II

BESCHREIBENDE STATISTIK

Darstellung und Analyse von Daten
ohne Zuhilfenahme von Stochastik

5. Aufbereitung von Daten

7.X.'2

Erhebungen und Versuche liefern oft eine große Anzahl von Daten, die zu überschaubarer Information zusammengefasst werden sollen:

Tabellen, Kennzahlen, Diagramme

5.1 Diskrete Größen

Solche Größen können höchstens abzählbar viele verschiedene Werte Annehmen, die sich nicht häuf

Für Artmerkmale Strichlisten, Balkendiagramme Kreisdiagramme:

Für n Beobachtungen mit den möglichen Ausprägungen $A_1, A_2, \ldots, A_m$ bestimmt man die absoluten Häufigkeiten $H_n(A_j)$, $j = 1(1)m$

$H_n(A_j)$ = Anzahl der Beobachtungen mit Ausprägung A_j

und die relativen Häufigkeiten $h_n(A_j)$, $j = 1(1)$

$$h_n(A_j) = \frac{H_n(A_j)}{n} \quad \forall\, j = 1(1)m$$

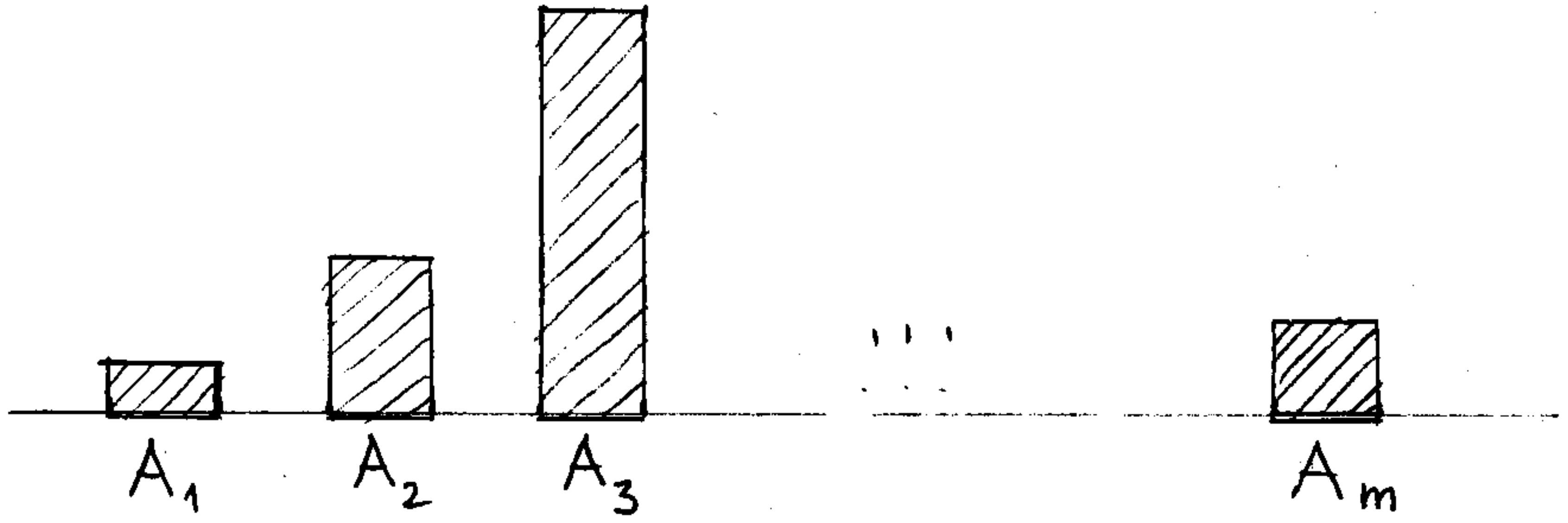
Die Ermittlung der Häufigkeiten erfolgt mit sog Strichlisten.

7.X.'2

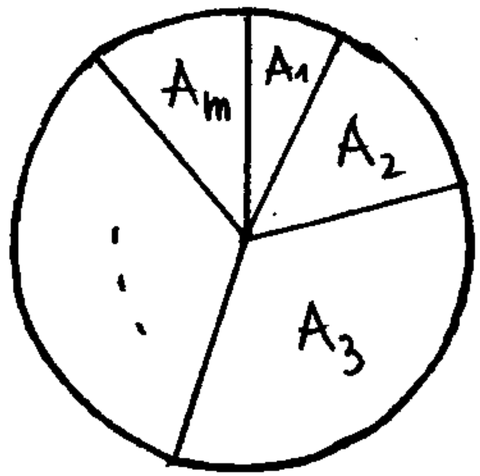
Häufigkeitstabelle

Merkmal-Ausprägung	Strichliste	Absolute Häufigkeit	Relative Häufigkeit
A_1	𝍸 𝍸 II …	$H_n(A_1)$	$h_n(A_1)$
A_2	𝍸 𝍸 …	$H_n(A_2)$	$h_n(A_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_j	𝍸 𝍸 𝍸 II …	$H_n(A_j)$	$h_n(A_j)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
A_m	III …	$H_n(A_m)$	$h_n(A_m)$
	Summe	n	1

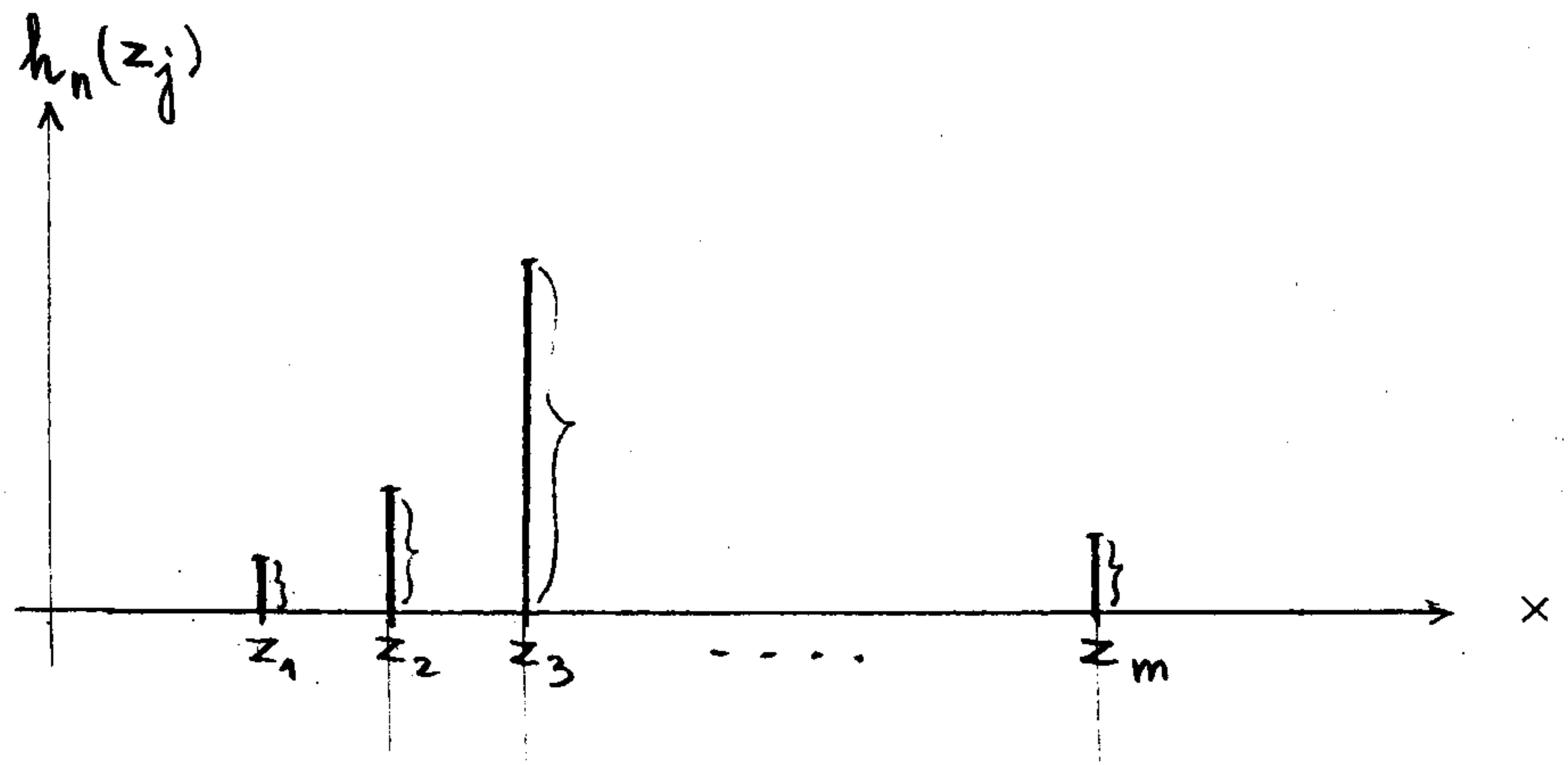
Balkendiagramm: Höhen der Balken proportional zu den Häufigkeiten



Kreisdiagramm: Sektorenflächen proportional zu den Häufigkeiten



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

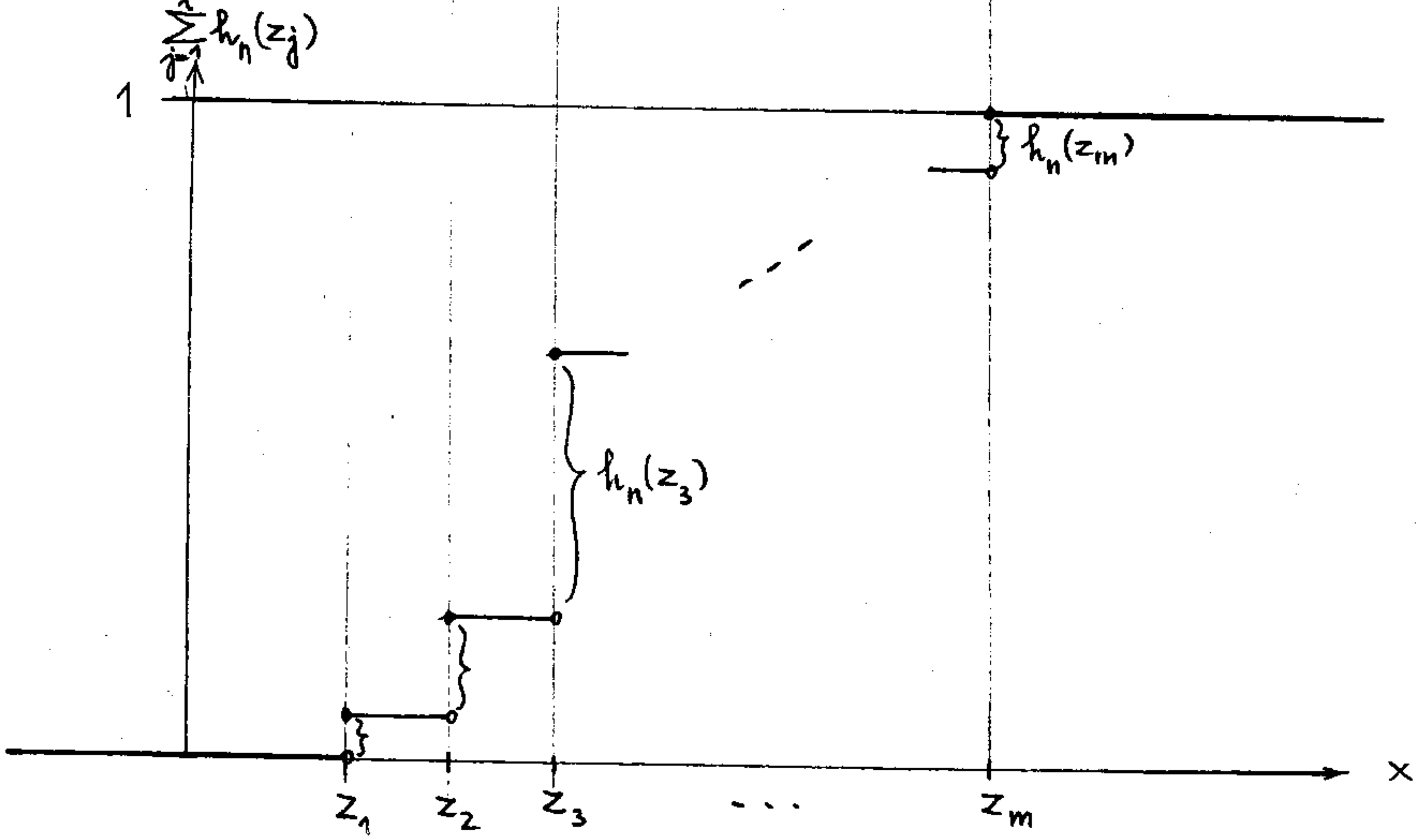


Manchmal werden Summenhäufigkeiten

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

gewünscht und als reelle Funktion dargestellt:

<u>Summenkurve</u>



5.2 Kontinuierliche 1-dim. Größen

Können alle Werte (Zahlen) aus einem Intervall $\langle a,b \rangle$ annehmen.

<u>Für kleinere Beobachtungszahlen</u> $(n < 30)$ und $(-\infty, x]$: $H_n((-\infty, x])$ und $h_n((-\infty, x])$

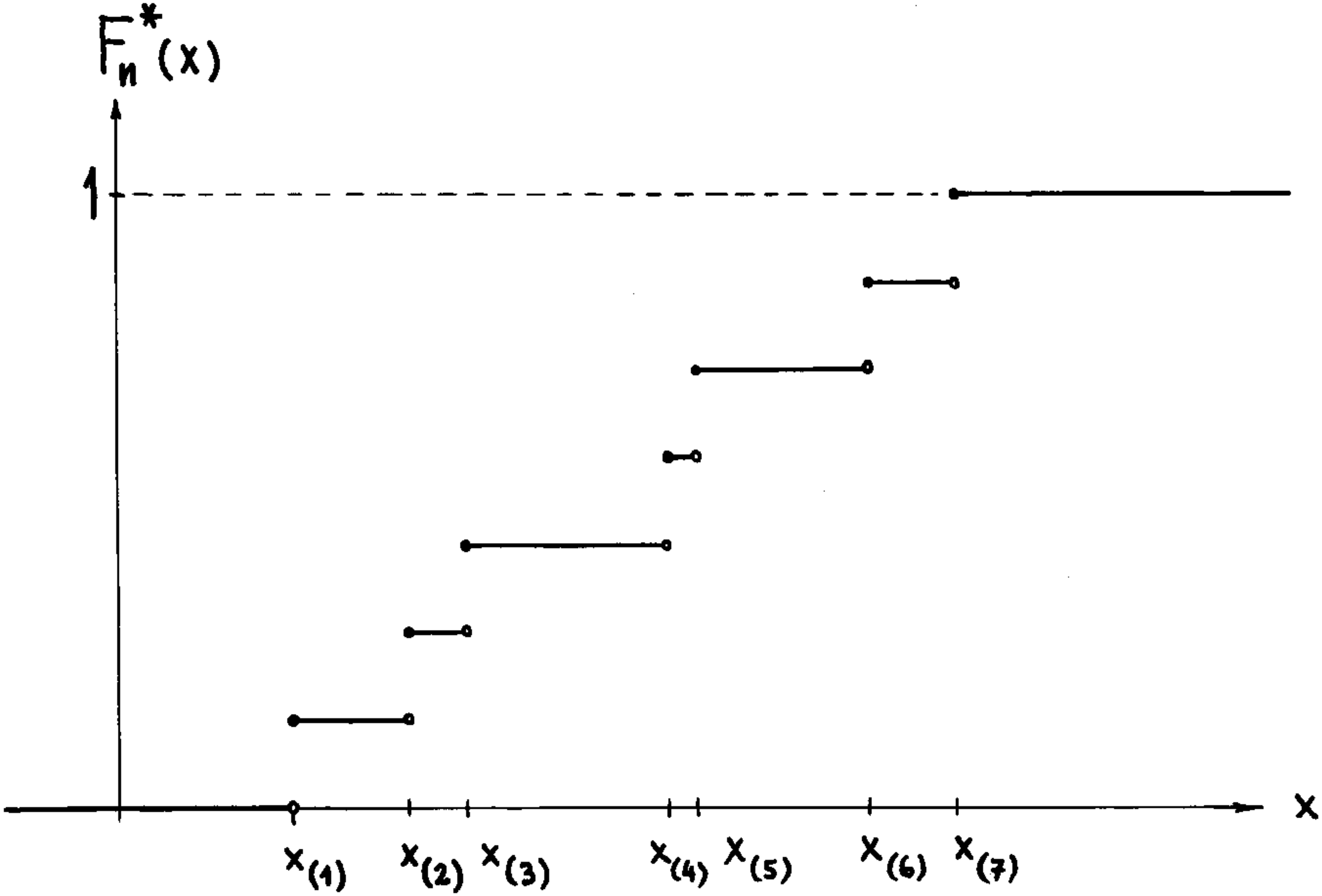
<u>Bem.</u>: Für Beobn. $x_1, \dots, x_n$ hinreichend

$$H_n((-\infty, x_i]) \quad \forall\, i = 1(1)n$$

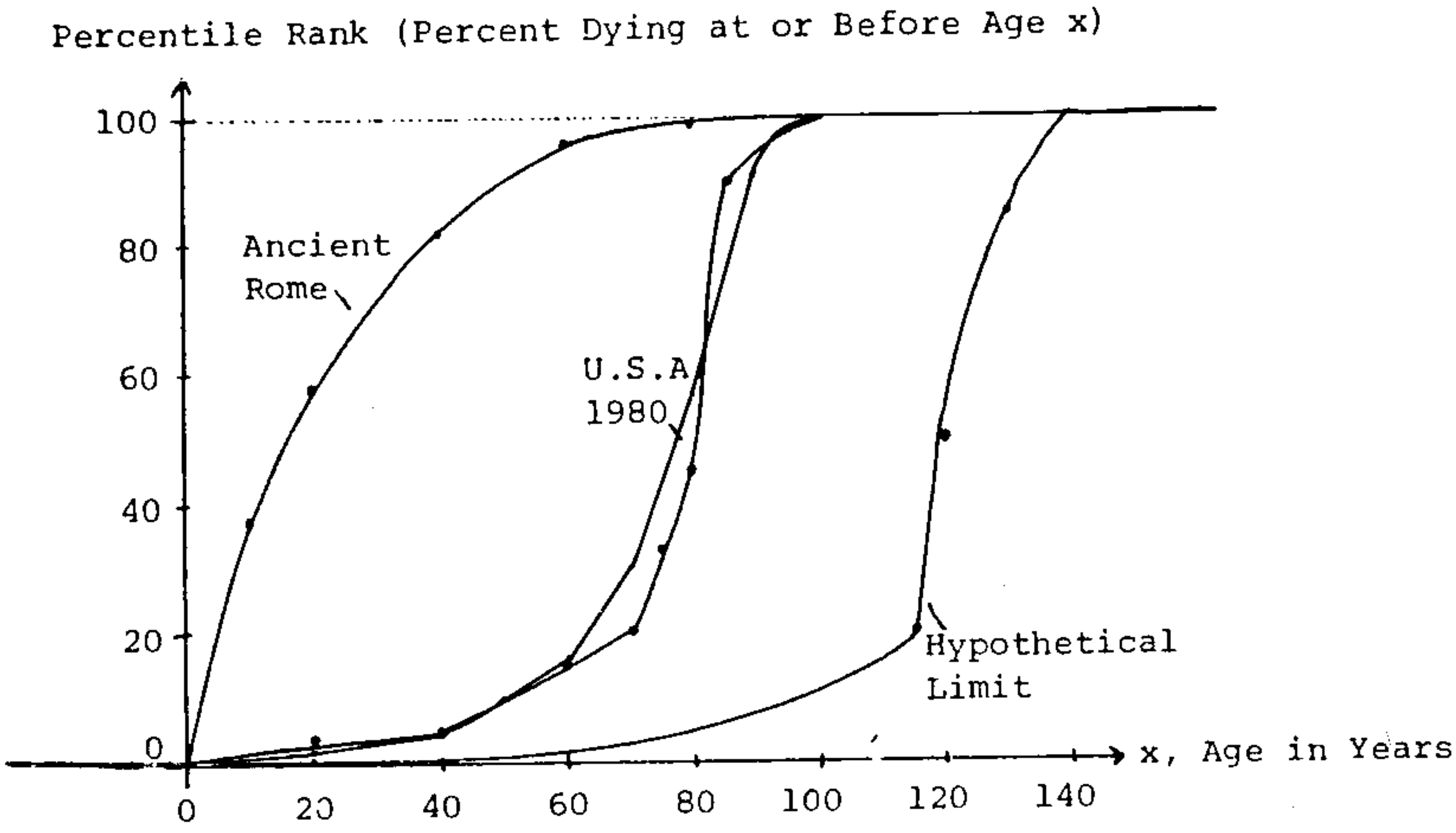
Grafische Darstellung $\quad x \to h_n((-\infty, x])$

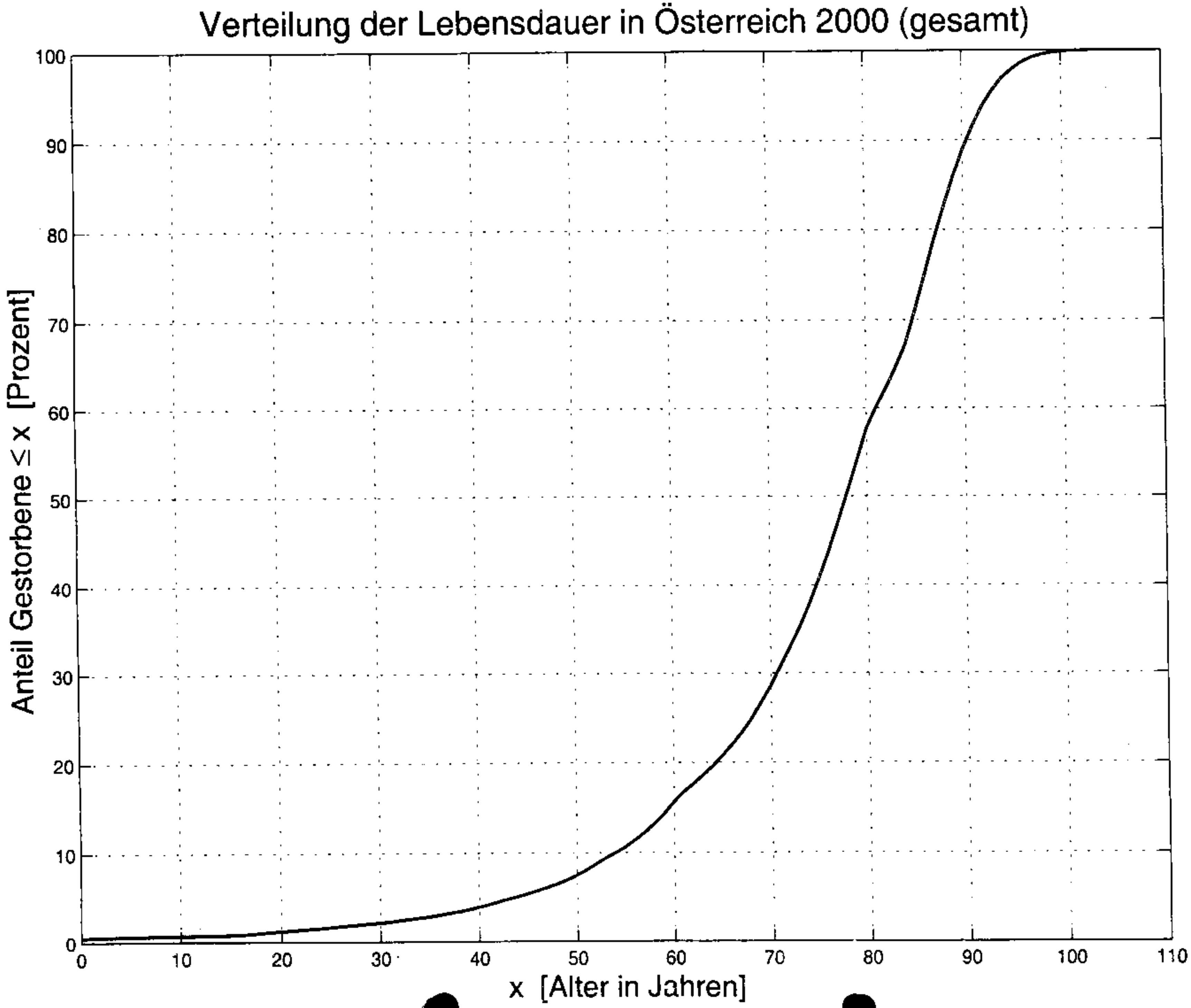
$$F_n^*: \mathbb{R} \longrightarrow [0,1]$$

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

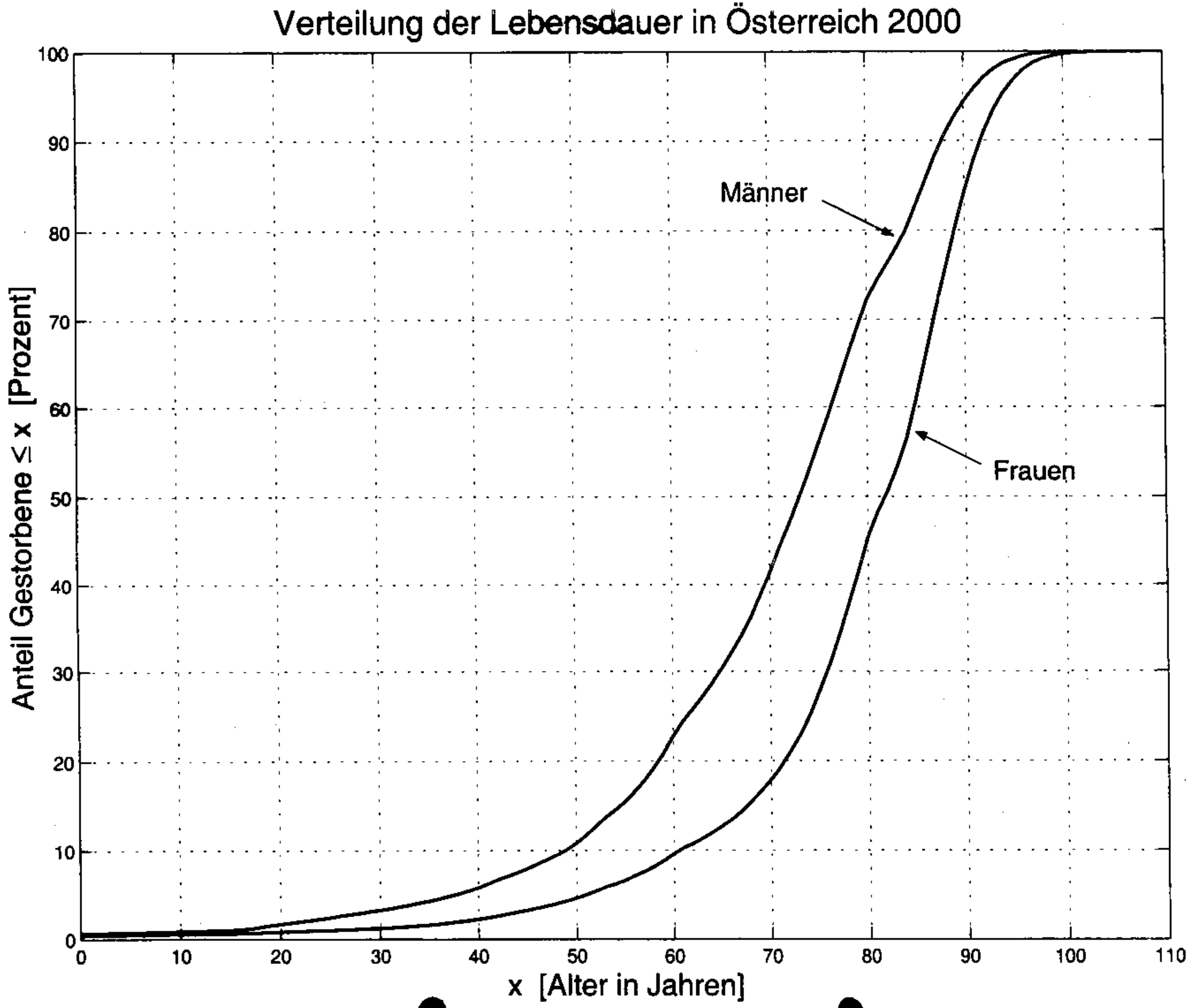


5.8

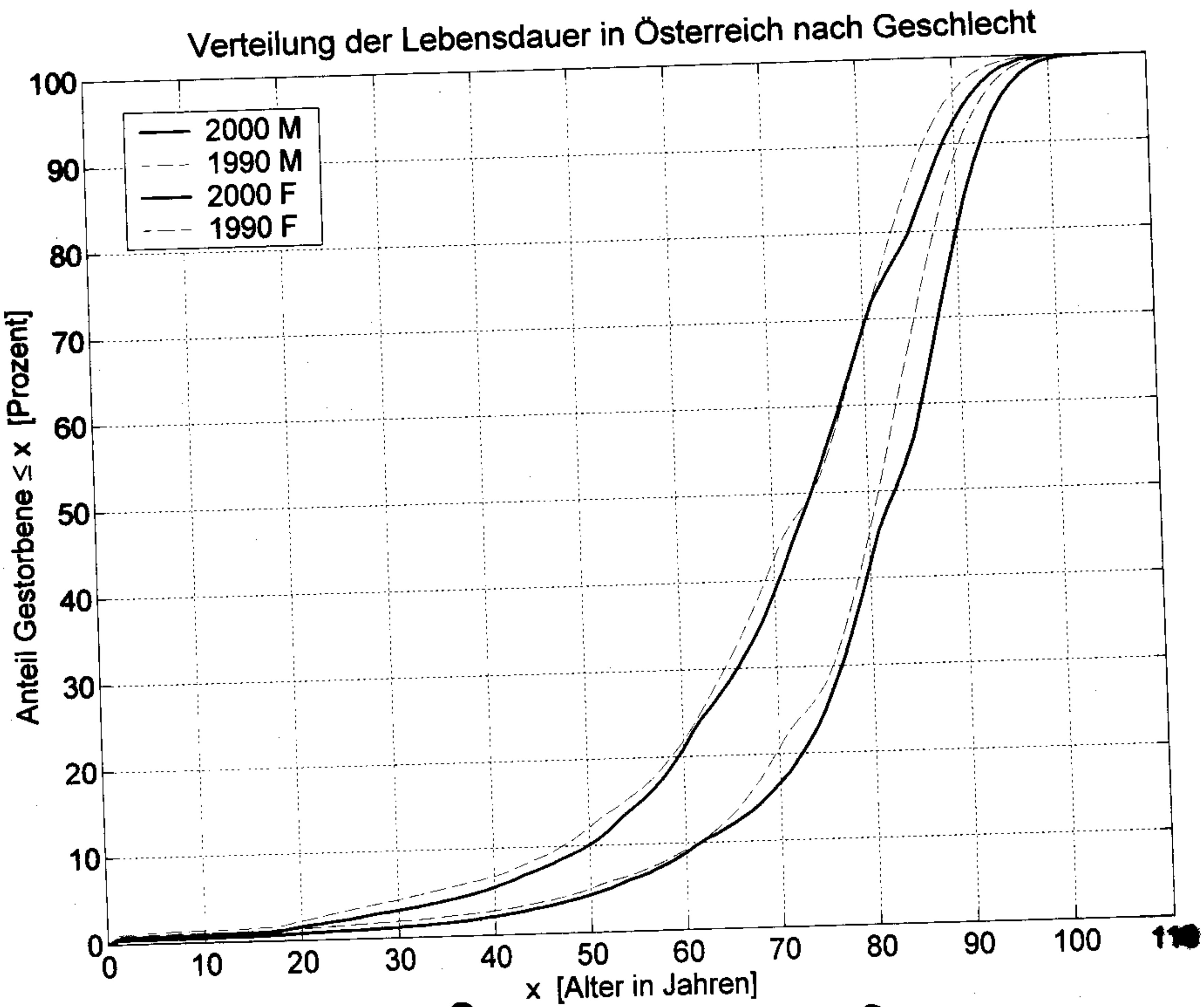

Percentile Rank (Percent Dying at or Before Age x)
100
80
60
40
20
0
0
20
40
60
80
100
120
140
x, Age in Years
Ancient
Rome
U.S.A
1980
Hypothetical
Limit



Verteilung der Lebensdauer in Österreich 2000 (gesamt)
Anteil Gestorbene ≤ x [Prozent]
x [Alter in Jahren]
0
10
20
30
40
50
60
70
80
90
100
110



Verteilung der Lebensdauer in Österreich 2000
Anteil Gestorbene ≤ x [Prozent]
x [Alter in Jahren]
Männer
Frauen
0
10
20
30
40
50
60
70
80
90
100
110



Verteilung der Lebensdauer in Österreich nach Geschlecht
2000 M
1990 M
2000 F
1990 F
Anteil Gestorbene ≤ x [Prozent]
x [Alter in Jahren]
0
10
20
30
40
50
60
70
80
90
100
110

Für größere Beobachtungszahlen n meist <u>Klasseneinteilungen</u> des sog. Merkmalraumes

$$\begin{array}{ccccc} [c_1, c_2], & (c_2, c_3], & (c_3, c_4], & \dots, & (c_m, c_{m+1}] \\ K_1 & K_2 & K_3 & \dots & K_m \end{array}$$

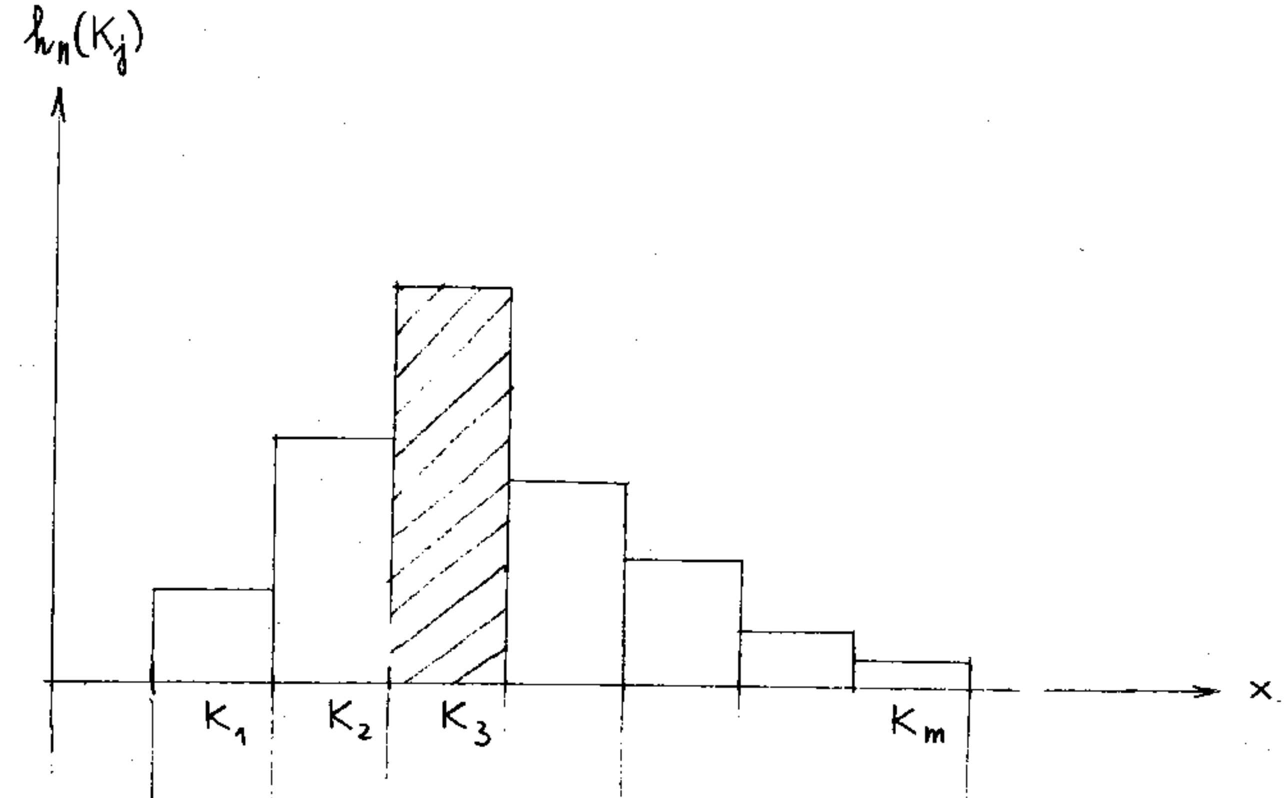
Häufigkeiten der <u>Klassen</u> K_j, $j = 1(1)m$

Analog zu Balkendiagrammen und Stabdiagrammen ermittelt man sog. <u>Histogramme</u>. (Abb.)

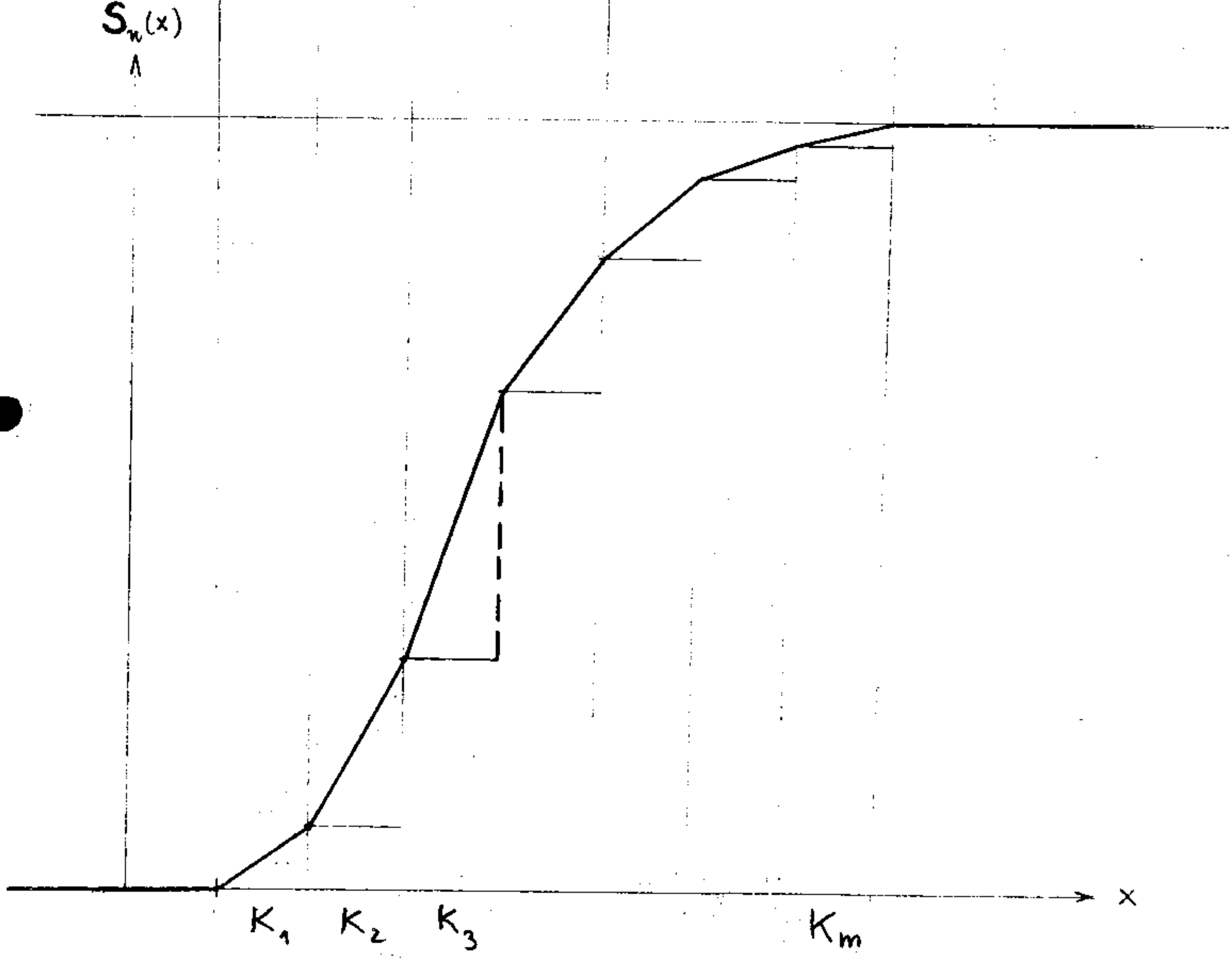
Analog zu Summenkurven ermittelt man sog. <u>Summenpolygone</u>. (Abb.)

Kontinuierliche Merkmale

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

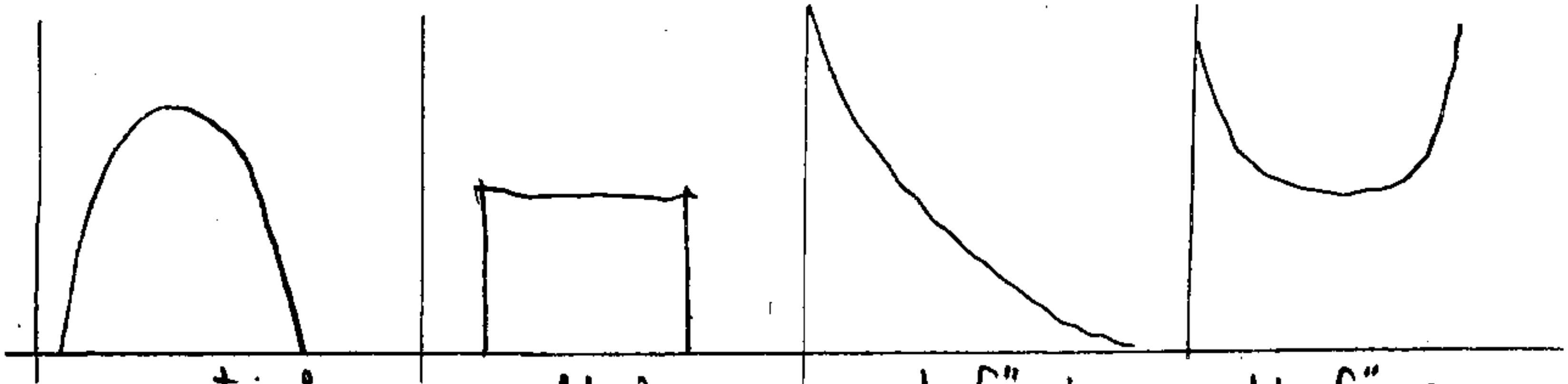


$h_n(K_j)$
x
K_1
K_2
K_3
K_m

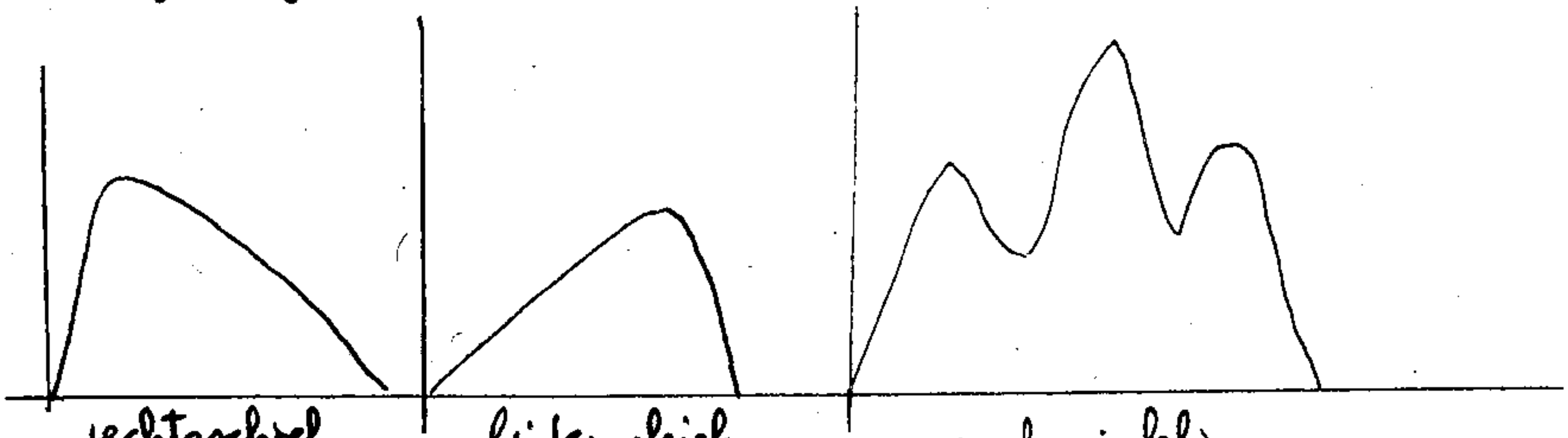


$S_n(x)$
x
K_1
K_2
K_3
K_m

Typen von Histogrammen (Verteilungen)

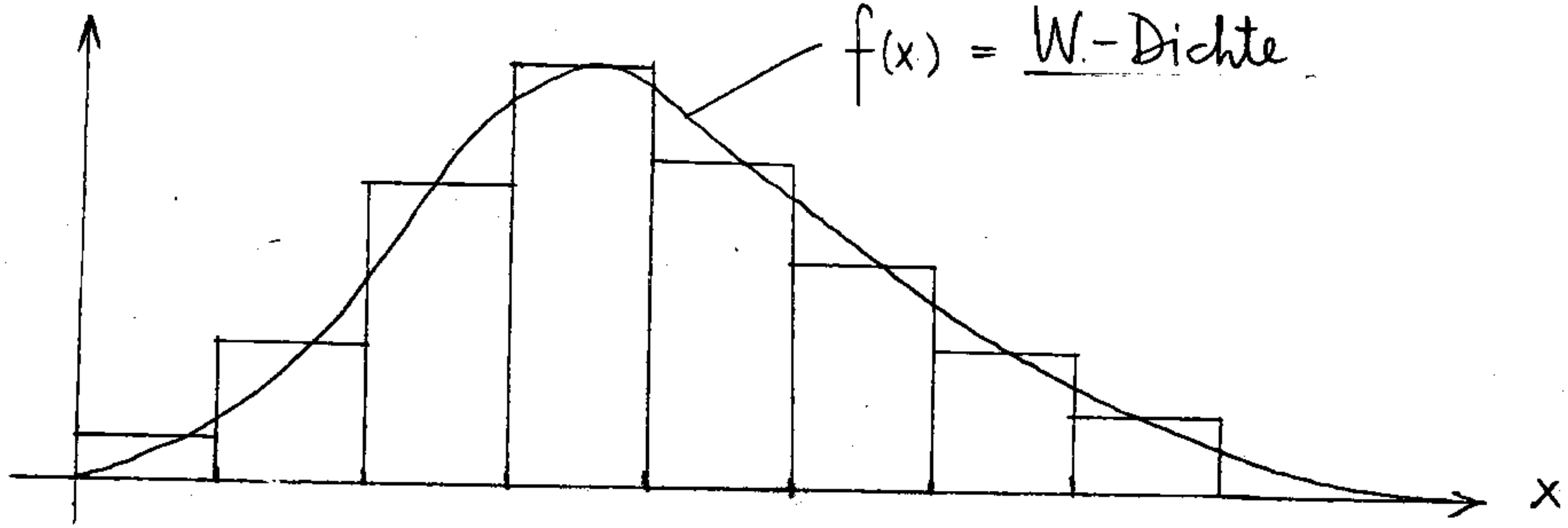


symmetrisch eingipfelig | rechteckig | J-förmig | U-förmig



rechtsschief linkssteil | linksschief rechtssteil | mehrgipfelig

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



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