Statistik und Wahrscheinlichkeitstheorie

Begleitmaterial zur Vorlesung

WS 2002/2003

O.Univ.Prof. Dipl.-Ing. Dr.techn. Reinhard VIERTL
Institut für Statistik und Wahrscheinlichkeitstheorie
Technische Universität Wien

STATISTIK UND WAHRSCHEINLICHKEITSTHEORIE

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
Wahrscheinlichkeitstheorie

Stochastik

<u>Wozu Statistik u. WR</u> ?

„Quantitative Erfassung von Massenphänomenen und Beschreibung von nichtdeterministischen Vorgängen"

1. Historisches und Grundsätzliches

<u>Stat. Erhebungen</u>: 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)

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

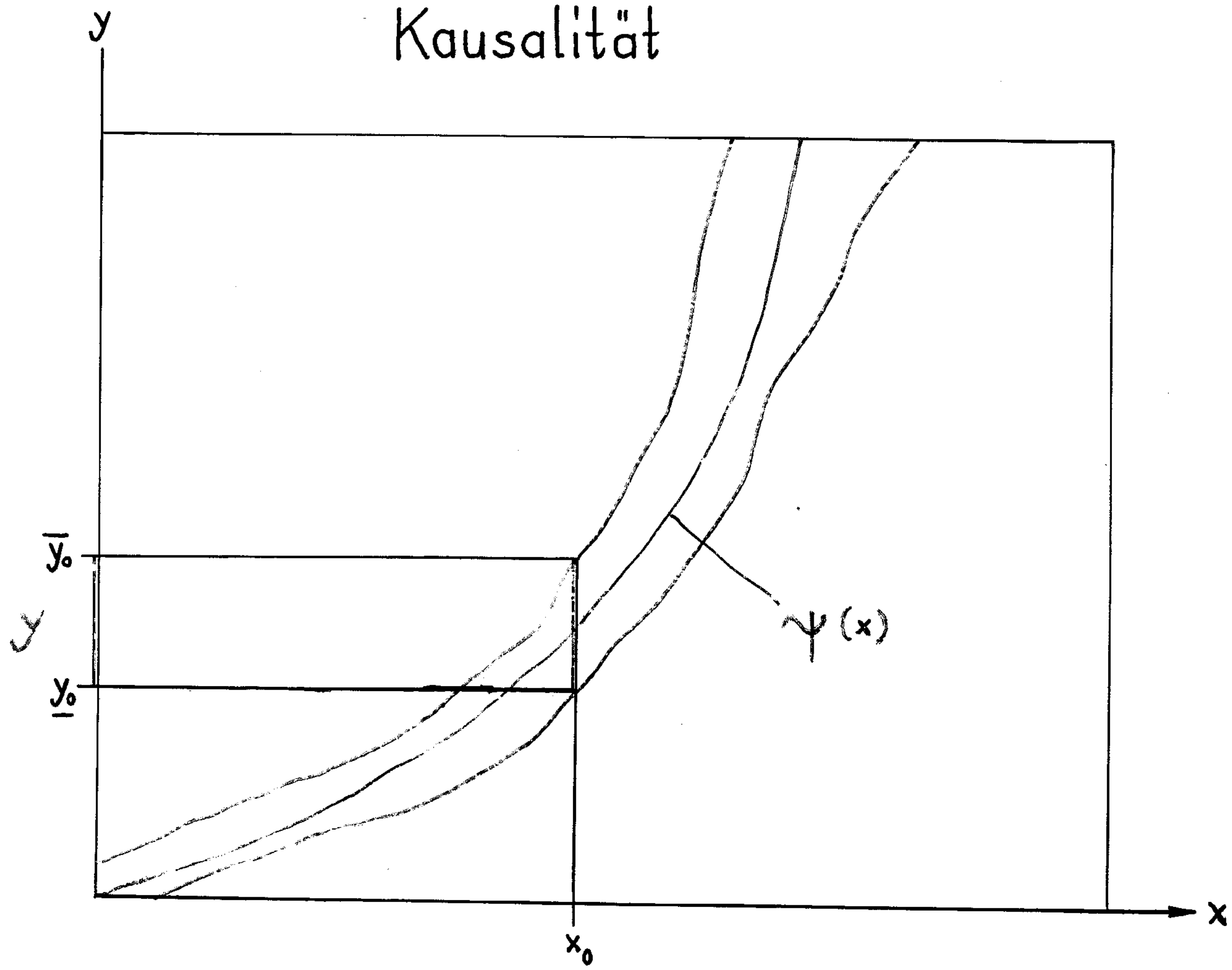
<u>Wahrscheinlichkeitsrechnung</u>: Im 16. Jhdt.
Beschr. von Glücksspielen
Beschr. von Unsicherheit

verschiedene W-Begriffe : objektivistisch
subjektivistisch
axiomatisch

Stochastik: στοχαστικός
στοχαζεσθαι

Kausalität : deterministisch, stochastisch, fuzzy

Kausalität



Begleitende **B**eispiele:

<u>B1</u>: Stat. Qualitätskontrolle

N … Losumfang, A … Anzahl schlechter Stücke

n … Stichprobenumfang, a —"— in der SP

Problem: Rückschluss auf A

<u>B2</u>: Lebensdauer eines Produktes

<u>B3</u>: Wartezeit bei einer Bedienstelle

<u>B4</u>: Abhängigkeit einer stochastischen Größe von einer Kovariablen

2. Beschreibende Statistik

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

2.1.2 <u>Ordnungsmerkmale</u>

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

2.3

2.2 <u>Kontinuierliche Merkmale</u>

können alle Zahlen aus einem Intervall annehmen

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

Darstellung der Häufigkeitsverteilung

<u>Empirische Verteilungsfunktion</u> $F_n^*(\cdot)$

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

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

2.4

7. X. '12

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

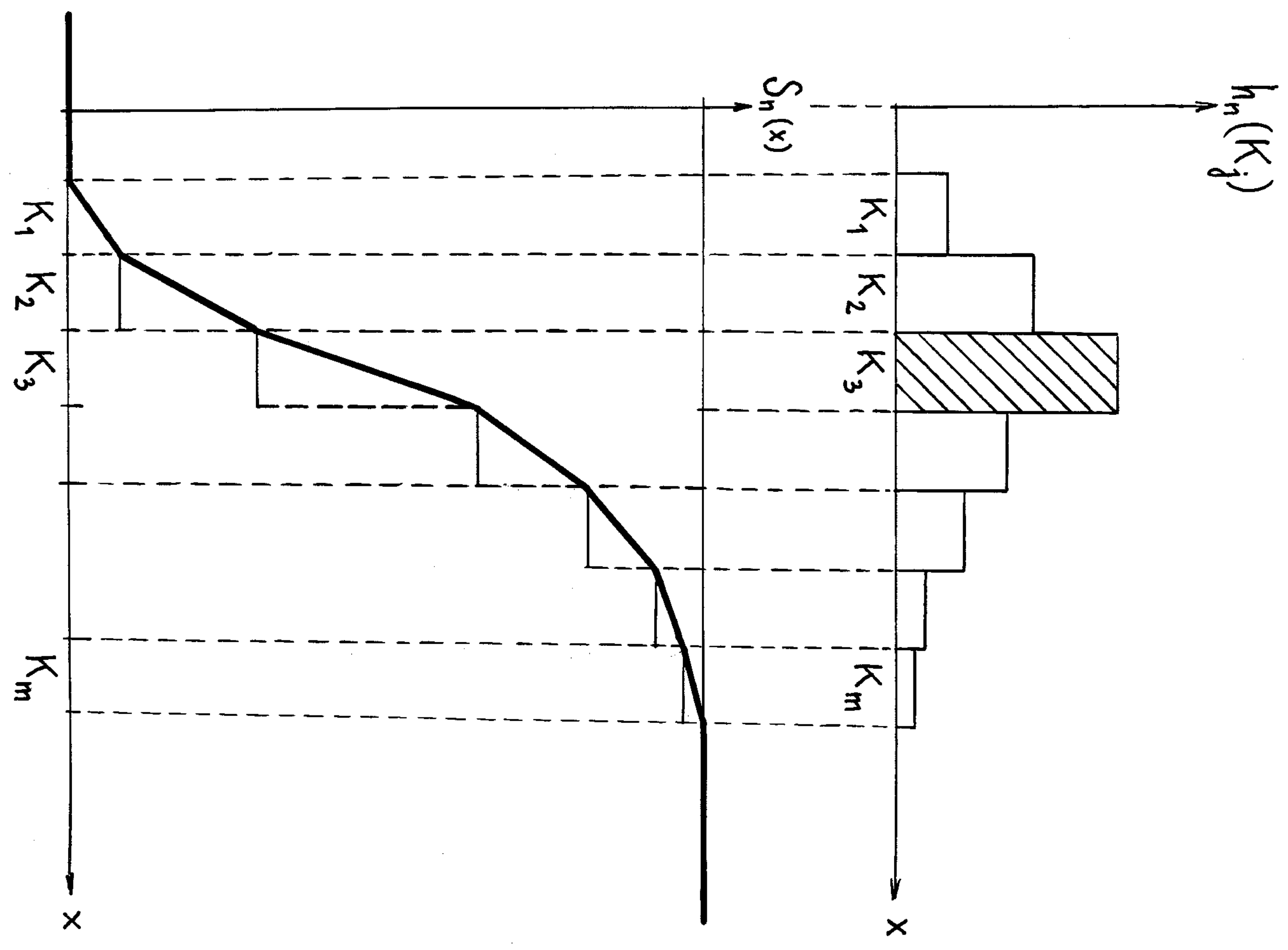
relative Häufigkeiten der Klassen

Darstellung: <u>Histogramm</u>

<u>Summenpolygon</u>

Theoretisches Modell: <u>Wahrscheinlichkeitsdichte</u>

<u>Verteilungsfunktion</u>



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 $A = \{a_{i_1}, \dots, a_{i_g}\}$, g Elemente

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

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

Anwendung: Qualitätskontrolle

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

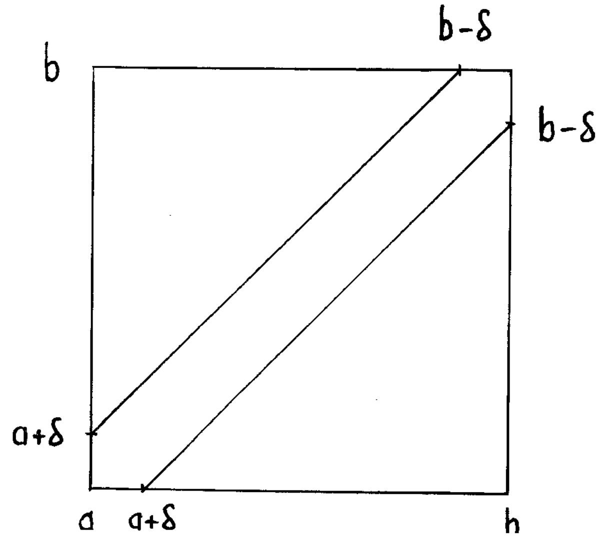
$$W(A) := \frac{I(A)}{I(M)}$$

Bem.: $k=1$ Länge
$k=2$ Fläche
$k=3$ Volumen

3.3

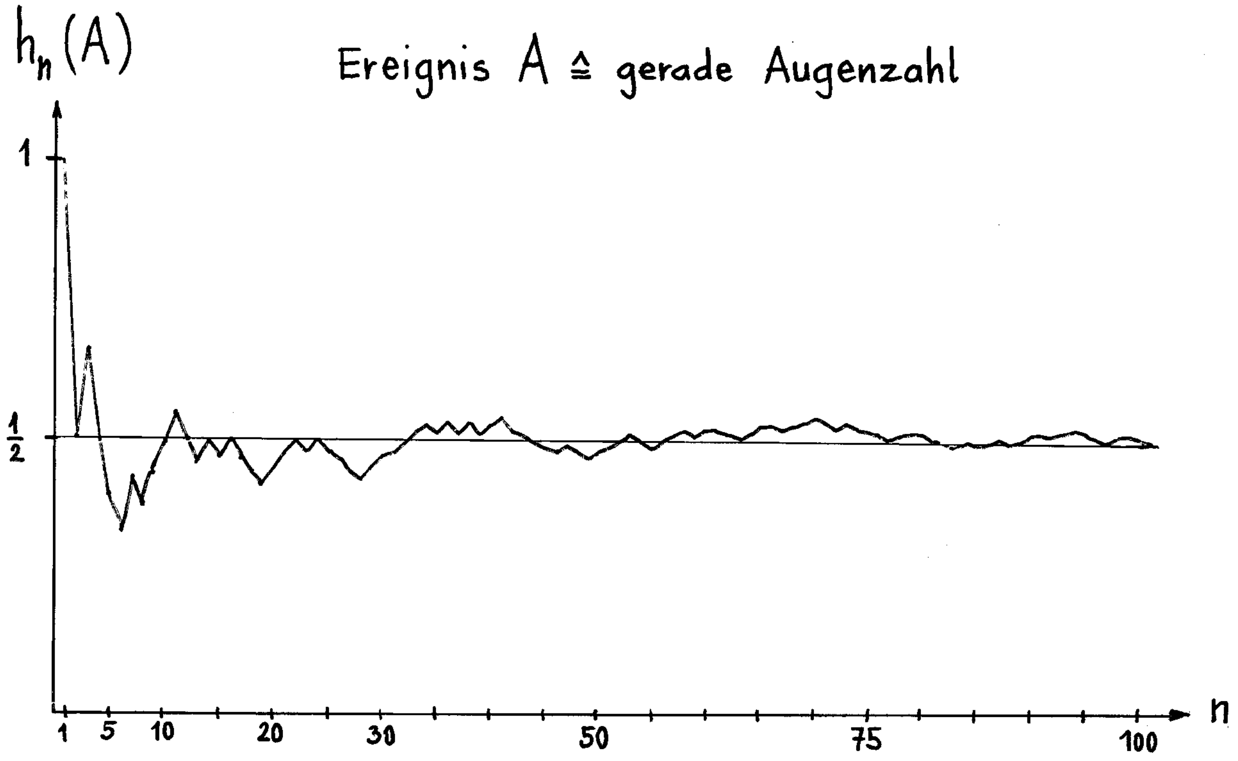
Bsp.: Zufällige Ankunft in $[a,b]$
Teilintervall $[a_1,b_1] \subset [a,b]$

Bsp.: Rendezvous-Problem
2 Ankünfte in $[a,b]$, unabhängig



$$|t_1 - t_2| \leq \delta$$

3.4



3.3 Grenzwert von Häufigkeiten

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

Empirisches Gesetz der großen Zahlen

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

$\mathcal{O}$ unmögliches Ereignis

e sicheres Ereignis

$A^\perp$ das zu A komplementäre Ereignis

$A \wedge A^\perp = \mathcal{O}$ und $A \vee A^\perp = e$

$A_1, A_2, A_3, \dots$ 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 $\mathcal{A}$

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

$W(\mathcal{O}) = 0$ und $W(e) = 1$

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

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

Bem.: Endliche Additivität folgt daraus

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

3.5 Subjektive Wahrscheinlichkeiten

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

Ereignisse: A, B ; es muss gelten:

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

(2) Falls A und B einander ausschließen, gilt

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

(3) Falls neue Information verfügbar:

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

II

STOCHASTISCHE GRUNDBEGRIFFE

4. Wahrscheinlichkeitsräume

Spezialfall axiom. W. , M Merkmalraum
Spezialfall von $\mathcal{A}$: <u>Ereignisfeld</u> $\mathcal{E}$ 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

Wahrscheinlichkeitsverteilung W auf $\mathcal{E}$:

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

$W(M) = 1$

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

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

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

4.1 Diskrete W-Räume

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

reelle Zahlen $p_1, p_2, \dots$ mit $p_i \geqq 0$, $p_i = W(a_i)$

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

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

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

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

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

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

$B = (A \cap B) \cup (A^c \cap B)$

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

(2) vollständige Induktion nach n

4.5 Bedingte Wahrscheinlichkeit

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

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

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

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

4.6 Multiplikationstheorem für Wahrsch.

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

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

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

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

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

(2) vollständige Induktion nach n.

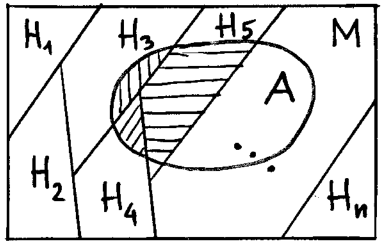
4.7 Satz von der vollständigen Wahrscheinl. und Bayes'sche Formel

Ist $H_1, \dots, H_n$ eine Zerlegung von M in Ereignisse, so gilt für jedes Ereignis A:

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

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

Beweis: (1)



$$A = A \cap M = A \cap \left[\bigcup_{k=1}^{n} H_k\right] = \bigcup_{k=1}^{n} (A \cap H_k)$$

da $H_1, \dots, H_n$ paarweise disjunkt

$\Rightarrow A \cap H_1, \dots, A \cap H_n$ paarweise disjunkt

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

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

(1) heißt Satz v. d. vollst. Wahrsch.

(2) $$W(H_i|A) = \frac{W(A \cap H_i)}{W(A)} = \frac{W(A|H_i) \cdot W(H_i)}{\sum_{k=1}^{n} W(A|H_k) \cdot W(H_k)}$$

Bayes'sche Formel

Bem.: $W(H_k)$ ... A-priori-Wahrsch. der H_k

$W(H_k|A)$ ... A-posteriori-Wahrsch. der H_k (nach dem Eintritt von A)

Anwendung: QK, Warensendung, gleiche Produkte von n Maschinen mit versch. Ausschusshäufigkeiten (Wahrsch.)

H_k ... k-te Maschine, $k = 1(1)n$

A ... Ausschussstück

$W(A|H_k)$

Zusammengesetzte Warensendung

$W(H_k)$ = Wahrsch. Stück von Maschine k

Frage 1: Wahrsch., dass ein Stück aus der Warensendung Ausschuss ist?

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

Frage 2: Wahrsch., dass ein Ausschussstück von der i-ten Maschine stammt?

$W(H_i|A) =$ Bayes'sche Formel

5. Stochastische Unabhängigkeit

Beispiel: QK, mit Zurücklegen
ohne — " —

Zusammengesetzte Versuche

$(M_1, \mathcal{E}_1, W_1)$, $(M_2, \mathcal{E}_2, W_2)$, $A_i \in \mathcal{E}_i$

$(M_1 \times M_2, \mathcal{E}_1 \otimes \mathcal{E}_2, W_1 \otimes W_2)$

$W_1 \otimes W_2 (A_1 \times A_2) := W_1(A_1) \cdot W_2(A_2)$

$\exists 1$ W-Vtlg. $W_1 \otimes W_2$ auf $\mathcal{E}_1 \otimes \mathcal{E}_2$

Def.: Zwei Ereignisse A und B eines W-Raumes heißen unabhängig, i.Z. $A \perp\!\!\!\perp B$, wenn gilt:

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

Es gilt:

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

Bew.: 1) $A \perp\!\!\!\perp B \Rightarrow W(A \cap B) = W(A) \cdot W(B)$

$\Rightarrow W(A|B) = W(A)$; zu $W(A|B^c)$:

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

$\Rightarrow W(A \cap B^c) = W(A)[1 - W(B)] = W(A) \cdot W(B^c)$

2) Umkehrung: Ü

Für mehrere Ereignisse: $A_1, \dots, A_n$ heiße unabhängig, wenn $\forall$ Teilfamilien $A_{j_1}, \dots, A_{j_k}$ gilt:

$$W\left(\bigcap_{i=1}^{k} A_{j_i}\right) = \prod_{i=1}^{k} W(A_{j_i}), \quad j_i \in \{1, \dots, n\}$$

B1: Stichprobe mit Zurücklegen aus N
n Stücke, A schlechte Stücke

$W(\{(x_1, \dots, x_n)\}) = W(\{x_1^\circ\} \cap \dots \cap \{x_n^\circ\}) = \prod_{i=1}^{n} W_e(\{x_i\})$

Sind genau k Einser im n-Tupel $\Rightarrow$

$W(\{(x_1, \dots, x_n)\}) = p^k (1-p)^{n-k}$, $k = 0(1)n$

$p = \frac{A}{N}$

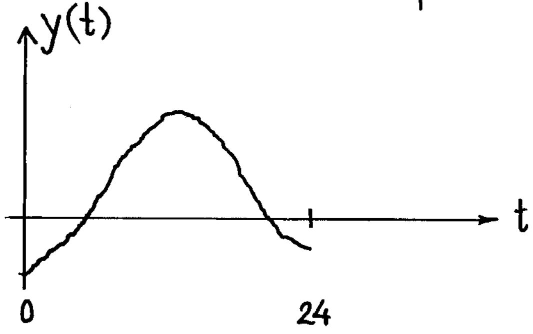
Bem.: Wesentlich ist die Anzahl schlechter Stücke

$$\sum_{i=1}^{n} x_i$$

6. Stochastische Größen

Alle nichtdeterministischen Größen

Beispiel: Minimaltemperatur innerhalb der nächsten 24 Stunden, Funktion des Temperaturverlaufes $y(t)$



$$X = \min_{0 \le t \le 24} y(t)$$

2-dimensionale stochastische Größe

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} \max y(t) \\ \min y(t) \end{pmatrix} = \underline{X} \quad \text{stoch. Vektor}$$

B1: X = Anzahl schlechter Stücke in den n ausgewählten

$$M_X = \{0, 1, 2, \ldots, n\}$$

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

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

Bem.: Stoch. Größen als Funktionen auf W-Räumen aufgefasst

Definition: X auf $(M, \mathcal{E}, W)$: $\mathcal{B}_k$ Borel-

$$X : M \longrightarrow \mathbb{R}^k \quad \text{mit } X^{-1}(B) \in \mathcal{E} \quad \forall B \in \mathcal{B}$$

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

Es gilt: Im Falle 1-dim. st. Gr. X, d.h. $k = 1$, gilt (hinreichend)

$$X^{-1}((-\infty, a]) = \{\omega : X(\omega) \le a\} \in \mathcal{E} \quad \forall a \in \mathbb{R}$$

W-Vtlg. von X:

$$W_X\{X \in B\} := W(X^{-1}(B)) \quad \forall B \in \mathcal{B}$$

B1: Punktwahrscheinlichkeiten

$$W\{X = k\} = W_{prod}(X^{-1}(\{k\})) =$$

$$= W_{pr.}\left(\left\{(x_1, \ldots, x_n) : \sum_{i=1}^{n} x_i = k\right\}\right) = W_{pr.}\left(\bigcup_{\substack{(x_1, \ldots, x_n): \\ \sum x_i = k}} \{(x_1, \ldots, x_n)\}\right)$$

$$= \sum_{\substack{(x_1, \ldots, x_n): \\ \sum x_i = k}} \underbrace{W_{prod}(\{(x_1, \ldots, x_n)\})}_{= p^k (1-p)^{n-k}} = \binom{n}{k} p^k (1-p)^{n-k}$$

für $k = 0(1)n$

III EINDIMENSIONALE VERTEILUNGEN

7. Verteilungsfunktionen

<u>Def.</u>: Die Verteilungsfunktion von X, $F(\cdot$

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

<u>Bem.</u>: $\{X \leq x\} := X^{-1}((-\infty, x])$

7.1 <u>Eigenschaften von VF</u> $F(\cdot)$

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

(2) $\quad x_1 < x_2 \implies F(x_1) \leq F(x_2)$

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

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

Rechtsstetigkeit

<u>Es gilt</u>: Hat X die VF $F(\cdot)$, so folgt

(a) $\quad W\{X = x\} = F(x) - \lim\limits_{h \downarrow 0} F(x-h)$

(b) $\quad W\{a < X \leq b\} = F(b) - F(a)$

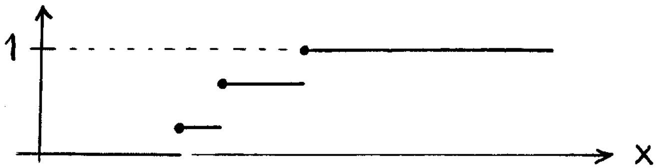
Begr. von (b): Ereignis $(a, b] = (-\infty, b] \setminus (-\infty, a]$

$$\Rightarrow W((a,b]) = W\big((-\infty, b] \setminus (-\infty, a]\big)$$

$$= W((-\infty, b]) - W((-\infty, a]) = F(b) - F(a) \qquad //$$

7.2 Typen von VF

a) treppenförmig



b) stetige, stückw. diffb.:

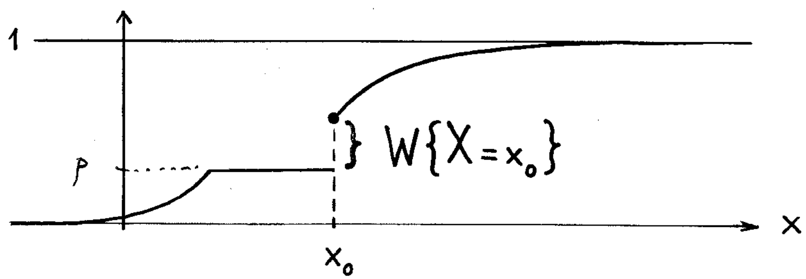
$$F(x) = \int_{-\infty}^{x} f(t)\,dt \qquad \forall x \in \mathbb{R}$$

c) Mischungen aus a) und b):

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

$$p \in (0,1)$$

$F_1(\cdot)$ treppenf., $F_2(\cdot)$ stetig

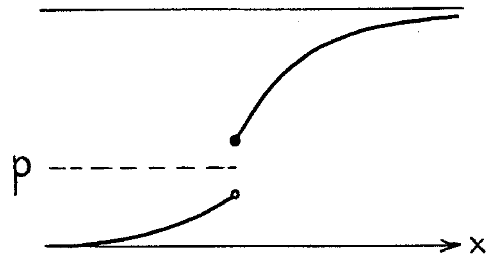


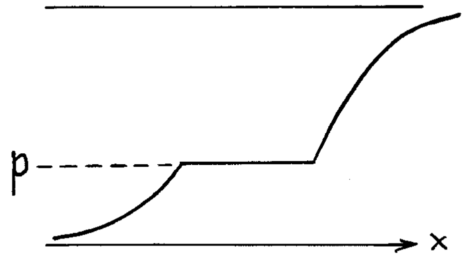
Def.: p-Fraktil x_p zur VF $F(\cdot)$:

$$F(x_p) = p \quad \text{mit} \quad p \in (0,1)$$

Bem.: 1) $X \sim F(\cdot) \Rightarrow W\{X \leq x_p\} = p$

2) p-Fraktile müssen weder existieren, noch eindeutig sein





21.III.'11

8. Diskrete 1-dim. Verteilungen von stoch. Größen X

M_X höchstens abzählbar, $M_X \subset \mathbb{R}$
ohne Häufungspunkt

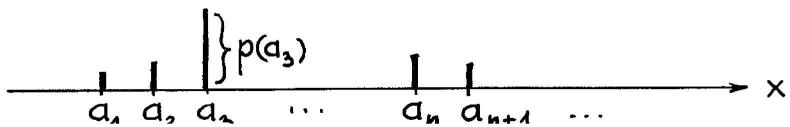
$$M_X = \{a_i : i \in I \subseteq \mathbb{N}\}$$

$$W\{X = a_i\} = W(\{a_i\}) = p(a_i) \geq 0$$

$$\text{mit} \quad \sum_{i \in I} p(a_i) = 1$$

$p(a_i)$ ... Punktwahrscheinlichkeit von a_i

Stabdiagramm:



Bem.: VF dazu ist eine Treppenfunktion
$W(\{x\}) = 0 \quad \forall x \notin M_X$

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

$X \sim \delta_\mu :\Leftrightarrow W\{X=\mu\} = 1$ Determinismus

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

$M = \{1, 2, \dots, m\}$

$p(k) = \frac{1}{m} \quad \forall k \in M$

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

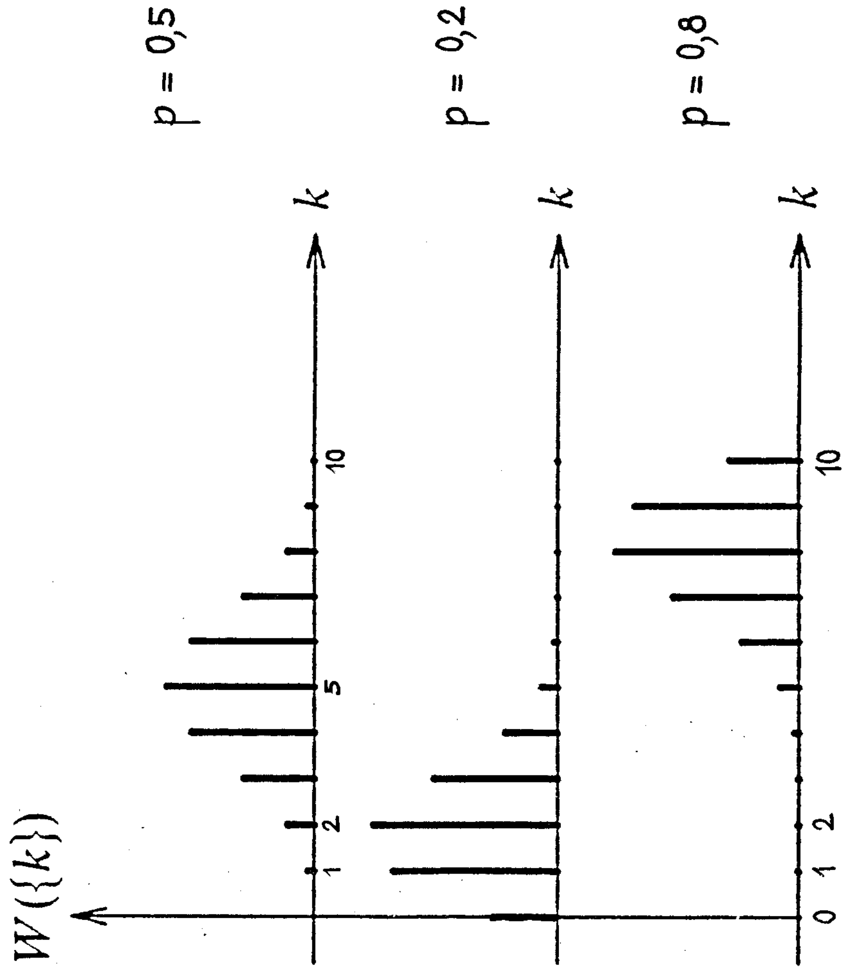
$M = \{0, 1\}$

$W\{X=1\} = p$
$W\{X=0\} = 1-p$ $\quad \Leftrightarrow: \quad X \sim A_p$

8.4 Binomialverteilung $B_{n,p}$

Ziehungen mit Zurücklegen
n unabhängige Alternativversuche
X = Anzahl der Einser $\Rightarrow X \sim B_{n,p}$

$M_X = \{0, 1, 2, \dots, n\}$
$W\{X=k\} = \binom{n}{k} p^k (1-p)^{n-k} \quad \forall k \in M_X$



Stabdiagramme verschiedener Binomialverteilungen $B_{10,p}$

8.5 Hypergeometrische Vtlg. $H_{N,A,n}$

Ziehungen ohne Zurücklegen
N Elemente, A Ausschussstücke
n Stück gezogen (Stichprobenumfang $n < N$)
X Anzahl der Ausschussstücke in der SP

$$M_X = \{k_1, k_1+1, k_1+2, \dots, k_2\}$$

mit $k_1 = \max(0, n-(N-A))$
$k_2 = \min(n, A)$

$$W\{X=a\} = \frac{\binom{A}{a}\binom{N-A}{n-a}}{\binom{N}{n}} \quad \text{für } a = k_1(1)k_2$$

Begründung: klass. W-Def., $\binom{N}{n}$ mögliche Auswahlen ohne Wiederholung
$\binom{A}{a}$ Auswahlen von a schlechten Stücken
$\binom{N-A}{n-a}$ Auswahlen von n−a guten Stücken

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

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

Übung: Vergleich der Punktwahrscheinl.

8.6 Poisson-Vtlg. P_μ, $\mu > 0$

$$X \sim P_\mu :\Leftrightarrow M_X = \mathbb{N}_0$$

$$W\{X=k\} = \frac{\mu^k e^{-\mu}}{k!} \quad \text{für } k = 0(1)\infty$$

Bem.: $0! := 1$, Exponentialreihe $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

Approximationen

$$B_{n,p} \approx P_{n \cdot p} \quad \text{für } n > 50,\ p < \frac{1}{10}$$

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

8.7 Geometrische Vtlg. G_p, $0 < p < 1$

Alternativversuch so oft unabhängig durchgeführt bis erstmals „1" kommt

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

Bem.: Geometrische Reihe

$$\sum_{n=1}^{\infty} p(1-p)^{n-1} = 1$$

9. Kontinuierliche 1-dim. Verteilungen bzw. kontin. stoch. Größen

Kontin. stoch. Größen X nehmen alle Werte eines Intervalls an, (stetige stoch. Gr.)
Ihre W-Vtlg. wird durch eine sog. Dichtefunktion

$$f: \mathbb{R} \longrightarrow [0,\infty)$$

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

bestimmt.

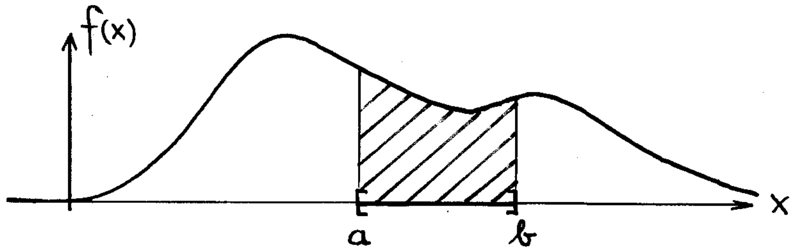
$$W\{a \leq X \leq b\} := \int_a^b f(x)\,dx$$

zugehörige Verteilungsfunktion $F(\cdot)$

Es gilt: 1) $F(x) = \int_{-\infty}^{x} f(t)\,dt$

2) $f(x) = \frac{d}{dx} F(x) \quad \forall x$ wo $F(\cdot)$ differenzierbar

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

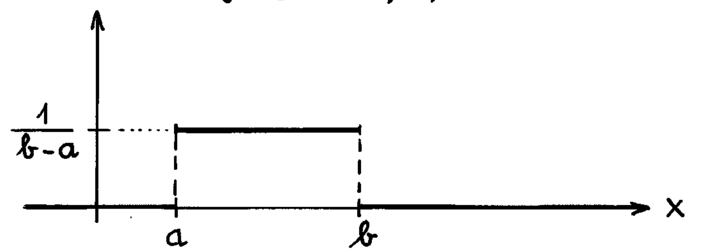


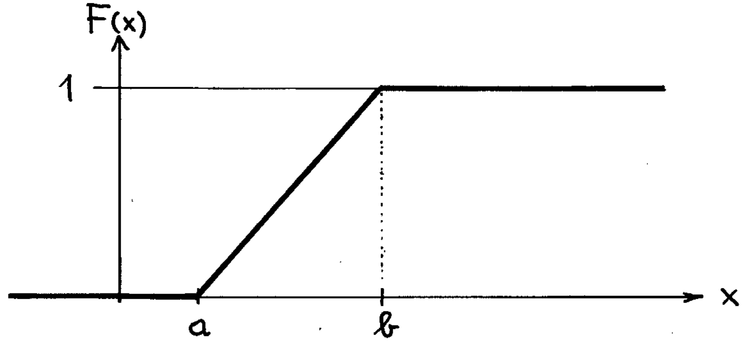
Definition: Die Indikatorfunktion $I_A(\cdot)$ einer Teilmenge $A \subseteq M$ ist

$$I_A(x) = \begin{cases} 1 & \text{für } x \in A \\ 0 & \text{für } x \notin A \end{cases}$$

9.1 Uniforme Vtlg. $U_{a,b}$, $a < b \in \mathbb{R}$

$$f(x) = \frac{1}{b-a} I_{(a,b)}(x) \qquad \forall x \in \mathbb{R}$$





9.2 Exponentialverteilung Ex_τ , $\tau > 0$

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

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

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

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

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

Es gilt: $\Phi(-x) = 1 - \Phi(x) \quad \forall x > 0$

→ Abb.

9.4 Allgem. Normalvtlg. $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}$, $\sigma^2 > 0$

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

→ Abb.

Satz 9.1: Für die VF $F(\cdot|\mu,\sigma^2)$ der $N(\mu,\sigma^2)$ gilt:

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

Beweis: Integral $\int_{-\infty}^{x} f(t)\,dt$ transformieren

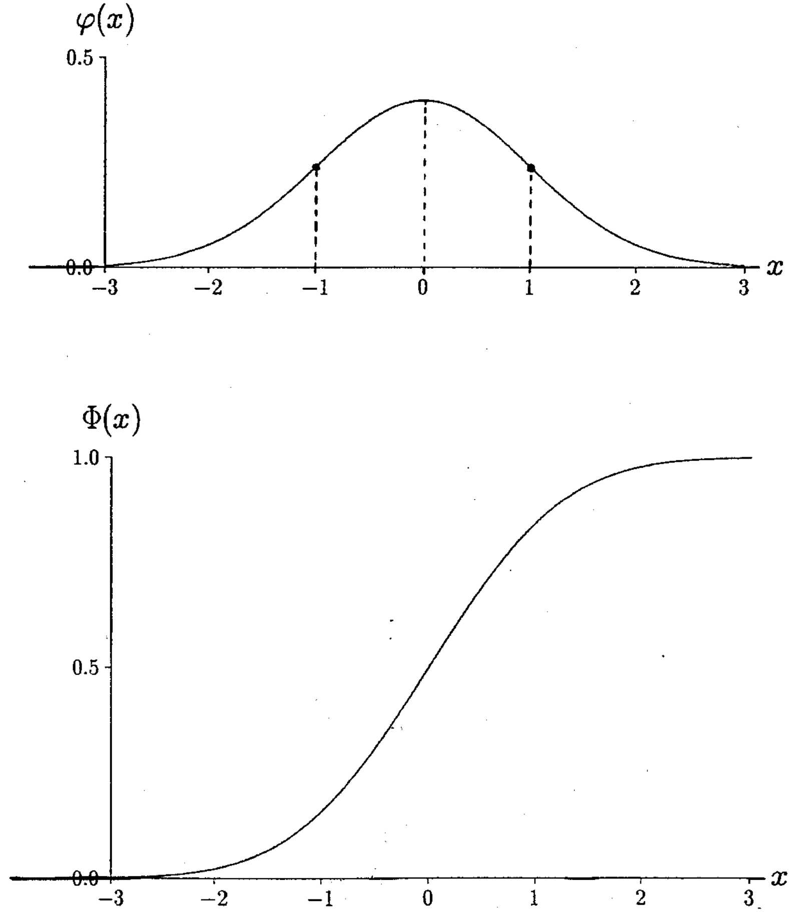
Substitution $\frac{t-\mu}{\sigma} = u$, $\quad dt = \sigma\,du$

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

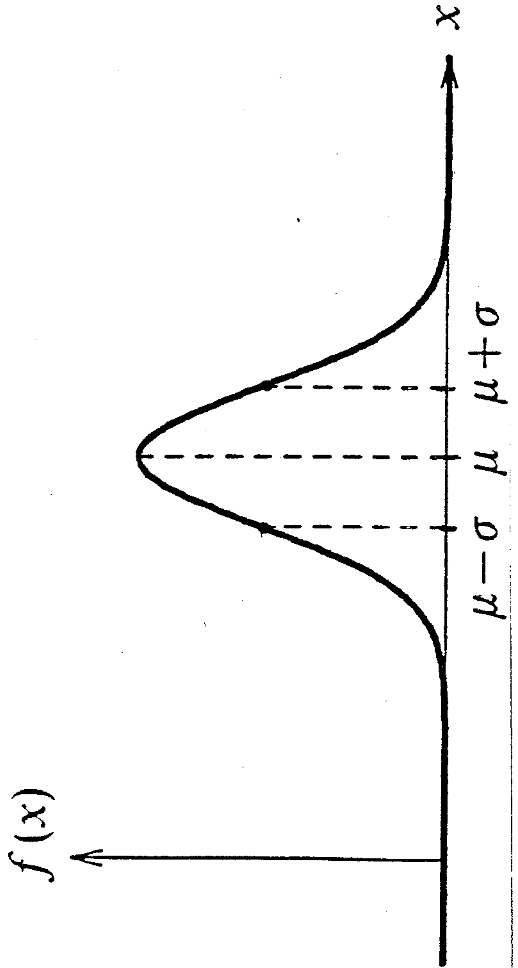
$$= \Phi\left(\frac{x-\mu}{\sigma}\right) \quad //$$

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

N(0,1)-Verteilung



Normalverteilung



9.5 Logarithmische Normalverteilung

$LN(\mu,\sigma^2)$, $\mu\in\mathbb{R}$, $\sigma^2>0$

$$X \sim LN(\mu,\sigma^2) :\Longleftrightarrow \begin{array}{l} M_X = (0,\infty) \\ \ln X \sim N(\mu,\sigma^2) \end{array}$$

Verteilungsfunktion

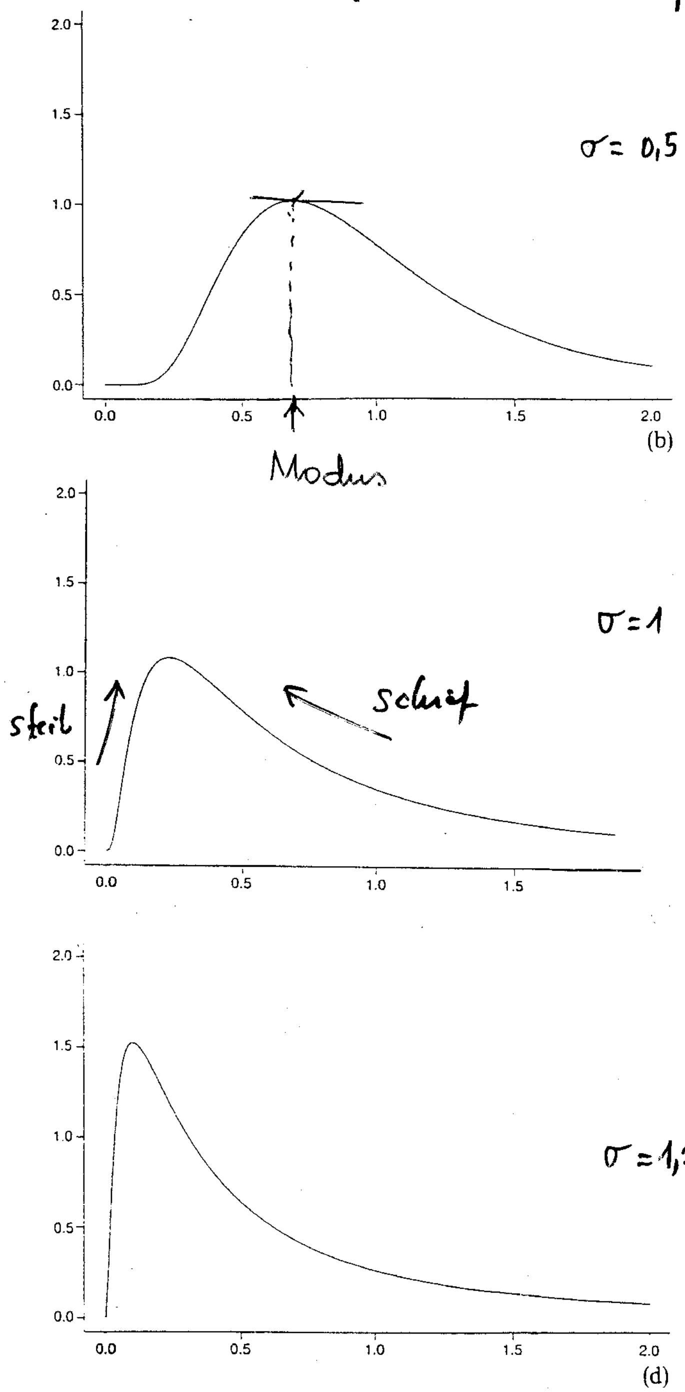
$$F_X(x) = W\{X \le x\} = W\{\ln X \le \ln x\}$$

$$= F_{N(\mu,\sigma^2)}(\ln x) = \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \quad \forall x > 0$$

Dichtefunktion $\quad f(x) = F'(x)$

→ Diagramm

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

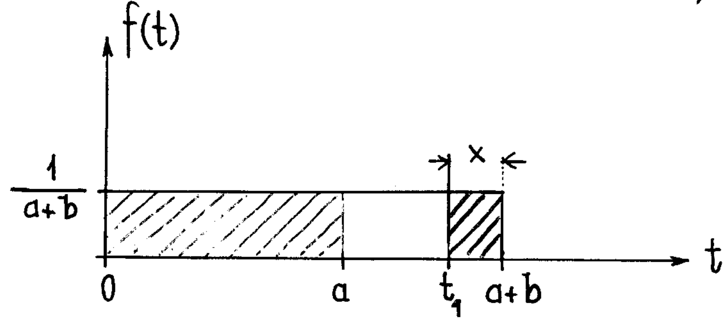


10. Gemischte Verteilungen

Beispiel: Wartezeit, Ampel

Grünphase a [sec.], Rotphase b [sec.]

Ankunft zufällig ... $T \sim U_{0,\, a+b}$



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

W-Vtlg. von X :

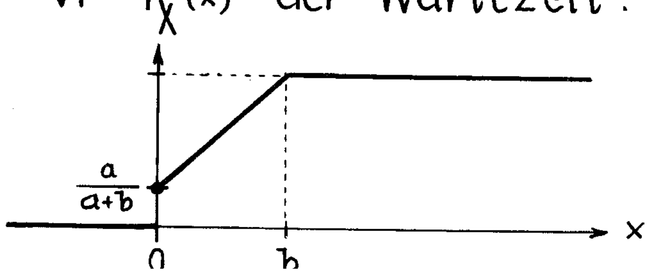
$W\{X=0\} = W\{T \in [0,a)\} = \frac{a}{a+b} > 0$

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

für $0 < x < b$

$\Rightarrow W\{X \leq x\} = \frac{a}{a+b} + \frac{x}{a+b}$ für $0 \leq x \leq b$

VF $F_X(x)$ der Wartezeit:



Bem.: modifizierte Dichte

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

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

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

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

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., $\quad \psi: \mathbb{R} \to \mathbb{R}$ (messbar)

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

Wert von $Y :=$ ψ(Wert von X)

$\Rightarrow \quad Y$ Stoch. Gr.

Frage: $\mathbb{E}Y$

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

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

12. FUNKTIONEN VON STOCH. GRÖSZEN – ERWARTUNGSWERT

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

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

Wert von $Y :=$ ψ(Wert von X)

$\Rightarrow \quad Y$ Stoch. Gr.

Frage: $\mathbb{E}Y$

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

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

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

Varianz von X bzw. Varianz der Verteilung

Berechnung mittels des SvuStat.

Für gemischte Verteilungen:

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

Bsp.: Wartezeit aus Abschnitt 10

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

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

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

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

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

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

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

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

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

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

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 <u>Streuung von X</u>

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

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

Es gilt: (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 $\mathbb{E}\left(\frac{1}{\sigma}(X-\mu)\right) = \dots$

13. GAMMAFUNKTION UND ZWEI VERTEILUNGEN

$n \in \mathbb{N} \;\Rightarrow\; n! := n(n-1)(n-2)\ldots 3\cdot 2\cdot 1$

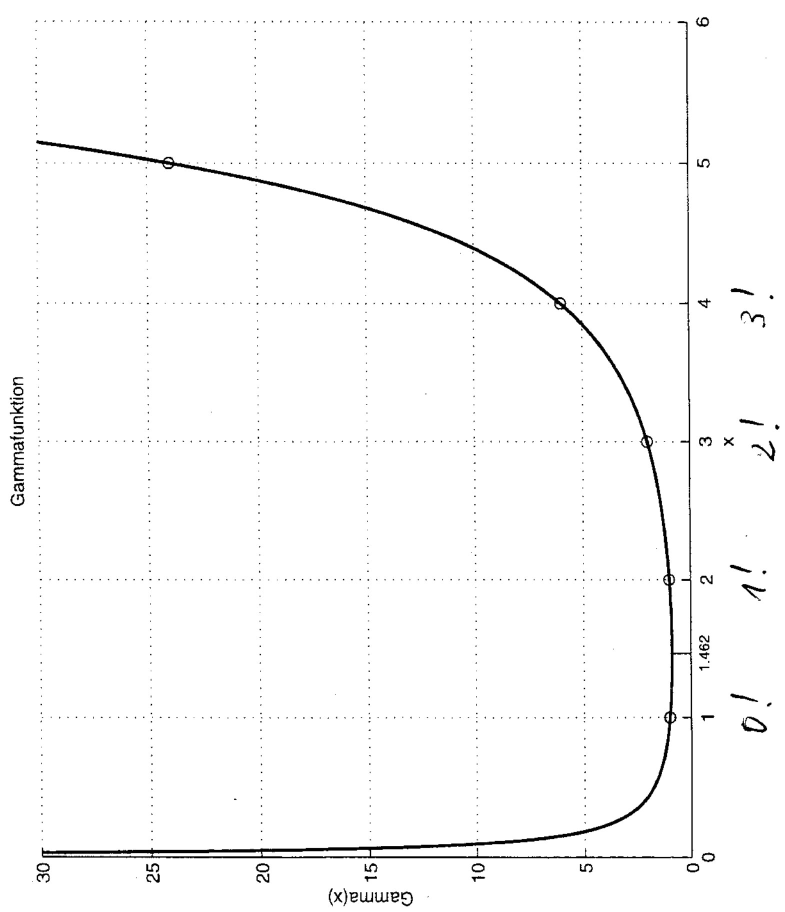
Fortsetzung dieser Funktion $n \to n!$ zu einer Funktion auf $(-1, \infty)$:

$$x \to x! := \int_0^\infty t^x e^{-t}\,dt$$

Bew.: partielle Integration

13.1 Gammafunktion

$$x \to \Gamma(x) := \int_0^\infty t^{x-1} e^{-t}\,dt \quad \text{für} \quad x > 0$$



<u>Es gilt</u>: 1) $\Gamma(x) = (x-1)\Gamma(x-1)$

2) $\Gamma(n) = (n-1)! \quad \forall n \in \mathbb{N}$

3) $\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$

13.2 Gammaverteilung $Gam(\alpha,\beta)$, $\alpha>0, \beta>0$

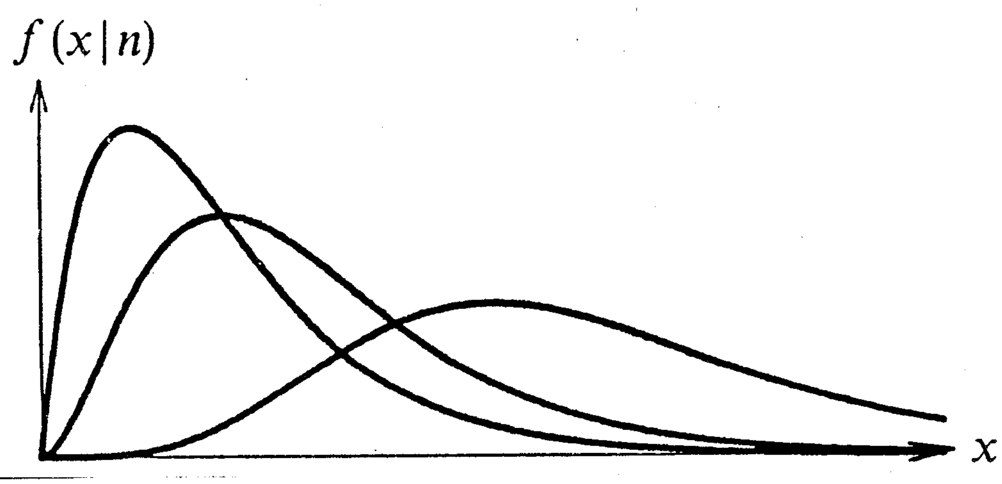
Dichte $f(x) = \dfrac{x^{\alpha-1}\, e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\cdot\beta^{\alpha}}\, I_{(0,\infty)}(x)$

Es gilt: $X \sim Gam(\alpha,\beta) \Rightarrow \mathbb{E}X = \alpha\beta$

$\operatorname{Var} X = \alpha\beta^2$

$Gam(\tfrac{n}{2},2)$ … χ^2_n Chiquadrat-Vtlg.

Dichtefunktionen von χ^2_n-Verteilungen



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

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

<u>Es gilt</u>: $X \sim t_n \Rightarrow \mathbb{E}X = 0 \quad \forall n > 1$

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

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

14. 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

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

14.2 Satz: Gilt zusätzlich zu den Vorauss. von Satz 14.1 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: $G(x) = W\{X^2 \leq x\}$ (Bem. zu Satz)

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

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

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

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} x^{-\frac{1}{2}} I_{(0,\infty)}(x) \quad \hat{=} \quad \chi_1^2$$

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

14.3 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 14.2 //

14.4 **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 $y \in [0,1]$: $G(y) = W\{F(X) \le y\} = W\{X \le F^{-1}(y)\}$

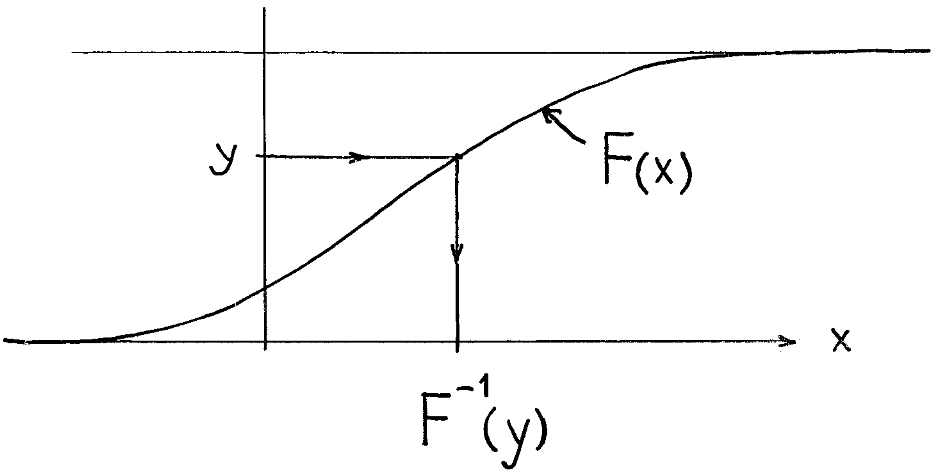
$= F(F^{-1}(y)) = y \Rightarrow G(\cdot)$ ist die VF der kontin. Gleichverteilung $U_{0;1}$

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 $U_{0;1}$

15. TSCHEBYSCHEFFsche UNGLEICHUNG

Abschätzung zum Beweis des Gesetzes der großen Zahlen

15.1 Tsch.-U.: Für jede SG X mit $\exists \operatorname{Var} X$ gilt $\forall \varepsilon > 0$:

$$W\{|X - \mathbb{E}X| \geq \varepsilon\} \leq \frac{\operatorname{Var} X}{\varepsilon^2}$$

Bew.: (für kontin. vt. SG) $X \sim f(\cdot)$, $\mu = \mathbb{E}X$, $\sigma^2 = \operatorname{Var} X$

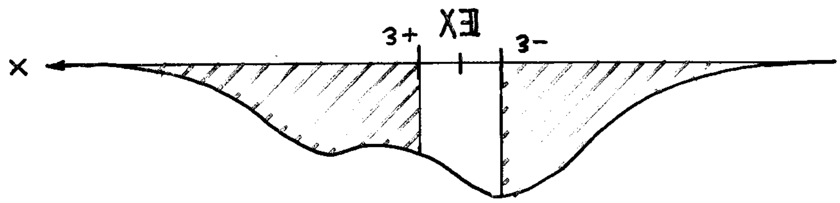
$$\frac{\sigma^2}{\varepsilon^2} = \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx = \frac{1}{\varepsilon^2}\int_{\{x:|x-\mu|\geq\varepsilon\}}(x-\mu)^2 f(x)\,dx \; +$$

15.2

$$+\frac{1}{\varepsilon^2}\int_{\{x:|x-\mu|<\varepsilon\}}(x-\mu)^2 f(x)\,dx \geq \frac{1}{\varepsilon^2}\int_{\{x:|x-\mu|\geq\varepsilon\}}(x-\mu)^2 f(x)\,dx =$$

$$= \int_{\{x:|x-\mu|\geq\varepsilon\}}\left(\frac{x-\mu}{\varepsilon}\right)^2 f(x)\,dx \geq \int_{\{x:|x-\mu|\geq\varepsilon\}} f(x)\,dx = W\{|X-\mu| \geq \varepsilon\}$$

da $|x-\mu| \geq \varepsilon \Leftrightarrow \left(\frac{x-\mu}{\varepsilon}\right)^2 \geq 1$



Bem. Die Tsch. U. ist sehr grob

IV

MEHRDIMENSIONALE VERTEILUNGEN

16. 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 <u>m-dim. Stochastischer Vektor</u>

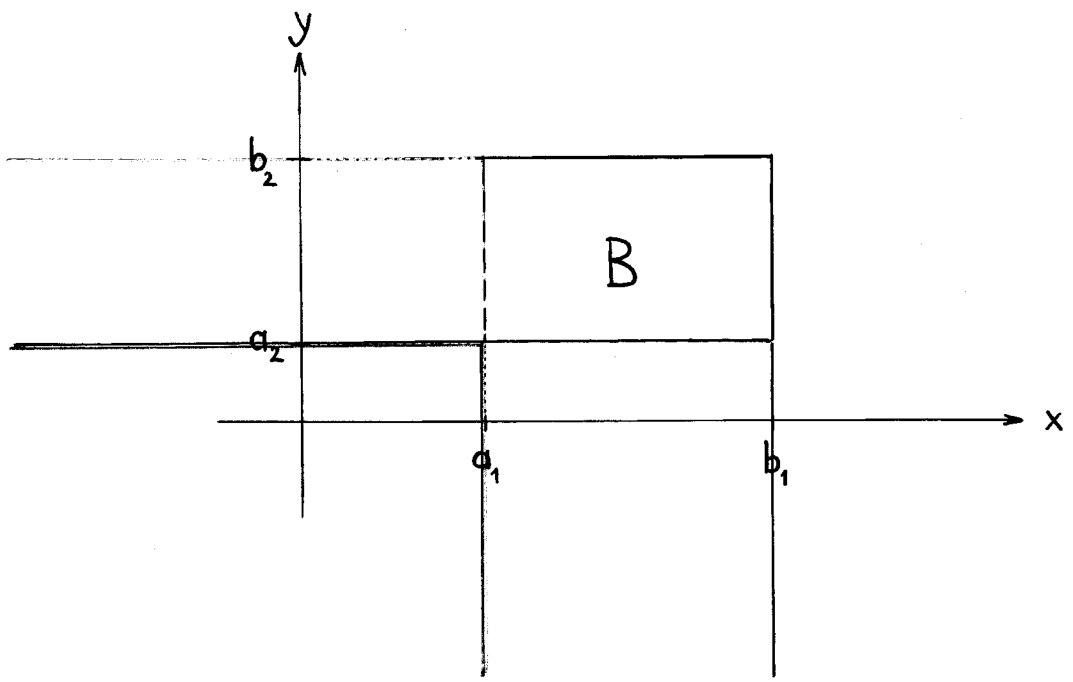
<u>Beispiel</u>: System aus m Komponenten
X_i Lebensdauer der i-ten Komponente

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

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

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

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



16.1 Randverteilungen

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

<u>Randverteilungen</u>.

<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. VFn $F_1(\cdot)$ von X und $F_2(\cdot)$ von Y sind

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

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

17. 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

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

Beispiel: 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	$p(a_1,b_1)$	$p(a_1,b_2)$	$\cdots$	$p(a_1,b_j)$	$\cdots$	$p(a_1,b_\ell)$
a_2	$p(a_2,b_1)$	$p(a_2,b_2)$	$\cdots$	$p(a_2,b_j)$	$\cdots$	$p(a_2,b_\ell)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
a_i	$p(a_i,b_1)$	$p(a_i,b_2)$	$\cdots$	$p(a_i,b_j)$	$\cdots$	$p(a_i,b_\ell)$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
a_k	$p(a_k,b_1)$	$p(a_k,b_2)$	$\cdots$	$p(a_k,b_j)$	$\cdots$	$p(a_k,b_\ell)$

17.1 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 <u>Randverteilung</u> siehe Tabelle

Für höherdimensionale SV_n gibt es mehrdimensionale Randverteilungen

17.2

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

18. MEHRDIMENSIONALE KONTIN. VTLGEN

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

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

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

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

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

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

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

Im Spezialfall $B = I_m = \times_{i=1}^{m} (a_i, b_i]$ 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>: <u>m-dim. Normalverteilung</u> $N(\mu, \Sigma)$

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

m-dim. Dichtefunktion mittels der inversen Matrix

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

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

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

18.1 2-dim. Normalverteilung $N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \varrho)$

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

Dichtefunktion $f(x,y) =$

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

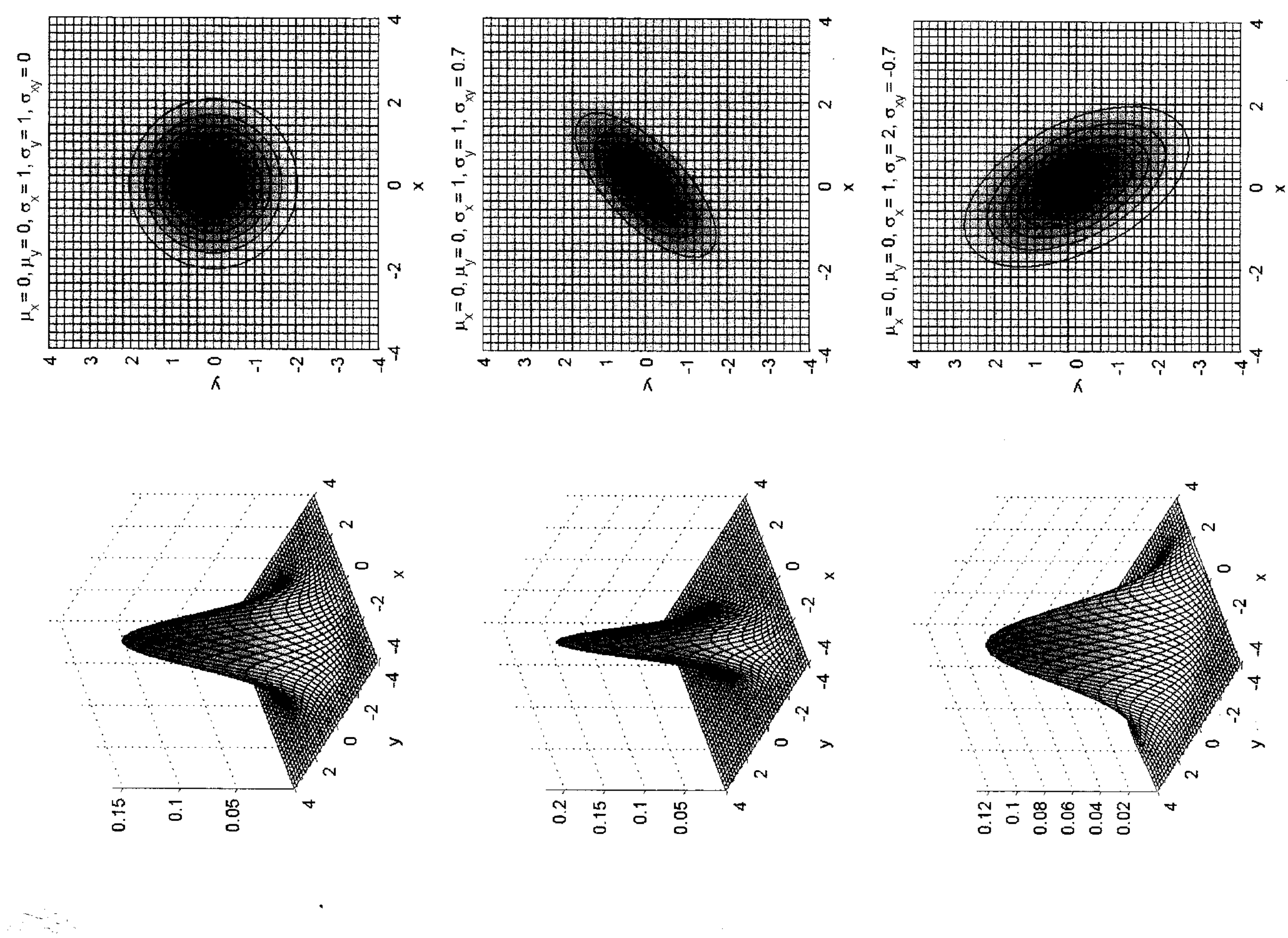
$$\forall (x,y) \in \mathbb{R}^2$$

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

ϱ ist ein Maß für den Zusammenhang von X und Y

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

$\varrho = 0$ stoch. unabhängig



18.2 Randdichten

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

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

$$= \int_{-\infty}^{\infty} \Big[\int_A f(x,y)\,dx \Big] dy = \int_A \Big[\underbrace{\int_{-\infty}^{\infty} f(x,y)\,dy}_{f_1(x) :=} \Big] dx$$

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

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

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

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

18.3 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.

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

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

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

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

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

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

(3) Im gemischten Fall eine Kombination

19.2 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.

<u>Bem.</u>: $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$

20. Kovarianz, Korrelation und Unabhängigkeit

Beschreibung von Zusammenhängen zwischen SGn

20.1 Kovarianz

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

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

<u>Bem.</u>: Berechnung mittels S.v.u.Stat.
$\mathrm{Cov}(X,Y)$ Maß für die Abhängigkeit von X und Y

<u>Beispiel</u>: $(X,Y) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \varrho) \Rightarrow \mathrm{Cov}(X,Y) = \varrho\sigma_x\sigma_y$

$(X_1, \ldots, X_m) \sim N(\mu, \Sigma) \Rightarrow \mathrm{Cov}(X_k, X_\ell) = \sigma_{k\ell}$

20.2 Satz: Falls $\exists\, \mathrm{Cov}(X,Y)$, so gilt:

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

Bew.: $\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-\mathbb{E}X)(Y-\mathbb{E}Y)\right]$
$= \mathbb{E}\left[X\cdot Y - Y\cdot\mathbb{E}X - X\cdot\mathbb{E}Y + (\mathbb{E}X)\cdot(\mathbb{E}Y)\right]$
$= \mathbb{E}(X\cdot Y) - \mathbb{E}X\cdot\mathbb{E}Y - \cancel{\mathbb{E}X\cdot\mathbb{E}Y} + \cancel{\mathbb{E}X\cdot\mathbb{E}Y}$ //

20.3 Satz: Sind $X_1, \ldots, X_n$ SGn mit $\exists\, \mathrm{Var}\, X_i \;\forall i = 1(1)n$ und $c_1, \ldots, c_n$ reelle Konstanten, so gilt:

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

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

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

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

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

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

20.4 Korrelationskoeffizient

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

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

<u>Es gilt</u>: 1) $-1 \leq \varrho_{X,Y} \leq 1$

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

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

20.5 Satz: Für paarweise unkorrelierte SGn gilt:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}\,X_i \qquad (\text{vgl. Satz } 20.3)$$

20.6 Stochastische Unabhängigkeit SGn

SGn X und Y sind <u>unabhängig</u>, 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 $\varrho = 0$.

20.7 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. $\Rightarrow$

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

Bew.: $\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] =$ Satz 20.7
$= \mathbb{E}[X - \mathbb{E}X] \cdot \mathbb{E}[Y - \mathbb{E}Y] = [\mathbb{E}X - \mathbb{E}X] \cdot [\mathbb{E}Y - \mathbb{E}Y] = 0$
$\Rightarrow \varrho_{X,Y} = 0$ //

21. BEDINGTE VERTEILUNGEN

(X,Y) 2-dim. SV

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

21.1 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 $Y|X=x$

$X|Y=y$

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

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

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

21.2 Kontinuierliche Vtlgen.

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

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

Lösung: <u>Bedingte Dichten</u> von $Y|X=x$

$X|Y=y$

$f_1(\cdot)$ Randdichte von X

$f_2(\cdot)$ —"— " Y

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

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

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

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

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

Definition: Für (X,Y) heißt die Funktion

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

auch genannt Regressionsfunktion von Y bezüglich X.

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

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

22. 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) $X_1, \dots, X_m$ SGn, $Z = \min(X_1, \dots, X_m)$

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

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

22.1 Lineartransformation von Normalverteilungen

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

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

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

(ohne Beweis)

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

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

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

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

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

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

22.3 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 <u>Ordnungsstatistiken</u>.

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

22.4 Faltung von W.-Vtlgen.

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

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

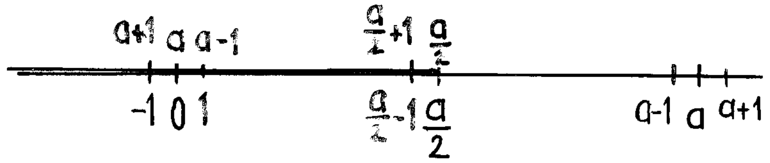
Beweis: (a) Diskrete Verteilungen $X \sim p_1(\cdot)$, $Y \sim p_2(\cdot)$

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

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

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

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

Bem.: 

„Faltung"

Anwendung

Additionstheorem für Poisson-Verteilungen:

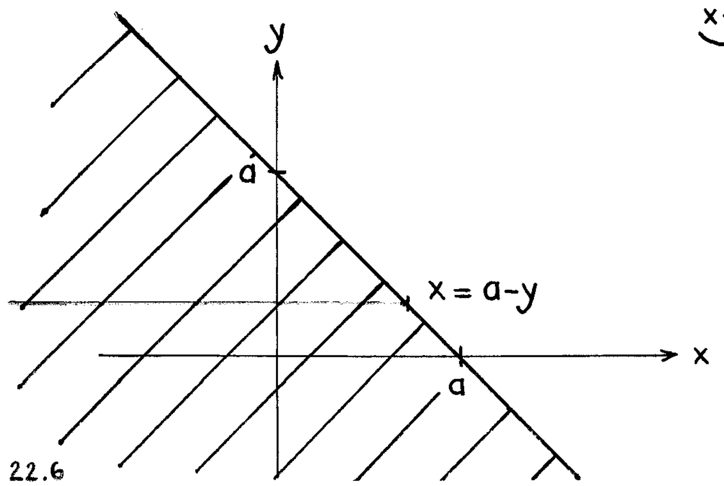
$$X \sim P_\mu,\ Y \sim P_\xi,\ X \perp\!\!\!\perp Y \Rightarrow X+Y \sim P_{\mu+\xi}$$

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

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

(b) Kontinuierliche Verteilungen $X \sim f(\cdot),\ Y \sim g(\cdot)$

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



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

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

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 $g(\cdot)$

Anwendung: 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

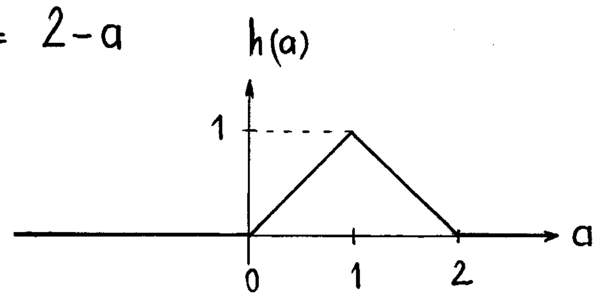
$[a-1 < y < a] \wedge [0 < y < 1]$.

Daher gilt

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

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

sonst: $h(a) = 0$



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

22.9

Bem.: Sonderfall χ_n^2

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

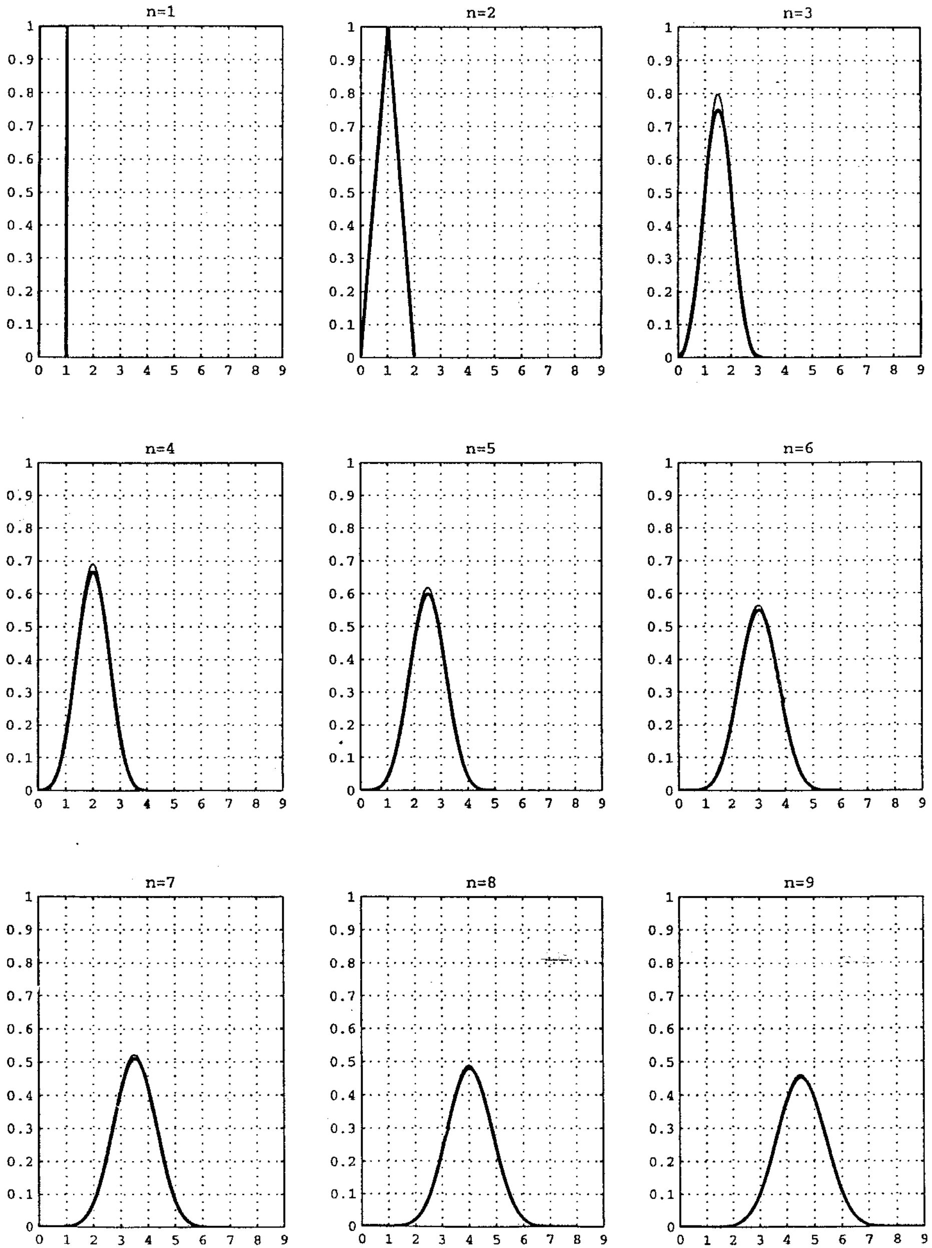
Sind $X_1, \dots, X_n$ unabhängig exponentialverteilt mit Dichte $f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$, so hat die SG $\sum_{i=1}^{n} X_i$ eine Erlang-Verteilung mit Dichtefunktion

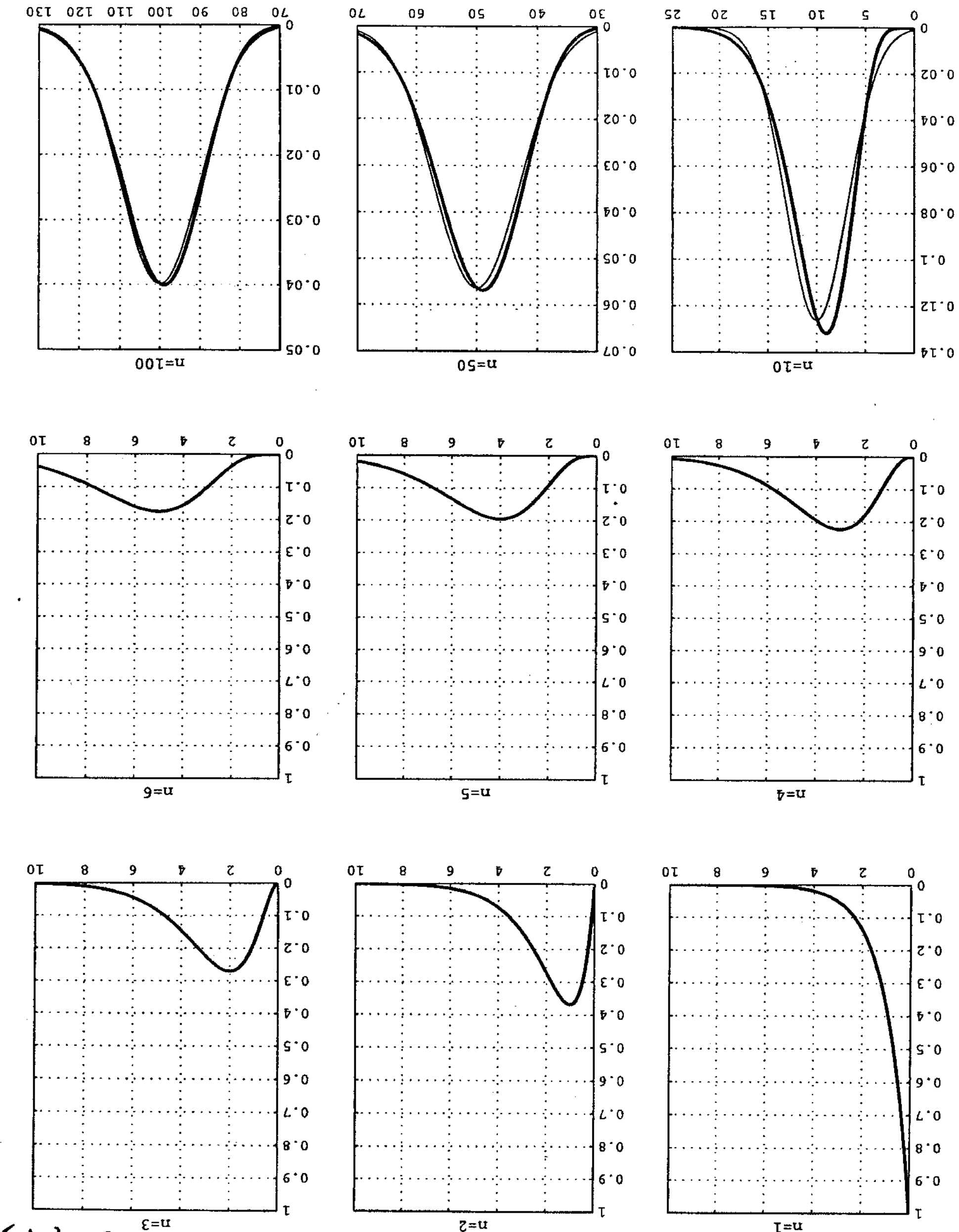
$$f(x|n,\lambda) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} I_{(0,\infty)}(x)$$

Beweis: Faltung und Induktion

Anwendung: Warteschlangen

22.10





Er (2,1)

Er (3,1)

V

FOLGEN STOCH. GRÖSZEN

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

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

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

23. GESETZ DER GROSZEN ZAHLEN

Konvergenz von Folgen SGn

<u>Bem.</u>: 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)

<u>Def.</u>: Eine Folge $(X_n, n \in \mathbb{N})$ von SG$_n$ X_n konvergiert in der Wahrscheinlichkeit gegen eine SG X_0, i.Z.

$X_n \xrightarrow[n \to \infty]{W} X_0$, wenn gilt:

$$\lim_{n \to \infty} W\{|X_n - X_0| \geq \varepsilon\} = 0 \quad \forall \varepsilon > 0$$

23.1 Gesetz der großen Zahlen: Ist $(X_n ; n \in \mathbb{N})$ eine UIV-Folge mit $\exists \operatorname{Var} X_i = \sigma^2 < \infty$, so gilt:

$$\overline{X}_n \xrightarrow[n \to \infty]{W} \mathbb{E}X_i \equiv \mu \in \mathbb{R}$$

Beweis: $\exists \operatorname{Var} X_i \Rightarrow \exists \mathbb{E}X_i = \mu$ (endlich), mit $S_n := X_1 + \dots + X_n$ folgt $\mathbb{E}S_n = n\mu$ und wegen

der Unabhängigkeit der $X_1, \dots, X_n$ $\operatorname{Var} S_n = n\sigma^2$

$$\Rightarrow \mathbb{E}\overline{X}_n = \mathbb{E}\left[\tfrac{1}{n} S_n\right] = \tfrac{1}{n}\mathbb{E}S_n = \tfrac{1}{n} \cdot n \cdot \mu = \mu .$$

$$\operatorname{Var} \overline{X}_n = \operatorname{Var}\left(\tfrac{1}{n} S_n\right) = \frac{1}{n^2} \operatorname{Var} S_n = \frac{1}{n^{\cancel{2}}} \cdot \cancel{n} \cdot \sigma^2 = \frac{\sigma^2}{n}$$

Aus der Tschebyscheff'schen Ungl. folgt

$$W\{|\overline{X}_n - \mu| \geq \varepsilon\} \leq \frac{\operatorname{Var} \overline{X}_n}{\varepsilon^2} \quad \forall \varepsilon > 0 .$$

$$\lim_{n \to \infty} W\{|\overline{X}_n - \mu| \geq \varepsilon\} \leq \lim_{n \to \infty} \frac{\sigma^2/n}{\varepsilon^2} = \lim_{n \to \infty} \frac{\sigma^2}{n\varepsilon^2} = 0 \quad //$$

<u>Bem.</u>: Zum empir. GgZ, $X_i := I_B \Rightarrow X_i \sim A_{W(B)}$

$$\Rightarrow h_n(B) = \overline{X}_n \xrightarrow[n \to \infty]{W} \mathbb{E}X_i = W(B)$$

24. ZENTRALER GRENZVERTEILUNGSSATZ

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

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

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

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

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

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

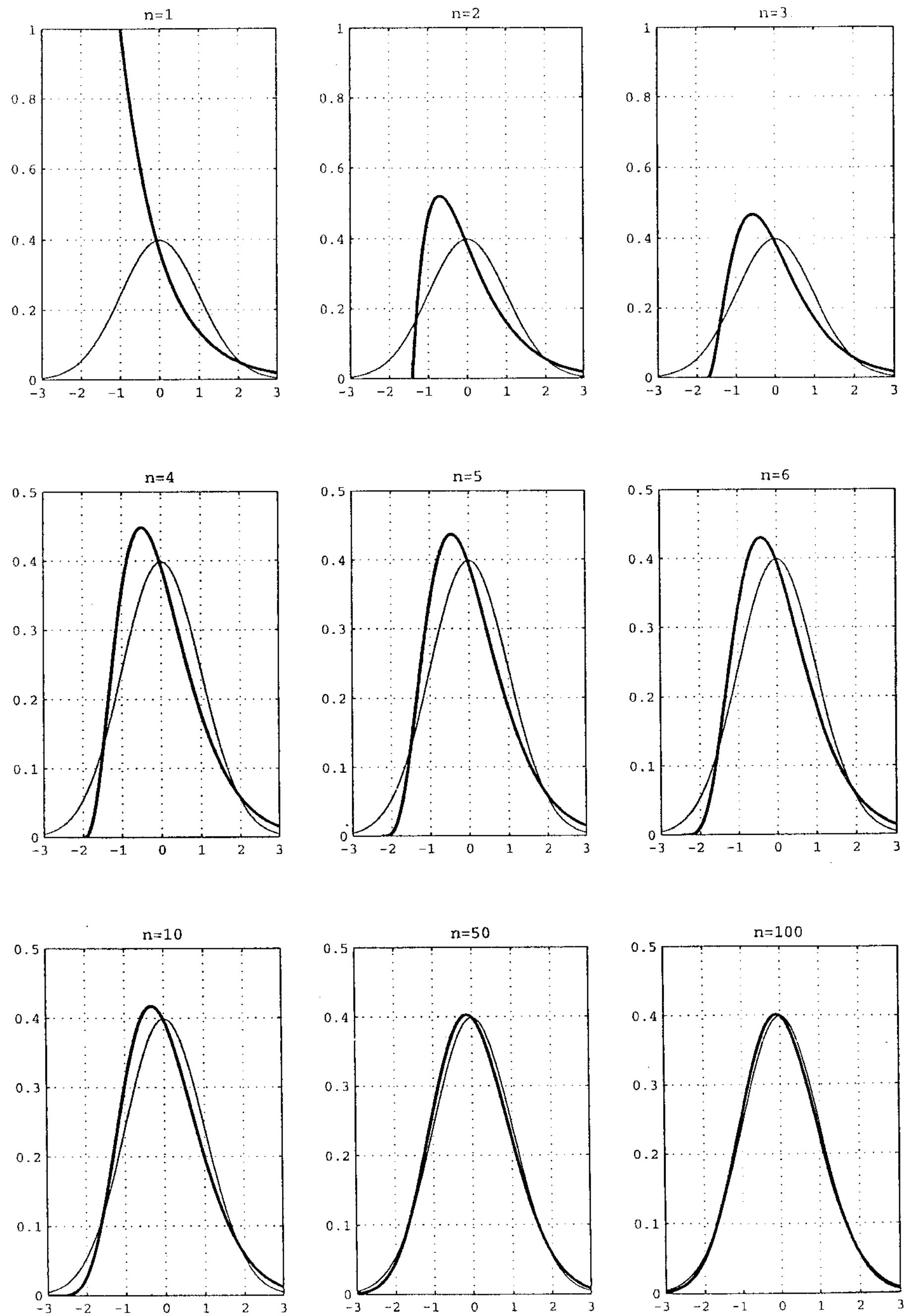
$$L_n(\varepsilon) := \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{\{x : |x-\mu_k| \geq \varepsilon s_n\}} (x-\mu_k)^2 \, dF_k(x) \xrightarrow[n \to \infty]{} 0 \quad \forall \varepsilon > 0$$

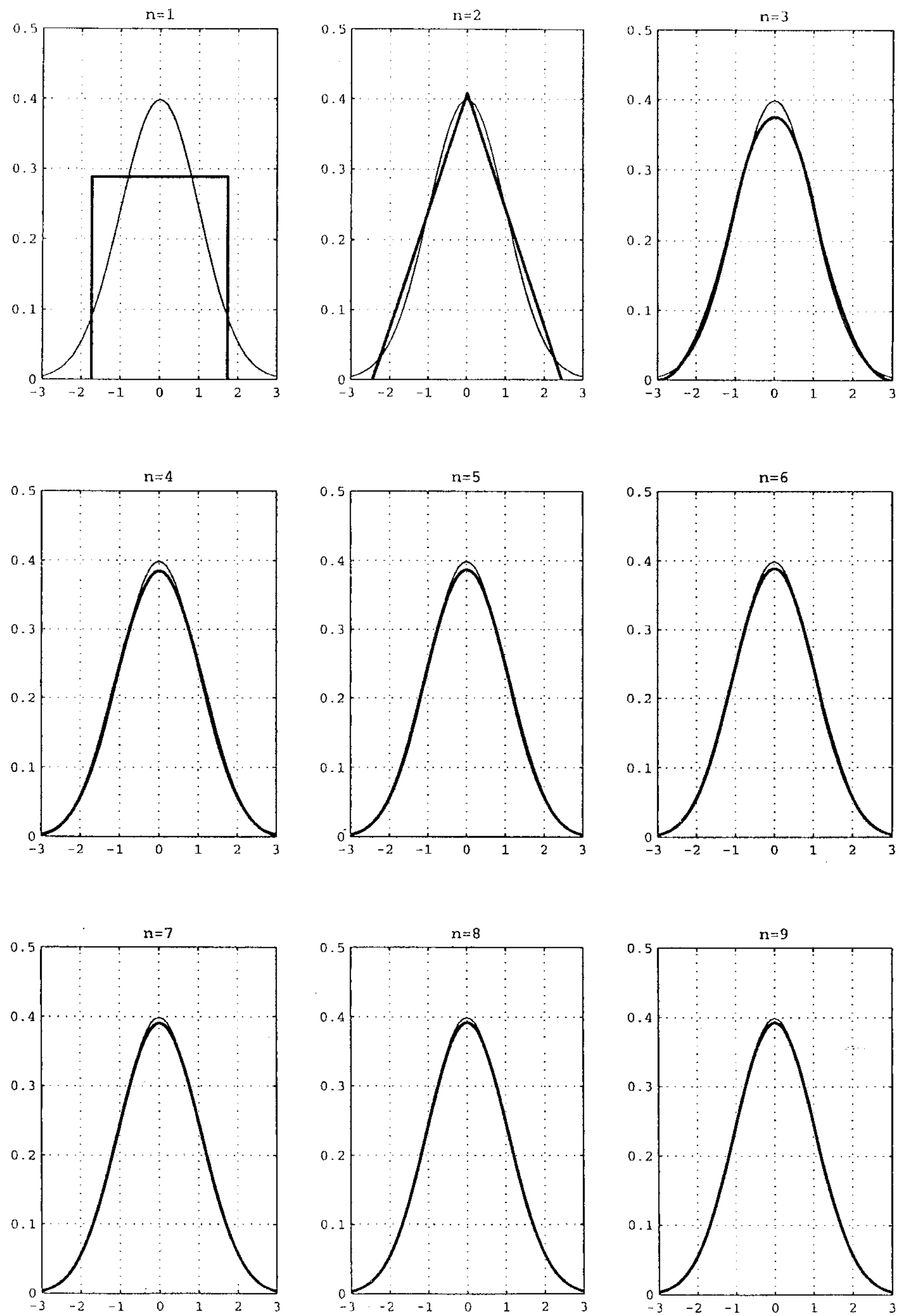
Dies bedeutet, dass keine einzelne Vtlg. $F_k(\cdot)$ dominierenden Einfluss auf die GV hat.

2) Für UIV-Folgen ist die LB erfüllt.

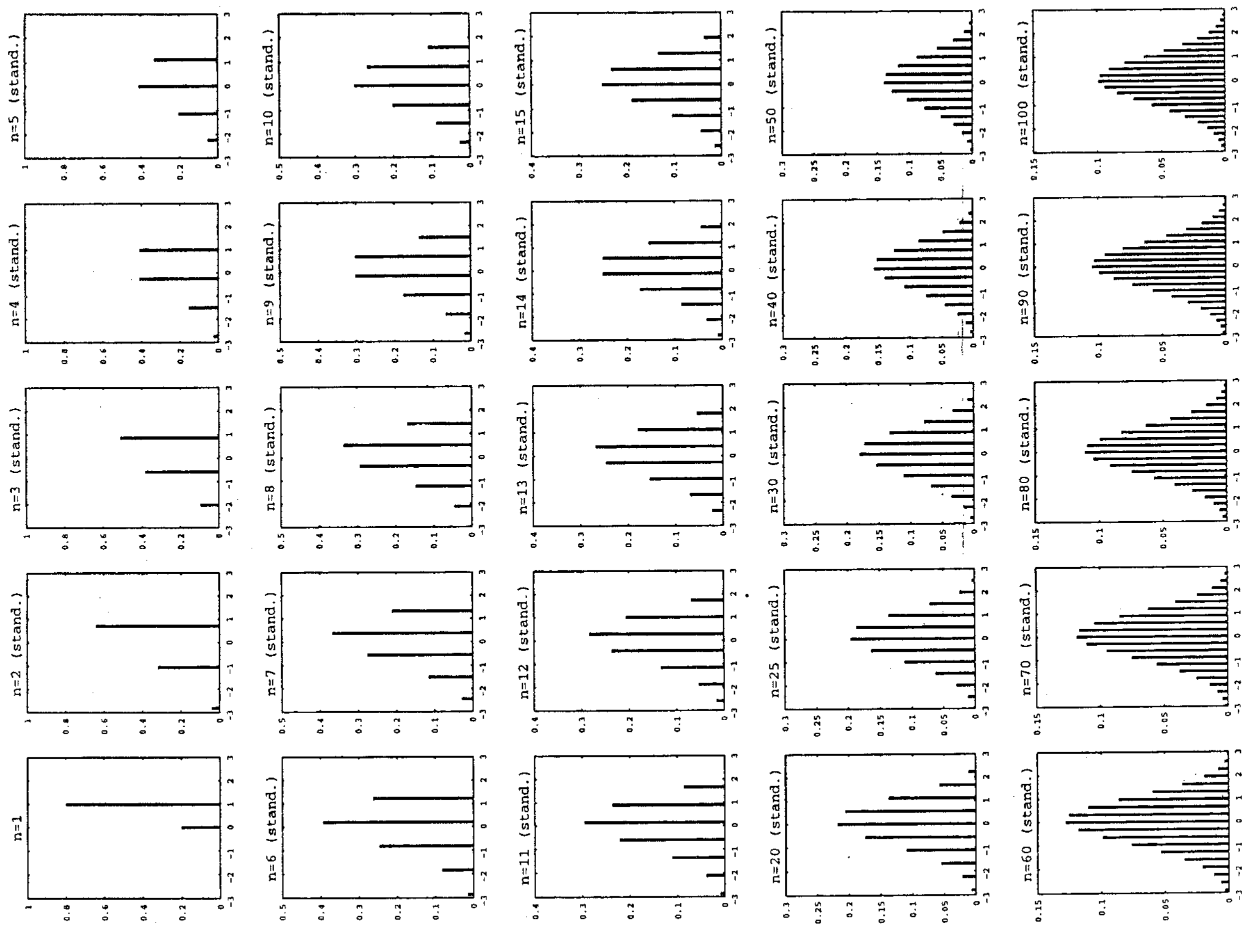
3) Für glm. beschränkte SG_n $|X_k| \leq C \ \forall k \in \mathbb{N}$ mit

$$\operatorname{Var}\left(\sum_{k=1}^{n} X_k\right) \xrightarrow[n \to \infty]{} \infty$$ ist die LB erfüllt.





24.3



24.2 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 $Y \sim N(\mu, \sigma^2)$:

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

25. 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 $F(x)$, 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. $F_n^*(\cdot)$

25.1 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) $F_n^*(x; X_1, \dots, X_n) \xrightarrow[n\to\infty]{W} F(x)$

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

$= W\{X \le x\}$

$= F(x)$

$= \frac{1}{n}\cdot n \cdot F(x) = F(x)$

(b) $\dfrac{I_{(-\infty,x]}(X_1) + \dots + I_{(-\infty,x]}(X_n)}{n} \;(GgZ) \xrightarrow[n\to\infty]{W} \mathbb{E}\, I_{(-\infty,x]}(X_i)$

also $F_n^*(x) \xrightarrow[n\to\infty]{W} F(x)$ //

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

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

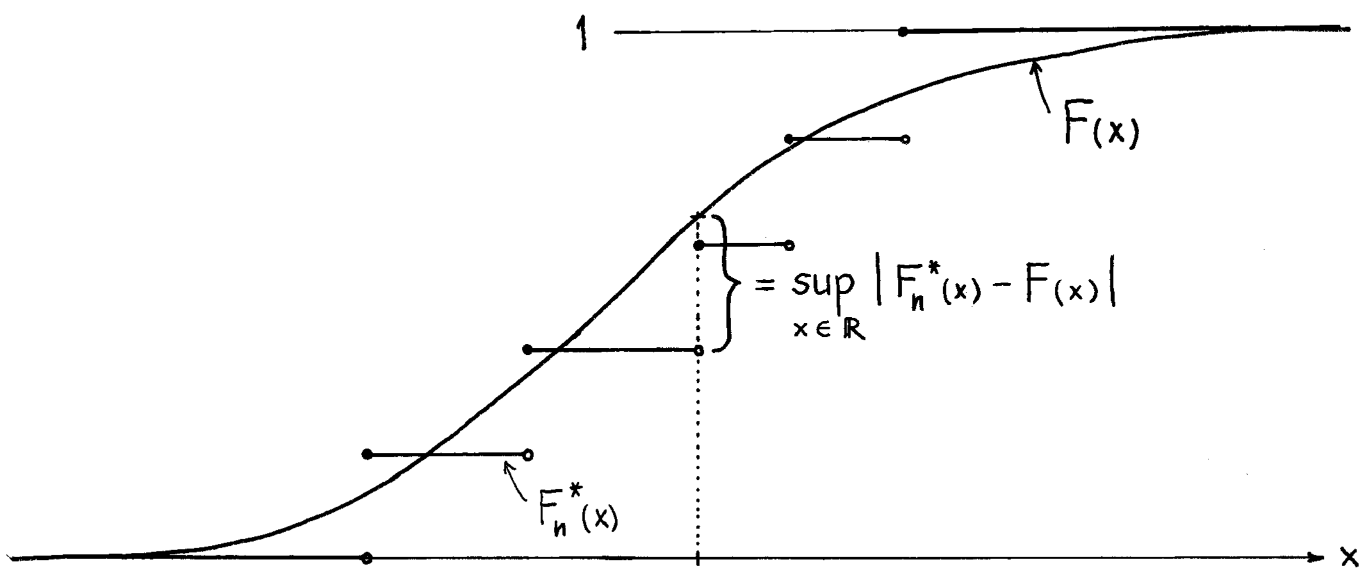
(ohne Beweis)

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

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

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

25.3



26. STICHPROBEN UND STATISTIKEN

Gesucht: Verteilung von SG_n 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) <u>Stichprobe von X</u>.

Nach Beobachtung von $X_1 = x_1, \dots, X_n = x_n$ nennt man $x_1, \dots, x_n$ <u>konkrete Stichprobe</u> 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.) <u>Statistik</u>, 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) $\psi(X_1, \dots, X_n) = \frac{1}{n}\sum_{i=1}^n X_i$ <u>Stichprobenmittel</u> $\overline{X}_n$

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

26.1 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 <u>Stichprobenraum</u>.

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

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

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

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

<ins>Beispiele</ins>: 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, $\theta = \mu$, $\Theta = (0, \infty)$

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

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

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

Es gibt 3 große Zweige der schließenden Statistik

a) Klassische objektivistische Statistik

b) Bayes'sche Statistik

c) Statistik bei unscharfer Information
(Fuzzy Information)

<u>Bem.</u>: Entscheidung bei Unsicherheit

Fuzzy Wahrscheinlichkeiten

26.5

VI

KLASSISCHE SCHLIESZENDE STATISTIK

27. KLASSISCHE PUNKTSCHÄTZUNGEN

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

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

$\vartheta(X_1, \dots, X_n)$ SG

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

Für geraffte Parameter $\tau(\theta) \in \mathbb{R}$
Schätzfunktion $t(X_1, \dots, X_n)$ mit $t : M_X^n \to \mathbb{R}$

Verschiedene <u>Güteeigenschaften</u> von Schätzfunktionen bzw. Schätzwerten.

27.1 Unverzerrtheit

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

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

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

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

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

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

<u>Bem.</u>: Für geraffte Parameter $\tau(\theta)$ für Unverzerrtheit:

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

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

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

27.3 Effizienz

Für unverzerrte SFn möglichst kleine Varianz

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

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

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

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

$$= \sum_{i=1}^{n} (\alpha_i - \tfrac{1}{n})^2 + \tfrac{2}{n}(\underbrace{\sum_{i=1}^{n} \alpha_i}_{=1} - 1) + \tfrac{1}{n} \geq 0$$

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

27.5 Konsistenz

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

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

konsistent, wenn $t_n(X_1, \dots, X) \xrightarrow[n\to\infty]{W} \xi$

d.h. wenn gilt:

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

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

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

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

27.6 Plausibilität

Eigenschaft von Schätzwerten für Parameter

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

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

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

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

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

Bem.: $\dfrac{\partial \ell(\theta; x_1, \dots, x_n)}{\partial \theta} = 0$ bzw. $\dfrac{\partial \ln \ell(\theta; x_1, \dots, x_n)}{\partial \theta} = 0$

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

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

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

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

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

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

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

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. $x_1, \dots, x_n$

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

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

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

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

$\Rightarrow$ 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 $\hat{\tau} = \overline{x}_n$ 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.

28. 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 KB_n mittels sogenannter Pivot-Größen

28.1 Pivot-Größen

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

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

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

(ohne Beweis)

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

28.3 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}

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

$$\forall(\mu,\sigma^2)$$

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

29. STAT. HYPOTHESEN UND TESTS

Problem: Welche Vtlg. hat eine SG X ?

Def.: Eine Stat. Hypothese ist eine Aussage über die Verteilung von X.

Bsp.: $\mathcal{H}$: X ist normalverteilt

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

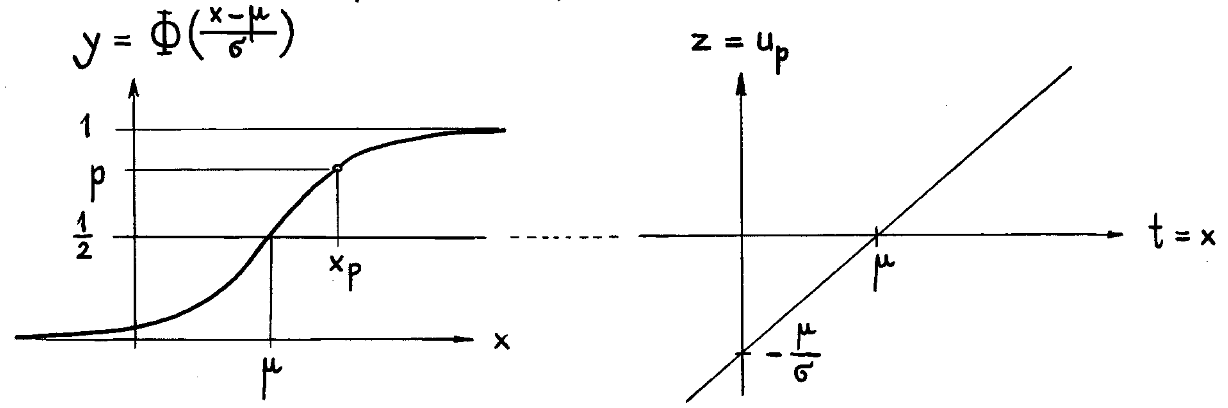
Def.: Ein Statistischer Test ist ein Verfahren zur Entscheidung, ob eine Stat. Hypothese angenommen oder verworfen wird.

29.1 Wahrscheinlichkeitspapiere

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

<u>Beispiel</u>: <u>Normalverteilungsnetz</u>

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



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

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

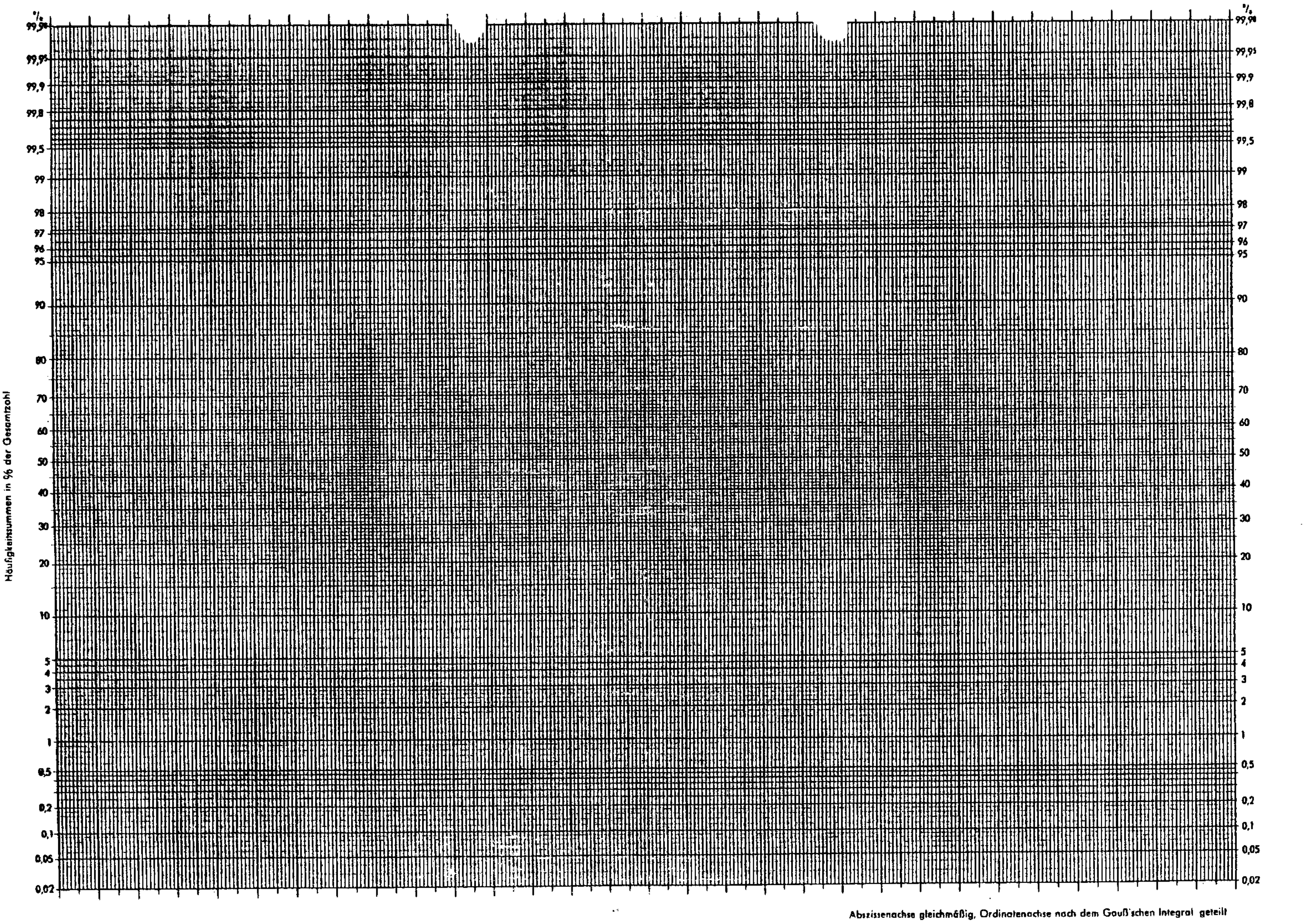
x_p .... p-Fraktiles der $N(\mu, \sigma^2)$

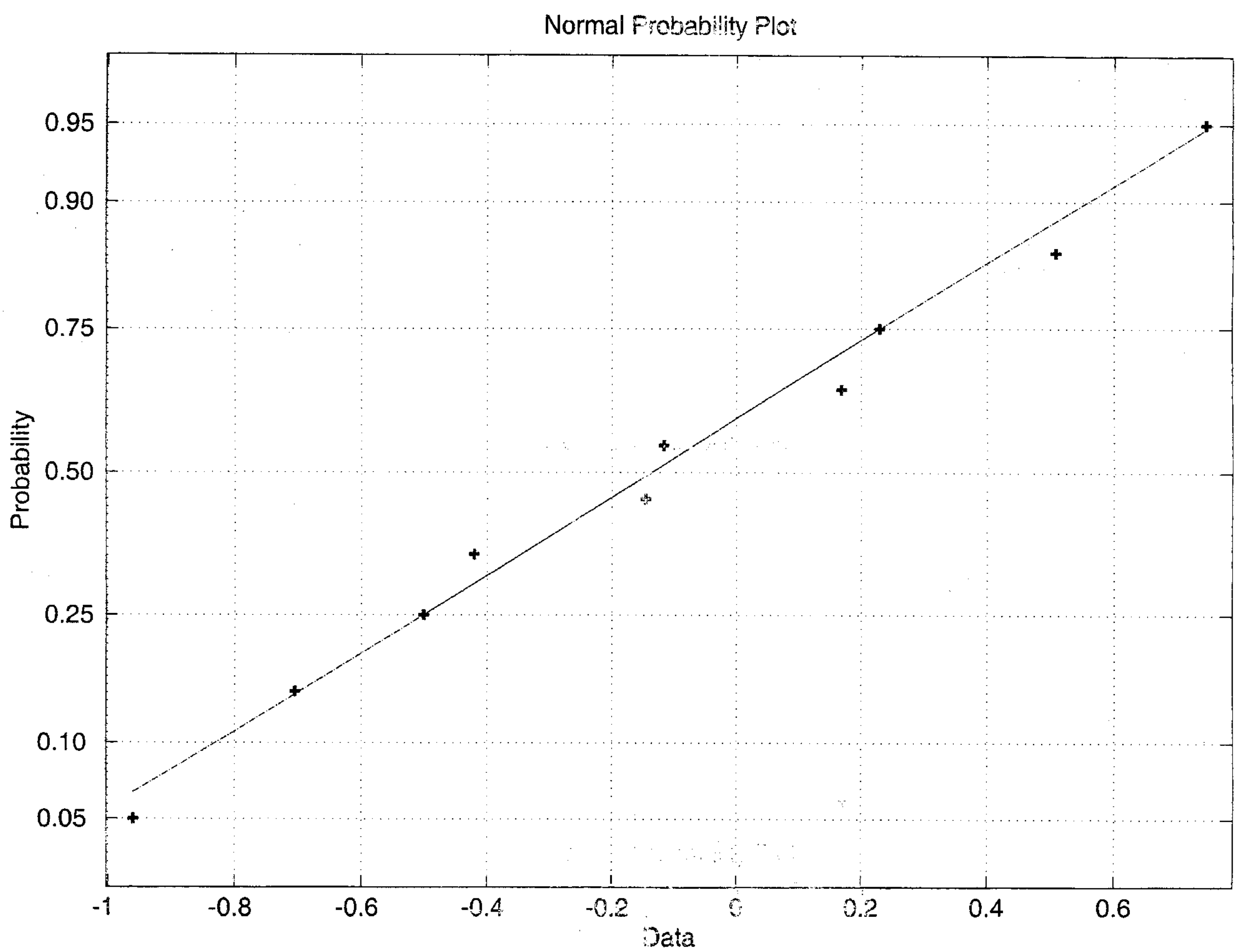
$$u_p = \frac{x_p - \mu}{\sigma} \qquad \forall\, p \in (0,1)$$

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

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

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


%
99,98
99,95
99,9
99,8
99,5
99
98
97
96
95
90
80
70
60
50
40
30
20
10
5
4
3
2
1
0,5
0,2
0,1
0,05
0,02
Häufigkeitssummen in % der Gesamtzahl
Abszissenachse gleichmäßig, Ordinatenachse nach dem Gauß'schen Integral geteilt


Normal Probability Plot
Probability
0.95
0.90
0.75
0.50
0.25
0.10
0.05
Data
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6

29.2 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.

29.3 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.

<u>Beispiel</u>: $X \sim N(\mu, \sigma_0^2)$, 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 (<u>Irrtumswahrscheinlichkeit</u>) 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. <u>Teststatistik</u> T ab: Es gilt, falls $\mathcal{H}_0$ richtig,

$$W\left\{ u_{\frac{\alpha}{2}} \leq T \leq u_{1-\frac{\alpha}{2}} \right\} = 1 - \alpha$$

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

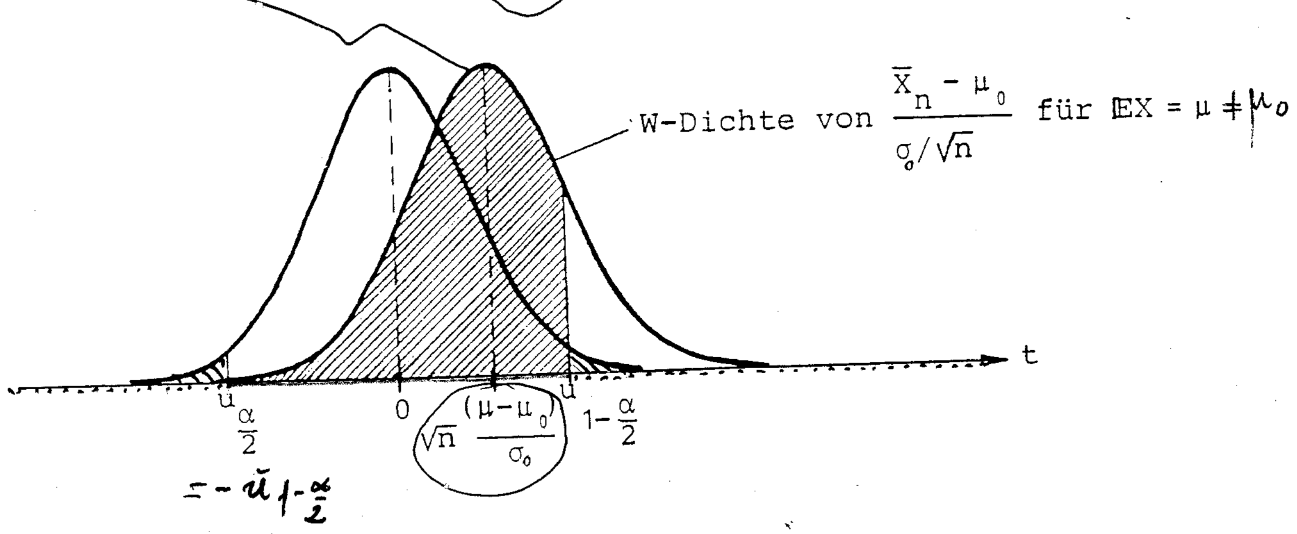
$$V = \left\{ (x_1, \dots, x_n) : \left| \frac{\overline{x}_n - \mu_0}{\sigma_0 / \sqrt{n}} \right| \geq u_{1-\frac{\alpha}{2}} \right\}$$

<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 <u>kritischer Bereich</u> für die Teststatistik

Zur Fehlerwahrscheinlichkeit zweiter Art:

$$T = \frac{\overline{X}_n - \mu_0}{\sigma_0/\sqrt{n}} \sim N\left(\frac{\mu-\mu_0}{\sigma_0/\sqrt{n}}, 1\right)$$



30. TESTS FÜR NORMALVERTEILUNGEN

Alle betrachteten SG als normalverteilt vorausges. Parameter unbekannt

30.1 t-Test für den Erwartungswert

Ist $X \sim N(\mu, \sigma^2)$ mit μ und σ^2 unbekannt und $x_1, \ldots, x_n$ 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:

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

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

Begründung: Satz 28.2 $\Rightarrow T = \dfrac{\sqrt{n}\,(\overline{X}_n - \mu_0)}{S_n} \sim t_{n-1}$

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

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

30.2 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:

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

Beweis: Satz 28.2 $\Rightarrow$ Vtlg. von $\dfrac{(n-1)\,S_n^2}{\sigma_0^2} \sim \chi^2_{n-1}$

Rest analog zum vorigen Beweis.

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

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

F-Test für die Gleichheit der Varianzen

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

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

$F_{m,n;p}$ ... 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 SPn

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

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

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

<u>Entscheidungsregel</u>: $\mathcal{H}_0$ verwerfen, falls für 2 konkrete SP_n $x_1, \dots, x_m$ und $y_1, \dots, y_n$ gilt

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

<u>Bem.</u>: Für mehr als 2 Messreihen Bartlett-Test

31. DER CHIQUADRAT-ANPASSUNGSTEST

Ein asymptotisches Verfahren bei größeren SP_n

<u>Problem</u>: Hat eine empirisch gegebene SG eine bestimmte Verteilung bzw. bestimmten Verteilungstyp

31.1 Einfache Hypothesen

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

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

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

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

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:

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

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

31.2 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:

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

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

31.4 (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_τ, $\tau \in (0,1)$ haben

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

32. KLASSISCHE REGRESSIONSRECHNUNG

Beschreibung kausaler, nicht deterministischer Zusammenhänge mit <u>Stochastischen Modellen</u>

<u>B4</u>: Abhängigkeit des Bremsweges von der Geschwindigkeit (Kovariable, Einflussgröße)

x ... Geschwindigkeit (—"—)

Y_x ... Bremsweg (SG) abhängige Größe

Stoch. Modell $Y_x = \underbrace{\psi(x)}_{\text{det.}} + \underbrace{U_x}_{\text{stoch.}}$

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

32.1 Lineare Regressionsfunktionen

$Y_x \sim W_\theta$ mit $\theta \in \mathbb{R}$ 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 <u>lineare Regressionsfunktion</u>, wenn sie als Funktion der unbekannten Parameter linear ist.

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

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

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

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.

32.2 Regressionsgeraden

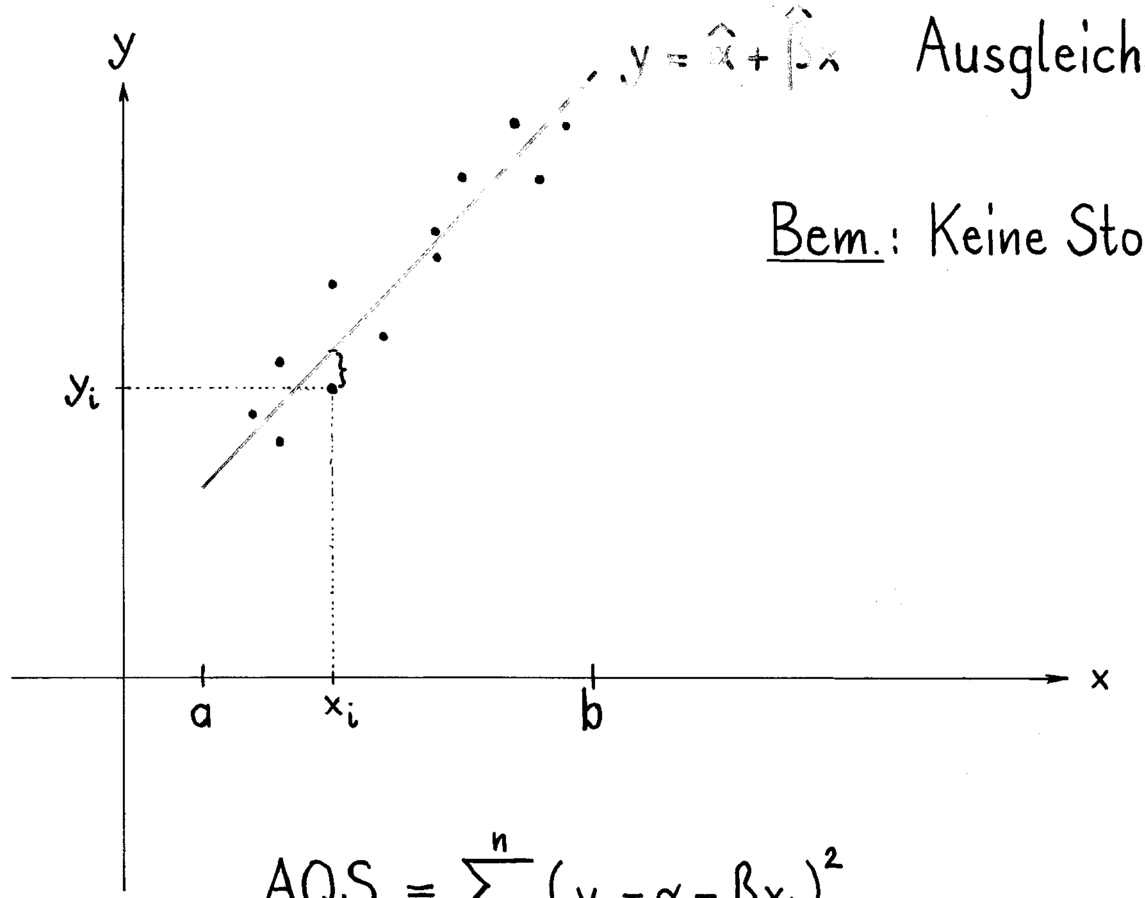
$x \in \mathbb{R}, \quad \psi(x,\alpha,\beta) = \alpha + \beta x \quad \forall\, x \in (a,b)$

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

Ausgleichsgerade



Bem.: Keine Stoch.

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

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

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

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}

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) $\operatorname{Var} Y_i \equiv \sigma^2$, 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

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

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

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.

Für die SFn A und B gilt:

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

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

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

Eine unverzerrte SF für σ^2 ist

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

(o. Bew.)

Bem.: Ein analoger Satz gilt für multiple lineare Regressionsmodelle

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 <u>A-priori-Verteilung</u> von $\tilde{\theta}$

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

33. 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 <u>A-posteriori-Verteilung</u>

33.1 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 :

$\{\theta_1\}, \{\theta_2\}, \ldots, \{\theta_k\}$ Zerlegung von $\Theta = \{\theta_1, \ldots, \theta_k\}$

$$\pi(\theta_j \mid x_1, \ldots, x_n) = W\{\tilde{\theta} = \theta_j \mid X_1 = x_1, \ldots, X_n = x_n\}$$

$$= \frac{W\{X_1 = x_1, \ldots, X_n = x_n, \tilde{\theta} = \theta_j\}}{W\{X_1 = x_1, \ldots, X_n = x_n\}}$$

wegen der Def. von bedingten Wahrsch. diskreter Größen gilt

$$W\{X_1 = x_1, \ldots, X_n = x_n, \tilde{\theta} = \theta_j\} = w(x_1, \ldots, x_n \mid \theta_j) W\{\tilde{\theta} = \theta_j\}$$

und

$$W\{X_1 = x_1, \ldots, X_n = x_n\} = \sum_{j=1}^{n} w(x_1, \ldots, x_n \mid \theta_j) \pi(\theta_j)$$

33.2

Daher gilt das sog. Bayes'sche Theorem (diskret):

$$\pi(\theta_j \mid x_1, \ldots, x_n) = \frac{w(x_1, \ldots, x_n \mid \theta_j) \pi(\theta_j)}{\sum_{j=1}^{k} w(x_1, \ldots, x_n \mid \theta_j) \pi(\theta_j)} \quad \forall j = 1(1)k$$

Im Fall einfacher SP_n gilt wegen der Unabhängigkeit

$$w(x_1, \ldots, x_n \mid \theta_j) = \prod_{i=1}^{n} p(x_i \mid \theta_j)$$

<u>Bem.</u>: Plausibilitätsfunktion $\ell(\theta; x_1, \ldots, x_n)$

$$\Rightarrow \quad w(x_1, \ldots, x_n \mid \theta_j) \pi(\theta_j) = \pi(\theta_j) \ell(\theta; x_1, \ldots, x_n)$$

Da $\sum_{j=1}^{k} \pi(\theta_j) \ell(\theta_j; x_1, \ldots, x_n)$ nach Beobachtung der konkreten SP $x_1, \ldots, x_n$ eine Konstante ist

33.3

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:

$$\pi(\theta_j | x_1, \dots, x_n) \propto \pi(\theta_j) \ell(\theta_j; x_1, \dots, x_n) \quad \forall \theta_j \in \Theta$$

33.2 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 $\tilde{\theta} \,|\, X_1 = x_1, \dots, X_n = x_n$

Lösung: Bedingte Dichte $\pi(\cdot | x_1, \dots, x_n)$

Gemeinsame Dichte von $(X_1, \dots, X_n, \tilde{\theta})$ ist

$$g(x_1, \dots, x_n, \theta) = \left[\prod_{i=1}^{n} f(x_i | \theta)\right] \pi(\theta)$$

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

$$\pi(\theta \mid x_1, \dots, x_n) = \frac{\pi(\theta)\, \ell(\theta; x_1, \dots, x_n)}{\int_\Theta \pi(\theta)\, \ell(\theta; x_1, \dots, x_n)\, d\theta} = \frac{1}{C}\, \pi(\theta)\, \ell(\theta; x_1, \dots, x_n)$$

In Kurzschreibweise:

$$\underline{\pi(\theta \mid x_1, \dots, x_n) \propto \pi(\theta)\, \ell(\theta; x_1, \dots, x_n)} \qquad \theta \in \Theta$$

<u>Bem.</u>: Für allgemeinere Daten D gilt auch

$$\boxed{\pi(\theta \mid D) \propto \pi(\theta)\, \ell(\theta; D)} \qquad \theta \in \Theta$$

Kurzform des Bayes'schen Theorems

34. SUFFIZIENZ UND KONJUGIERTE VERTEILUNGSFAMILIEN

Welche Information aus den beobachteten Daten D ist hinreichend (suffizient) zur Berechnung der A-posteriori-Verteilung ?

Kann man die Berechnung der ApoV vereinfachen ?

34.1 Suffiziente Statistiken

<u>B3</u>: Wartezeit $X \sim Ex_\tau$, AprV (Dichte) $\pi(\cdot)$,
Daten $x_1, \dots, x_n$, $\quad \ell(\tau; x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{1}{\tau} e^{-\frac{x_i}{\tau}}$

$\Rightarrow$ ApoD $\pi(\tau|x_1,\dots,x_n) \propto \pi(\tau)\frac{1}{\tau^n} e^{-\frac{1}{\tau}\sum_{i=1}^n x_i}$

$\Rightarrow \sum_{i=1}^n x_i$ hinreichend

<u>Def.</u>: Ist $X \sim f(\cdot|\theta)$, $\theta \in \Theta$ SM mit kontin. Parameter und $X_1,\dots,X_n$ SP von X, so heißt eine Statistik $S = s(X_1,\dots,X_n)$ <u>suffizient für θ</u>, falls die ApoD von den Daten $x_1,\dots,x_n$ nur über $s(x_1,\dots,x_n)$ abhängt.

<u>Beispiele</u>: 1) s.o.

2) $X \sim N(\mu,\sigma^2) \Rightarrow (\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$ suffizient

3) $X \sim A_\theta \Rightarrow \sum_{i=1}^n X_i$ suffizient für θ (s.u. 34.2)

34.2 Konjugierte Verteilungsfamilien

<u>B1</u>: Anteil θ, die Familie der sog. <u>Betaverteilungen</u> hat folgende Eigenschaft: Für $\pi(\cdot) \in \mathcal{F}$ und bel. Beob. $x_1,\dots,x_n$ ist auch $\pi(\cdot|x_1,\dots,x_n) \in \mathcal{F}$

<u>Def.</u>: $T \sim Be(a,b)$ Betaverteilung, falls $M_T = [0,1]$ und die Dichtefunktion ist

$$f(t) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} t^{a-1}(1-t)^{b-1} I_{[0,1]}(t)$$

<u>Fortsetzung von B1</u>: Für $X \sim A_\theta \mathrel{\hat{=}} p(\cdot|\theta)$ gilt

$$p(x|\theta) = \theta^x (1-\theta)^{1-x} I_{\{0,1\}}(x)$$

$$\ell(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} I_{\{0,1\}}(x_i) I_{[0,1]}(\theta)$$
$$= \theta^{\sum_1^n x_i} (1-\theta)^{n-\sum_1^n x_i} I_{\{0,1\}}(x_i) I_{[0,1]}(\theta)$$

Für $\pi(\cdot) \triangleq Be(a,b)$ folgt daher wegen $x_i \in \{0,1\}$

$$\pi(\theta | x_1, \dots, x_n) \propto \pi(\theta) \ell(\theta; x_1, \dots, x_n) =$$
$$= \theta^{a-1} (1-\theta)^{b-1} I_{[0,1]}(\theta)\, \theta^{\sum_1^n x_i} (1-\theta)^{n-\sum_1^n x_i}$$
$$= \theta^{a+\sum_1^n x_i - 1} (1-\theta)^{b+n-\sum_1^n x_i - 1} I_{[0,1]}(\theta)$$

Dichtef. der $Be(a+\sum_{i=1}^n x_i, b+n-\sum_{i=1}^n x_i)$

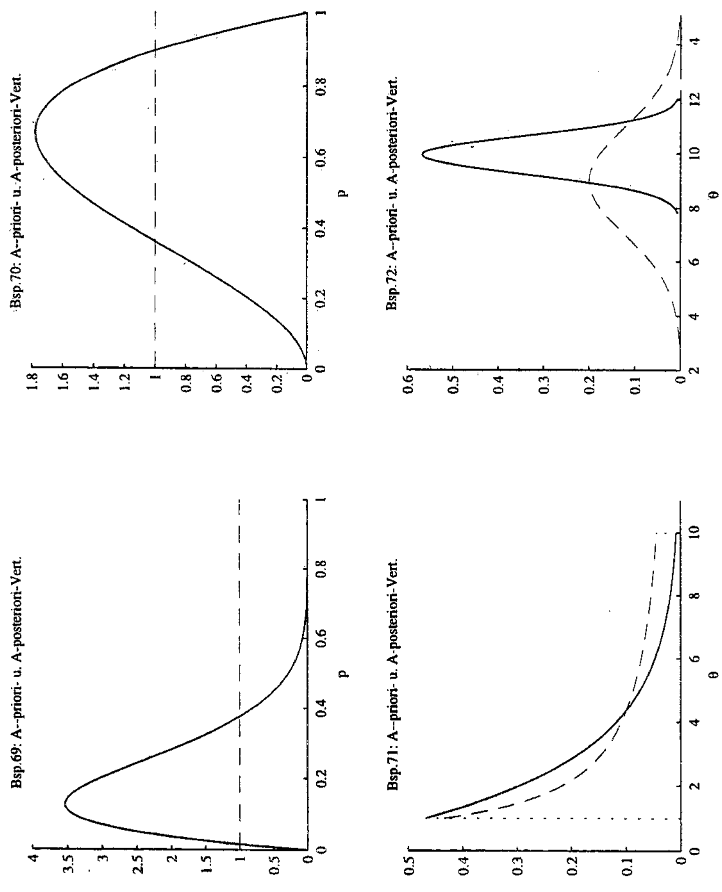
34.4

Def.: Eine Familie $\mathcal{F}$ von W-Vtlgen für den Parameter θ eines SM $X \sim W_\theta$; $\theta \in \Theta$ heißt <u>zum Modell konjugierte A-priori-Familie</u>, falls für jede AprV $\in \mathcal{F}$ und beliebige Daten auch die ApoV $\in \mathcal{F}$.

Bsp.: $X \sim P_\mu$, $\mu > 0$, die Gammavtlgen $Gam(a, \frac{1}{b})$ bilden eine dazu konjugierte A-priori-Familie

Ü

34.5



35. VERWENDUNG DER A-POSTERIORI-VTLG.

In der ApoV alle wesentliche Information enthalten

$X \sim f(\cdot|\theta)$, $\theta \in \Theta$, $\pi(\cdot)$, Daten D, $\pi(\cdot|D)$

35.1 Prädiktivverteilungen

Prognose für $X|D$: Randverteilung von X aus der gemeinsamen Vtlg. von $(X,\tilde{\theta}) \sim f(x|\theta)\pi(\theta|D)$

$$f(x|D) = \int_{\Theta} f(x|\theta)\pi(\theta|D)\,d\theta \qquad \forall x \in M_X$$

$f(\cdot|D)$ heißt <u>Prädiktivdichte</u> für X

35.2 A-posteriori-Bayes-Schätzer

Schätzwert $\hat{\theta}$ für $\theta \in \mathbb{R}$

<u>Def.</u>: Der <u>A-posteriori-Bayes-Schätzer</u> ist, falls er existiert, der Erwartungswert von $\tilde{\theta}|D$,

$$\hat{\theta} = \int_{\Theta} \theta \, \pi(\theta|D) \, d\theta$$

Für Vektorparameter $\theta = (\theta_1, \dots, \theta_k)$ und gerafften Parameter $\tau(\theta) \in \mathbb{R}$ ist der <u>ApoBS</u> für $\tau(\theta)$

$$\widehat{\tau(\theta)} := \mathbb{E}_D \tau(\tilde{\theta}) = \int_{\Theta} \tau(\theta) \pi(\theta|D) \, d\theta$$

falls $\exists$

<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

$$\pi(\theta|x_1, \dots, x_n) \propto \theta^{\sum_1^n x_i} (1-\theta)^{n - \sum_1^n x_i} I_{[0,1]}(\theta)$$

$$\hat{\theta} = \frac{\int_0^1 \theta \cdot \theta^{\Sigma x_i} (1-\theta)^{n-\Sigma x_i} \, d\theta}{\int_0^1 \theta^{\Sigma x_i} (1-\theta)^{n-\Sigma x_i} \, d\theta} = \dots = \frac{\sum_{i=1}^n x_i + 1}{n+2}$$

35.3 HPD - Bereiche

$\theta \in \Theta \subseteq \mathbb{R}^k$, $D = x_1, \dots, x_n \Rightarrow \pi(\cdot | D)$

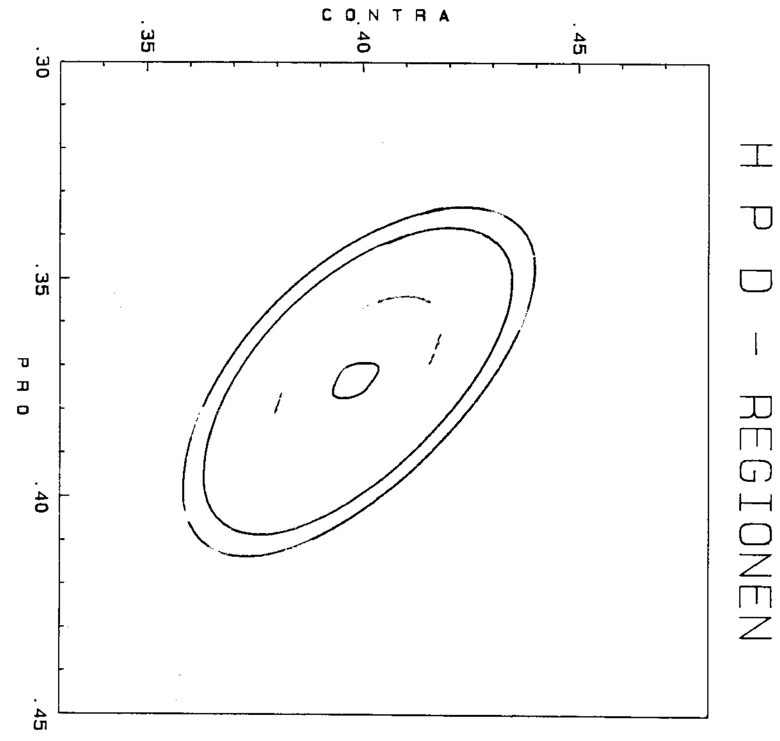
vorgegebene Wahrscheinlichkeit $1-\alpha$ (Sicherheit)

<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



35.3 HPD - Bereiche

$\theta \in \Theta \subseteq \mathbb{R}^k$, $D = x_1, \dots, x_n \Rightarrow \pi(\cdot|D)$

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

35.4 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>: <u>Relative Plausibilität</u> $\frac{\alpha_1}{\alpha_0}$

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

$$\frac{\alpha_1}{\alpha_0} = \frac{\int_{\Theta_1} \pi(\theta|D)\,d\theta}{\int_{\Theta_0} \pi(\theta|D)\,d\theta} = \frac{\int_{\Theta_1} \pi(\theta)\,\ell(\theta; D)\,d\theta}{\int_{\Theta_0} \pi(\theta)\,\ell(\theta; D)\,d\theta} \qquad \text{Ü}$$

36. 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 $D = (x_1, \dots, x_n) \Rightarrow \pi(\cdot \mid D)$ von $\tilde{\theta}$

Entscheidungskriterium: Zu erwartender Verlust

$$\mathbb{E}_{\pi(\cdot \mid D)} L(\tilde{\theta}, d_j)$$

<u>Def.</u>: Die <u>Bayes'sche Entscheidung</u> ist jene, die kleineren a-posteriori zu erwartenden Verlust hat.

<u>Beispiel</u>: Ü

VIII
ERGÄNZUNGEN

37. UNSCHARFE INFORMATION

Sowohl Beobachtungen SG_n als auch A-priori-Verteilungen sind oft nicht exakt, also unscharf (engl. fuzzy).

37.1 Unscharfe Zahlen

Jede exakte reelle Zahl x_0 ist eindeutig charakterisiert durch die Indikatorfunktion $I_{\{x_0\}}(\cdot)$.

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) $\exists x_0 \in \mathbb{R} : \ \xi(x_0) = 1$

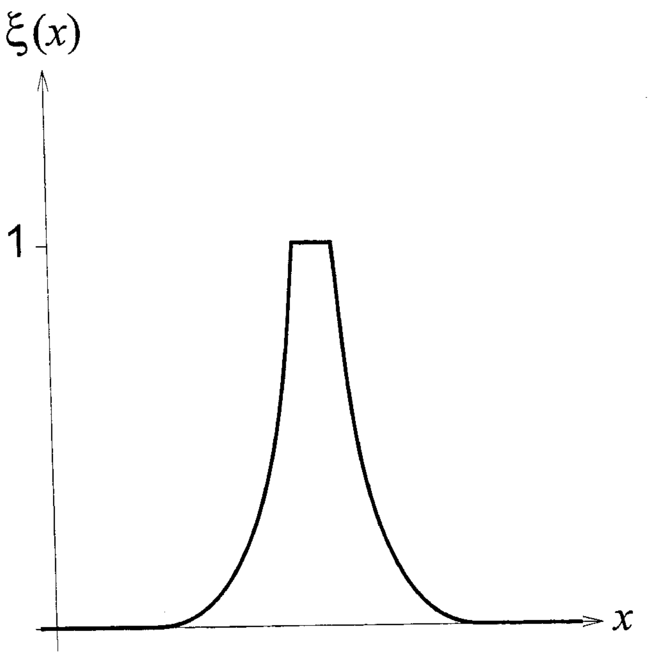
(3) $\forall \delta \in (0,1]$ ist $B_\delta(x^*) := \{x \in \mathbb{R} : \xi(x) \geq \delta\}$ ein abgeschlossenes endliches Intervall $[a_\delta, b_\delta]$ genannt δ-Schnitt

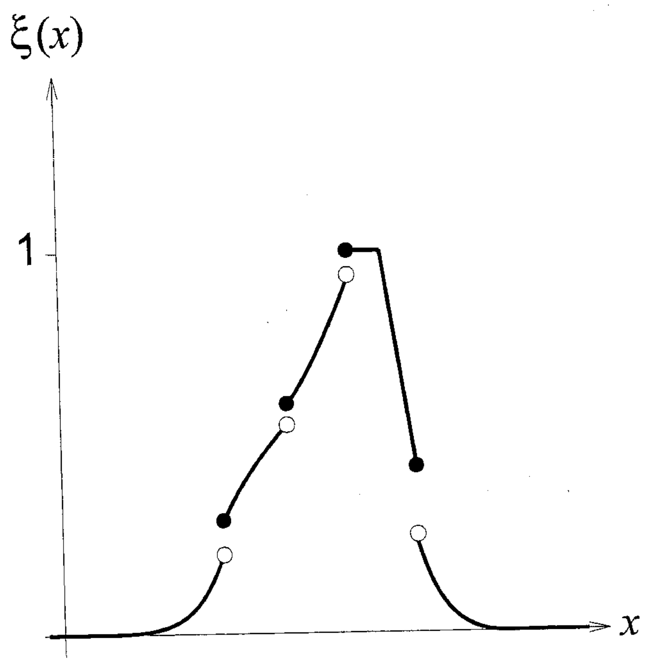
Bem.: Unscharfe Zahlen sind Spezialfälle sog. unscharfer Mengen (= fuzzy sets, ensembles flous)

Bem.: Char. F_n sind etwas anderes als Dichten

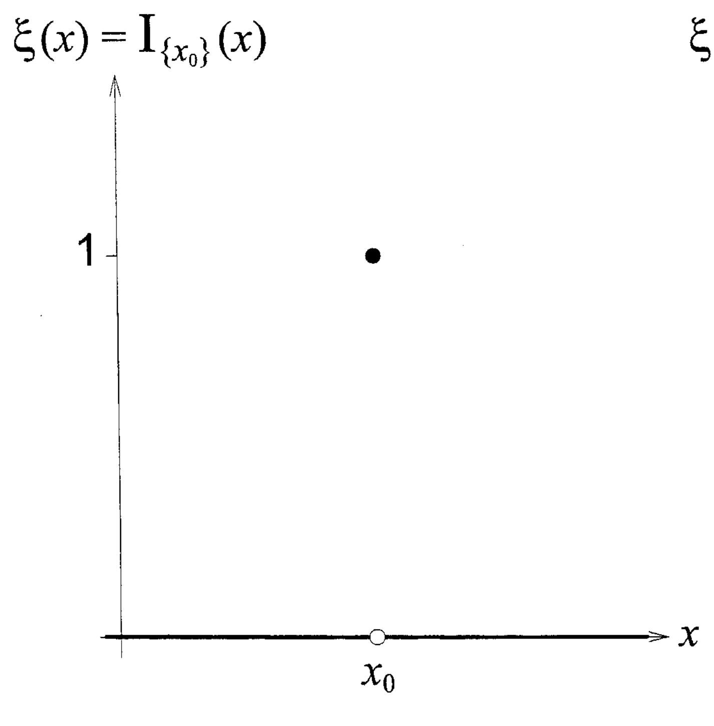
$\int_{\mathbb{R}} \xi(x)dx$ ist das Maß der Unschärfe

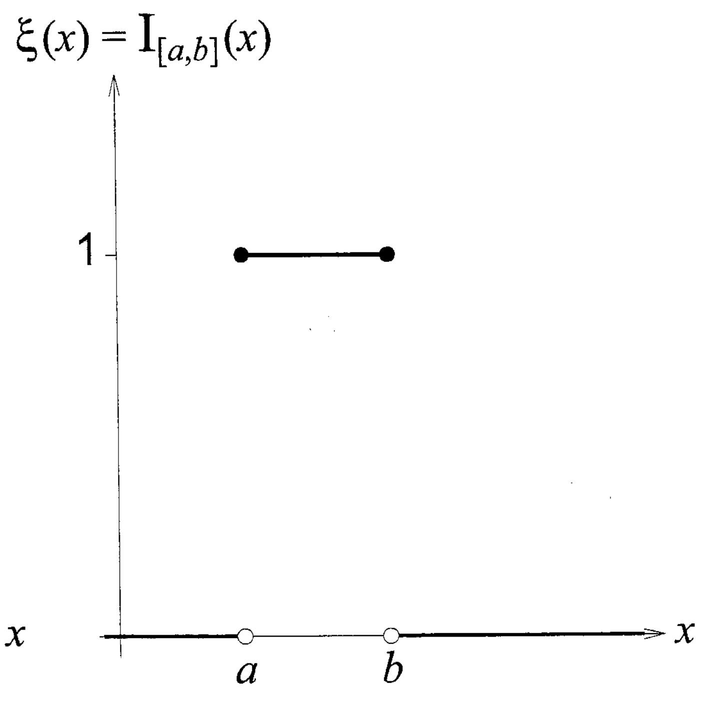
Charakterisierende Funktionen





Spezialfälle charakterisierender Funktionen





37.2 Unscharfe Stichproben

Konkrete SP_n kontinuierlicher SG_n 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 $\zeta : \mathbb{R}^n \to [0,1]$ festgelegt.

Die Kombination von $\xi_1(\cdot), \dots, \xi_n(\cdot)$ erfolgt folgendermaßen:

$$\zeta(x_1, \dots, x_n) := \min_{i=1(1)n} \xi_i(x_i) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

Bem.: δ-Schnitte von $\zeta(\cdot, \dots, \cdot)$:

$$C_\delta(\underline{x}^*) = \bigtimes_{i=1}^{n} C_\delta(x_i^*) \quad \forall \delta \in (0,1]$$

37.3 Verallgemeinerte Schätzungen

Schätzfunktion $\vartheta : M_X^n \longrightarrow \Theta$

bei exakten Beobachtungen $x_1, \cdots, x_n$

Schätzwert $\hat{\theta} = \vartheta(x_1, \cdots, x_n)$

Für unscharfe Beobachtungen $x_1^*, \cdots, x_n^*$
und kombin. unscharfen Vektor $\underline{x}^*$ mit vcF $\xi(\cdot, \cdots, \cdot)$
ergibt $\vartheta(x_1^*, \cdots, x_n^*)$ einen unscharfen Wert $\hat{\theta}^*$

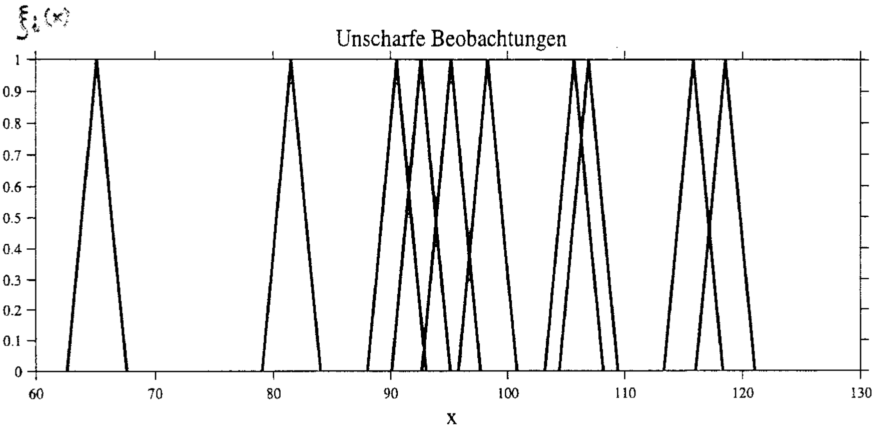
Die char. F. von $\hat{\theta}^*$ erhält man mittels des sog. <u>Erweiterungsprinzips</u> aus der FST :

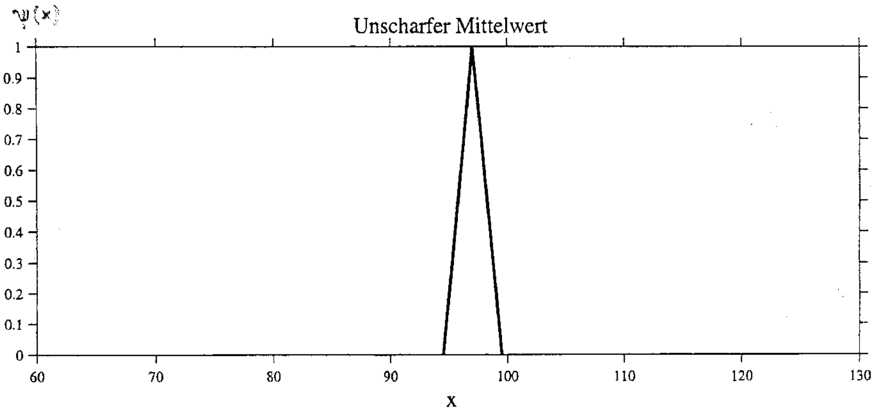
Für die c.F. $\psi(\cdot)$ von $\hat{\theta}^*$ gilt mit $\underline{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$

$$\psi(\theta) = \begin{cases} \sup\{\xi(\underline{x}) : \vartheta(\underline{x}) = \theta\} & \text{falls } \vartheta^{-1}(\theta) \neq \emptyset \\ 0 & \text{falls } \vartheta^{-1}(\theta) = \emptyset \end{cases} \quad \forall\, \theta \in \Theta$$

<u>Bem.</u>: 1) Für $\theta \in \mathbb{R}$ und stetige Funktion $\vartheta(\cdot, \cdots, \cdot)$ ist obiges $\psi(\cdot)$ eine c.F. einer unscharfen Zahl.

2) Für Vektorparameter erhält man ein unscharfes Element (unscharfe Teilmenge) von Θ.


$\xi_i(x)$
Unscharfe Beobachtungen
x


$\psi(x)$
Unscharfer Mittelwert
x