
Beweis: (a) Diskrete Verteilungen $X \sim p_1(\cdot)$, $Y \sim p_2(\cdot)$

$$q(a) = W\{X+Y=a\} = W\Big(\bigcup_{\substack{(x,y):\\ x+y=a}} \{X=x, Y=y\}\Big)$$

$$= W\Big(\bigcup_{\substack{(x,y):\\ x+y=a}} [\{X=x\}\cap\{Y=y\}]\Big) = \sum_{\substack{(x,y):\\ x+y=a}} W(\{X=x\}\cap\{Y=y\})$$

$$= \sum_{\substack{(x,y):\\ x+y=a}} W\{X=x\}\cdot W\{Y=y\} = \sum_{x\in M_X} W\{X=x\}\cdot W\{Y=a-x\}$$

$$= \sum_{x\in M_X} p_1(x)\cdot p_2(a-x)$$

Bem.: 

„Faltung"

Anwendung:

Additionstheorem für Poisson-Verteilungen:

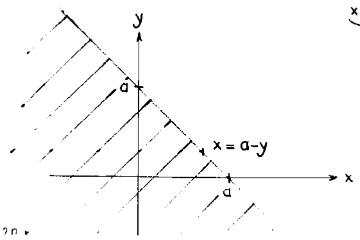
$X \sim P_\mu$, $Y \sim P_\xi$, $X \perp\!\!\!\perp Y \Rightarrow X+Y \sim P_{\mu+\xi}$

Bew.: $q(a) = \sum_{x=0}^{\infty} p_1(x)\cdot p_2(a-x) = \sum_{x=0}^{a} \frac{\mu^x e^{-\mu}}{x!}\cdot\frac{\xi^{a-x}\, e^{-\xi}}{(a-x)!}$

$$= \frac{e^{-(\mu+\xi)}}{a!} \underbrace{\sum_{x=0}^{a} \binom{a}{x} \mu^x \xi^{a-x}}_{=(\mu+\xi)^a} = \frac{(\mu+\xi)^a\, e^{-(\mu+\xi)}}{a!} \;\hat{=}\; P_{\mu+\xi}$$

(b) Kontinuierliche Verteilungen $X \sim \overset{F(\cdot)}{f(\cdot)}$, $Y \sim \overset{G(\cdot)}{g(\cdot)}$, $X \perp\!\!\!\perp Y$

$$H(a) = W\{X+Y \le a\} = \iint_{x+y\le a} f(x)g(y)\,dx\,dy$$



$$= \int_{-\infty}^{\infty}\Big(\int_{-\infty}^{a-y} f(x)\,dx\Big) g(y)\,dy$$

$$= \int_{-\infty}^{\infty} F(a-y)\,g(y)\,dy$$

Differentiation von $H(\cdot)$ ergibt die Dichte $h(\cdot)$

<u>Es gilt</u>: $h(a) = \frac{d}{da}\int_{-\infty}^{\infty} F(a-y)g(y)\,dy = \int_{-\infty}^{\infty} \frac{d}{da} F(a-y)g(y)\,dy$

$$= \int_{-\infty}^{\infty} f(a-y)g(y)\,dy$$ Faltprodukt der Dichten $f(\cdot)$ und $g(\cdot)$

<u>Anwendung</u>: Addition unabhängiger Zufallszahlen

$X \sim U_{0;1}$, $Y \sim U_{0;1}$, $X \perp\!\!\!\perp Y$, gesucht: Dichte von $X+Y$

$$h(a) = \int_{-\infty}^{\infty} f(a-y)g(y)\,dy = \int_{-\infty}^{\infty} I_{(0,1)}(a-y)\, I_{(0,1)}(y)\,dy$$

Der Integrand ist genau dann positiv $(=1)$, wenn

$$[0 < a-y < 1] \wedge [0 < y < 1],$$

dies gilt genau dann, wenn

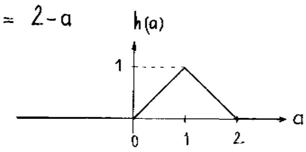
$$[a-1 < y < a] \wedge [0 < y < 1].$$

Daher gilt

für $0 \le a \le 1$: $\quad h(a) = \int_0^a 1\,dy = a$

für $1 \le a \le 2$: $\quad h(a) = \int_{a-1}^1 1\,dy = 2-a$

sonst: $\quad h(a) = 0$



Additionstheorem für Normalverteilungen:

Gilt $X_i \sim N(\mu_i, \sigma_i^2)$ für $i = 1(1)n$ und sind $X_1, \dots, X_n$ unabh.

$$\Rightarrow \sum_{i=1}^n X_i \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right)$$

Beweis: Faltung der Dichten und Induktion

Additionstheorem für Gammaverteilungen:

Gilt $X_i \sim Gam(\alpha_i, \lambda)$ für $i = 1(1)n$ und $X_1, \dots, X_n$ unabhängig

$$\Rightarrow \sum_{i=1}^n X_i \sim Gam\left(\sum_{i=1}^n \alpha_i, \lambda\right)$$

Beweis: Faltung und Induktion

<u>Bem.</u>: Sonderfall χ_n^2

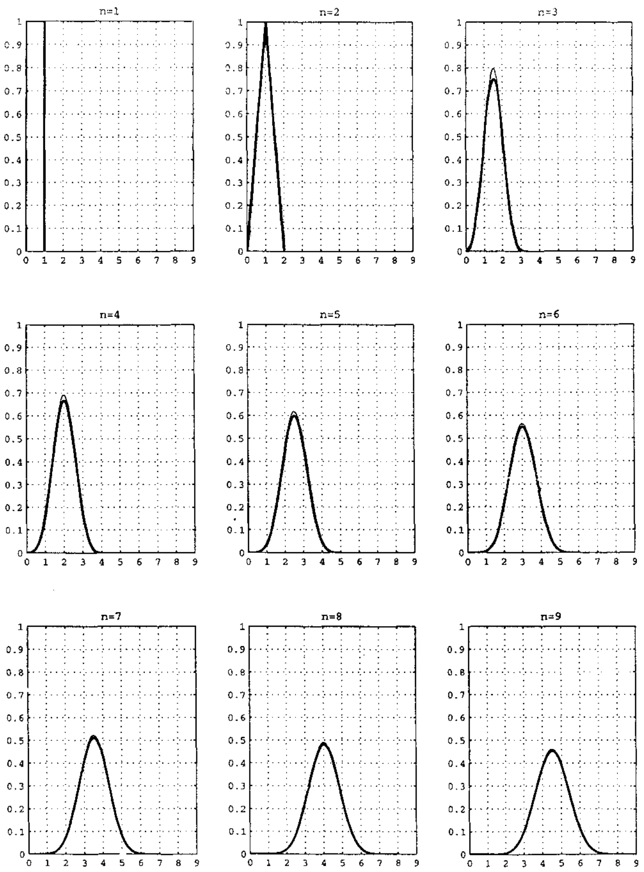
Erlang-Vtlg. $Er_{n,\lambda}$

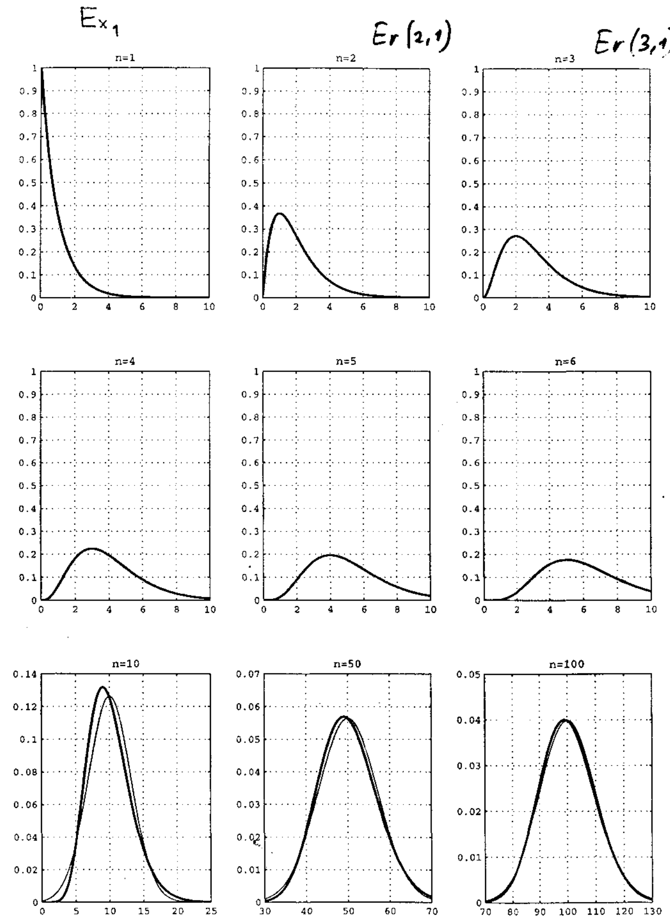
Sind $X_1, \dots, X_n$ unabhängig exponentialverteilt mit Dichte $f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$, so hat die SG $\sum_{i=1}^n X_i$ eine Erlang-Verteilung mit Dichtefunktion

$$f(x|n,\lambda) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} I_{(0,\infty)}(x)$$

Beweis: Faltung und Induktion

<u>Anwendung</u>: Warteschlangen





V

FOLGEN STOCH. GRÖSZEN

$X_1, X_2, \dots, X_n, \dots$ unendliche Folge von SGn X_i

Def.: Eine ∞ Folge von SGn heißt unabhängige Folge, wenn je endlich viele der SGn unabhängig sind.

Bem.: UIV-Folgen (engl. IID) als mathem. Beschreibung für beliebig oft wiederholbare stat. Versuche.

Es gibt verschiedene Konvergenzbegriffe in der Stochastik:

- fast sichere Konvergenz
- Verteilungskonvergenz
- fast sicher gleichmäßige Konvergenz

Es gibt noch andere Konvergenzarten:

Konv. i.d. W
Konv. i.q. M

21. GESETZ DER GROSZEN ZAHLEN

Konvergenz von Folgen SGn

Bem.: Für UIV-Folgen und das empirische GgZ gilt für ein festes Ereignis B nach n Durchführungen für die relativen Häufigkeiten $h_n(B)$ mit

$$X_i = \begin{cases} 1 & \text{falls B beim } i\text{-ten Versuch eintritt} \\ 0 & \text{falls B beim } i\text{-ten Versuch nicht eintritt} \end{cases}$$

$$h_n(B) = \frac{1}{n}\sum_{i=1}^{n} X_i =: \overline{X_n}$$

Bem.: $\overline{X}_n$ ist eine SG (vor Durchf. der Versuche)

Eine Folge $X_1, X_2, \dots$ von SGn X_n konvergiert fast sicher gegen eine SG X_0, wenn gilt:

$$W(\{\omega : \lim_{n\to\infty} X_n(\omega) = X_0(\omega)\}) = 1$$

Symbolisch: $X_n \xrightarrow[n\to\infty]{f.s.} X_0$

21.1 Gesetz der großen Zahlen:

Ist $(X_n;\, n\in\mathbb{N})$ eine UIV-Folge mit $\exists\, \mathbb{E}X_i = \mu$, so gilt für die Folge $(\overline{X_n};\, n\in\mathbb{N})$:

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{f.s.} \mu$$

Zum empirischen GgZ:

$X_i = I_B \;\Rightarrow\; X_i \sim A_{W(B)} \;\Rightarrow\; \mathbb{E}X_i = W(B)$

und aus dem GgZ 21.1 folgt daher

$$h_n(B) = \overline{X_n} \xrightarrow[n\to\infty]{f.s.} W(B)$$

D.h. die Folge der relativen Häufigkeiten konvergiert fast sicher gegen die Wahrscheinlichkeit jedes festen Ereignisses B.

22. ZENTRALER GRENZVERTEILUNGSSATZ

Verteilung von Summen von unabh. SGn, Faltung vgl. Abschn. 22

22.1 ZGS: Ist $X_1, X_2, \dots$ eine ∞ Folge von unabh. SGn mit $\exists\, \mathrm{Var}\, X_k \;\forall k\in\mathbb{N}$ und erfüllen die VFn $F_k(\cdot)$ der X_k die sog. Lindeberg-Bedingung, so gilt für die Folge $Z_n, n\in\mathbb{N}$ der standardisierten Summen

$$Z_n := \frac{\sum_{k=1}^{n} X_k - \mathbb{E}\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right)}}$$

$$\lim_{n\to\infty} W\{Z_n \le x\} = \Phi(x) \quad \forall x\in\mathbb{R}.$$

Bem.: 1) Die LB lautet mit den Bez. $\mu_k := \mathbf{E}X_k$,

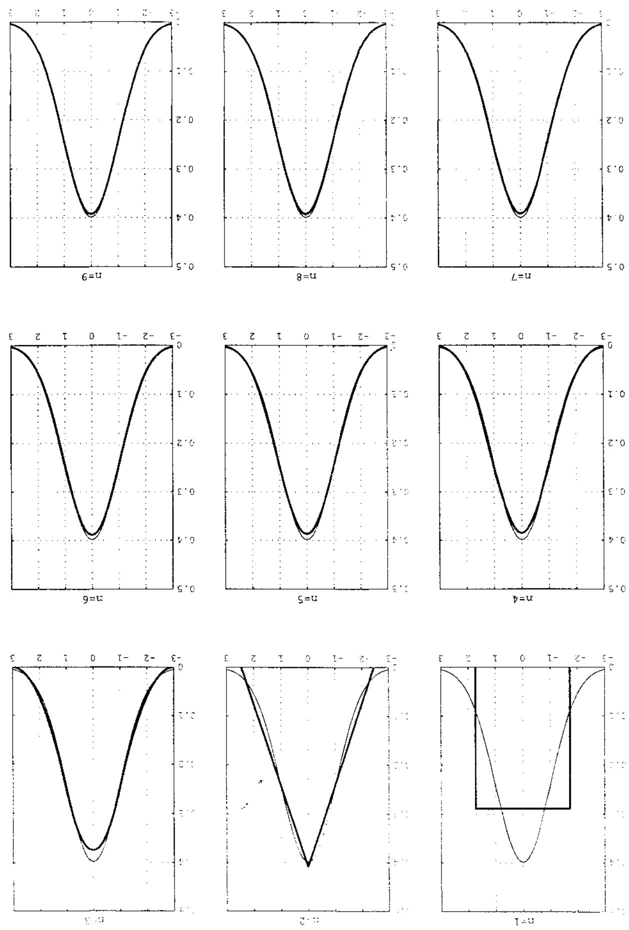
$s_n^2 := \mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \mathrm{Var}\,X_k$

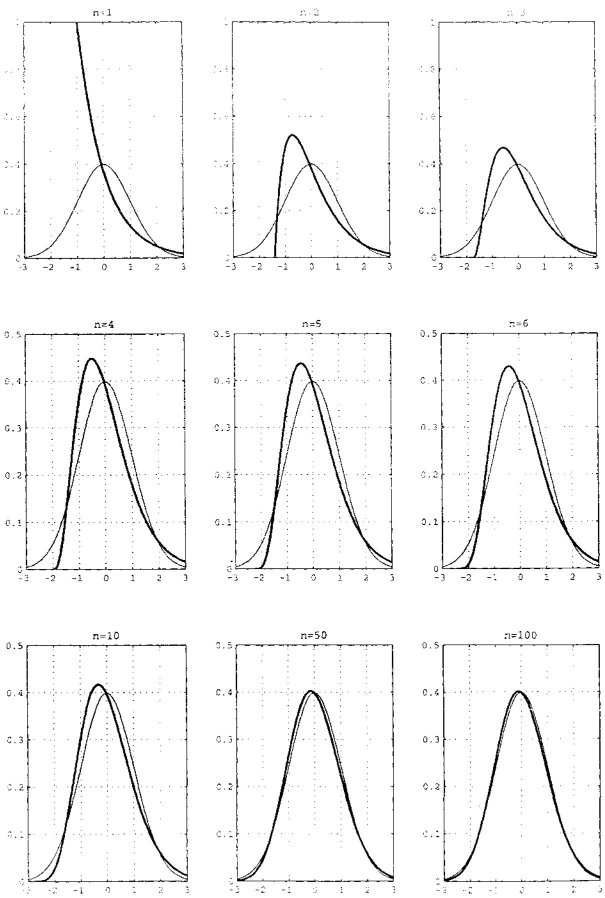
$$L_n(\varepsilon) := \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{\{x : |x-\mu_k| \geq \varepsilon s_n\}} (x-\mu_k)^2 \, dF_k(x) \xrightarrow[n \to \infty]{} 0 \quad \forall \varepsilon > 0$$

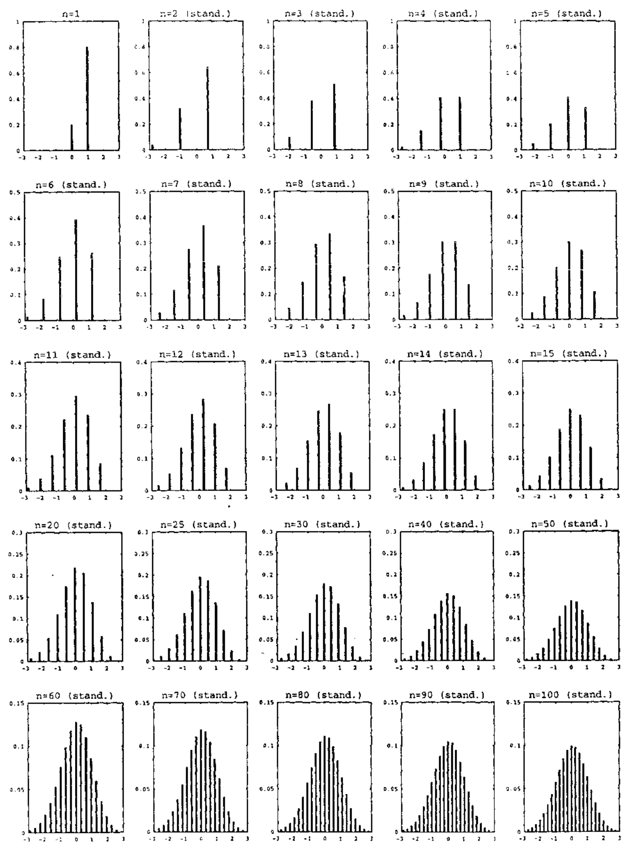
Dies bedeutet, dass keine einzelne Vtlg. $F_k(\cdot)$ dominierenden Einfluss auf die GV hat.

2) Für UIV-Folgen ist die LB erfüllt.

3) Für glm. beschränkte SGn $|X_k| \leq C \;\; \forall k \in \mathbb{N}$ mit $\mathrm{Var}\left(\sum_{k=1}^{n} X_k\right) \xrightarrow[n \to \infty]{} \infty$ ist die LB erfüllt.







22.2 Normalapproximation

$\sum_{k=1}^{n} X_k = S$ mit $X_1, \dots, X_n$ unabh. für größeres n

wegen $Z = \frac{S - \mathbb{E}S}{\sqrt{\operatorname{Var} S}} \overset{\cdot}{\sim} N(0,1)$ „approximativ"

$\Rightarrow S = \sqrt{\operatorname{Var} S} \cdot Z + \mathbb{E}S \overset{\cdot}{\sim} N(\mathbb{E}S, \operatorname{Var} S)$

<u>Bem.</u>: 1) Approximation für die VF

$$W\{S \leq x\} \approx \Phi\left(\frac{x - \mathbb{E}S}{\sqrt{\operatorname{Var} S}}\right)$$

2) Bei Approximation einer diskret vt. SG X durch $Y \sim N(\mu, \sigma^2)$:

$$W\{X = x\} \approx W\left\{x - \tfrac{1}{2} \leq Y \leq x + \tfrac{1}{2}\right\}$$

23. FUNDAMENTALSATZ DER STATISTIK

$X_1, X_2, \dots$ UIV-Folge mit $X_i \sim F(\cdot)$

Für $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ empir. VF $F_n^*(\cdot)$ aus Abschn. 2.2

Die SG $\frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i)$ liefert nach Beobachtung der $X_i = x_i$ $\forall x \in \mathbb{R}$ genau den Wert $F_n^*(x)$, d.h.

$$F_n^*(x; X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i)$$

Für $n \to \infty$ konvergiert die Folge $F_n^*(x)$ bzw. $F_n^*(\cdot)$

23.1 Satz:

Gilt $X_1, X_2, \dots$ UIV-Folge mit $X_i \sim F(\cdot)$, so gilt für jedes feste $x \in \mathbb{R}$

(a) $\mathbb{E}\left(F_n^*(x; X_1, \dots, X_n)\right) = F(x)$

(b) $F_n^*(x; X_1, \dots, X_n) \xrightarrow[n \to \infty]{\text{f.s.}} F(x)$

Beweis: (a) $\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} I_{(-\infty, x]}(X_i)\right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\, I_{(-\infty, x]}(X_i)$

$= \frac{1}{n} \cdot n \cdot F(x) = F(x)$ $\qquad = W\{X_i \leq x\} = F(x)$

(b) $\frac{I_{(-\infty, x]}(X_1) + \dots + I_{(-\infty, x]}(X_n)}{n}$ $(\text{GgZ}) \xrightarrow[n \to \infty]{\text{f.s.}} \mathbb{E}\, I_{(-\infty, x]}(X_i)$

also $F_n^*(x) \xrightarrow[n \to \infty]{\text{f.s.}} F(x)$ //

23.2 Fundamentalsatz:

Für UIV-Folgen mit $X_i \sim F(\cdot)$ gilt

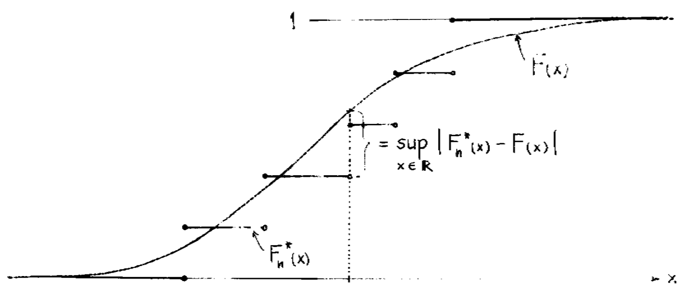
$$W\left\{\lim_{n \to \infty} \sup_{x \in \mathbb{R}} \left|F_n^*(x) - F(x)\right| = 0\right\} = 1$$

(ohne Beweis)

<u>Bem.</u>: 1) f.s. glm. Konvergenz

2) $\sup_{x \in \mathbb{R}} \left|F_n^*(x) - F(x)\right|$ da $\max_{x \in \mathbb{R}} \left|F_n^*(x) - F(x)\right|$ u.U. $\nexists$

3) $F_n^*(\cdot)$ Schätzung für $F(\cdot)$



24. STICHPROBEN UND STATISTIKEN

<u>Gesucht</u>: Verteilung von SG. X

gegeben: Beobachtungen von X

<u>Def.</u>: Sind $X_1, \dots, X_n$ UIV wie X, so heißen $X_1, \dots, X_n$ eine (mathematische) Stichprobe von X.

Nach Beobachtung von $X_1 = x_1, \dots, X_n = x_n$ nennt man $x_1, \dots, x_n$ konkrete Stichprobe von X. (kurz SP)

Ist $\psi : \mathbb{R}^n \to \mathbb{R}^k$ messbar und $X_1, \dots, X_n$ eine SP, so heißt $\psi(X_1, \dots, X_n)$ eine (k-dim.) Statistik, wenn $\psi(\cdot)$ nicht von unbekannten Parametern abhängt.

<u>Bsp.</u>: 1) $\psi(X_1, \dots, X_n) = \left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2\right)$

2) $\psi(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i$ Stichprobenmittel $\overline{X}_n$

3) $\psi(X_1, \dots, X_n) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2$ Stichprobenvarianz

24.1 Es gilt: Hat X den Merkmalraum M_X, so ist der Merkmalraum einer SP $(X_1, \dots, X_n)$ von X gleich M_X^n, genannt Stichprobenraum.

Für die W-Vtlg. einer SP gilt:

(a) Im diskreten Fall $X \sim (M_X, p(x))$

$$w(x_1, \dots, x_n) = \prod_{i=1}^n p(x_i) \quad \forall (x_1, \dots, x_n) \in M_X^n$$

(b) Im kontinuierlichen Fall $X \sim f(\cdot)$ gilt für die gemeinsame Dichte $g(x_1, \dots, x_n)$ der SP

$$g(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i) \quad \forall (x_1, \dots, x_n) \in M_X^n$$

<u>Definition</u>: Eine Statistik, die zur Schätzung einer charakteristischen Größe der W-Vtlg. von X dient, nennt man eine Schätzfunktion.

<u>Definition</u>: Hängt die W-vtlg. W_θ, $\theta \in \Theta$ von X von einem sog. Parameter θ ab, so heißt die Menge Θ der möglichen Werte des Parameters der Parameterraum.

Beispiele: 1) Ex_τ, $\theta = \tau$, $\Theta = (0,\infty)$

2) $N(\mu,\sigma^2)$, $\theta = (\mu,\sigma^2)$, $\Theta = \mathbb{R}\times(0,\infty)$

3) A_θ, $\Theta = [0,1]$

4) P_μ, $\theta = \mu$, $\Theta = (0,\infty)$

Bem.: Geraffter Parameter $\tau(\theta)$ mit $\tau : \Theta \to \mathbb{R}$

Bsp.: $N(\mu,\sigma^2)$, $\tau(\theta) = \mu$ oder $\tau(\theta) = \sigma$

Bem.: Falls die Vtlg. durch einen endlichdim. Parameter bestimmt ist, spricht man von parametrischer Statistik; sonst von nichtparametrischer St.

Es gibt 3 große Zweige der schließenden Statistik

a) Klassische objektivistische Statistik
b) Bayes'sche Statistik
c) Statistik bei unscharfer Information (Fuzzy Information)

Bem.: Entscheidung bei Unsicherheit

Fuzzy Wahrscheinlichkeiten

VI

KLASSISCHE SCHLIESZENDE STATISTIK

25. KLASSISCHE PUNKTSCHÄTZUNGEN

SG $X \sim (M_X, W_X)$, $X_1,\dots,X_n$ SP von X

$\vartheta : M_X^n \to \mathbb{R}^k$ Statistik

$\vartheta(X_1,\dots,X_n)$ SG für $k=1$, SV für $k>1$

Falls $\vartheta(X_1,\dots,X_n)$ Schätzfunktion und $x_1,\dots,x_n$ konkrete SP $\Rightarrow$ $\vartheta(x_1,\dots,x_n)$ Schätzwert

Für geraffte Parameter $\tau(\theta) \in \mathbb{R}$

Schätzfunktion $t(X_1,\dots,X_n)$ mit $t : M_X^n \to \mathbb{R}$

Verschiedene Güteeigenschaften von Schätzfunktionen bzw. Schätzwerten.

25.1 Unverzerrtheit 129

Def.: Eine SF $\vartheta(X_1, \dots, X_n)$ heißt unverzerrt (oder erwartungstreu) für eine char. Größe ξ, falls

$$\mathbb{E}\, \vartheta(X_1, \dots, X_n) = \xi \quad \forall \text{ mögl. Vtlgen von } X$$

Bem.: Für parametrische Modelle $X \sim W_\theta$; $\theta \in \Theta$ muss gelten

$$\mathbb{E}_\theta\, \vartheta(X_1, \dots, X_n) = \xi \quad \forall\ \theta \in \Theta$$

25.2

25.2 Satz: Für jede SG X und SP $X_1, \dots, X_n$ gilt:

(a) falls $\exists\, \mathbb{E}X \Rightarrow \overline{X}_n$ unverzerrte SF für $\mathbb{E}X$

(b) für $\exists\, \operatorname{Var} X \Rightarrow$ die Stichprobenvarianz S_n^2 ist unverzerrt für $\operatorname{Var} X$

Beweis: (a) Übung

(b) $$\mathbb{E}\, \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X}_n)^2 = \frac{1}{n-1} \sum_{i=1}^n \mathbb{E}\,(X_i^2 - 2X_i\overline{X}_n + \overline{X}_n^2)$$

$$= \dots = \mathbb{E}(X_i^2) - (\mathbb{E}X_i)^2 = \operatorname{Var} X_i = \operatorname{Var} X \quad //$$

Bem.: Für geraffte Parameter $\tau(\theta)$ für Unverzerrtheit:

$$\mathbb{E}_\theta\, t(X_1, \dots, X_n) = \tau(\theta) \qquad \forall\, \theta \in \Theta$$

25.3

Bsp.: $X \sim N(\mu, \sigma^2)$, $\tau(\theta) = \mu$, $t(X_1, \dots, X_n) = \overline{X}_n$

$\Rightarrow \mathbb{E}_\theta \overline{X}_n = \mu \quad \forall\, \theta = (\mu, \sigma^2) \in \Theta$

25.3 Effizienz 130

Für unverzerrte SFn möglichst kleine Varianz

$$\mathcal{T} = \{\, t(X_1, \dots, X_n) : T = t(X_1, \dots, X_n) \text{ unverzerrt} \,\}$$

Def.: Eine SF $t^*(X_1, \dots, X_n) \in \mathcal{T}$ heißt effizient (in der Klasse $\mathcal{T}$), falls

$$\operatorname{Var} t^*(X_1, \dots, X_n) = \min_{t \in \mathcal{T}} \operatorname{Var} t(X_1, \dots, X_n)$$

25.4

25.4 Satz: Ist X eine SG mit endlicher $\operatorname{Var} X$ und $X_1, \dots, X_n$ eine SP von X, so ist $\overline{X}_n$ die effiziente lineare SF für $\mathbb{E}X$.

Beweis: $\mathcal{T} = \{\, t(X_1, \dots, X_n) = \sum_{i=1}^n \alpha_i X_i \,\}$ mit $\sum_{i=1}^n \alpha_i = 1$

wegen $\mathbb{E}\left(\sum_{i=1}^n \alpha_i X_i\right) = \mathbb{E}X_i$ (Unverzerrtheit)

Es gilt $\operatorname{Var}\left(\sum_{i=1}^n \alpha_i X_i\right) = \sum_{i=1}^n \alpha_i^2 \operatorname{Var} X_i$ (wegen Ua)

$\Rightarrow$ Varianz der SF minimal für $\sum_{i=1}^n \alpha_i^2$ minimal.

Es gilt $$\sum_{i=1}^n \alpha_i^2 = \sum_{i=1}^n \left(\alpha_i - \tfrac{1}{n} + \tfrac{1}{n}\right)^2 = \sum_{i=1}^n \left[\left(\alpha_i - \tfrac{1}{n}\right)^2 + 2\left(\alpha_i - \tfrac{1}{n}\right)\tfrac{1}{n} + \left(\tfrac{1}{n}\right)^2\right] = \sum_{i=1}^n \left(\alpha_i - \tfrac{1}{n}\right)^2 + \sum_{i=1}^n 2\left(\alpha_i - \tfrac{1}{n}\right)\tfrac{1}{n} + \sum_{i=1}^n \tfrac{1}{n^2} = \cdot/.$$

25.5

$$= \sum_{i=1}^{n} (\alpha_i - \tfrac{1}{n})^2 + \tfrac{2}{n}\left(\sum_{i=1}^{n} \alpha_i - 1\right) + \tfrac{1}{n} \;\geq 0$$

Ausdruck minimal für $\alpha_i = \frac{1}{n} \quad \forall i = 1(1)n$ //

25.5 Konsistenz 131

Für Stichprobenumfang $n \to \infty$ Konvergenz von Folgen von Schätzfunktionen gegen gesuchte Werte (GgZ)

<u>Definition</u>: Ist $X_1, X_2, \cdots$ eine UIV-Folge, so heißt eine Folge $t_n(X_1, \cdots, X_n);\ n \in \mathbb{N}$ von Schätzfunktionen für eine char. Größe ξ der Verteilung von X_i

konsistent, wenn $\quad t_n(X_1, \cdots, X_n) \xrightarrow[n\to\infty]{W} \xi$

d.h. wenn gilt:

$$\lim_{n\to\infty} W\{\xi - \varepsilon < t_n(X_1, \cdots, X_n) < \xi + \varepsilon\} = 1 \quad \forall \varepsilon > 0$$

<u>Bem.</u>: $\xi \in \mathbb{R}$; ξ ist häufig ein (geraffter) Parameter

<u>Es gilt</u>: Aus dem GgZ folgt, dass $\overline{X}_n;\ n \in \mathbb{N}$ eine (stark) konsistente Schätzfolge für $\mathbf{E}X_i$ bildet (Vsn.)

<u>Beispiel</u>: Die Folge der Werte $F_n^*(x),\ n \in \mathbb{N}$ der empir. VFn an einer festen Stelle $x \in \mathbb{R}$ bildet für $n \to \infty$ eine konsistente Schätzfolge für den Wert $F(x)$

25.6 Plausibilität 132

Eigenschaft von Schätzwerten für Parameter

<u>Für diskret vt. SG$_n$</u> $X \sim p(\cdot|\theta),\ \theta \in \Theta$ gemeinsame Vtlg. einer SP $(X_1, \cdots, X_n)$ diskret mit Punktwahrscheinlichk.

$$w(x_1, \cdots, x_n|\theta) = \prod_{i=1}^{n} p(x_i|\theta) \qquad \forall (x_1, \cdots, x_n) \in M_X^n$$

Nach Beobachtung der konkreten SP $x_1, \cdots, x_n$ ist $w(x_1, \cdots, x_n|\theta)$ eine Funktion der Variablen θ, genannt Plausibilitätsfunktion $\ell(\theta; x_1, \cdots, x_n)$, wobei $x_1, \cdots, x_n$ Konstanten sind.

$$\ell : \Theta \longrightarrow [0, \infty)$$

Ist θ_0 der wahre Parameterwert, so ist ein möglichst guter Schätzwert $\hat{\theta}$ für θ_0 gesucht

<u>Definition</u>: Der plausible Schätzwert $\hat{\theta}$ für θ_0 ist jener Parameterwert, für den, falls er existiert, die Wahrsch. der erhaltenen konkreten SP $x_1, \cdots, x_n$ maximal ist, d.h.

$$\ell(\hat{\theta}; x_1, \cdots, x_n) = \max_{\theta \in \Theta} \ell(\theta; x_1, \cdots, x_n)$$

<u>Bem.</u>: $\dfrac{\partial \ell(\theta; x_1, \ldots, x_n)}{\partial \theta} = 0$ bzw. $\dfrac{\partial \ln \ell(\theta; x_1, \cdots, x_n)}{\partial \theta} = 0$

notwendige Bedingung falls $\ell(\theta; x_1, \cdots, x_n)$ diffb.

Beispiel: $X \sim A_\theta$, konkrete SP $x_1, \dots, x_n$

$$p(x|\theta) = \theta^x (1-\theta)^{1-x} \, I_{\{0,1\}}(x) \quad \text{mit } \theta \in [0,1]$$

$$\ell(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} \, I_{\{0,1\}}(x_i)$$

$$= \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} I_{\{0,1\}}(x_i)$$

log. PF $\ln \ell(\theta; x_1, \dots, x_n) \Rightarrow \dots = 0$ Ü

plausibler Schätzwert $\hat{\theta} = \bar{x}_n = \frac{x_1 + \dots + x_n}{n}$

Ähnlich: plaus. SW für μ in P_μ Ü

Für kontinuierliche Verteilungen $X \sim f(\cdot|\theta)$; $\theta \in \Theta$ ist die Plausibilitätsfunktion die gemeinsame Dichtef. der Stichprobe, aufgefasst als Funktion von θ, d.h.

$$\ell(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i|\theta) = g(x_1, \dots, x_n|\theta)$$

Def.: Der plausible Schätzwert $\hat{\theta}$ ist, falls er $\exists$, jener Parameterwert, der die PF maximiert

Bem.: Für Vektorparameter $\theta = (\theta_1, \dots, \theta_k)$

$$\frac{\partial \ell(\theta_1, \dots, \theta_k; x_1, \dots, x_n)}{\partial \theta_j} = 0 \quad \forall j = 1(1)k$$

oder analog für $\ln \ell(\theta; x_1, \dots, x_n)$

Bsp.: $X \sim N(\mu, \sigma^2)$, $\theta = (\mu, \sigma^2)$, $k = 2$, Beob. $x_1, \dots, x_n$

$$\ell^*(\theta, x_1, \dots, x_n) = \ln \ell(\mu, \sigma^2; x_1, \dots, x_n) =$$

$$= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2} \quad \text{Ü}$$

$$\frac{\partial \ell^*}{\partial \mu} = \frac{\sum_{i=1}^{n} (x_i - \mu)}{\sigma^2} = 0 \quad \Rightarrow \quad \hat{\mu} = \bar{x}_n$$

$$\frac{\partial \ell^*}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2(\sigma^2)^2} = 0 \quad \Rightarrow \quad \widehat{\sigma^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_n)^2}{n}$$

$\Rightarrow$ plaus. SW für $\theta = (\mu, \sigma^2)$ ist $\widehat{(\mu, \sigma^2)} = (\bar{x}_n, \frac{n-1}{n} s_n^2)$

Bem.: versch. Güteeigenschaften

Bsp.: $X \sim Ex_\tau$, $f(x|\tau) = \frac{1}{\tau} e^{-\frac{x}{\tau}} I_{(0,\infty)}(x)$ mit $\tau \in (0, \infty)$

$\Rightarrow$ plaus. SW $\hat{\tau} = \bar{x}_n$ Ü

Definition: Eine SF $\vartheta(X_1, \dots, X_n)$, die jeder konkreten SP $x_1, \dots, x_n$ den plausiblen SW $\vartheta(x_1, \dots, x_n)$ zuordnet, heißt plausible Schätzfunktion.

26. KONFIDENZBEREICHE

Quantitative Güteangaben für Parameterschätzungen

Def.: Ein KB mit ÜDW $1-\alpha$ für einen Parameter θ_0 eines Stoch. Modells $X \sim f(\cdot|\theta)$, $\theta \in \Theta$ ist eine Teilmenge $C_{1-\alpha} \subset \Theta$ für die gilt $W\{\theta_0 \in C_{1-\alpha}\} = 1-\alpha$.

Bem.: Konstruktion von KBn mittels sogenannter Pivot-Größen

26.1 Pivot-Größen

Def.: Eine SG, die eine Funktion $q(X_1, \dots, X_n, \theta)$ einer Stichprobe von $X \sim W_\theta$, $\theta \in \Theta$ und des Parameters θ ist, deren Verteilung nicht von θ abhängt, nennt man Pivot-Größe

26.2 Satz: Ist $X_1, \dots, X_n$ eine SP von $X \sim N(\mu, \sigma^2)$, so gilt:

$$\left.\begin{array}{l} \dfrac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1} \\[2ex] \dfrac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1} \end{array}\right\} \text{unabhängig}$$

(ohne Beweis)

Damit kann man sowohl 2-dim. KB als als auch Konfidenzintervalle für μ bzw. σ^2 konstruieren

26.3 Satz: Konfidenzintervalle mit ÜDW $1-\alpha$ für die Parameter der Normalverteilung sind:

für μ: $\left[\overline{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}} \,;\, \overline{X}_n + \frac{S_n}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}}\right]$

für σ^2: $\left[\frac{(n-1)S_n^2}{\chi^2_{n-1;1-\frac{\alpha}{2}}} \,;\, \frac{(n-1)S_n^2}{\chi^2_{n-1;\frac{\alpha}{2}}}\right]$

Beweis: (für μ) SP $X_1, \dots, X_n$ $\Rightarrow$

$\frac{\overline{X}_n - \mu}{S_n/\sqrt{n}}$ Pivot-Größe mit t-Vtlg. t_{n-1}

$t_{n-1;p}$ = p-Fraktiles der t_{n-1}

$$\Rightarrow W_{\mu,\sigma^2}\left\{t_{n-1;\frac{\alpha}{2}} \le \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \le t_{n-1;1-\frac{\alpha}{2}}\right\} = 1-\alpha \quad \forall(\mu,\sigma^2)$$

Das Ereignis $\{\cdots\}$ lässt sich beschreiben durch äquivalente Umformung der Doppelungleichung in

$$\overline{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{n-1;1-\frac{\alpha}{2}} \le \mu \le \overline{X}_n - \frac{S_n}{\sqrt{n}} \cdot t_{n-1;\frac{\alpha}{2}}$$

und wegen $t_{n-1;\frac{\alpha}{2}} = -t_{n-1;1-\frac{\alpha}{2}}$ folgt die Beh.

(für σ^2 analoge Umformung) Ü

27. STAT. HYPOTHESEN UND TESTS

<u>Problem</u>: Welche Vtlg. hat eine SG X ?

<u>Def.</u>: Eine Stat. Hypothese ist eine Aussage über die Verteilung von X.

<u>Bsp.</u>: $\mathcal{H}$: X ist normalverteilt

<u>Bem.</u>: Zu Stat. Hypothesen auch Gegenhypothesen betrachtet : $\mathcal{H}_0$ $\mathcal{H}_1$

<u>Def.</u>: Ein Statistischer Test ist ein Verfahren zur Entscheidung, ob eine Stat. Hypothese angenommen oder verworfen wird.

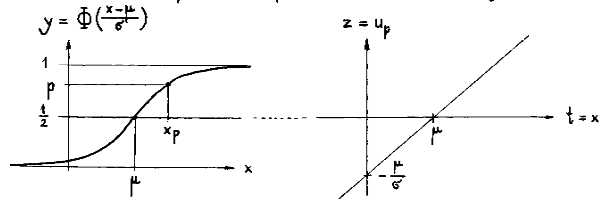
27..

27.1 Wahrscheinlichkeitspapiere

Transformation der $(x, F(x))$-Ebene auf (t, z)-Ebene dass die Bilder der VF$_n$ $F(\cdot)$ aus $\mathcal{H}_0$ Geraden sind.

<u>Beispiel</u>: Normalverteilungsnetz

$\mathcal{H}_0 : X \sim N(\mu, \sigma^2)$, μ und σ^2 beliebig



Transformation $t = x$, $p \to u_p$ (p-Fraktiles der N(0,1

27..

<u>Es gilt</u>: Zusammenhang zwischen u_p und x_p

x_p .... p-Fraktiles der $N(\mu, \sigma^2)$

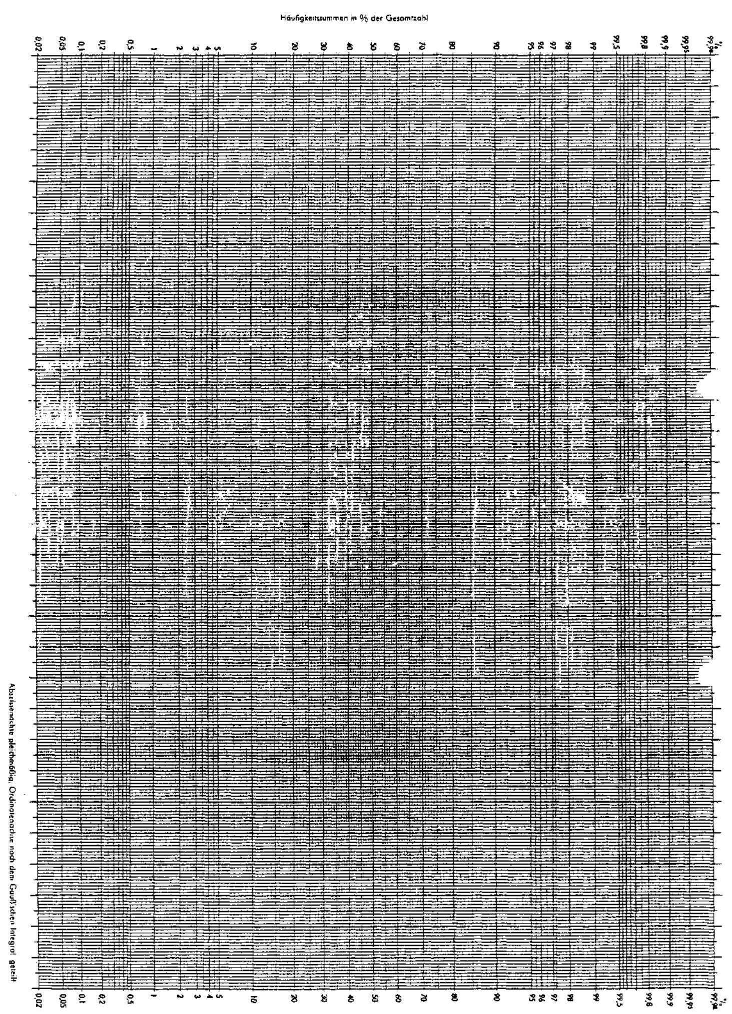
$$u_p = \frac{x_p - \mu}{\sigma} \qquad \forall\, p \in (0,1)$$

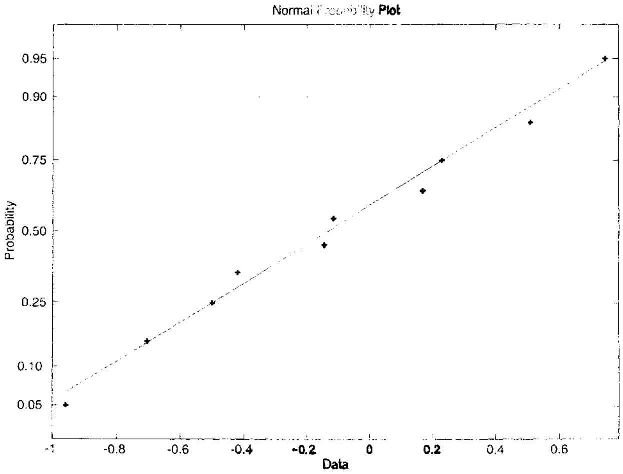
Beweis: $X \sim N(\mu, \sigma^2) \Rightarrow p = W\{X \leq x_p\} =$

$$= W\Big\{\underbrace{\frac{X-\mu}{\sigma}}_{\sim N(0,1)} \leq \underbrace{\frac{x_p - \mu}{\sigma}}_{= u_p}\Big\};$$ wegen der strikten Monotonie von $\Phi(\cdot) \Rightarrow$ Beh. //

<u>Test</u>: Gestalt des Bildes der empir. VF oder einer anderen Schätzung der VF.

27..





27.2 Fehlerarten und Fehlerwahrscheinlichkeiten

Eine Stat. Hypothese $\mathcal{H}_0$ kann richtig oder falsch sein.

Def.: Die Verwerfung einer richtigen Hypothese heißt ein Fehler erster Art.

Die Annahme einer falschen Hypothese heißt ein Fehler zweiter Art.

Bem.: Fehlerwahrscheinlichkeiten

$$\alpha = W\{\mathcal{H}_0 \text{ verworfen} \mid \mathcal{H}_0 \text{ richtig}\}$$

$$\beta = W\{\mathcal{H}_0 \text{ angenommen} \mid \mathcal{H}_0 \text{ falsch}\}$$

Ideal wären Tests, die beide Fehlerwahrscheinlichkeiten klein halten. → Plausibilitätsquotiententests

Die Konstruktion von Stat. Tests erfolgt über Fehlerwahrscheinlichkeiten erster Art.

27.3 Verwerfungsräume, Teststatistiken und kritische Bereiche

SP $X_1, \dots, X_n$ von X, Hypothese $\mathcal{H}$

Stichprobenraum M_X^n zerlegt: $M_X^n = V \cup A$

falls eine konkrete SP $x_1, \dots, x_n$ in V liegt, wird $\mathcal{H}$ verworfen, V heißt Verwerfungsraum

Beispiel: $X \sim N(\mu, \sigma_0^2)$, σ_0^2 bekannt, SP $X_1, \dots, X_n$

$\mathcal{H}_0: \mu = \mu_0$ mit gegebenem μ_0

Zur Konstruktion eines Tests gibt man eine kleine Wahrscheinlichkeit α eines Fehlers 1. Art (Irrtumswahrscheinlichkeit) vor.

Bei richtiger $\mathcal{H}_0$ gilt $T = \dfrac{\overline{X}_n - \mu_0}{\sigma_0 / \sqrt{n}} \sim N(0,1)$

Die Testentscheidung hängt vom Wert der sog. Teststatistik T ab: Es gilt, falls $\mathcal{H}_0$ richtig,

$$W\{u_{\frac{\alpha}{2}} \le T \le u_{1-\frac{\alpha}{2}}\} = 1 - \alpha$$

Wegen $u_{\frac{\alpha}{2}} = -u_{1-\frac{\alpha}{2}}$ nimmt man den Verwerfungsraum

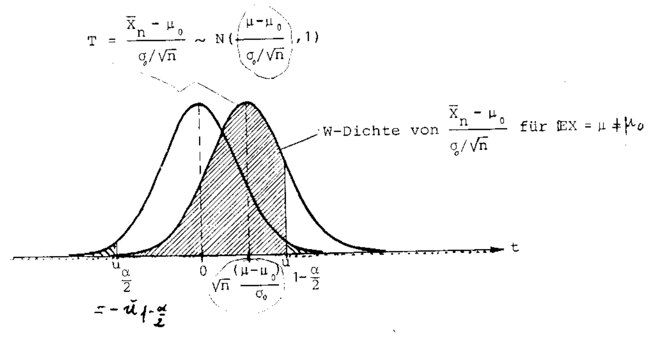
$$V = \left\{ (x_1, \dots, x_n) : \left| \frac{\overline{x}_n - \mu_0}{\sigma_0 / \sqrt{n}} \right| \geq u_{1-\frac{\alpha}{2}} \right\}$$

<u>Bem.</u>: Man nimmt $\mathcal{H}_0$ an oder verwirft $\mathcal{H}_0$, je nachdem ob der konkrete Wert t der Teststatistik T im Intervall $(-u_{1-\frac{\alpha}{2}}, u_{1-\frac{\alpha}{2}})$ liegt oder im Komplement $(-\infty, -u_{1-\frac{\alpha}{2}}] \cup [u_{1-\frac{\alpha}{2}}, \infty) = C$

C heißt kritischer Bereich für die Teststatistik

27.9

Zur Fehlerwahrscheinlichkeit zweiter Art:



27.10

28. TESTS FÜR NORMALVERTEILUNGEN

Alle betrachteten SG als normalverteilt vorausges. Parameter unbekannt

28.1 t-Test für den Erwartungswert

Ist $X \sim N(\mu, \sigma^2)$ mit μ und σ^2 unbekannt und $x_1, \dots, x_n$ eine konkrete SP von X, so ist ein Test für $\mathcal{H}_0 : \mu = \mu_0$ mit Wahrsch. α eines Fehlers 1. Art durch folgenden Verwerfungsraum V gegeben:

$$V = \left\{ (x_1, \dots, x_n) : \frac{\sqrt{n}\,|\overline{x}_n - \mu_0|}{s_n} \geq t_{n-1;\,1-\frac{\alpha}{2}} \right\}$$

mit $t_{n-1;\,p}$ ... p-Fraktiles der t_{n-1}-Vtlg.

28.1

Begründung: Satz 26.2 $\Rightarrow T = \frac{\sqrt{n}\,(\overline{X}_n - \mu_0)}{S_n} \sim t_{n-1}$

Dichtef. der t_{n-1} symmetrisch $\Rightarrow t_{n-1;\,\frac{\alpha}{2}} = -t_{n-1;\,1-\frac{\alpha}{2}}$

$$W\{\text{Fehler 1. Art}\} = W\left\{ \frac{\sqrt{n}\,|\overline{X}_n - \mu_0|}{S_n} \geq t_{n-1;\,1-\frac{\alpha}{2}} \right\} = \alpha \quad //$$

28.2 Test für die Varianz

Ist $X \sim N(\mu, \sigma^2)$ mit μ und σ^2 unbekannt, $x_1, \dots, x_n$ eine konkrete SP von X, so ist ein Test für die $\mathcal{H}_0 : \sigma^2 = \sigma_0^2$ mit Wahrsch. α eines Fehlers 1. Art durch folgenden Verwerfungsraum V bestimmt:

28.2

$$V = \left\{ (x_1, \dots, x_n) : \frac{(n-1)s_n^2}{\sigma_0^2} \notin \left(\chi^2_{n-1;\frac{\alpha}{2}} , \chi^2_{n-1;1-\frac{\alpha}{2}} \right) \right\}$$

Beweis : Satz 26.2 $\Rightarrow$ Vtlg. von $\frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi^2_{n-1}$

Rest analog zum vorigen Beweis.

Bem.: Kritischer Bereich $(-\infty, \chi^2_{n-1;\frac{\alpha}{2}}] \cup [\chi^2_{n-1;1-\frac{\alpha}{2}}, \infty)$

28.3 Zwei-Stichproben-Problem

Frage: Haben 2 SGn dieselbe Normalverteilung?

Gegeben: 2 Messreihen $x_1, \dots, x_m$
$y_1, \dots, y_n$

t-Test für die Gleichheit der Erwartungswerte

Ist $X_1, \dots, X_m$ eine SP einer $N(\mu_x, \sigma^2)$ und $Y_1, \dots, Y_n$ eine SP einer $N(\mu_y, \sigma^2)$ und sind die SPn unabhängig, so gilt

$$T = \frac{\overline{X}_m - \overline{Y}_n - \mu_x + \mu_y}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \frac{(m-1)S_m^2 + (n-1)S_n^2}{m+n-2}}} \sim t_{m+n-2}$$

(ohne Beweis)

Ein Test für $\mathcal{H}_0 : \mu_x = \mu_y$ mit Wahrsch. α eines Fehlers 1. Art:

verwerfen falls $t = \left| \frac{\overline{x}_m - \overline{y}_n}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right) \frac{(m-1)s_m^2 + (n-1)s_n^2}{m+n-2}}} \right| \geq t_{m+n-2;1-\frac{\alpha}{2}}$

F-Test für die Gleichheit der Varianzen S. 156

Def.: Die F-Verteilung mit m und n Freiheitsgraden, i.Z. $F_{m,n}$ hat folgende Dichtefunktion S. 56

$$f(x \mid m,n) = \frac{\Gamma\left(\frac{m+n}{2}\right) m^{\frac{m}{2}} n^{\frac{n}{2}} x^{\frac{m}{2}-1}}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right) (mx+n)^{\frac{m+n}{2}}} I_{(0,\infty)}(x)$$

$F_{m,n;p}$ ... p-Fraktiles der $F_{m,n}$ (tabelliert)

Es gilt: Ist $X_1, \dots, X_m$ SP von $X \sim N(\mu_x, \sigma_x^2)$ und $Y_1, \dots, Y_n$ SP von $Y \sim N(\mu_y, \sigma_y^2)$, unabhängige SP,

$$\Rightarrow \frac{S_x^2 / \sigma_x^2}{S_y^2 / \sigma_y^2} \sim F_{m-1, n-1}$$

Test: Ein Test für $\mathcal{H}_0 : \sigma_x^2 = \sigma_y^2$ mit Wahrsch. α eines Fehlers 1. Art durch die Teststatistik

$$T = \frac{S_x^2}{S_y^2}$$

Entscheidungsregel: $\mathcal{H}_0$ verwerfen, falls für 2 konkrete SPn $x_1, \dots, x_m$ und $y_1, \dots, y_n$ gilt

$$\frac{s_x^2}{s_y^2} \notin \left(F_{m-1,n-1;\frac{\alpha}{2}} , F_{m-1,n-1;1-\frac{\alpha}{2}} \right)$$

Bem.: Für mehr als 2 Messreihen Bartlett-Test

29. DER CHIQUADRAT-ANPASSUNGSTEST

Ein asymptotisches Verfahren bei größeren SP$_n$

Problem: Hat eine empirisch gegebene SG eine bestimmte Verteilung bzw. bestimmten Verteilungstyp

29.1 Einfache Hypothesen

Satz: Ist $X_1, X_2, \dots$ eine unbeschränkte SP von $X \sim W$ auf $(M_X, \mathfrak{C})$ und $A_1, \dots, A_r$ eine Zerlegung von M_X mit $A_j \in \mathfrak{C}$ $\forall j = 1(1)r$, so gilt mit den Bezeichnungen

$w_j = W(A_j)$ für $j = 1(1)r$

$Y_j^{(n)}$ = absolute Häufigkeit von A_j nach n Beobachtungen

$$\lim_{n\to\infty} W\left\{ Z_n = \sum_{j=1}^{r} \frac{[Y_j^{(n)} - nw_j]^2}{nw_j} \leq \chi^2_{r-1;p} \right\} = p \quad \forall p \in (0,1)$$

d.h. Z_n ist asymptotisch nach χ^2_{r-1} verteilt.

Test: Mit den voranstehenden Bezeichnungen ist ein Test für $\mathcal{H}_0: X \sim W$ durch den Verwerfungsraum

$$V = \{(x_1, \dots, x_n) : z_n \geqq \chi^2_{r-1;1-\alpha}\} \quad \text{gegeben,}$$

dessen Wahrscheinlichkeit eines Fehlers 1. Art für $n \to \infty$ gegen α konvergiert.

Begründung: Bei richtiger $\mathcal{H}_0$ ist die Teststatistik Z_n asymptotisch χ^2_{r-1}-verteilt ($n \to \infty$)

Bem.: Bei Anwendung soll für $A_1, \dots, A_r$ gelten:

$nw_j \geqq 5 \quad \forall j = 1(1)r$

Beispiel: Es soll getestet werden, ob 100 „zufällig" aus $[0,1]$ gewählte Zahlen $x_1, \dots, x_{100}$ nach $U_{0,1}$ verteilt angesehen werden können. Ü

29.2 Zusammengesetzte Parameterhypothesen

Problem: Hat eine empir. geg. SG einen bestimmten parametrischen Verteilungstyp

$X_1, \dots, X_n, \dots$ SP von X, $X \sim W_\theta, \theta \in \Theta$

$\mathcal{H}_0: X \sim W_\theta$, $\theta = (\theta_1, \dots, \theta_s)$ s-dim Parameter

Satz: Ist $X_1, \dots, X_n, \dots$ eine unbeschr. SP von $X \sim W_\theta$ mit $\theta \in \Theta \subseteq \mathbb{R}^s$, wobei Θ ein offenes, nicht ausgeartetes s-dimensionales Intervall ist und $A_1, \dots, A_r$ eine Zerlegung von M_X in Ereignisse, wobei $r > s+1$,

$w_j(\theta) := W_\theta(A_j)$, $j = 1(1)r$

$Y_j^{(n)}$:= absolute Häuf. von A_j nach n Beob. von X und

$\exists$ die plausiblen Schätzwerte $\hat{\theta}^{(n)}$ für θ, so gilt:

$$\lim_{n\to\infty} W\left\{ Z_n = \sum_{j=1}^{r} \frac{[Y_j^{(n)} - nw_j(\hat{\theta}^{(n)})]^2}{nw_j(\hat{\theta}^{(n)})} \leq \chi^2_{r-s-1;p} \right\} = p \quad \forall p \in (0,1)$$

d.h. Z_n ist asymptotisch nach χ^2_{r-s-1} verteilt

(ohne Beweis, 1900)

Chiquadrattest für zusammengesetzte Hypothesen:

Ein Test für die zusammengesetzte $\mathcal{H}_0: X \sim W_\theta,\ \theta \in \Theta$ dessen Wahrsch. eines Fehlers 1. Art mit $n \to \infty$ gegen α konvergiert, ist durch folgenden Verwerfungsraum V gegeben:

$$V = \{(x_1, \dots, x_n) :\ z_n(x_1, \dots, x_n) \geq \chi^2_{r-s-1;\,1-\alpha}\}$$

<u>B3</u>: Testen, ob Wartezeiten eine $Ex_\tau,\ \tau \in (0,\infty)$ haben

<u>Beispiel</u>: Testen, ob die Anzahl von Ausfällen eines Computertyps in einem bestimmten Zeitraum eine Poisson-Verteilung hat

29.5

30. KLASSISCHE REGRESSIONSRECHNUNG

Beschreibung kausaler, nicht deterministischer Zusammenhänge mit Stochastischen Modellen

<u>B5</u>: Abhängigkeit des Bremsweges von der Geschwindigkeit (Kovariable, Einflussgröße)

x ... Geschwindigkeit (—"—)

Y_x ... Bremsweg (SG) abhängige Größe

Stoch. Modell $\quad Y_x = \psi(x) + U_x$

$\mathbb{E}Y_x = \psi(x)$ Regressionsfunktion 2. Art

30.1

30.1 Lineare Regressionsfunktionen

$Y_x \sim W_\theta$ mit $\theta \in R$ oder $\theta \in \mathbb{R}^k$, d.h.

$$\theta = \begin{pmatrix}\theta_1\\ \vdots \\ \theta_k\end{pmatrix}$$

$$\psi(x) = \mathbb{E}_\theta Y_x = \psi(x, \theta)$$

Auch die Einflussgröße x kann ein Vektor $\underline{x} = \begin{pmatrix}x_1\\ \vdots \\ x_k\end{pmatrix}$ sein

<u>Def.</u>: Eine Regressionsfunktion heißt lineare Regressionsfunktion, wenn sie als Funktion der unbekannten Parameter linear ist.

30..

<u>Beispiele</u>: 1) $\theta = \begin{pmatrix}\alpha\\ \beta\end{pmatrix}$, $\psi(x,\theta) = \alpha + \beta x$

2) $\theta = \begin{pmatrix}\theta_0\\ \theta_1\\ \vdots \\ \theta_k\end{pmatrix}$, $\psi(x,\theta) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_k x^k$

polynomische Regression

3) für vektorielle Einflussgrößen $\underline{x} = (x_1, \dots, x_k)$ und Parametervektor $\theta = \begin{pmatrix}\theta_1\\ \vdots \\ \theta_k\end{pmatrix}$

$$\psi(\underline{x}, \theta) = \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_k x_k = \underline{x}\,\theta$$

multiple Regression

30..

Bem.: Die Regressionsparameter müssen aus Beobachtungen (x_i, y_i), $i = 1(1)n$ bzw. bei Vektorgrößen $\underline{x}$ aus Beobachtungen $(x_{i1}, \dots, x_{ik}, y_i)$, $i = 1(1)n$ geschätzt werden.

30.2 Regressionsgeraden

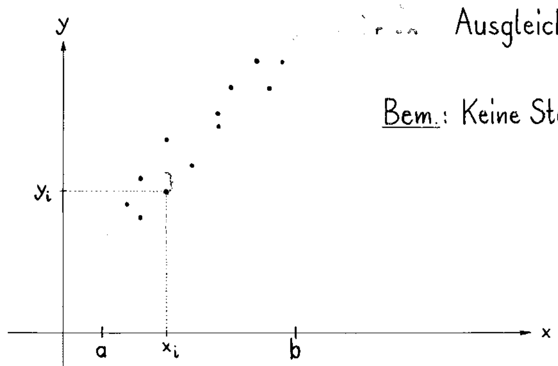
$x \in \mathbb{R}$, $\psi(x, \alpha, \beta) = \alpha + \beta x \quad \forall x \in (a,b)$

Voraussetzung: $Y_x = \alpha + \beta x + U_x$ mit konst. Varianz

$\mathrm{Var}\, U_x \equiv \sigma^2 \quad \forall x \in (a,b)$ ... Homoskedastizität

Daten (x_i, y_i), $i = 1(1)n$, mindestens 2 verschiedene x_i

30.4



Ausgleichsgerade

Bem.: Keine Stoch.

$$AQS = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

30.5

Frage: Liefert die Ausgleichsrechnung, d.h. suche jene Werte $\hat{\alpha}$ und $\hat{\beta}$, sodass

$$\sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = \underset{\alpha,\beta}{\mathrm{Min}} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

auch statistisch sinnvolle Schätzungen für α und β?

Def.: Schätzfunktionen A und B für die Parameter α und β der Regressionsgeraden heißen linear, wenn sie folgende Form haben:

$$A = \sum_{i=1}^{n} c_i Y_i \quad \text{und} \quad B = \sum_{i=1}^{n} d_i Y_i$$

wobei Y_i steht für Y_{x_i}

30.6

164

Satz von Gauß-Markoff: Mit den Bezeichnungen dieses Abschnitts mit einer SP (x_i, Y_i), $i = 1(1)n$ und unter folgenden Voraussetzungen (1) bis (4)

(1) $x_1, \dots, x_n$ sind nicht alle gleich

(2) $Y_1, \dots, Y_n$ sind unkorreliert

(3) $\mathrm{Var}\, Y_i \equiv \sigma^2$, d.h. alle Varianzen sind gleich

(4) $\mathbb{E} Y_i = \alpha + \beta x_i \quad \forall i = 1(1)n$ mit reellen Konstanten α, β wobei α, β und σ^2 unbekannt sind,

gilt:

Die effizienten linearen Schätzfunktionen A und B für α und β sind folgende

30.7

$$A = \frac{\left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{i=1}^{n} Y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} x_i Y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

$$B = \frac{n\left(\sum_{i=1}^{n} x_i Y_i\right) - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}$$

und diese ergeben für konkrete Beobachtungspaare (x_i, y_i), $i = 1(1)n$ die Werte $\hat{\alpha}$ und $\hat{\beta}$ der Ausgleichsgeraden $y = \hat{\alpha} + \hat{\beta}x$, Schätzwerte für α und β.

30.8

Für die SF$_n$ A und B gilt:

$$\operatorname{Var} A = \frac{\sum_{i=1}^{n} x_i^2}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}\,\sigma^2$$

$$\operatorname{Var} B = \frac{n}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}\,\sigma^2$$

$$\operatorname{Cov}(A,B) = -\frac{\sum_{i=1}^{n} x_i}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2}\,\sigma^2$$

30.9

Eine unverzerrte SF für σ^2 ist

$$S^2 := \frac{\sum_{i=1}^{n} (Y_i - A - Bx_i)^2}{n-2}$$

(o. Bew.)

<u>Bem.</u>: Ein analoger Satz gilt für multiple lineare Regressionsmodelle

30.10

VII

ELEMENTE DER BAYES-STATISTIK

Hier alle unbekannten Größen durch SG$_n$ beschrieben, auch Parameter θ von Stoch. Modellen $X \sim W_\theta$, $\theta \in \Theta$

$\tilde{\theta} \sim \pi(\cdot)$ auf Θ vor Erhebung von Daten D

$\pi(\cdot)$ heißt A-priori-Verteilung von $\tilde{\theta}$

Nach Beobachtung von D neue Verteilung $\pi(\cdot \mid D)$

31. BAYES'sches THEOREM

$X \sim f(\cdot|\theta),\ \theta \in \Theta$, $\pi(\cdot)$ A-priori-Vtlg. von $\tilde{\theta}$

SP $x_1, \dots, x_n$

Die durch $X_1 = x_1, \dots, X_n = x_n$ bedingte Vtlg. von $\tilde{\theta}$, i.Z. $\pi(\cdot|x_1, \dots, x_n)$ heißt A-posteriori-Verteilung

31.1 Diskreter Fall

$X \sim p(\cdot|\theta),\ \theta \in \{\theta_1, \dots, \theta_k\}$, a-priori $\pi(\theta_j),\ j = 1(1)k$

Daten $x_1, \dots, x_n$ von $X_1, \dots, X_n$ SP von X

A-posteriori-Wahrscheinlichkeiten der θ_j:

31.1

$\{\theta_1\}, \{\theta_2\}, \dots, \{\theta_k\}$ Zerlegung von $\Theta = \{\theta_1, \dots, \theta_k\}$

$$\pi(\theta_j|x_1, \dots, x_n) = W\{\tilde{\theta} = \theta_j | X_1 = x_1, \dots, X_n = x_n\}$$

$$= \frac{W\{X_1 = x_1, \dots, X_n = x_n, \tilde{\theta} = \theta_j\}}{W\{X_1 = x_1, \dots, X_n = x_n\}}$$

wegen der Def. von bedingten Wahrsch. diskreter Größen gilt

$$W\{X_1 = x_1, \dots, X_n = x_n, \tilde{\theta} = \theta_j\} = w(x_1, \dots, x_n|\theta_j)\,W\{\tilde{\theta} = \theta_j\}$$

und

$$W\{X_1 = x_1, \dots, X_n = x_n\} = \sum_{j=1}^{k} w(x_1, \dots, x_n|\theta_j)\,\pi(\theta_j)$$

31.2

Daher gilt das sog. Bayes'sche Theorem (diskret):

$$\pi(\theta_j|x_1, \dots, x_n) = \frac{w(x_1, \dots, x_n|\theta_j)\,\pi(\theta_j)}{\sum_{i=1}^{k} w(x_1, \dots, x_n|\theta_i)\,\pi(\theta_i)} \quad \forall j = 1(1)k$$

Im Fall einfacher SP_n gilt wegen der Unabhängigkeit

$$w(x_1, \dots, x_n|\theta_j) = \prod_{i=1}^{n} p(x_i|\theta_j)$$

Bem.: Plausibilitätsfunktion $\ell(\theta; x_1, \dots, x_n)$

$$\Rightarrow\ w(x_1, \dots, x_n|\theta_j)\,\pi(\theta_j) = \pi(\theta_j)\,\ell(\theta_j; x_1, \dots, x_n)$$

Da $\sum_{j=1}^{k} \pi(\theta_j)\,\ell(\theta_j; x_1, \dots, x_n)$ nach Beobachtung der konkreten SP $x_1, \dots, x_n$ eine Konstante ist

31.3

gilt $\pi(\theta_j|x_1, \dots, x_n) = \frac{1}{C}\,\pi(\theta_j)\,\ell(\theta_j; x_1, \dots, x_n) \quad \forall j = 1(1)k$

In Kurzschreibweise:

$$\pi(\theta_j|x_1, \dots, x_n) \propto \pi(\theta_j)\,\ell(\theta_j; x_1, \dots, x_n) \quad \forall \theta_j \in \Theta$$

31.2 Kontinuierlicher Fall

$X \sim f(\cdot|\theta),\ \theta \in \Theta$, $\tilde{\theta} \sim \pi(\cdot)$ A-priori-Dichte

SP $X_1, \dots, X_n$; für gegebenen Parameter θ ist

$g(x_1, \dots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta)$ Dichte der SP

Gesucht: Verteilung von $\tilde{\theta} \mid X_1 = x_1, \dots, X_n = x_n$

31.4

<u>Lösung</u>: Bedingte Dichte $\pi(\cdot \mid x_1, \dots, x_n)$

Gemeinsame Dichte von $(X_1, \dots, X_n, \tilde{\theta})$ ist

$$g(x_1, \dots, x_n, \theta) = \left[\prod_{i=1}^{n} f(x_i \mid \theta)\right]\pi(\theta)$$

Randdichte von $X_1, \dots, X_n$ ist

$$\int_{\Theta} g(x_1, \dots, x_n, \theta)\, d\theta = \int_{\Theta} \pi(\theta)\,\ell(\theta; x_1, \dots, x_n)\, d\theta \,,$$ dies ist

nach Beobachtung von konkreten $x_1, \dots, x_n$ eine Konstante. Daher gilt für die A-posteriori-Dichte des Parameters

31.5

$$\pi(\theta \mid x_1, \dots, x_n) = \frac{\pi(\theta)\,\ell(\theta; x_1, \dots, x_n)}{\int_{\Theta} \pi(\theta)\,\ell(\theta; x_1, \dots, x_n)\, d\theta} = \frac{1}{C}\,\pi(\theta)\,\ell(\theta; x_1, \dots, x_n)$$

In Kurzschreibweise:

$$\pi(\theta \mid x_1, \dots, x_n) \propto \pi(\theta)\,\ell(\theta; x_1, \dots, x_n) \qquad \theta \in \Theta$$

<u>Bem.</u>: Für allgemeinere Daten D gilt auch

$$\pi(\theta \mid D) \propto \pi(\theta)\,\ell(\theta; D) \qquad \theta \in \Theta$$

Kurzform des Bayes'schen Theorems

31.6

32. VERWENDUNG DER A-POSTERIORI-VTLG.

In der ApoV alle wesentliche Information enthalten

$X \sim f(\cdot \mid \theta)$, $\theta \in \Theta$, $\pi(\cdot)$, Daten D, $\pi(\cdot \mid D)$

32.1 Prädiktivverteilungen

Prognose für $X \mid D$: Randverteilung von X aus der gemeinsamen Vtlg. von $(X, \tilde{\theta}) \sim f(x \mid \theta)\pi(\theta \mid D)$

$$f(x \mid D) = \int_{\Theta} f(x \mid \theta)\,\pi(\theta \mid D)\, d\theta \qquad \forall x \in M_X$$

$f(\cdot \mid D)$ heißt Prädiktivdichte für X

32.1

32.2 A-posteriori-Bayes-Schätzer

Schätzwert $\hat{\theta}$ für $\theta \in \mathbb{R}$

<u>Def.</u>: Der A-posteriori-Bayes-Schätzer ist, falls er existiert, der Erwartungswert von $\tilde{\theta} \mid D$,

$$\hat{\theta} = \int_{\Theta} \theta\,\pi(\theta \mid D)\, d\theta$$

Für Vektorparameter $\theta = (\theta_1, \dots, \theta_k)$ und gerafften Parameter $\tau(\theta) \in \mathbb{R}$ ist der ApoBS für $\tau(\theta)$

$$\widehat{\tau(\theta)} := \mathbb{E}_D\,\tau(\tilde{\theta}) = \int_{\Theta} \tau(\theta)\,\pi(\theta \mid D)\, d\theta$$

falls $\exists$

32.2

<u>B1</u>: Für $\pi(\cdot) \mathrel{\hat{=}} U_{0,1}$ ist der ApoBS für den Anteil θ nach Beob. von $x_1, \dots, x_n$ folgendermaßen:

$$\pi(\theta|x_1,\dots,x_n) \propto \theta^{\sum_1^n x_i}(1-\theta)^{n-\sum_1^n x_i}\, I_{[0,1]}(\theta)$$

$$\hat{\theta} = \frac{\int_0^1 \theta \cdot \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}\,d\theta}{\int_0^1 \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}\,d\theta} = \dots = \frac{\sum_{i=1}^n x_i + 1}{n+2}$$

32.3 HPD-Bereiche

180

$\theta \in \Theta \subseteq \mathbb{R}^k$, $D = x_1, \dots, x_n \Rightarrow \pi(\cdot|D)$

vorgegebene Wahrscheinlichkeit $1-\alpha$ (Sicherheit)

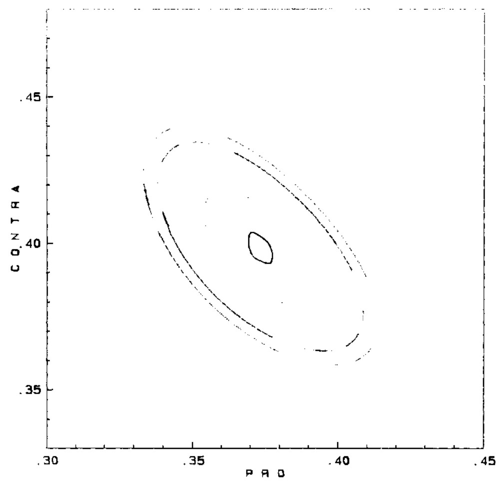
<u>Def.</u>: Ein HPD-Vertrauensbereich ist eine Teilmenge $\Theta^* \subseteq \Theta$, mit

$$\int_{\Theta^*} \pi(\theta|D)\,d\theta = 1-\alpha$$

und auf Θ^* ist $\pi(\theta|D)$ größtmöglich, d.h.

$\pi(\theta|D) \geq C \quad \forall \theta \in \Theta^*$, wobei C die größtmögliche Konstante ist

H P D - REGIONEN



32.4 A-posteriori-Wahrscheinlichkeiten von statistischen Hypothesen

$X \sim W_\theta$, $\theta \in \Theta$, $\pi(\cdot)$, $(x_1, \dots, x_n) = D$, $\pi(\cdot|D)$

Parameterhypothesen $\mathcal{H}_0: \theta \in \Theta_0 \subset \Theta$

$\mathcal{H}_1: \theta \in \Theta_1 \subset \Theta \setminus \Theta_0$

ApoW der Hypothesen

$$\alpha_0 := W\{\tilde{\theta} \in \Theta_0 | D\}$$

$$\alpha_1 := W\{\tilde{\theta} \in \Theta_1 | D\}$$

<u>Def.</u>: Relative Plausibilität $\frac{\alpha_1}{\alpha_0}$

Bem.: Für kontinuierlichen Parameterraum Θ gilt

$$W\{\Theta_j | D\} = \int_{\Theta_j} \pi(\theta|D)\,d\theta \quad \text{für } j = 0 \text{ bzw. } 1$$

und

$$\frac{\alpha_1}{\alpha_0} = \frac{\int_{\Theta_1} \pi(\theta|D)\,d\theta}{\int_{\Theta_0} \pi(\theta|D)\,d\theta} = \frac{\int_{\Theta_1} \pi(\theta)\,\ell(\theta; D)\,d\theta}{\int_{\Theta_0} \pi(\theta)\,\ell(\theta; D)\,d\theta} \qquad \text{Ü}$$

33. BAYES'sche ENTSCHEIDUNGEN

Entscheidungen oft mit Nutzen bzw. Verlust verbunden.

B1: Warenlieferung N Stück, θ Schlechtanteil

G Gewinn für jedes gute Stück

K Verlust für jedes schlechte Stück

d_0 Entscheidung Annahme der Lieferung

d_1 Entscheidung Ablehnung der Lieferung

$L(\theta, d_j)$ Verlust für Schlechtanteil und Entscheidung d_j

$$\Rightarrow \quad L(\theta, d_0) = \theta \cdot N \cdot K - (1-\theta) \cdot N \cdot G$$
$$L(\theta, d_1) = (1-\theta) \cdot N \cdot G$$

θ unbekannt, beschrieben durch SG $\tilde{\theta} \sim \pi(\cdot)$

Erhebung einer SP $D = (x_1, \cdots, x_n) \Rightarrow \pi(\cdot|D)$ von $\tilde{\theta}$

Entscheidungskriterium: Zu erwartender Verlust

$$\mathbb{E}_{\pi(\cdot|D)}\, L(\tilde{\theta}, d_j)$$

Def.: Die Bayes'sche Entscheidung ist jene, die kleineren a-posteriori zu erwartenden Verlust hat.

Bemerkung: Bei Wirtschaftsentscheidungen betrachtet man Nutzenfunktionen $U(\theta, d)$

Entscheidungskriterium ist dann die Maximierung des zu erwartenden Nutzens:

$$\overline{U}(d) = \mathbb{E}_{\pi(\theta|D)}\, U(\tilde{\theta}, d)$$

Für die optimale (Bayes'sche) Entscheidung d_{opt} muss gelten:

$$\overline{U}(d_{opt}) = \max\{\overline{U}(d) : d \in \mathcal{D}\}$$

wobei $\mathcal{D}$ die Menge der möglichen Entscheidungen ist

Bemerkung: In B1 gilt $\mathcal{D} = \{d_0, d_1\}$

33.1 <u>Beispiel</u>: Autohändler

Problem: Zahl der zu bestellenden Autos

Pro verkauftem Auto 5 GE Gewinn

Falls ein Auto nicht verkauft wird 3 GE Verlust

Stoch. Modell:

θ ... Anzahl der verkauften Autos, sG $\tilde{\theta}$

d ... Anzahl der bestellten Autos (Entscheidung)

P ... W-Vtlg. von $\tilde{\theta}$ auf $\mathbb{N}_0$

$$U(\theta, d_j) = \begin{cases} \theta \cdot 5 - (d_j - \theta)\cdot 3 & \text{für } \theta \leq d_j \\ d_j \cdot 5 & \text{sonst} \end{cases}$$

Mögliche Werte für d_j: $0, 1, 2, \dots, m$

Tabelle der Nutzenfunktion $U(\theta, d_j)$

nachgefragte Autos	Anzahl der bestellten Autos: 0	1	2	3	……	m
0	0	-3	-6	-9	⋯	$-3m$
1	0	5	2	-1	⋯	$5-3(m-1)$
2	0	5	10	7	⋯	$10-3(m-2)$
3	0	5	10	15	⋯	$15-3(m-3)$
⋮	⋮	⋮	⋮	⋮		⋮
k	0	5	10	15	⋯	$5k-3(m-k)$

Die W-Vtlg. P von $\tilde{\theta}$ ist aus der Erfahrung des Händlers zu finden (A-priori-Information).

Die optimale (Bayes'sche) Entscheidung ist jene Anzahl d_j, für die

$$\overline{U}(d_j) = \sum_{i=0}^{k} U(i, d_j)\,p(i) \quad \text{maximal ist.}$$

Beispiel: Übungen

33.2 Bayes-Schätzer bezüglich Verlust (S. 185)

Optimale Schätzwerte für Parameter θ von stoch. Modellen $X \sim W_\theta$, $\theta \in \Theta$ bei

Verlustfunktion $L(\theta, \hat{\theta})$

$\hat{\theta} = \vartheta(x_1, \dots, x_n)$ Schätzwert für $\theta \in \mathbb{R}$

<u>Beispiele</u> von Verlustfunktionen

$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$

$L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$

<u>Definition</u>: Der Bayes-Schätzer für den 1-dim. Parameter θ bezüglich einer Verlustfunktion $L(\cdot,\cdot)$ ist jener, der den Erwartungswert $\mathbb{E}_{\pi(\cdot|D)} L(\tilde{\theta}, \hat{\theta})$ des Verlusts minimiert.

Satz 33.1: Der Bayes-Schätzer $\hat{\theta}_B$ für einen 1-dim. Parameter θ ist unter quadratischer Verlustfunktion gleich dem Erwartungswert der A-posteriori-Verteilung, d.h.

$$\hat{\theta}_B = \mathbb{E}_{\pi(\cdot|D)}\,\tilde{\theta}$$

(ohne Beweis)

Satz 33.2: Der Bayes-Schätzer für $\theta \in \mathbb{R}$ bei Verlustfunktion $L(\theta,\hat{\theta}) = |\theta - \hat{\theta}|$ ist gleich dem Median der A-posteriori-Verteilung von $\tilde{\theta}$.

(ohne Beweis)

33.3 Bayes-Tests (S. 186)

Mit Hilfe der A-posteriori-Verteilung von $\tilde{\theta}$ kann man Wahrscheinlichkeiten von Parameterhypothesen berechnen:

$$\mathcal{H}: \theta \in \Theta_0 \;, \quad W\{\tilde{\theta} \in \Theta_0 \,|\, D\}$$

Tests mittels Quotienten von Wahrscheinlichkeiten von Hypothesen.

Außerdem kann man mit Hilfe von Verlustfunktionen sogenannte Bayes-Tests als Bayes'sche Entscheidungsregeln konstruieren.

VIII ERGÄNZUNGEN

34. UNSCHARFE INFORMATION 189 f.

Sowohl Beobachtungen SGn als auch A-priori-Verteilungen sind oft nicht exakt, also unscharf (engl. fuzzy).

34.1 Unscharfe Zahlen 9 f.

Jede exakte reelle Zahl x_0 ist eindeutig charakterisiert durch die Indikatorfunktion $I_{\{x_0\}}(\cdot)$.

34.1 Def.: Eine unscharfe Zahl x^* ist bestimmt durch ihre sog. charakterisierende Funktion $\xi(\cdot)$, die folgende Eigenschaften haben muss:

(1) $\xi : \mathbb{R} \to [0,1]$

(2) $\exists x_0 \in \mathbb{R} : \ \xi(x_0) = 1$

(3) $\forall \delta \in (0,1]$ ist $B_\delta(x^*) := \{x \in \mathbb{R} : \xi(x) \geq \delta\}$ ein abgeschlossenes endliches Intervall $[a_\delta, b_\delta]$ genannt δ-Schnitt

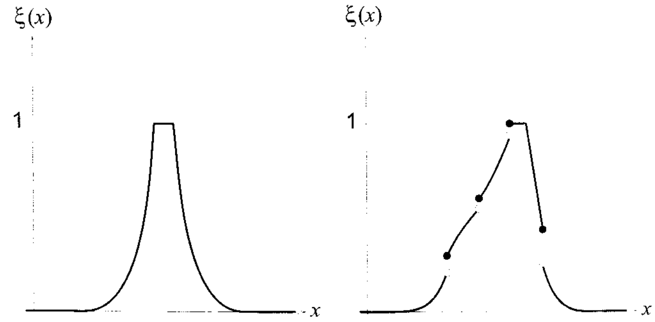
Bem.: Unscharfe Zahlen sind Spezialfälle sog. unscharfer Mengen (= fuzzy sets, ensembles flous)

Bem.: Char. Fn sind etwas anderes als Dichten

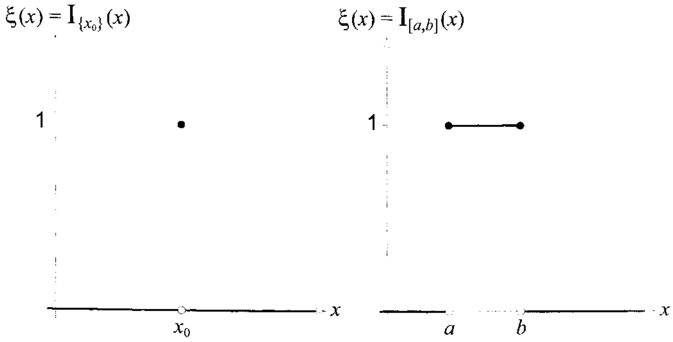
$\int_{\mathbb{R}} \xi(x)\,dx$ ist das Maß der Unschärfe

34.-

Charakterisierende Funktionen



Spezialfälle charakterisierender Funktionen



34.4

34.2 Unscharfe Stichproben

Konkrete SPn kontinuierlicher SGn sind endliche Folgen $x_1^*, \dots, x_n^*$ von unscharfen Zahlen.

Um Verfahren der schließenden Statistik zu adaptieren, ist die Kombination von $x_1^*, \dots, x_n^*$ zu einem sog. unscharfen Vektor im Stichprobenraum notwendig.

X mit Merkmalraum M_X, Stichprobenraum M_X^n

Spezialfall $M_X \subseteq \mathbb{R}$, $M_X^n \subseteq \mathbb{R}^n$

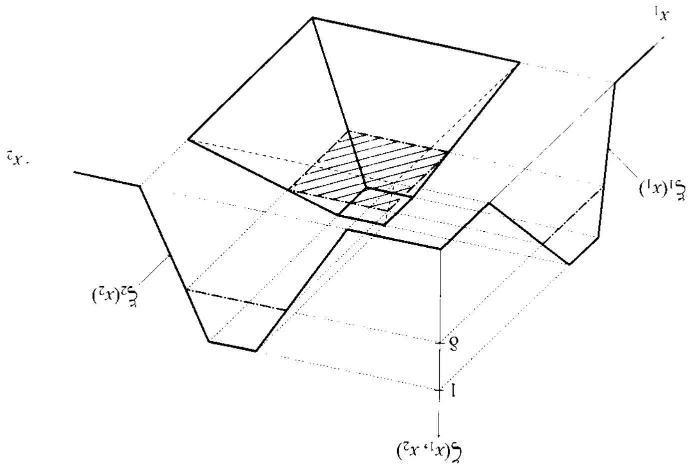
Ein unscharfer Vektor $\underline{x}^*$ ist durch seine sog. vektorcharakterisierende Funktion $\xi : \mathbb{R}^n \to [0,1]$ festgelegt.

Die Kombination von $\xi_1(\cdot), \dots, \xi_n(\cdot)$ erfolgt folgendermaßen:

$$\xi(x_1, \dots, x_n) := \min_{i=1(1)n} \xi_i(x_i) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

Bem.: δ-Schnitte von $\xi(\cdot, \dots, \cdot)$:

$$C_\delta(\underline{x}^*) = \mathop{\times}_{i=1}^{n} C_\delta(x_i^*) \quad \forall \delta \in (0,1]$$



34.3 Verallgemeinerte Schätzungen

Schätzfunktion $\vartheta : M_X^n \to \Theta$

bei exakten Beobachtungen $x_1, \dots, x_n$

Schätzwert $\hat{\theta} = \vartheta(x_1, \dots, x_n)$

Für unscharfe Beobachtungen $x_1^*, \dots, x_n^*$ und kombin. unscharfen Vektor $\underline{x}^*$ mit vcF $\xi(\cdot, \dots, \cdot)$ ergibt $\vartheta(x_1^*, \dots, x_n^*)$ einen unscharfen Wert $\hat{\theta}^*$

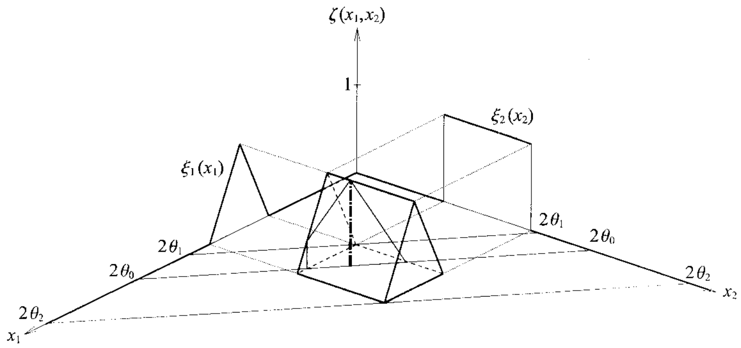
Die char. F. von $\hat{\theta}^*$ erhält man mittels des sog. Erweiterungsprinzips aus der FST:

Für die c.F. $\psi(\cdot)$ von $\hat{\theta}^*$ gilt mit $\underline{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$

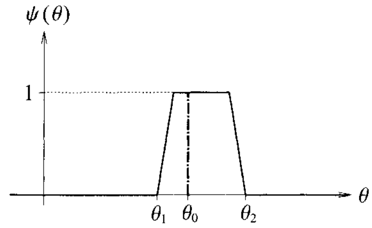
$$\psi(\theta) = \begin{cases} \sup\{\xi(\underline{x}) : \vartheta(\underline{x}) = \theta\} & \text{falls } \vartheta^{-1}(\theta) \neq \emptyset \\ 0 & \text{falls } \vartheta^{-1}(\theta) = \emptyset \end{cases} \quad \forall \theta \in \Theta$$

Bem.: 1) Für $\theta \in \mathbb{R}$ und stetige Funktion $\vartheta(\cdot, \dots, \cdot)$ ist obiges $\psi(\cdot)$ eine c.F. einer unscharfen Zahl.

2) Für Vektorparameter erhält man ein unscharfes Element (unscharfe Teilmenge) von Θ.



$$\hat{\theta}^* = \frac{x_1^* \oplus x_2^*}{2}$$



Verallgemeinerte Addition

$x^* \triangleq \xi(\cdot)\,,\ y^* \triangleq \eta(\cdot) \ \Rightarrow\ x^* \oplus y^* \triangleq \psi(\cdot)$

$$\psi(z) := \begin{cases} \sup\limits_{(x,y)} \{ \min[\xi(x), \eta(y)] : x+y=z \} \\ 0 \quad \text{falls } \nexists (x,y) : x+y=z \end{cases} \quad \forall z \in \mathbb{R}$$

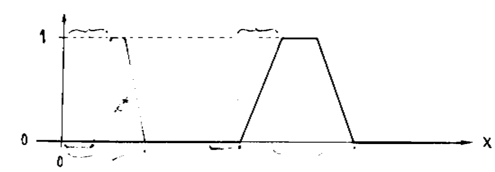
Bemerkung: Berechnung auch über δ-Schnitte möglich

Multiplikation mit einer Konstanten

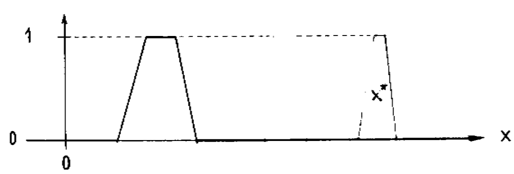
$x^* \triangleq \xi(\cdot)\,,\ c \in \mathbb{R} \ \Rightarrow\ c \odot x^* \triangleq \eta(\cdot)$

$\eta(x) := \xi\left(\frac{x}{c}\right) \quad \forall\ x \in \mathbb{R} \quad \text{falls} \quad c \neq 0$

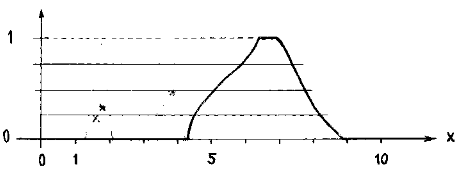
Addition $x^* \oplus y^*$

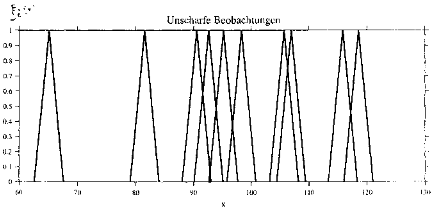


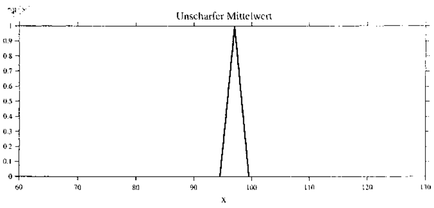
Subtraktion $x^* \ominus y^*$



Multiplikation $x^* \odot y^*$







Satz 34.1: Sind die c.F.n einer unscharfen SP alle von derselben symmetrischen Gestalt, nur mit verschiedenem Lagekoordinaten m_i, $i=1(1)n$, so ist die c.F. des unscharfen Stichprobenmittels auch von derselben Gestalt, wobei die Lagekoordinate von $\overline{x}_n^*$ durch

$$m = \frac{1}{n}\sum_{i=1}^{n} m_i \quad \text{gegeben ist.}$$

(ohne Beweis)