

FUZZY INFORMATION AND STATISTICS

Reinhard Viertl

Vienna University of Technology

www.statistik.tuwien.ac.at/public/viertl/

FUZZY INFORMATION

- Fuzzy Data
- Fuzzy a-priori Knowledge
- Fuzzy Probabilities
- Soft Computing ECSC

KINDS OF DATA UNCERTAINTY

Variability

Errors

Missing Values

Imprecision (Fuzzy Data)

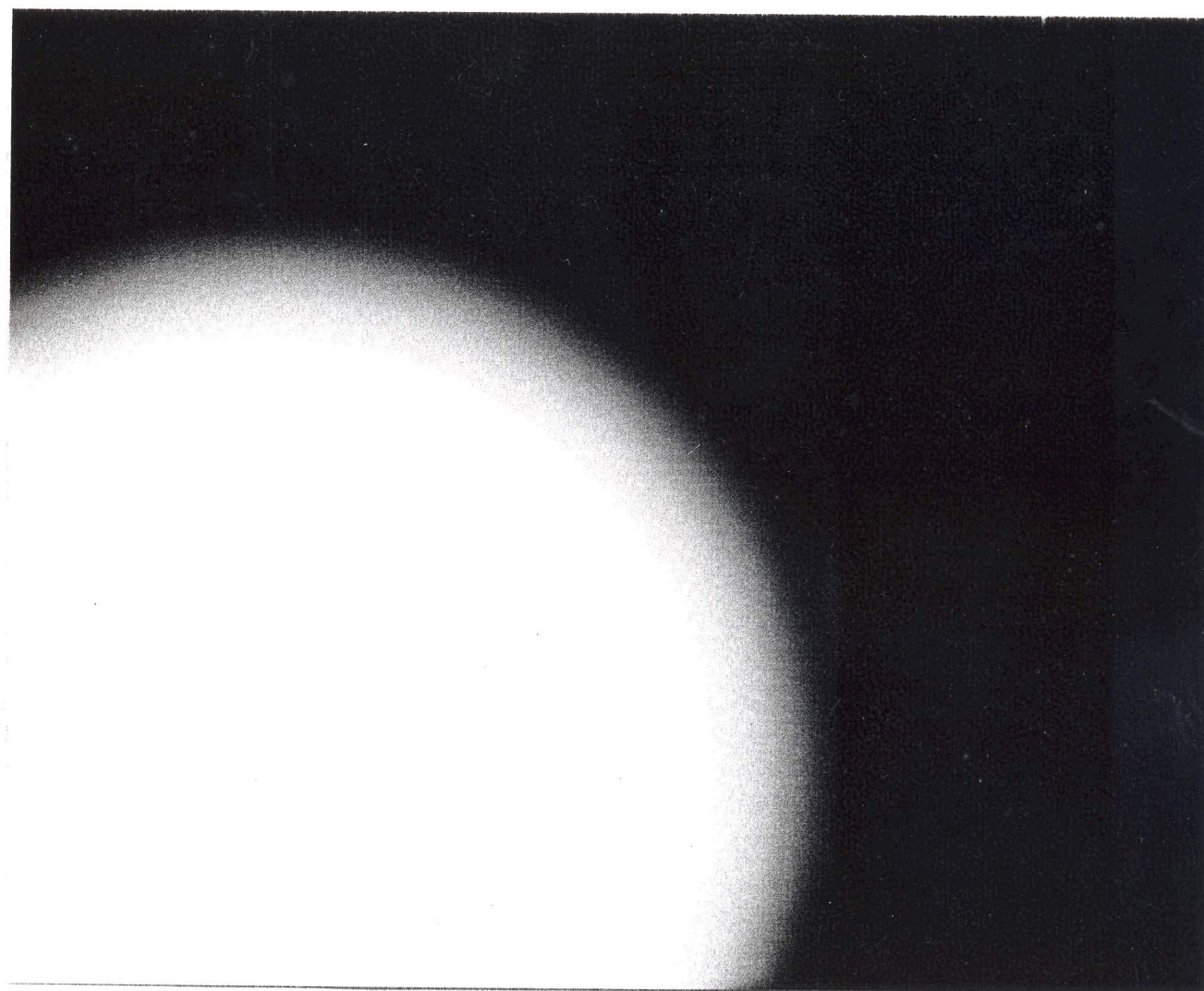
Description: Fuzzy Numbers

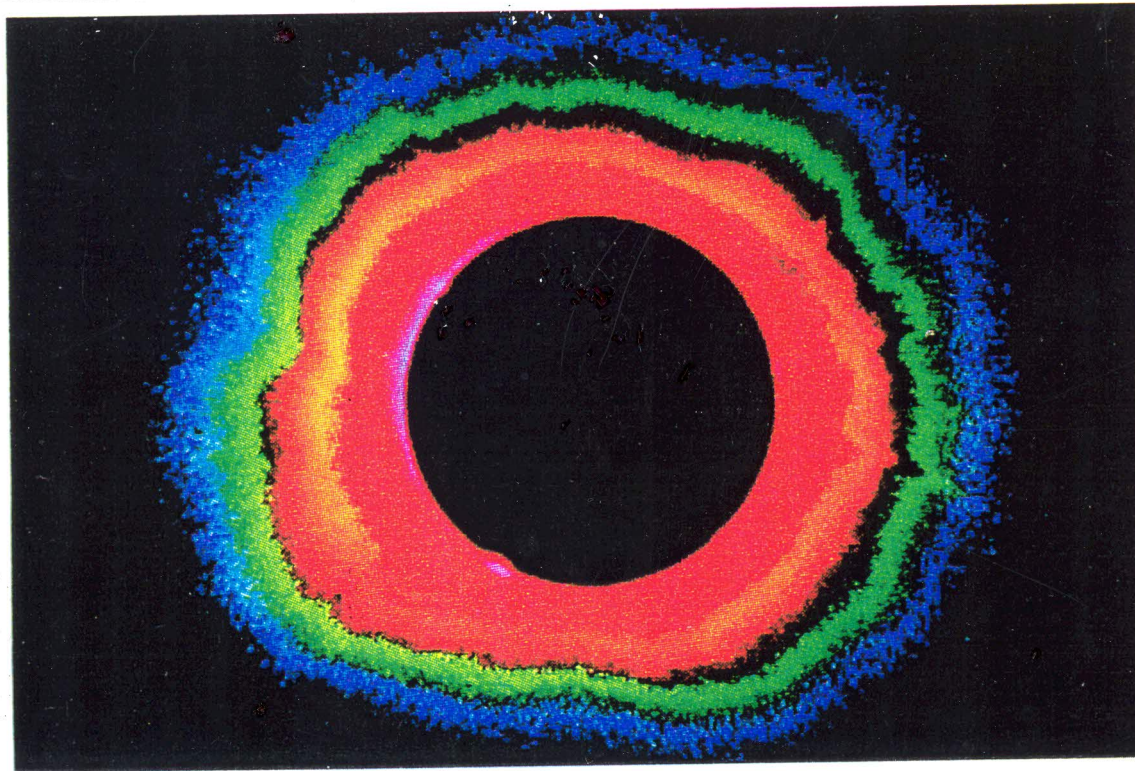
Fuzzy Vectors

Fuzzy Functions

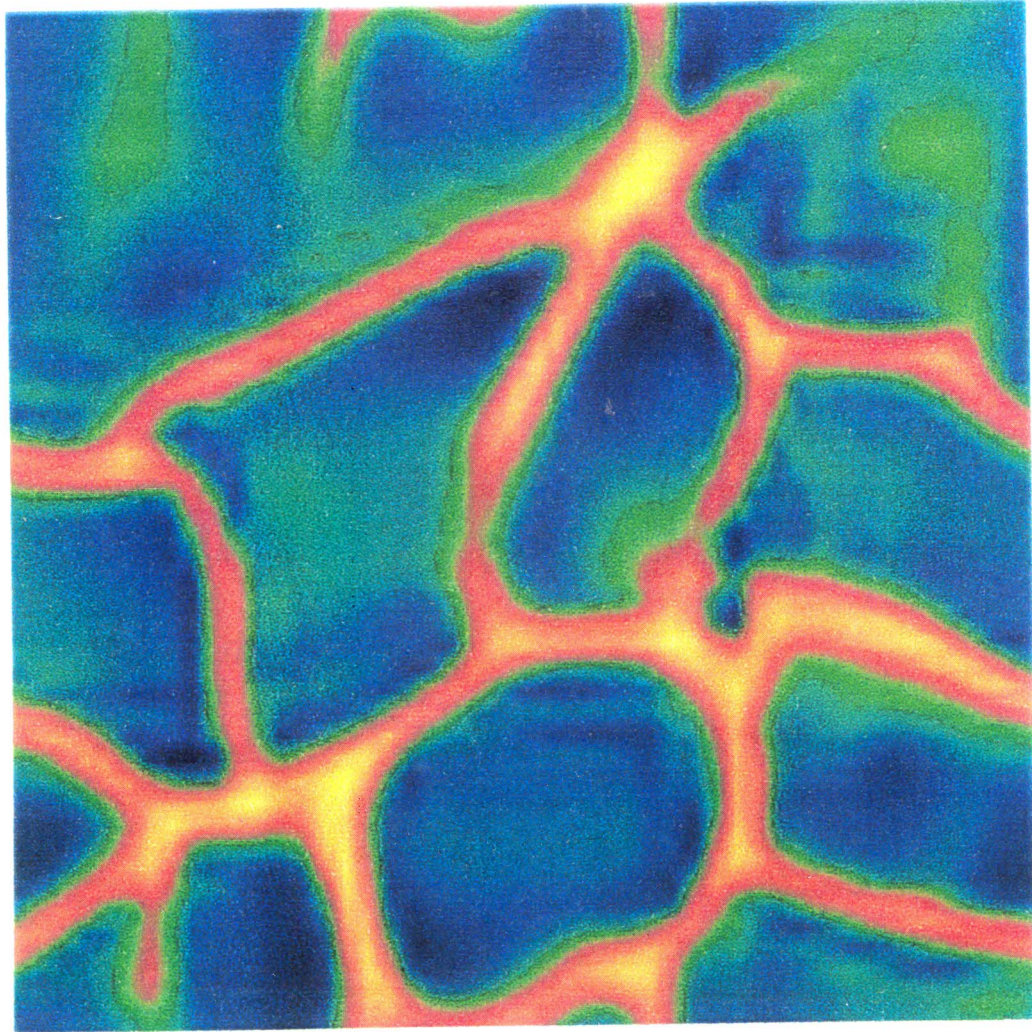
FUZZY DATA

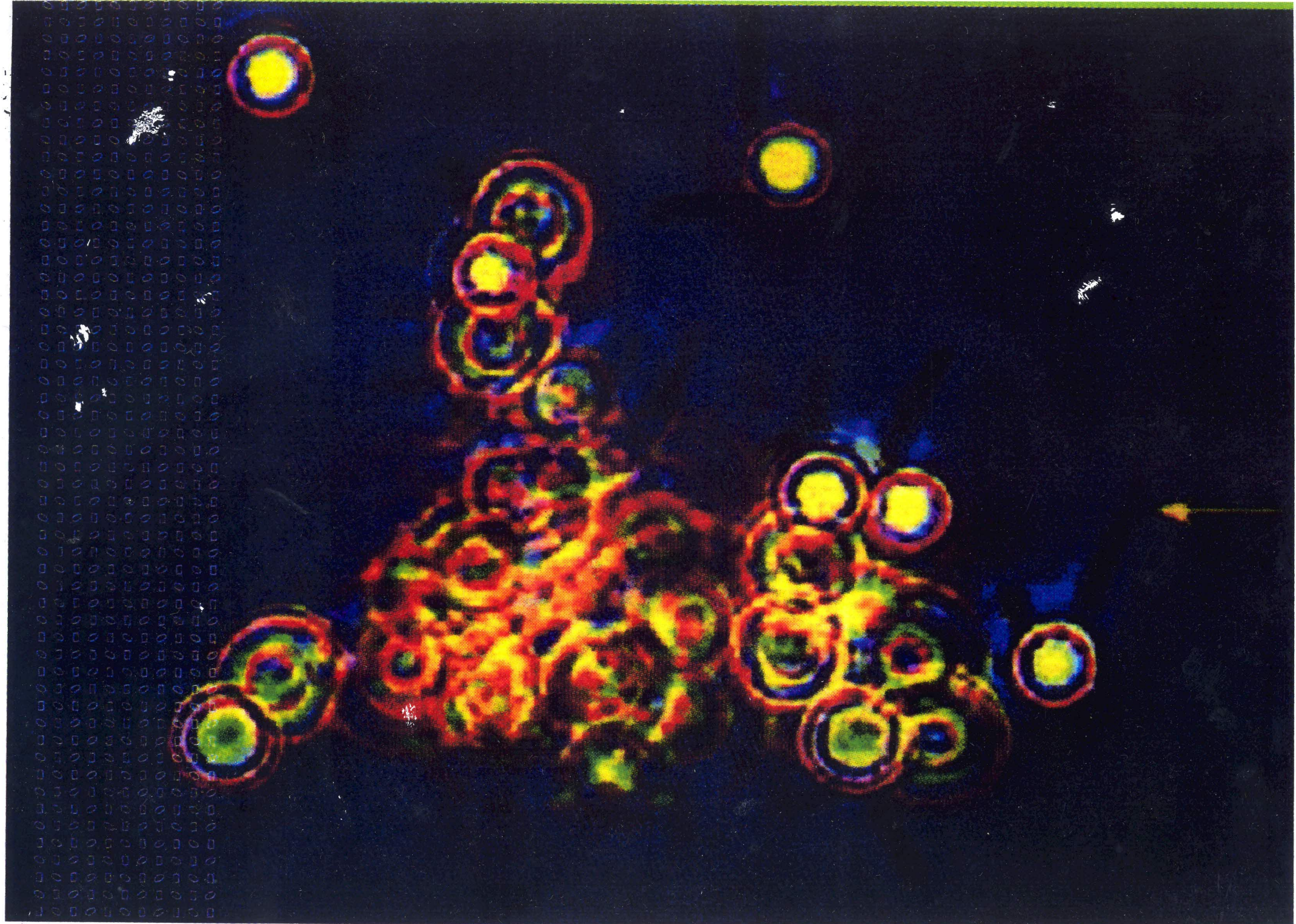
- Environmental Data
- Recovering Times
- Quality of Life Data
- Migration Data
- \vdots
- Precision Measurement Data

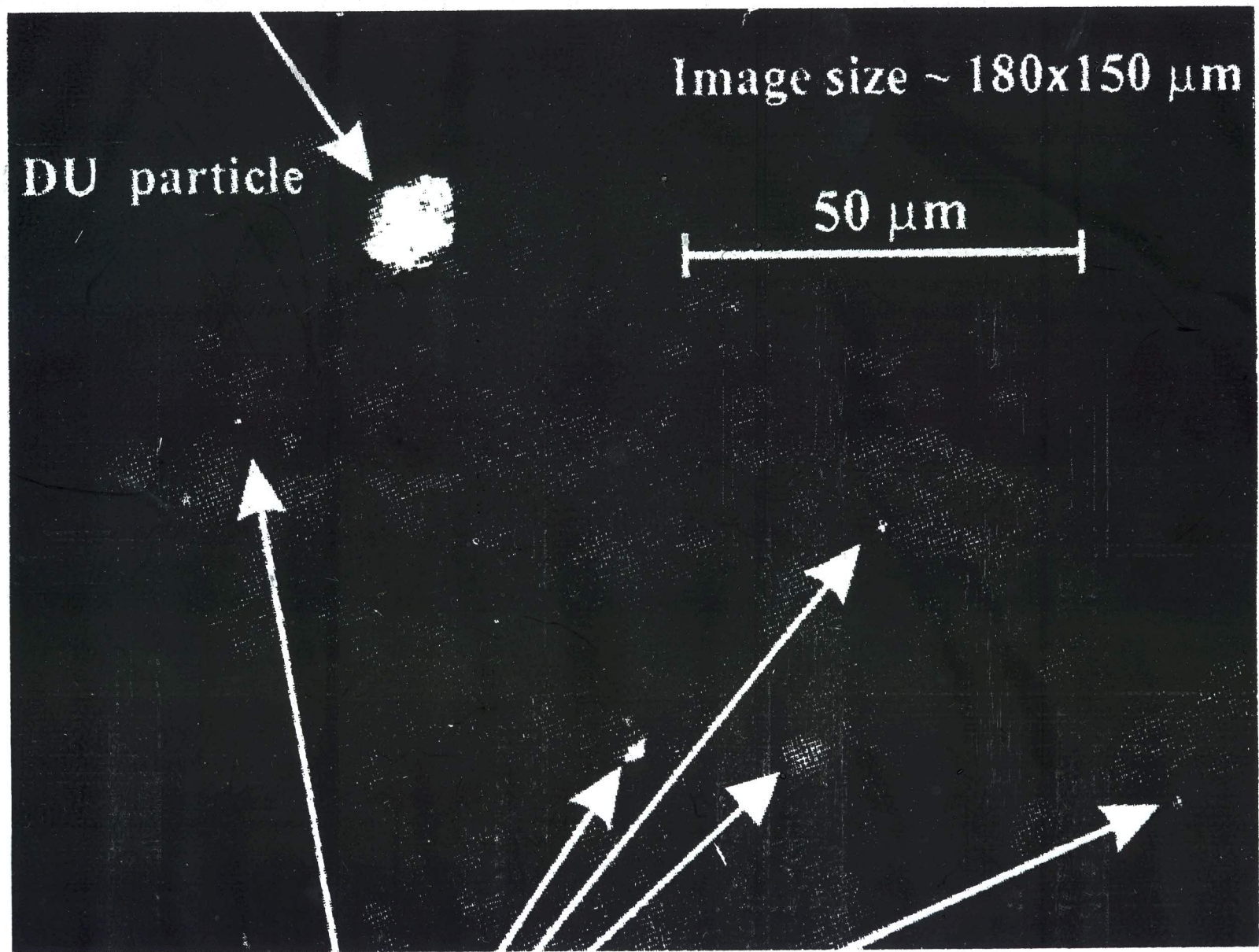


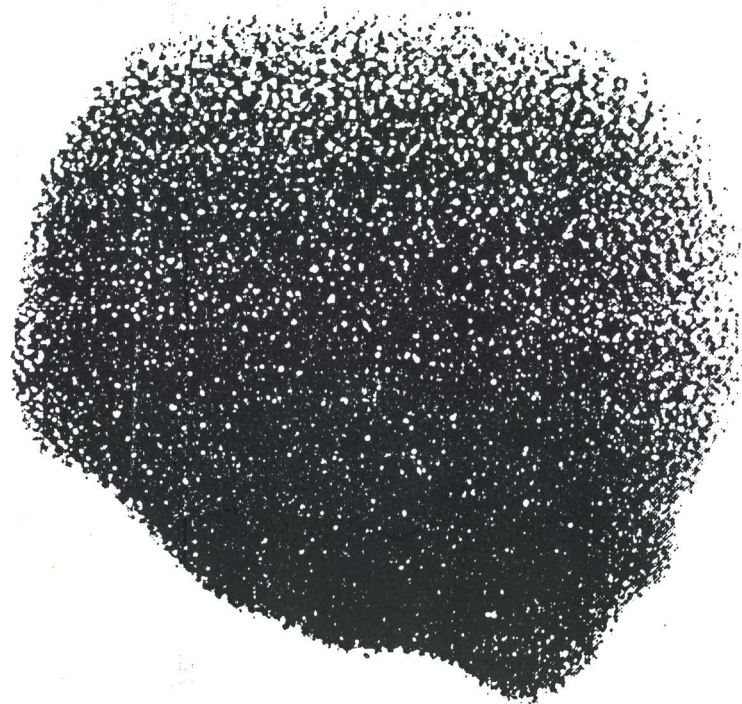


**False colour map of
solar corona,
showing contours of
equal brightness.**

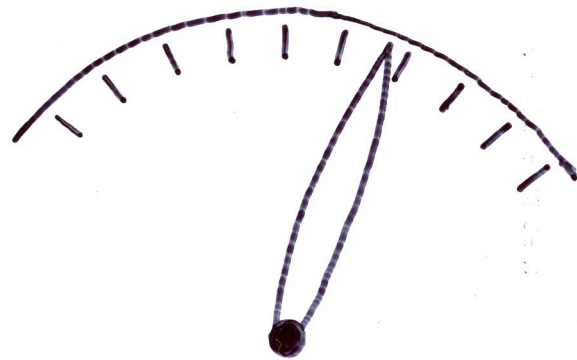








MEASUREMENTS



analog

4.823

digital

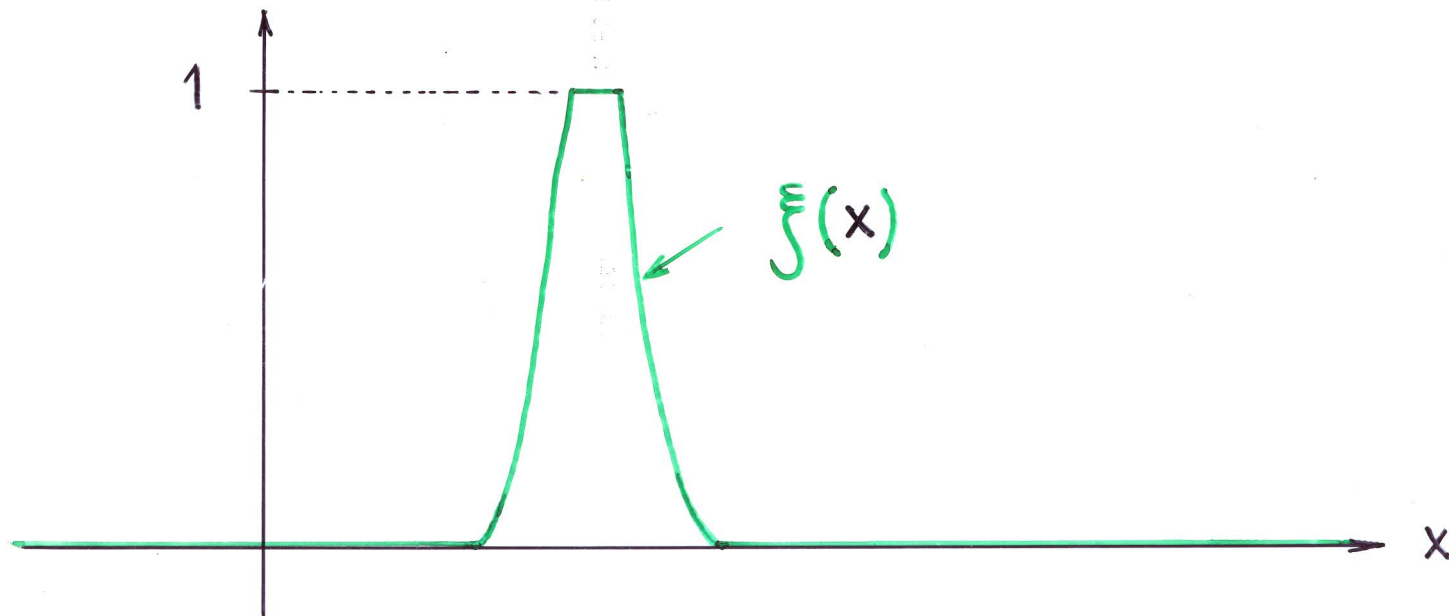
Results ?

MEASUREMENT RESULTS

Not precise numbers but more or less non-precise

Mathematical model: Fuzzy number x^*

Characterizing function $\xi(\cdot)$



Characterizing Function $\xi(\cdot)$

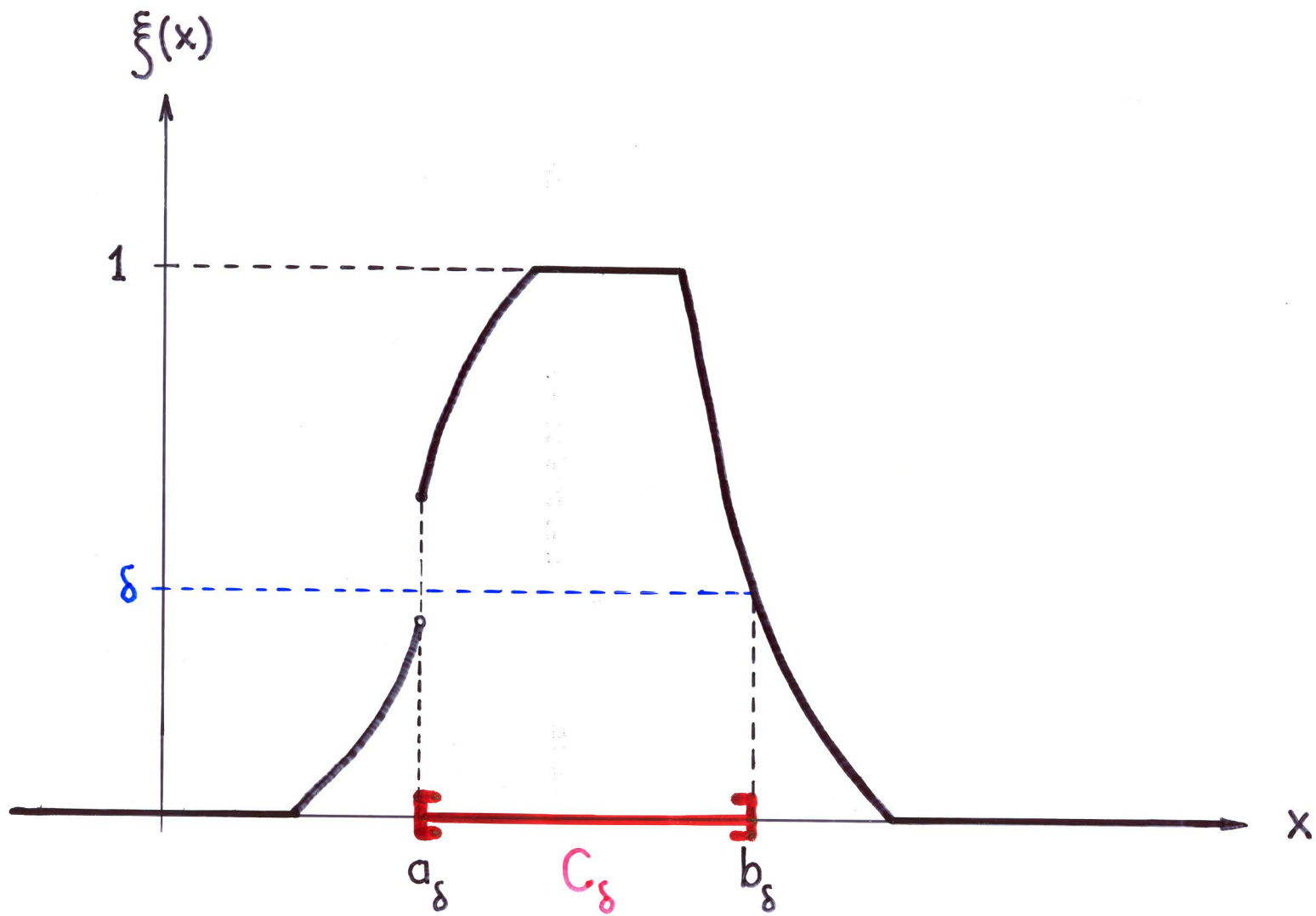
(1) $0 \leq \xi(x) \leq 1 \quad \forall x \in \mathbb{R}$

(2) support $[\xi(\cdot)]$ is bounded

(3) $\forall \delta \in (0, 1]$ the δ -Cut C_δ

$$C_\delta = \{x \in \mathbb{R} : \xi(x) \geq \delta\} \neq \emptyset$$

is a closed interval $[a_\delta; b_\delta]$



Lemma: For a characterizing function of a fuzzy number the following holds:

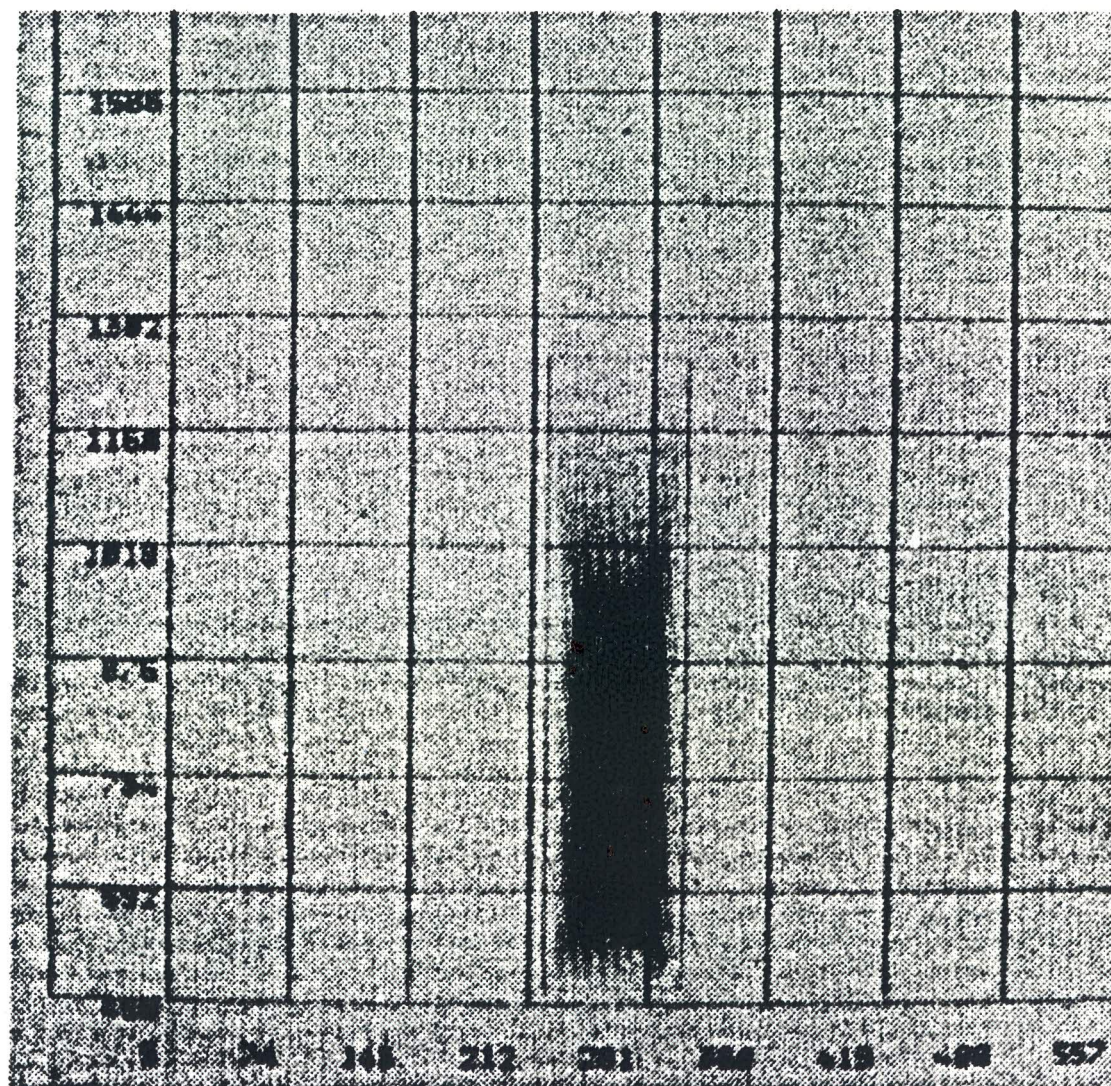
$$\xi(x) = \max \{ \delta \cdot \mathbb{1}_{C_\delta[\xi(\cdot)]}(x) : \delta \in [0, 1] \} \quad \forall x \in \mathbb{R}$$

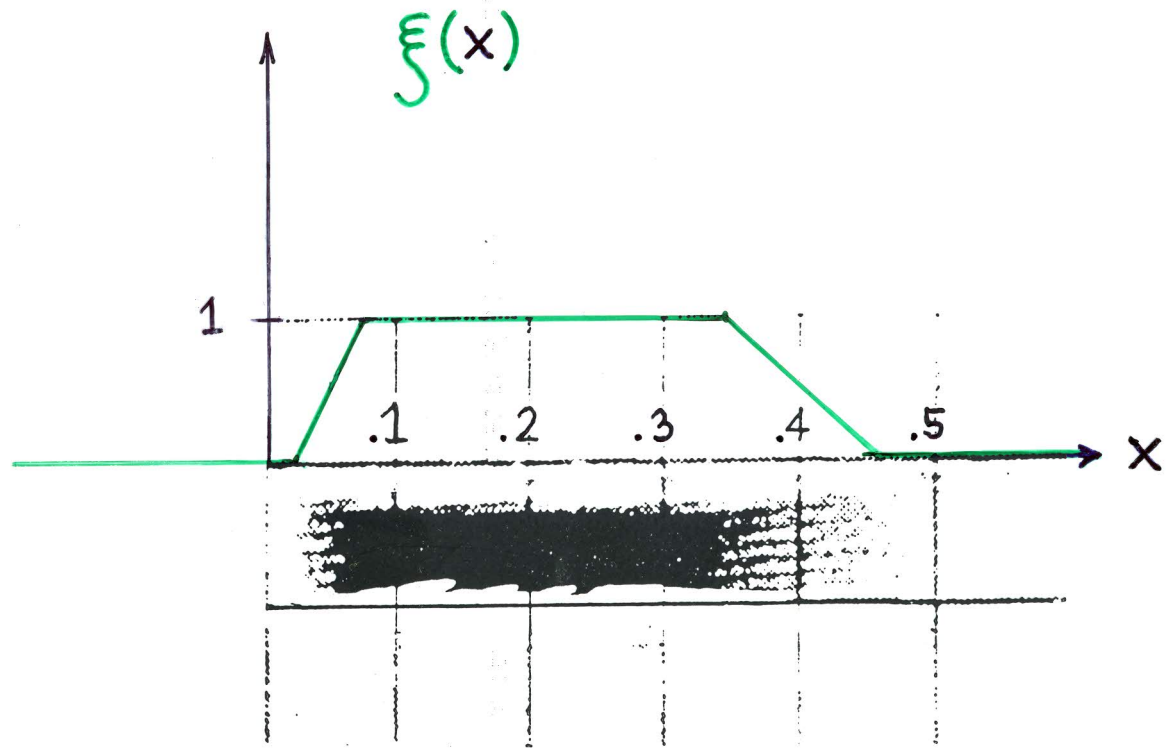
Remark: Not all families of nested closed intervals $[a_\delta; b_\delta]$ are the δ -cuts of a fuzzy number.

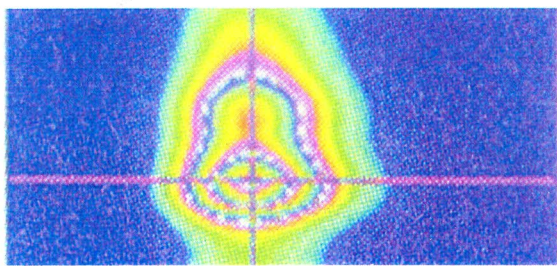
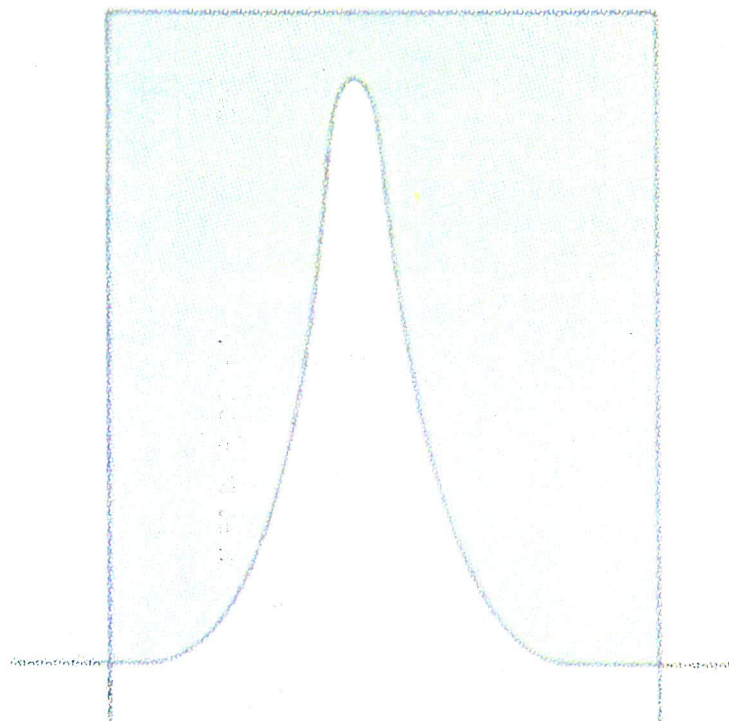
But the following definition yields a fuzzy number:

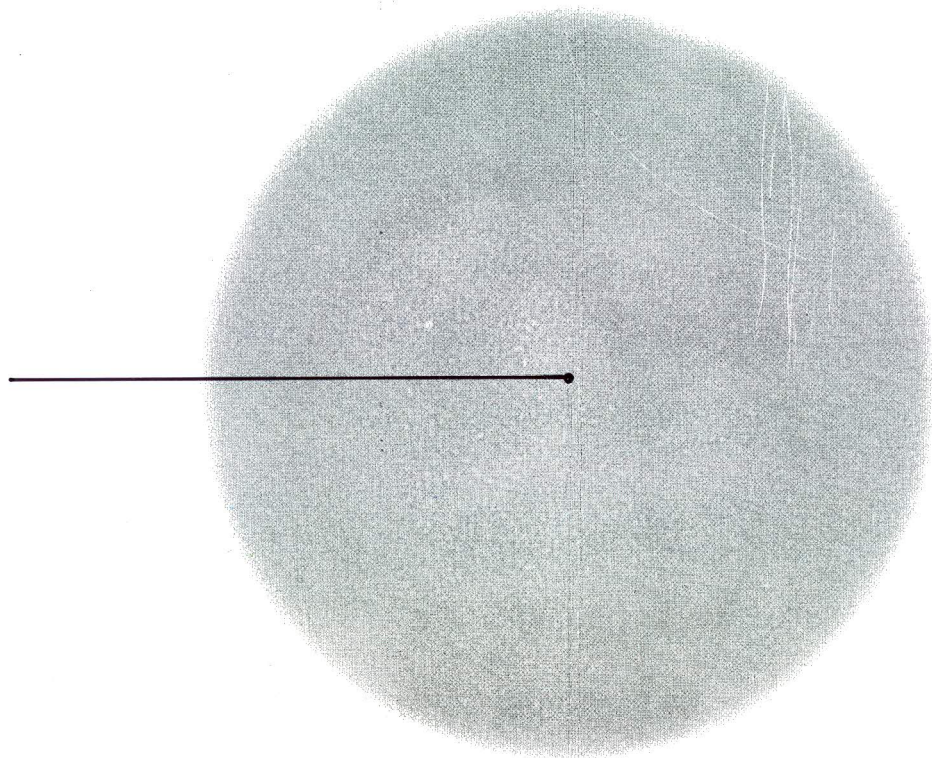
$$\xi(x) := \sup \{ \delta \cdot \mathbb{1}_{[a_\delta; b_\delta]}(x) : \delta \in [0, 1] \} \quad \forall x \in \mathbb{R}$$

one observation as presented by the screen









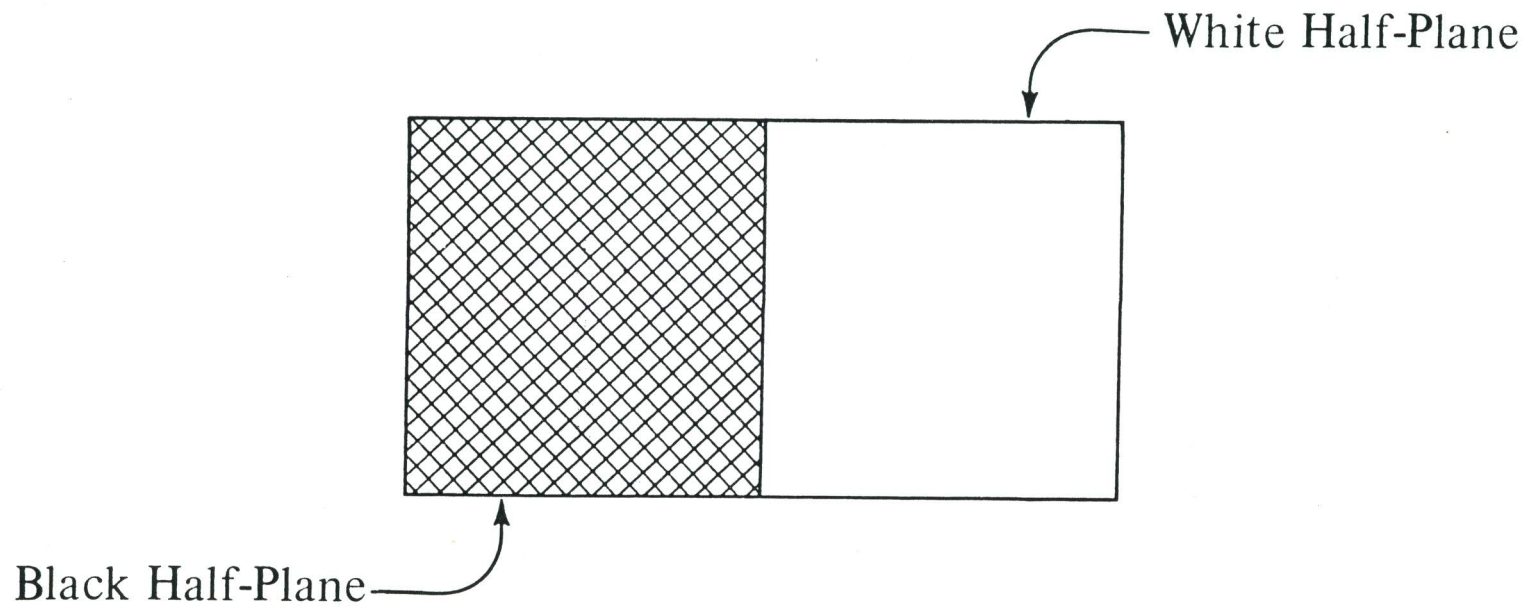
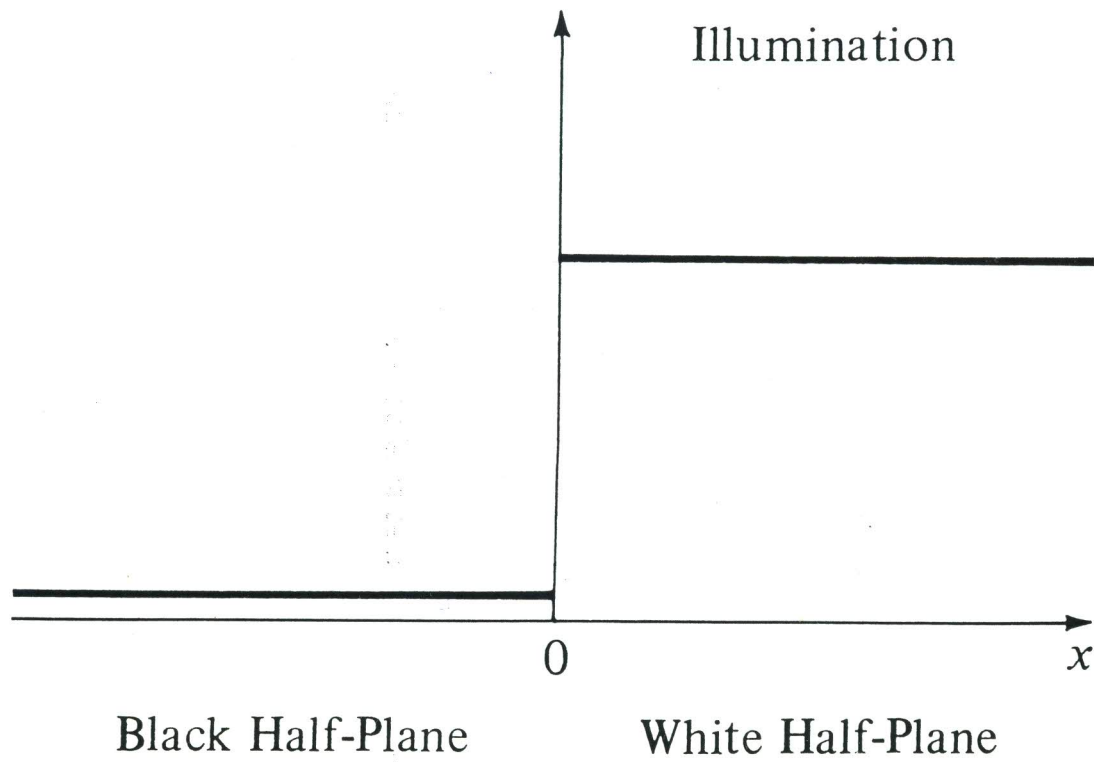


Figure Black and white half-planes.



Figure

Ideal illumination on a horizontal line

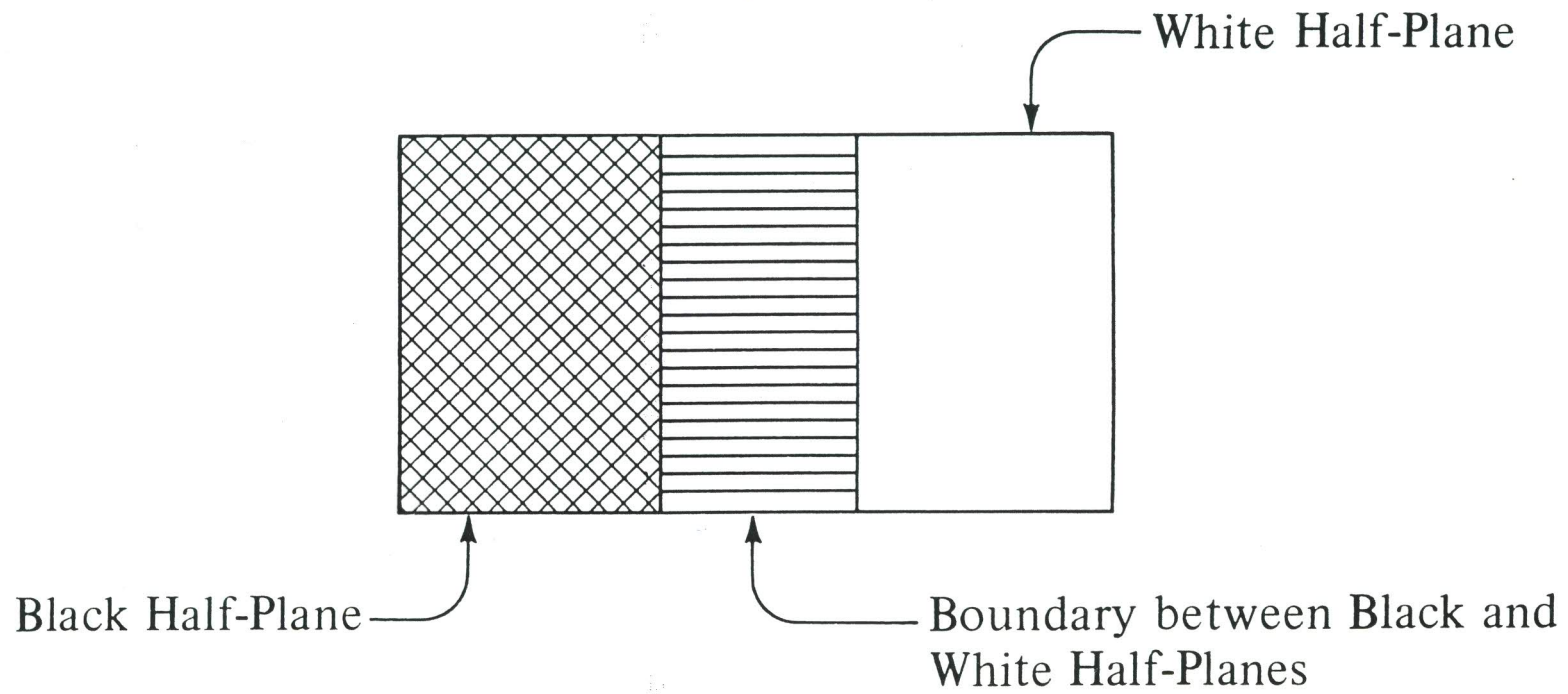
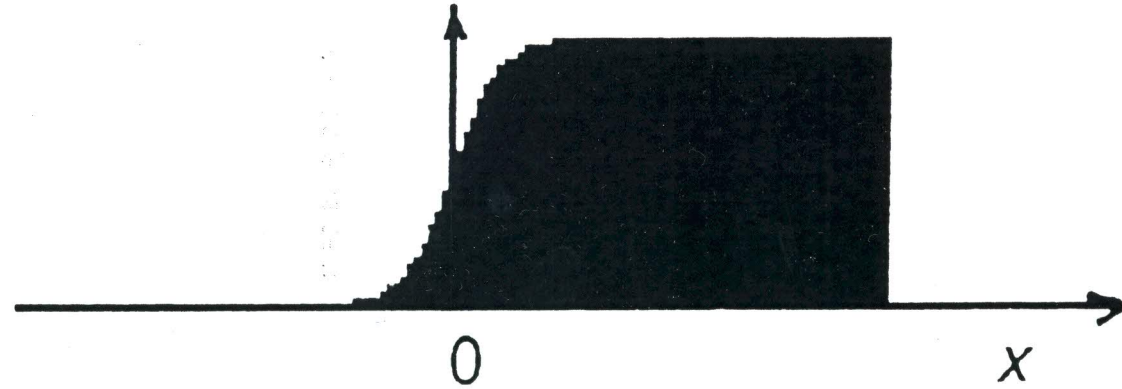


Figure Boundary between half-planes.

Illumination

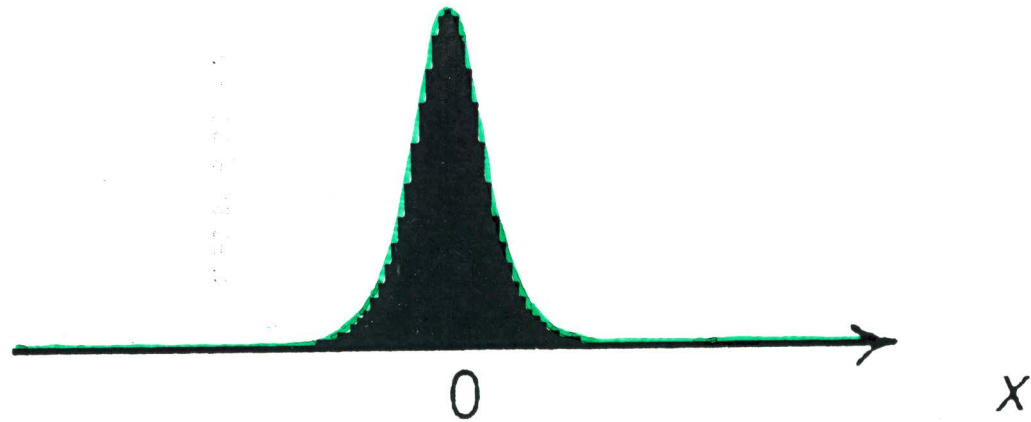


Black Halfplane

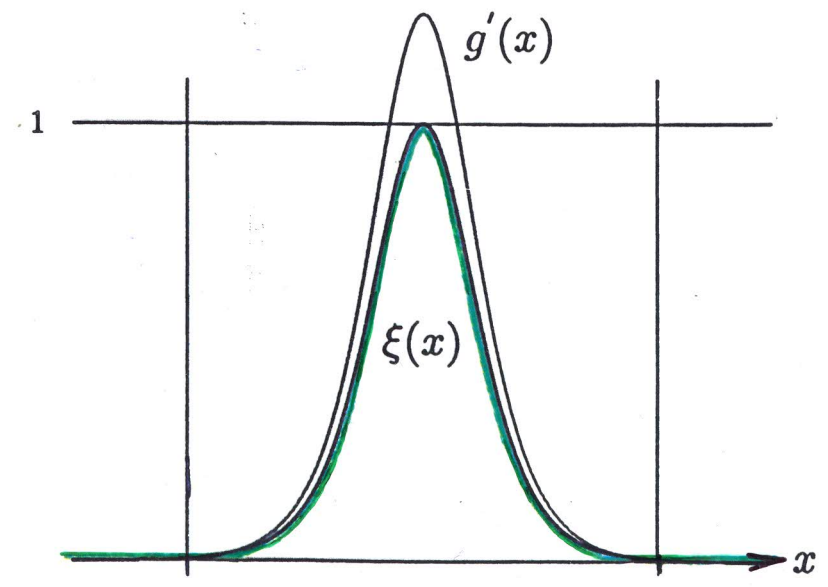
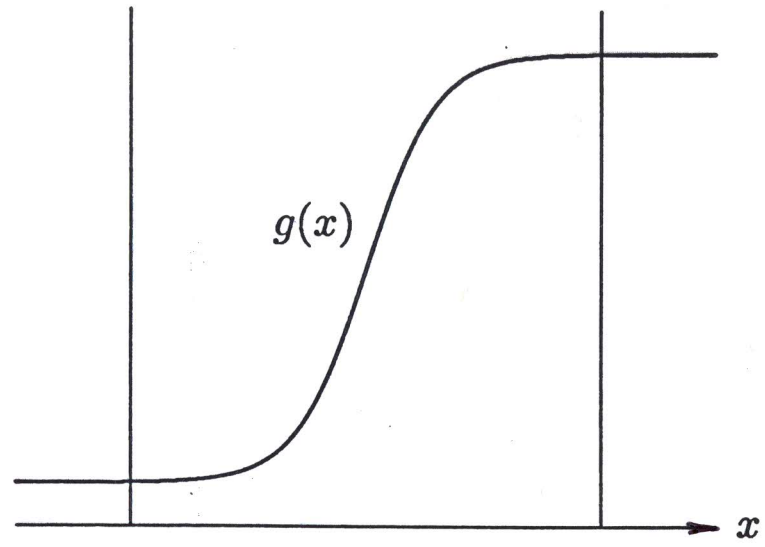
White Halfplane

Realistic illumination on a horizontal line

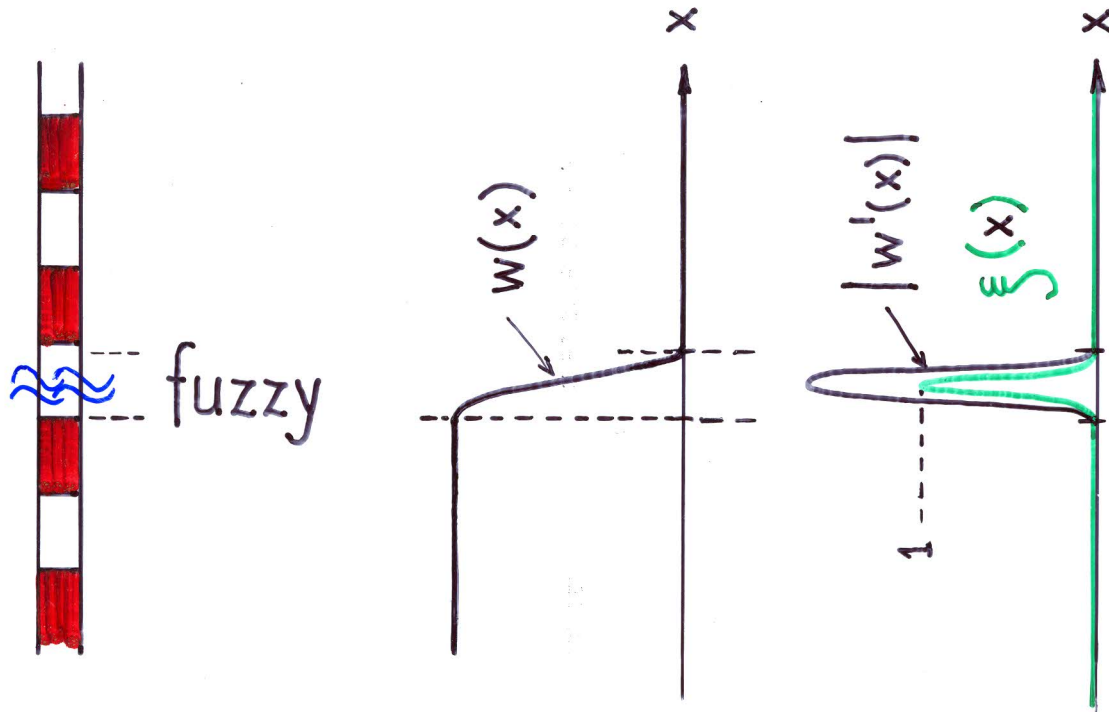
Scaled Rate of Change
of Illumination



Derivative of illumination function displayed

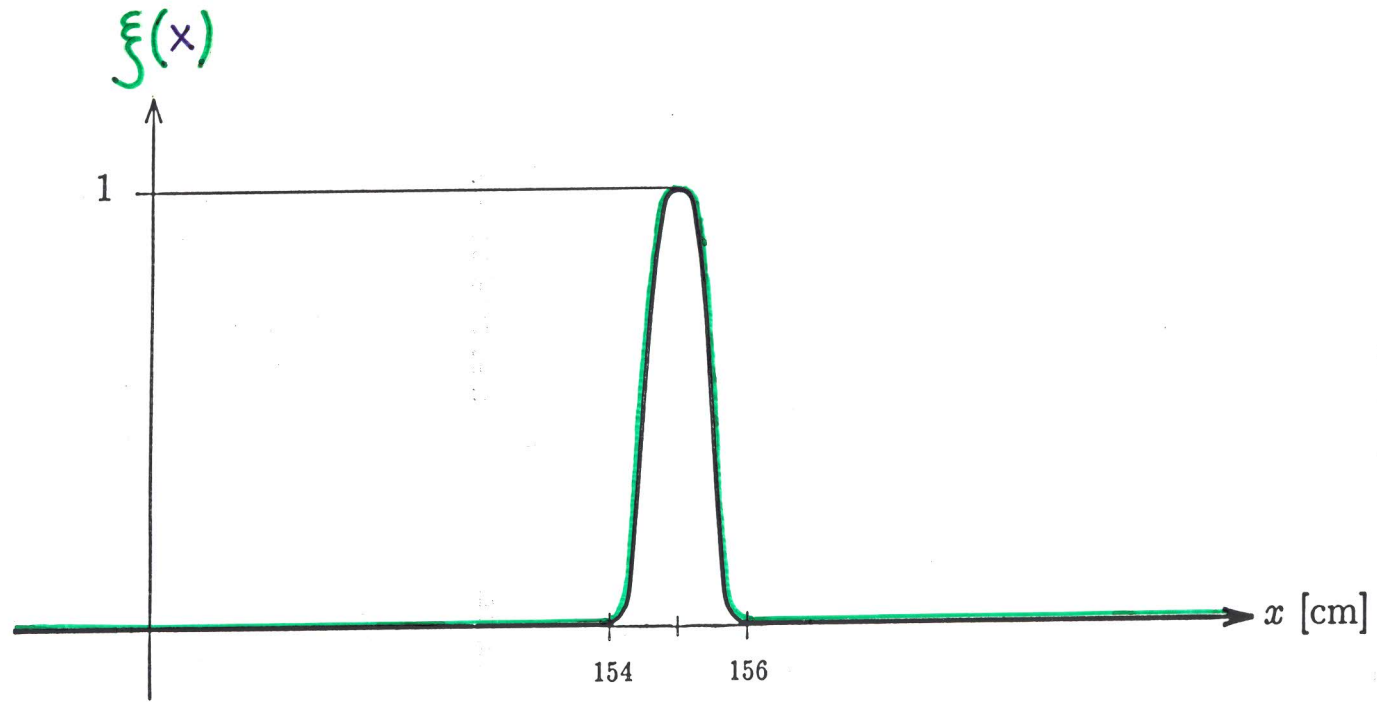


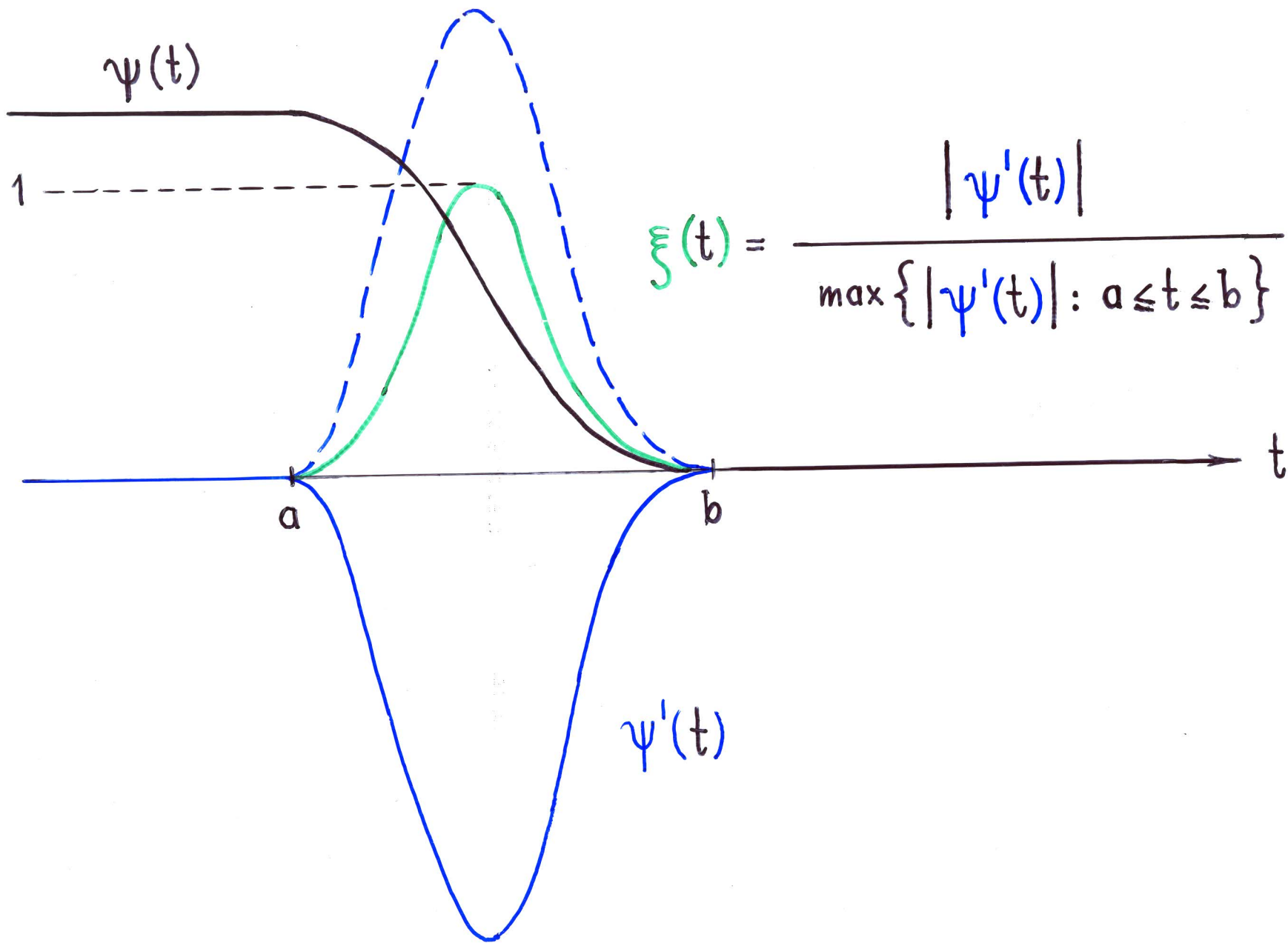
WATER LEVEL

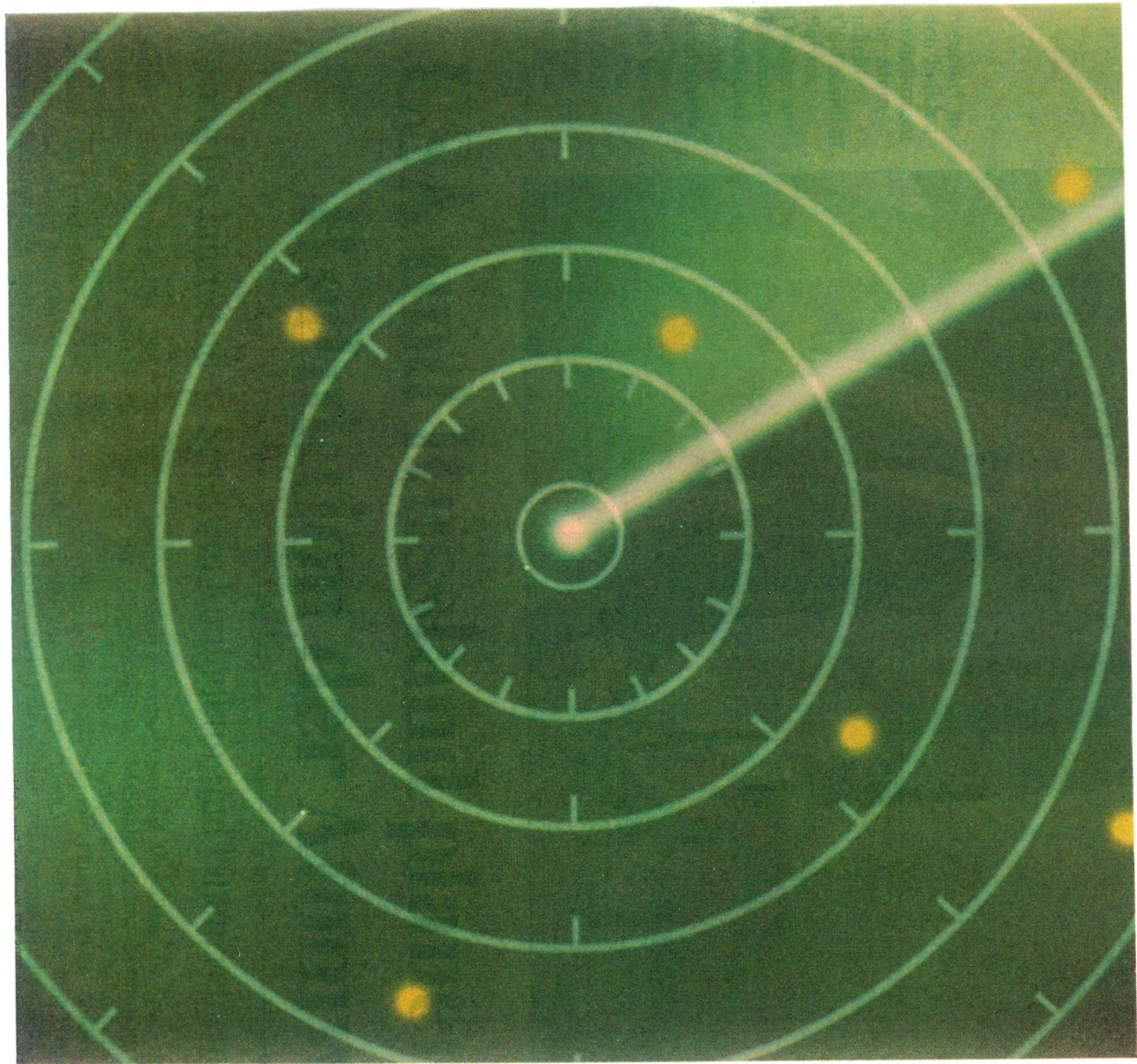


$$\xi(x) = \frac{|w'(x)|}{\max\{|w'(x)| : x \in \mathbb{R}\}} \quad \forall x \in \mathbb{R}$$

Figure : Water level







FUZZY VECTOR

vector-characterizing function $f(\cdot)$

$$f: \mathbb{R}^k \rightarrow [0, 1]$$

obeying

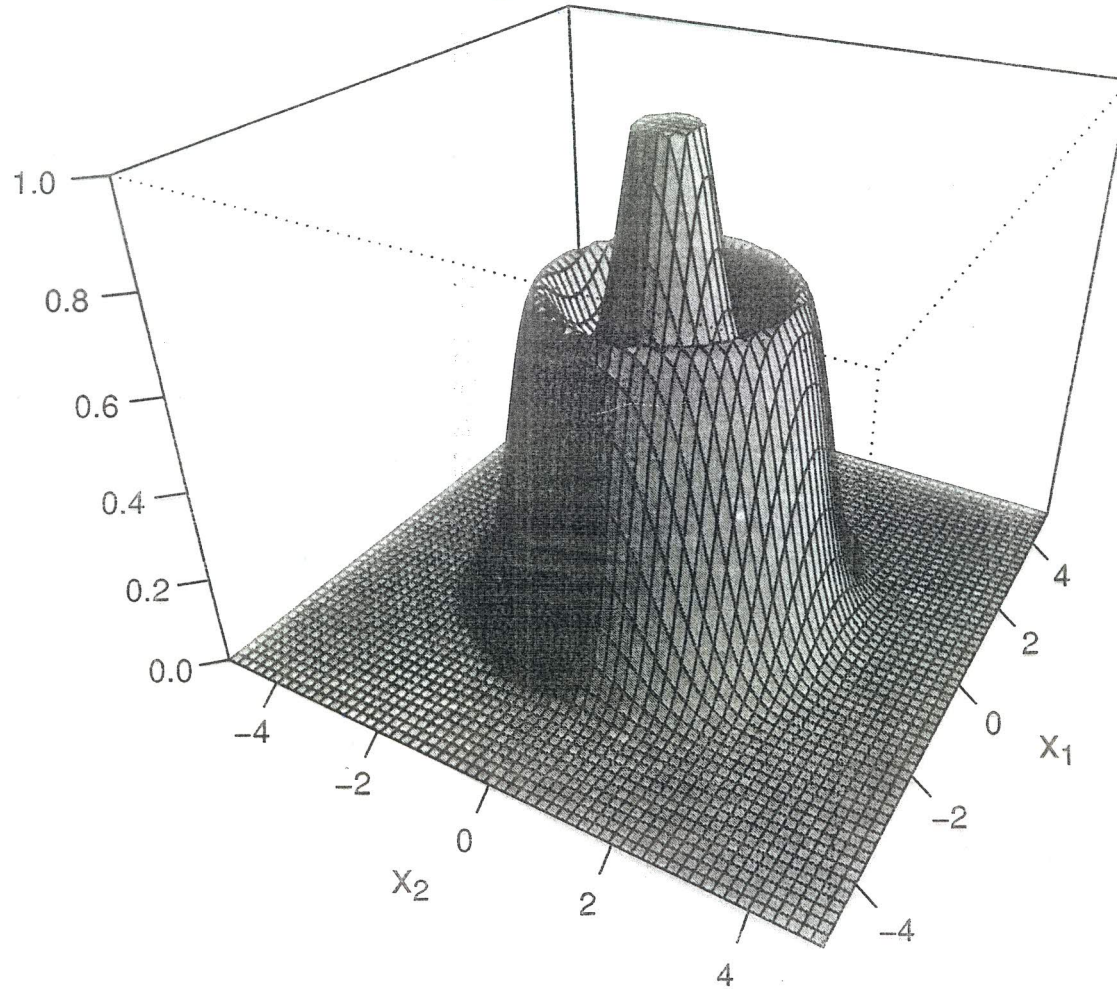
- $\forall \delta \in (0, 1]$ the so-called δ -cut

$$C_\delta [f(\cdot)] := \{ \underline{x} \in \mathbb{R}^k : f(\underline{x}) \geq \delta \} \neq \emptyset$$

is a finite union of simply connected closed sets

- $\text{Supp}[f(\cdot)]$ is bounded

$$\xi(x_1, x_2)$$



Measured Quantity

Variation

&

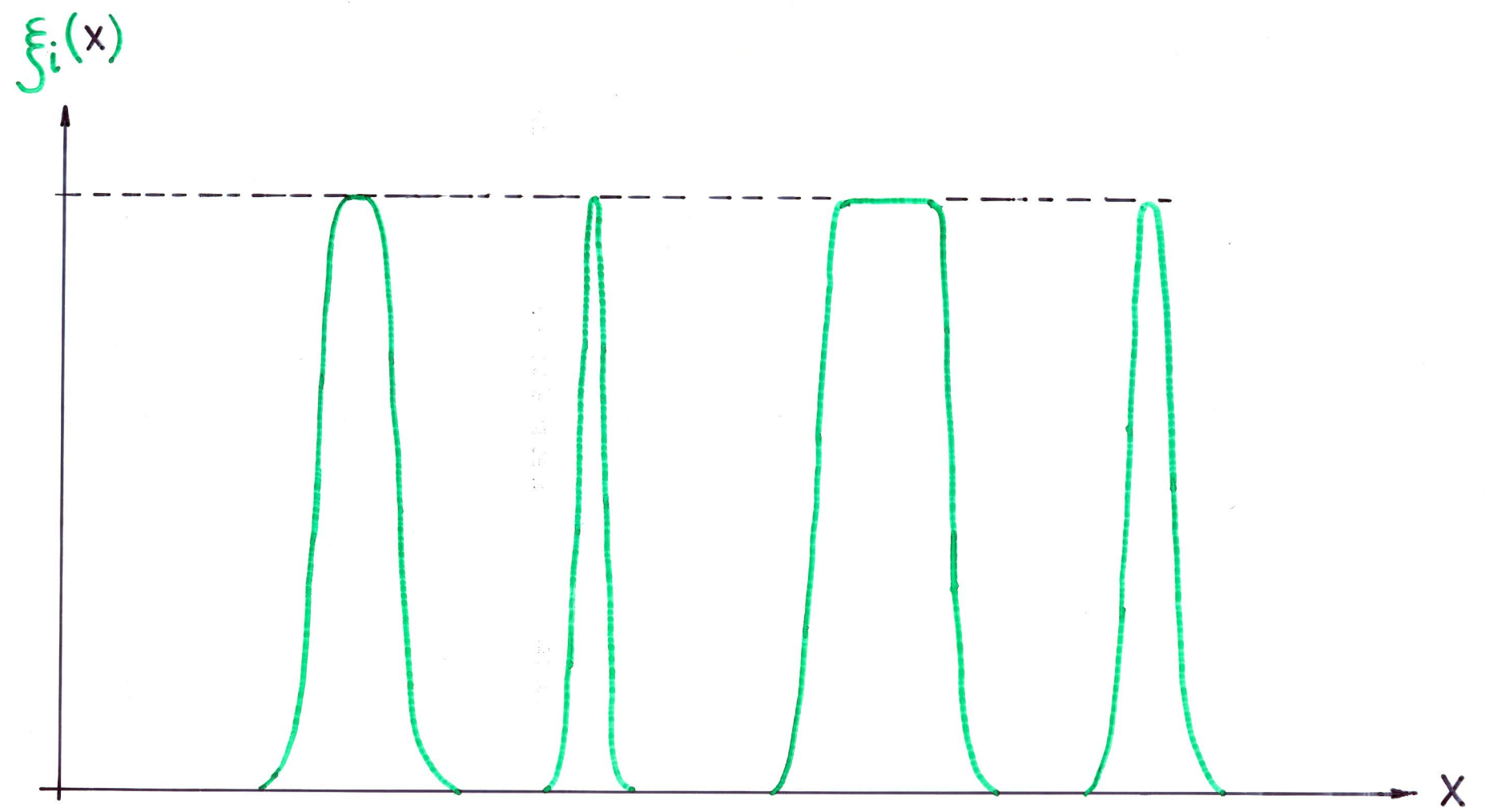
Imprecision

Stochastic Models

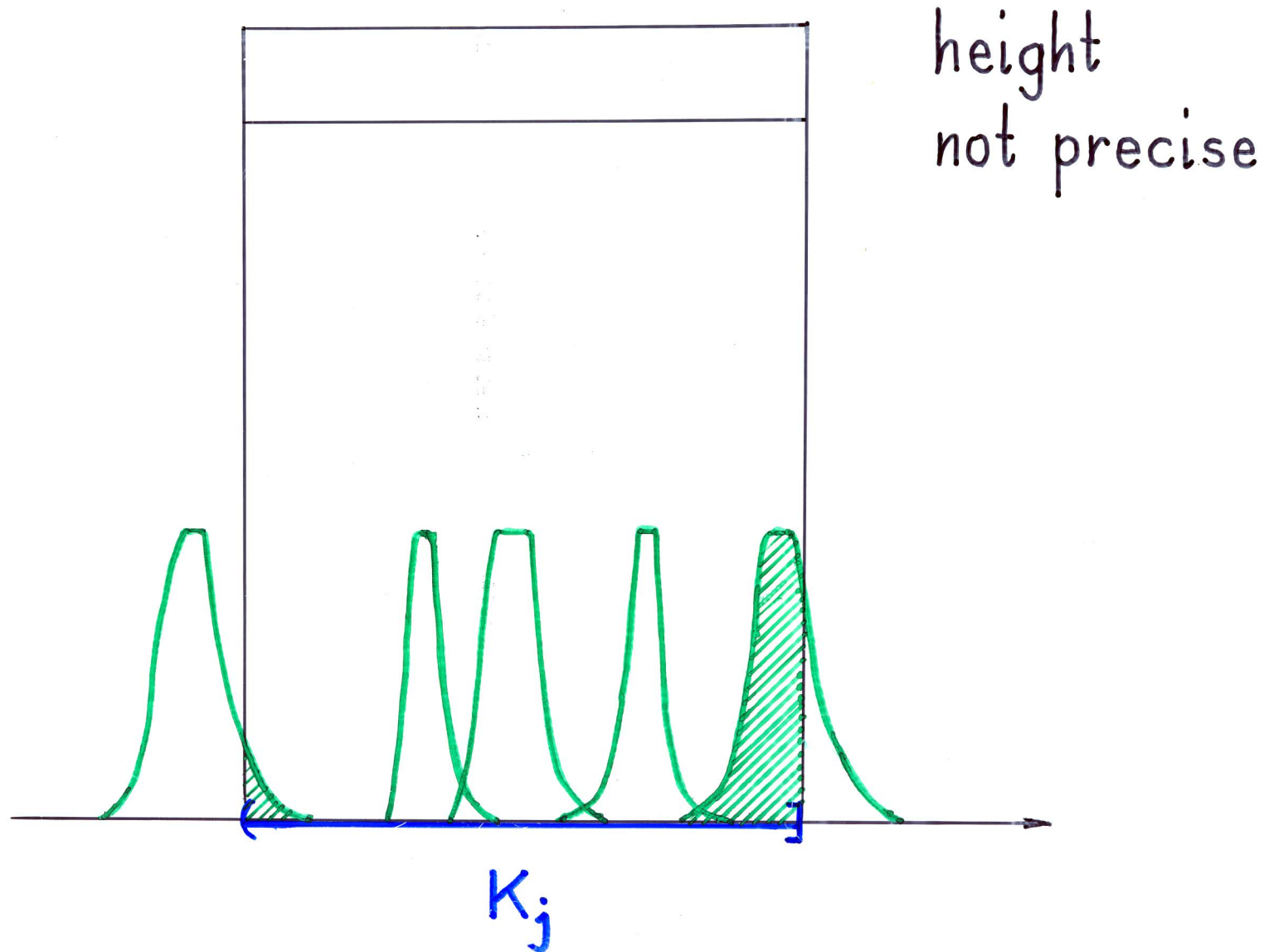
Fuzzy Models

Statistical Analysis of Fuzzy Data

FUZZY SAMPLE



FUZZY HISTOGRAMS



CONSTRUCTION LEMMA

Let $(A_\delta; \delta \in (0, 1])$ be a nested family of subsets of a set M . Then the membership function of the corresponding fuzzy subset of M is given by

$$\xi(x) = \sup \{ \delta \cdot \mathbb{1}_{A_\delta}(x) : \delta \in [0, 1] \} \quad \forall x \in M$$

with $A_0 := M$

FUZZY FREQUENCY

n_j^* fuzzy absolute frequency of class K_j

$$\delta\text{-Cuts } C_\delta(n_j^*) = [\underline{n}_\delta(K_j), \bar{n}_\delta(K_j)] \quad \forall \delta \in (0, 1]$$

where

$$\bar{n}_\delta(K_j) = \# \text{ observ. with } C_\delta(\xi_i(\cdot)) \cap K_j \neq \emptyset$$

$$\underline{n}_\delta(K_j) = \# \text{ observ. with } C_\delta(\xi_i(\cdot)) \subseteq K_j$$

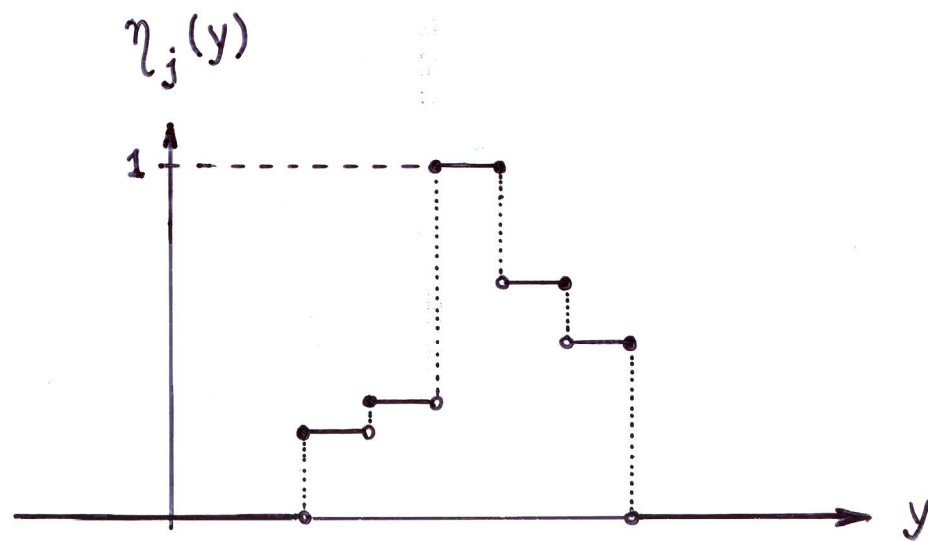
\Rightarrow char. f. $\psi_j(\cdot)$ of n_j^* given by its values

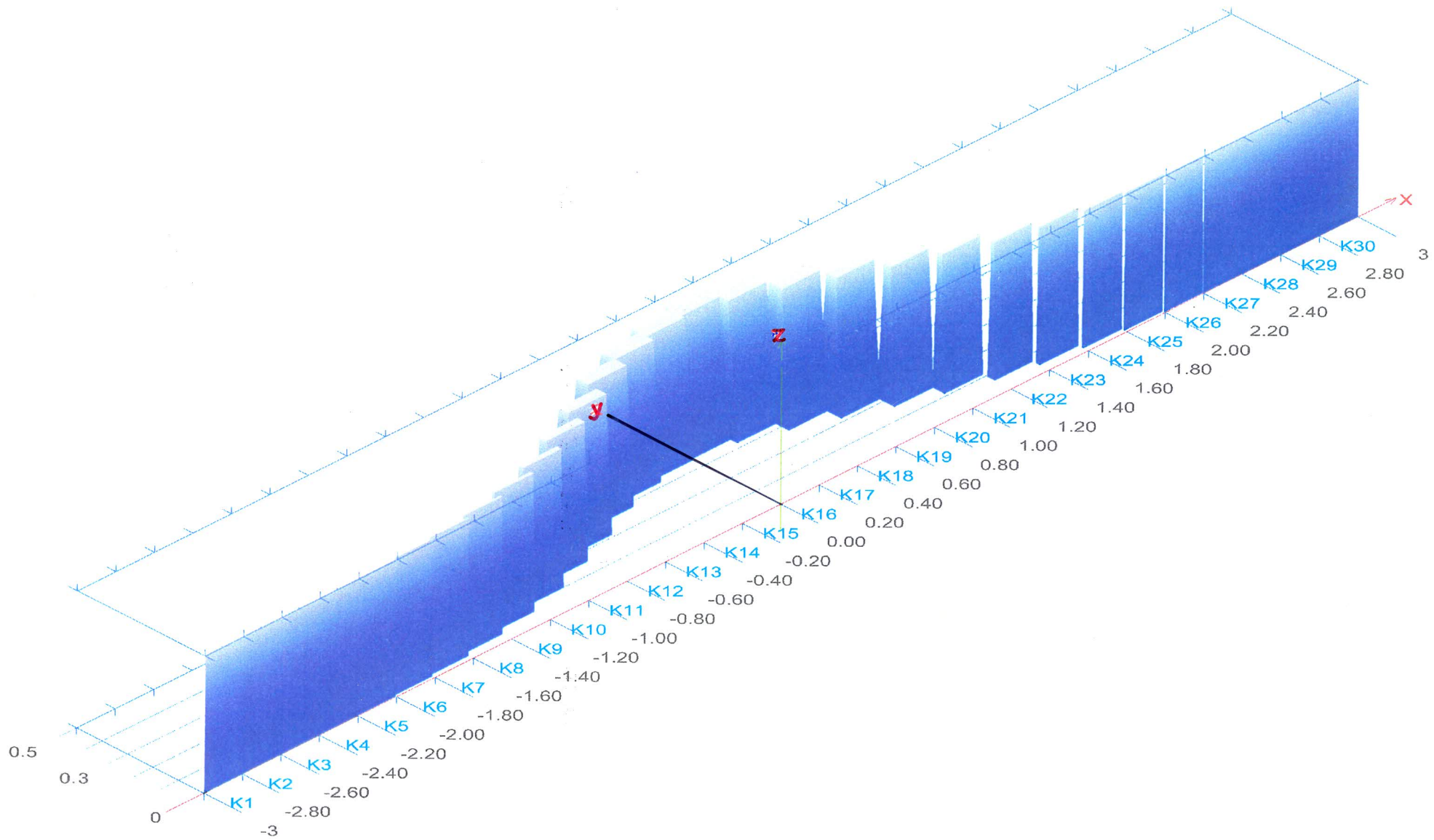
$$\psi_j(y) := \sup_{\delta \in [0, 1]} \delta \cdot I_{C_\delta(n_j^*)}(y) \quad \forall y \in \mathbb{R}$$

$h_j^* := \frac{n_j^*}{n}$ fuzzy relative frequency of class K_j

\Rightarrow char. f. $\eta_j(\cdot)$ of h_j^* is given by

$$\eta_j(y) = \psi_j(ny) \quad \forall y \in \mathbb{R}$$





CALCULATIONS

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

Averages \bar{x}_n^*

Indicators and Indexes I^*

$$I^* = f(x_1^*, \dots, x_n^*; w_1, \dots, w_n)$$

Functions of Fuzzy Variables

Extension Principle

STANDARD STATISTICAL INFERENCE

$X \sim P_\theta; \theta \in \Theta, M_X$ Observation Space

x_1, \dots, x_n Sample, $x_i \in M_X \Rightarrow (x_1, \dots, x_n) \in M_X^n$

M_X^n Sample Space

- Estimators $\mathcal{J}(x_1, \dots, x_n), \mathcal{J}: M_X^n \rightarrow \Theta$
- Confidence Regions $\kappa(x_1, \dots, x_n)$
- Test Statistics $t(x_1, \dots, x_n)$

Generalization for Fuzzy Data ?

COMBINED FUZZY SAMPLE

Sample x_1^*, \dots, x_n^*

$\xi_1(\cdot), \dots, \xi_n(\cdot)$

x_i^* Fuzzy Element of Observation Space M

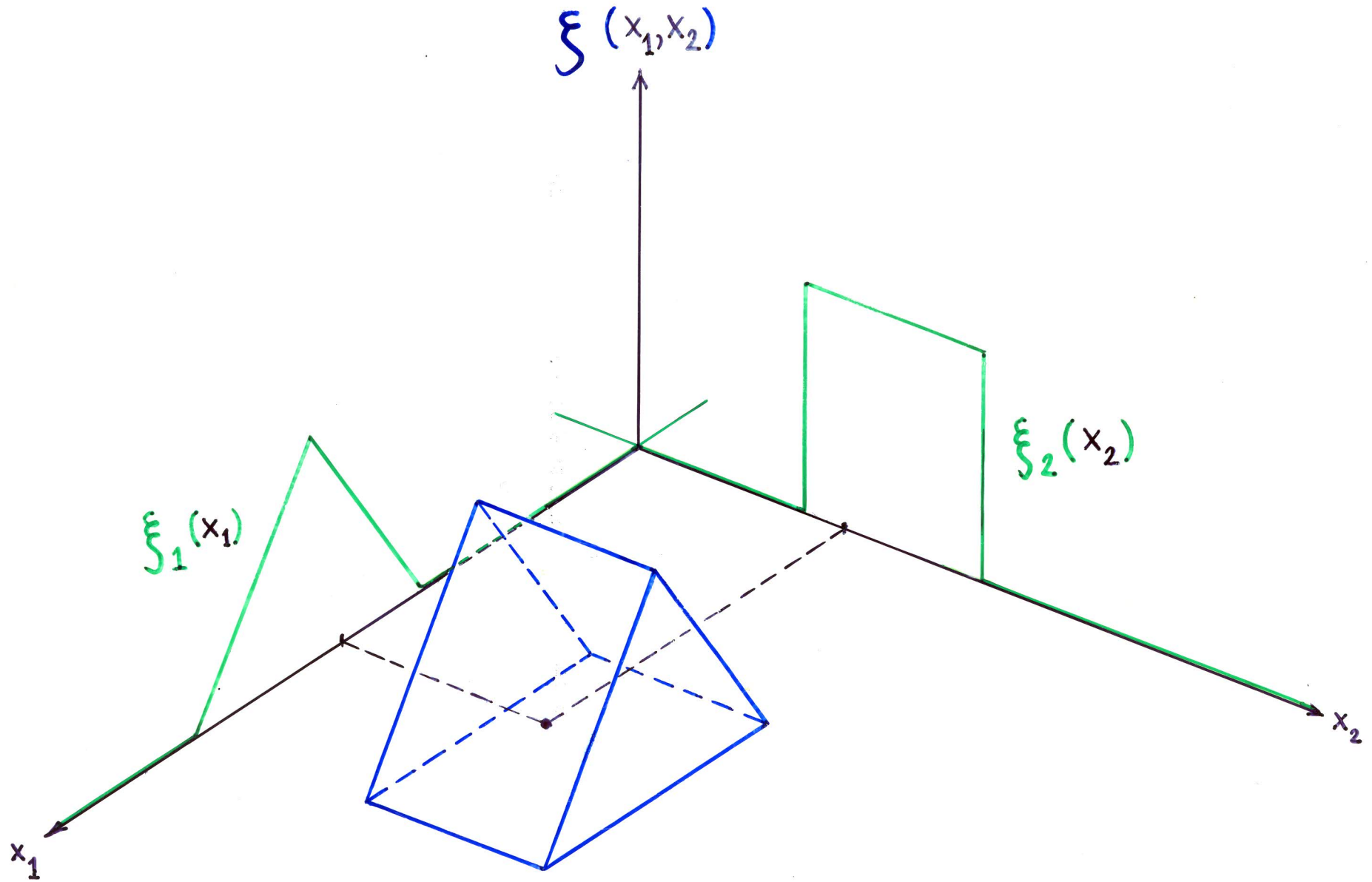
$M^n = \{ \underline{x} = (x_1, \dots, x_n) : x_i \in M \}$ Sample Space

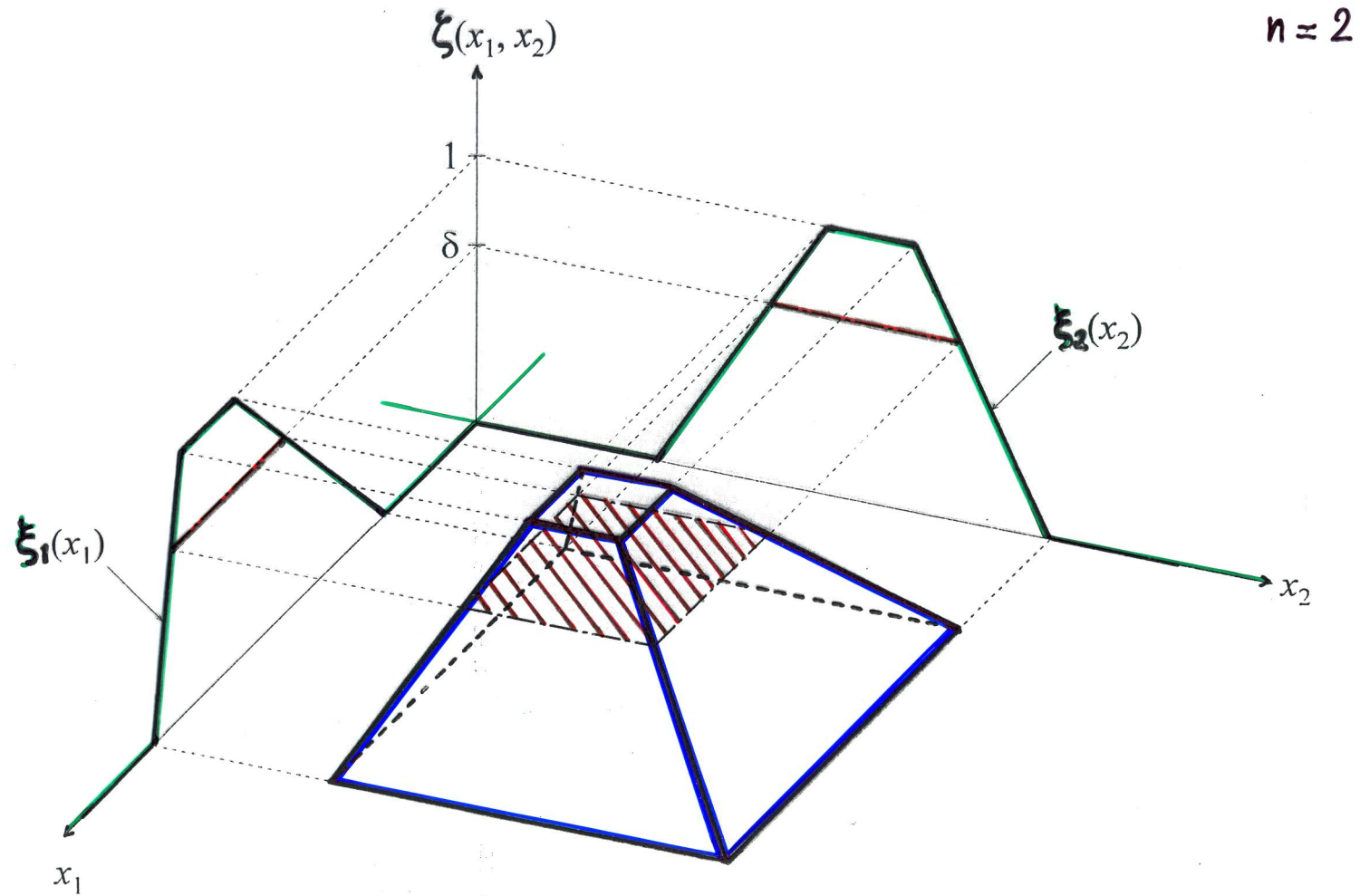
\underline{x}^* Fuzzy Element of M^n with VCF $\xi(\cdot)$

$$\xi(x_1, \dots, x_n) = T_n(\xi_1(x_1), \dots, \xi_n(x_n)) \quad \forall (x_1, \dots, x_n)$$

\underline{x}^* Combined Fuzzy Sample

$$n = 2$$





$$\xi(\underline{x}) := \min_{i=1(1)n} \xi_i(x_i)$$

GENERALISED ESTIMATORS

$\hat{\theta} = \mathcal{J}(x_1, \dots, x_n)$ Classical Estimator

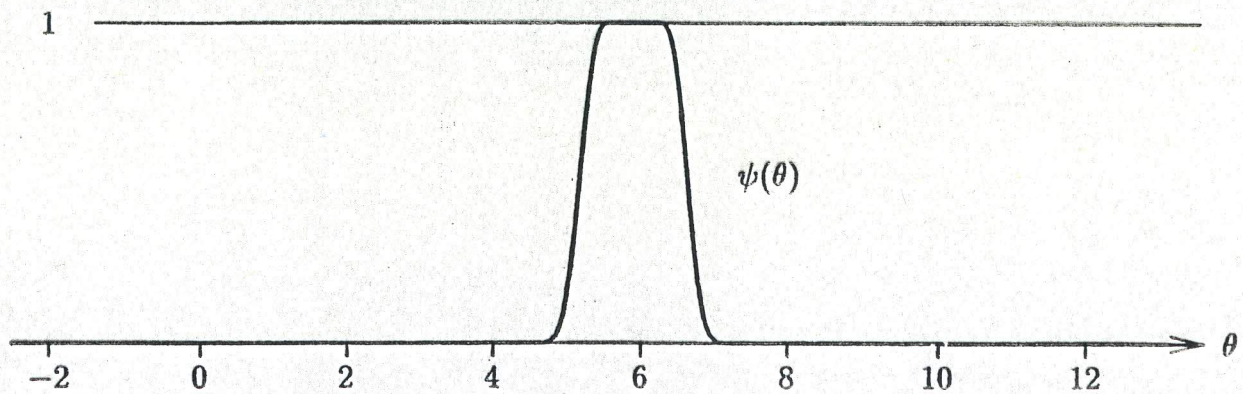
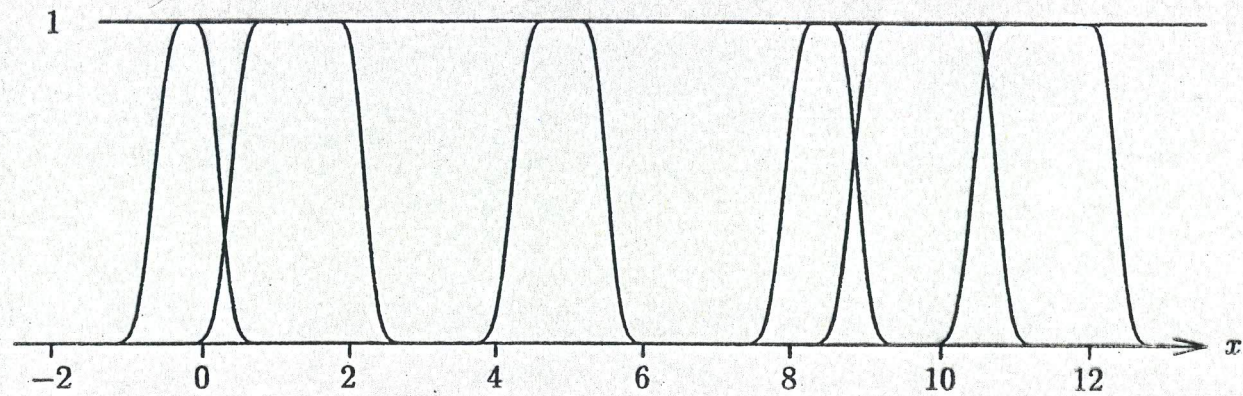
$\underline{x} = (x_1, \dots, x_n) \in M_x^n$, $\mathcal{J}: M_x^n \rightarrow \Theta$

\underline{x}^* Combined Fuzzy Sample $\hat{=} \mathcal{J}(\cdot, \dots, \cdot)$

$\hat{\theta}^*$ Fuzzy Estimator $\hat{=} \psi(\cdot)$

$\psi(\theta) := \sup \{ \mathcal{J}(\underline{x}) : \underline{x} \in \mathcal{J}^{-1}(\{\theta\}) \} \quad \forall \theta \in \Theta$

with $\mathcal{J}^{-1}(\{\theta\}) = \{ \underline{x} \in M_x^n : \mathcal{J}(\underline{x}) = \theta \}$



FUZZY CORRELATION

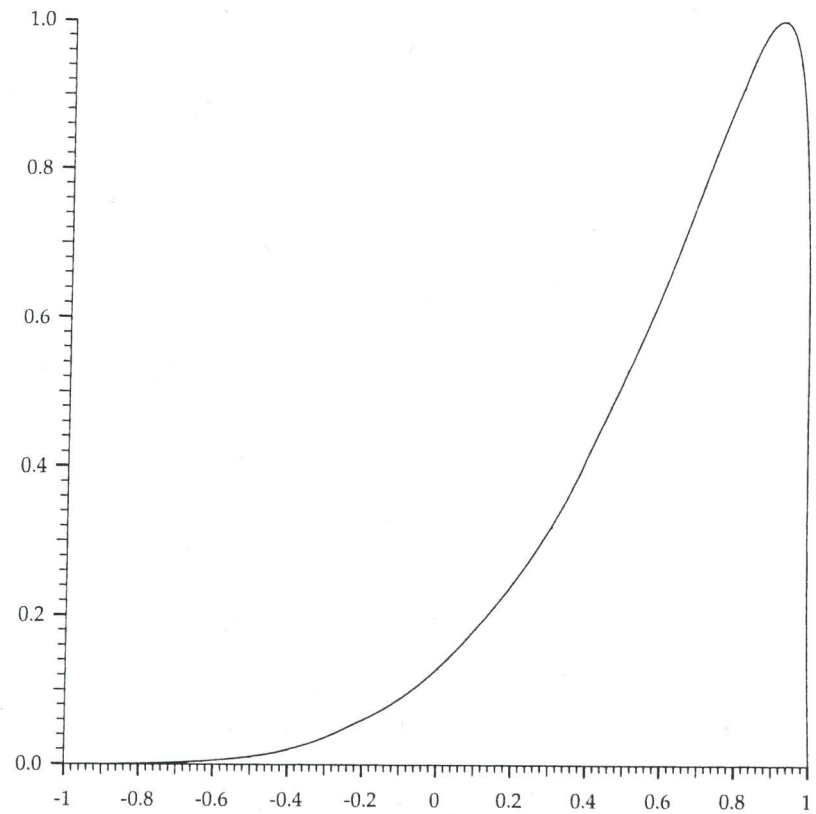
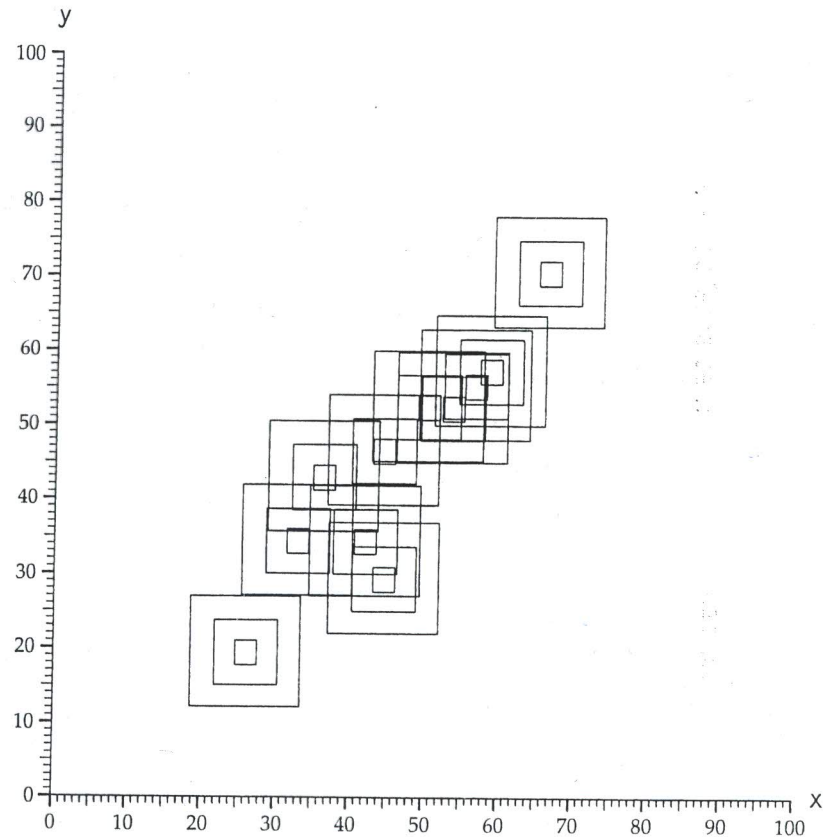


Figure Fuzzy correlation coefficient.

FUZZY CONFIDENCE REGIONS

$$X \sim P_{\theta}, \theta \in \Theta$$

$\kappa(\cdot)$ Confidence function

\underline{x}^* Combined fuzzy sample $\xi(\cdot)$ VCF

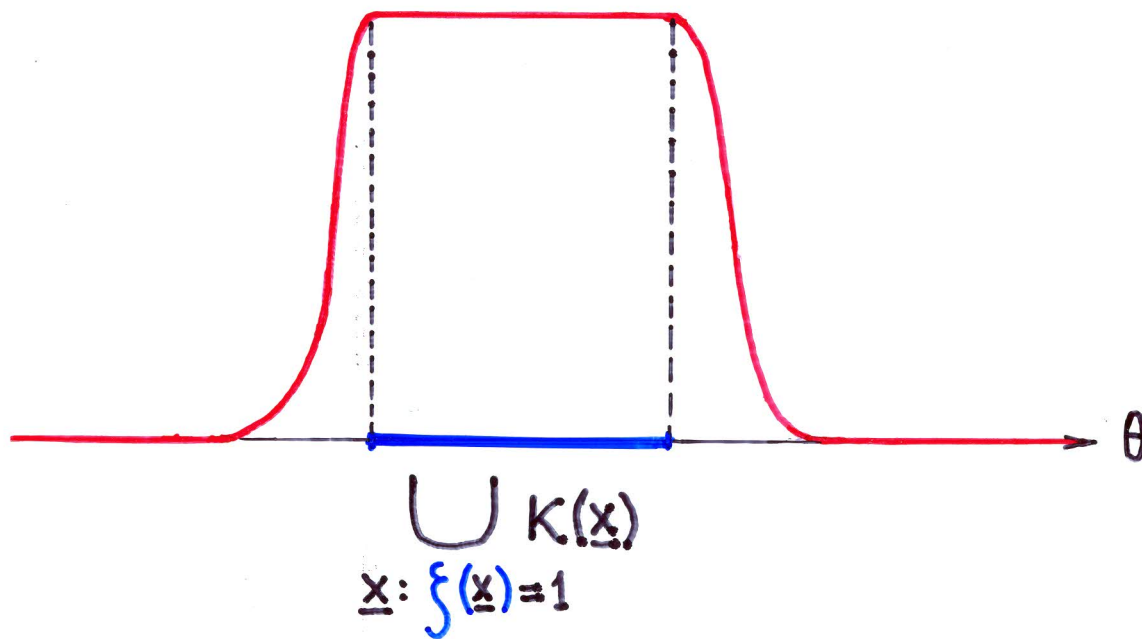
Generalized confidence set $\kappa(\underline{x}^*)$ is a fuzzy subset of Θ whose membership function $\varphi(\cdot)$ is defined by

$$\varphi(\theta) := \sup \{ \xi(\underline{x}) : \theta \in \kappa(\underline{x}) \} \quad \forall \theta \in \Theta$$

$$\underline{x} \in \text{support} [\xi(\cdot)]$$

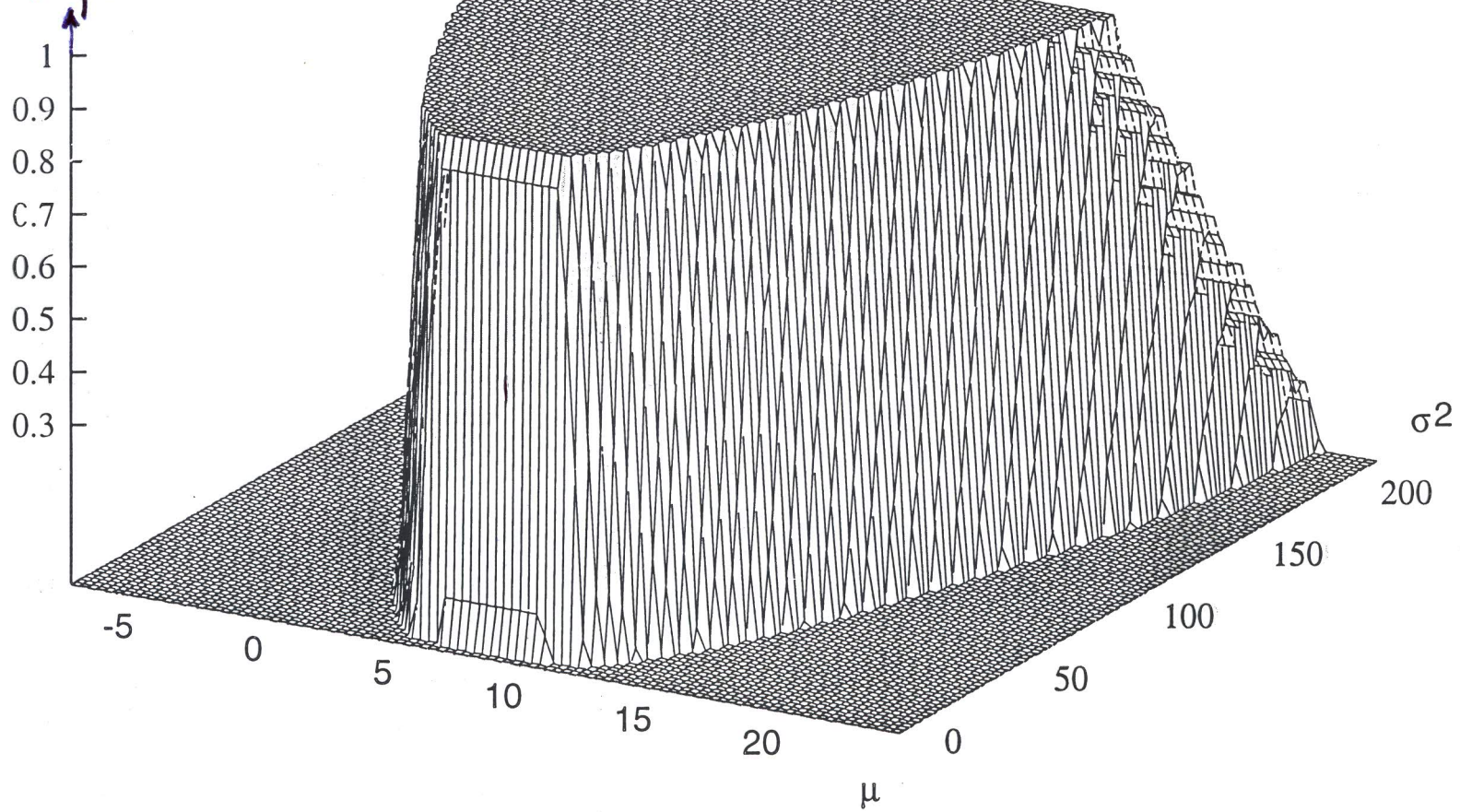
The following holds:

$$(1) \quad \varphi(\theta) = 1 \quad \forall \theta \in \bigcup_{\underline{x}: \xi(\underline{x})=1} K(\underline{x})$$

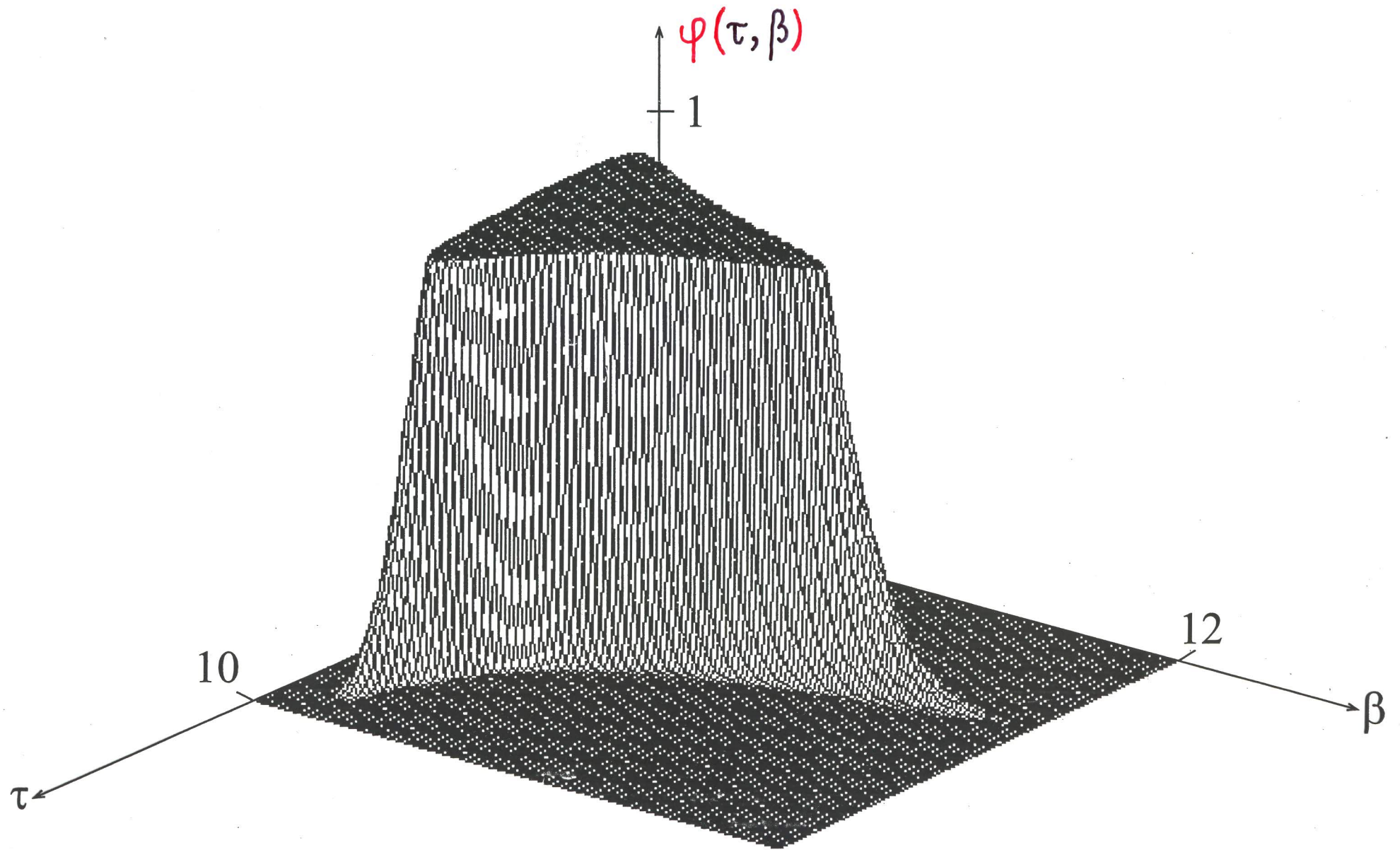


$$(2) \quad \text{For classical samples } \underline{x} = (x_1, \dots, x_n):$$
$$\varphi(\cdot) = \mathbb{1}_{K(\underline{x})}(\cdot)$$

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



Weibull (τ, β) , $F(t) = 1 - e^{-\left(\frac{t}{\tau}\right)^\beta}$



STATISTICAL TESTS

$T = t(X_1, \dots, X_n)$ Test Statistic

x_1^*, \dots, x_n^* Fuzzy Sample

$$t^* = t(x_1^*, \dots, x_n^*) \hat{=} \eta(\cdot)$$

Calculation of $\eta(\cdot)$ based on the combined fuzzy sample $\underline{x}^* \hat{=} \mathcal{F}(x_1, \dots, x_n)$ and the extension principle:

$$t^* = t(\underline{x}^*)$$

FIRST SOLUTION

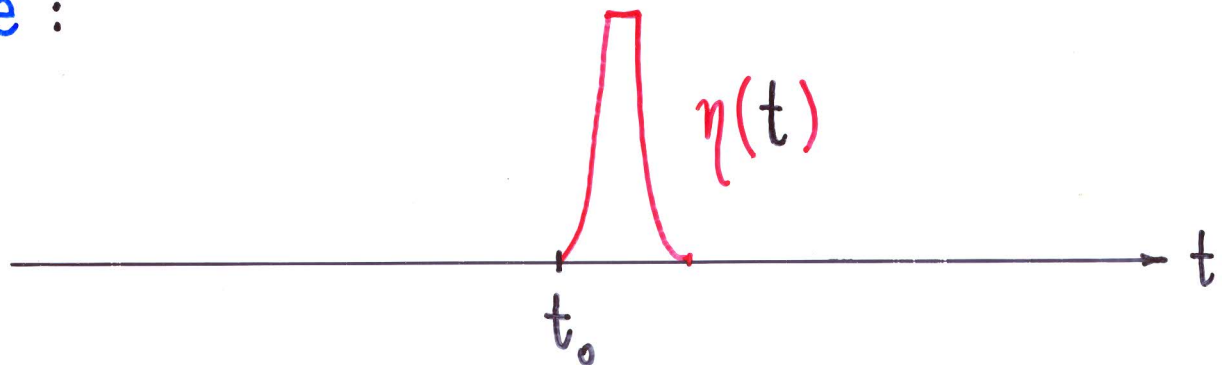
p-value approach

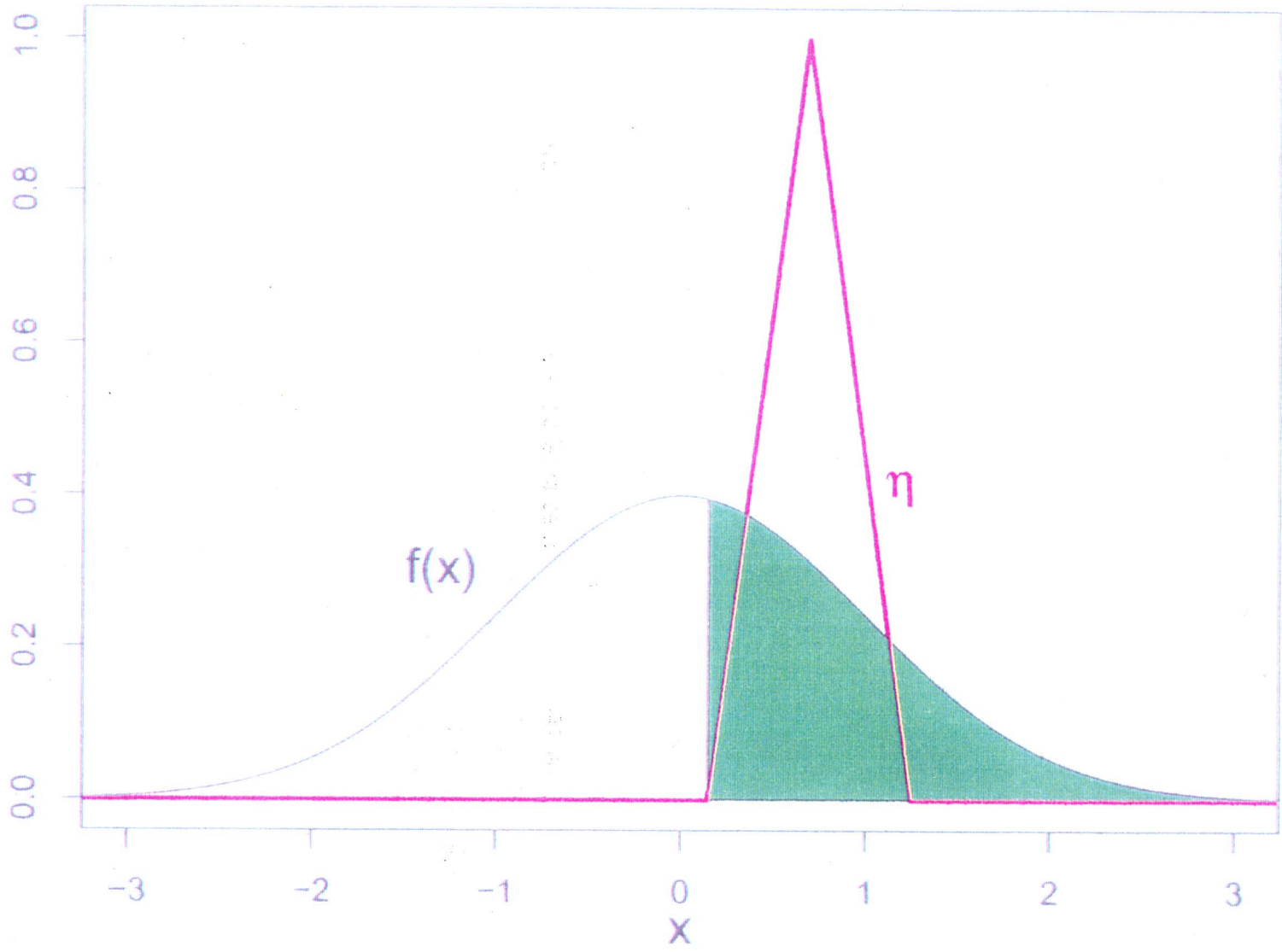
- For precise data $x_1, \dots, x_n \Rightarrow t = t(x_1, \dots, x_n)$

p-value : largest error probability for which a hypothesis is rejected for t

- For fuzzy data $x_1^*, \dots, x_n^* \Rightarrow t^* \triangleq \eta(\cdot)$

p-value :





FUZZY P-VALUES

t^* fuzzy value of the test statistic T

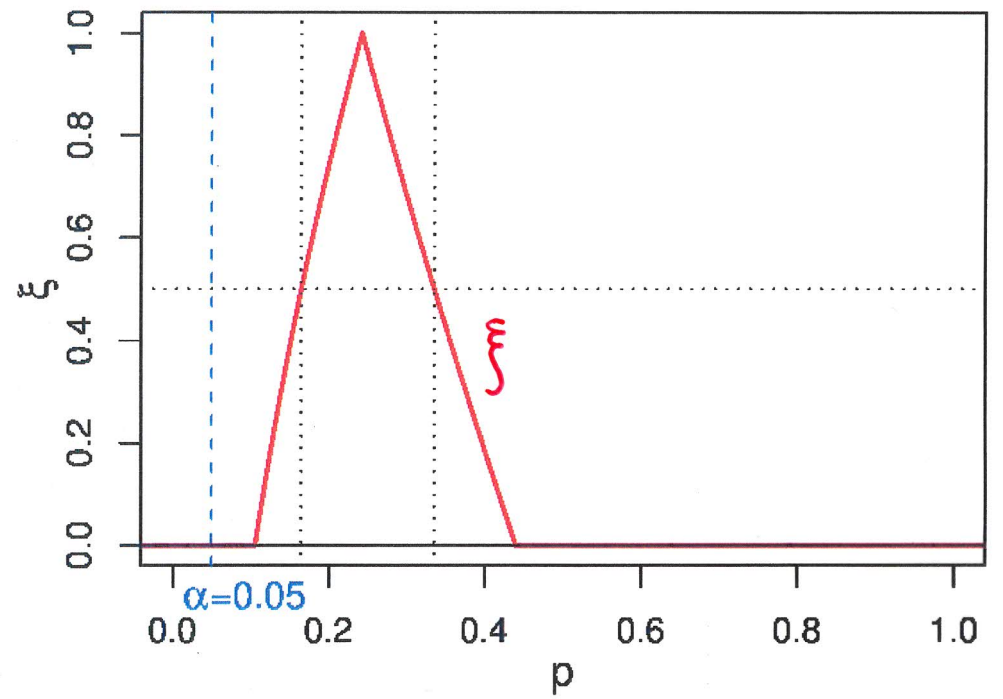
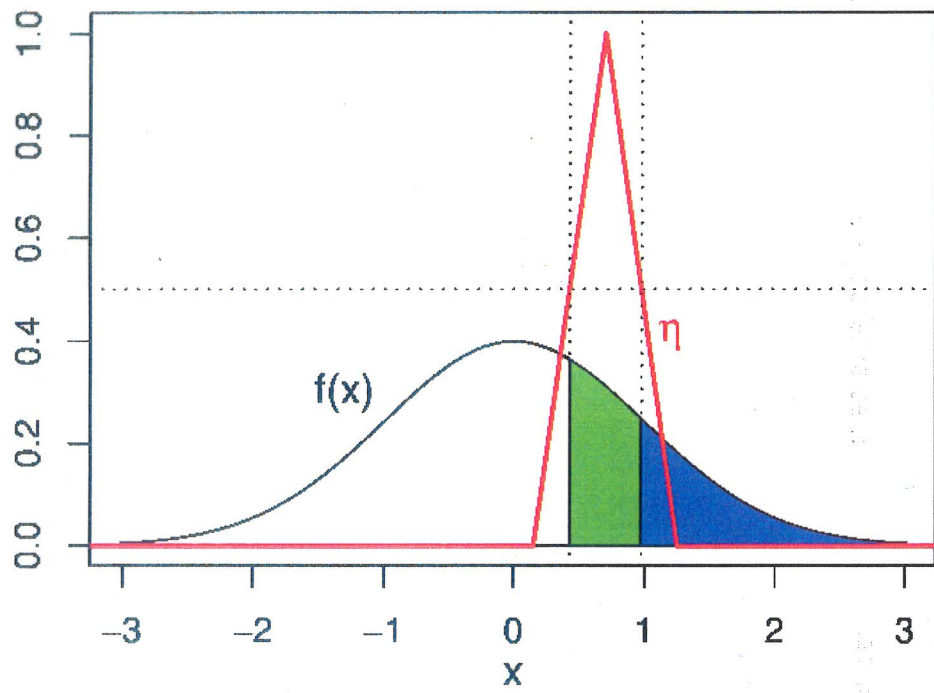
$$C_\delta(t^*) = [t_1(\delta), t_2(\delta)] \quad \forall \delta \in (0, 1]$$

p^* fuzzy p-value

For one-sided tests

$$A_\delta(p^*) = [\Pr(T \leq t_1(\delta)), \Pr(T \leq t_2(\delta))]$$

$$\xi(p) = \sup \{ \delta \cdot \mathbf{1}_{A_\delta}(p) : \delta \in [0, 1] \} \quad \forall p \in \mathbb{R}$$



BAYESIAN INFERENCE

$X \sim f(\cdot | \theta)$, $\theta \in \Theta$, $\tilde{\theta}$ Stochastic Qu.

$\pi(\cdot)$ a-priori distribution on Θ

x_1, \dots, x_n Sample information

Updating of the a-priori distribution

$$\pi(\theta | x_1, \dots, x_n) = \frac{\pi(\theta) \cdot l(\theta; x_1, \dots, x_n)}{\int_{\Theta} \pi(\theta) \cdot l(\theta; x_1, \dots, x_n) d\theta} \quad \forall \theta \in \Theta$$

a-posteriori distribution

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

FOR FUZZY DATA ?

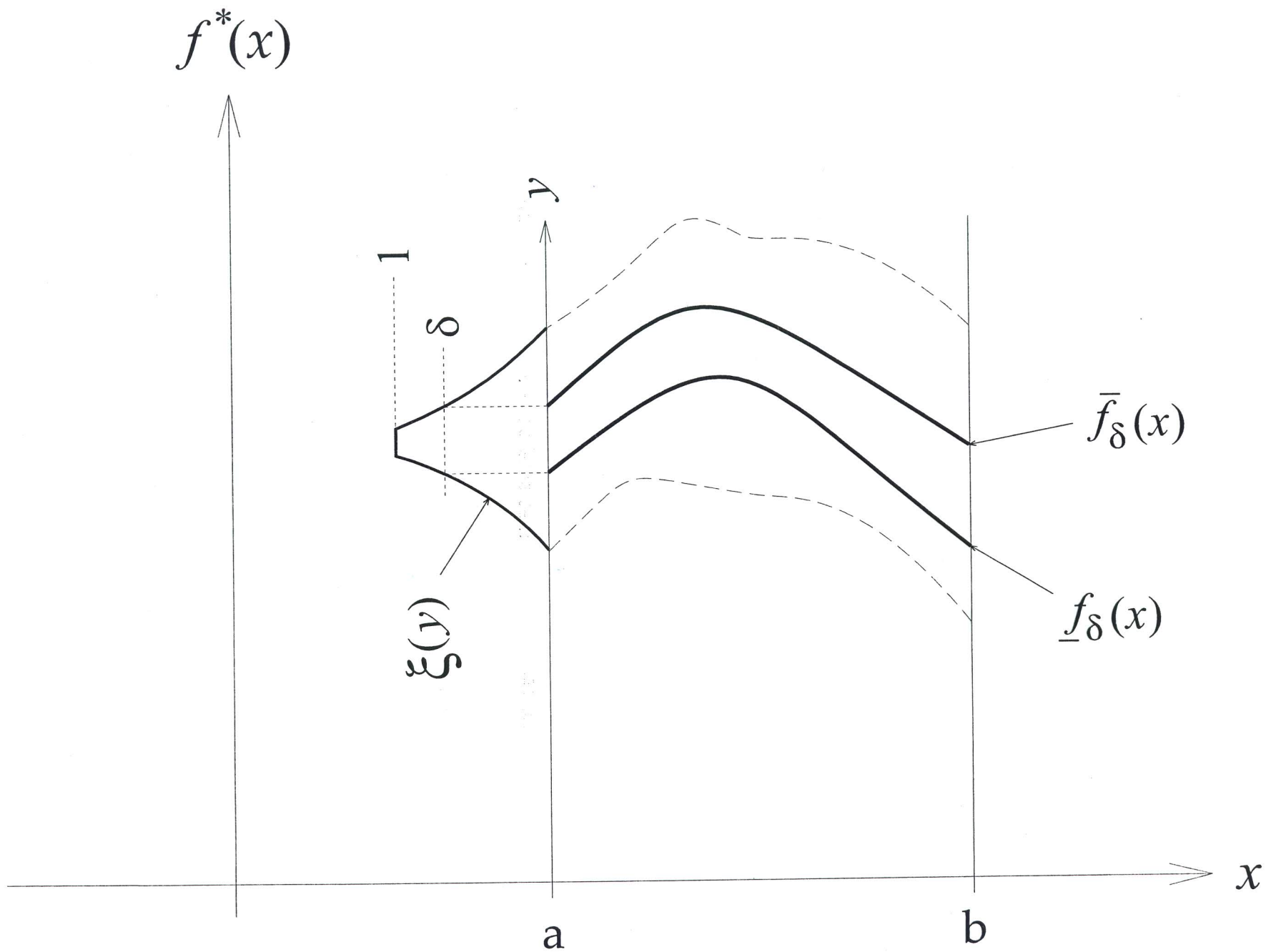
Fuzzy valued functions $f^*: M \rightarrow \mathcal{F}_I(\mathbb{R})$

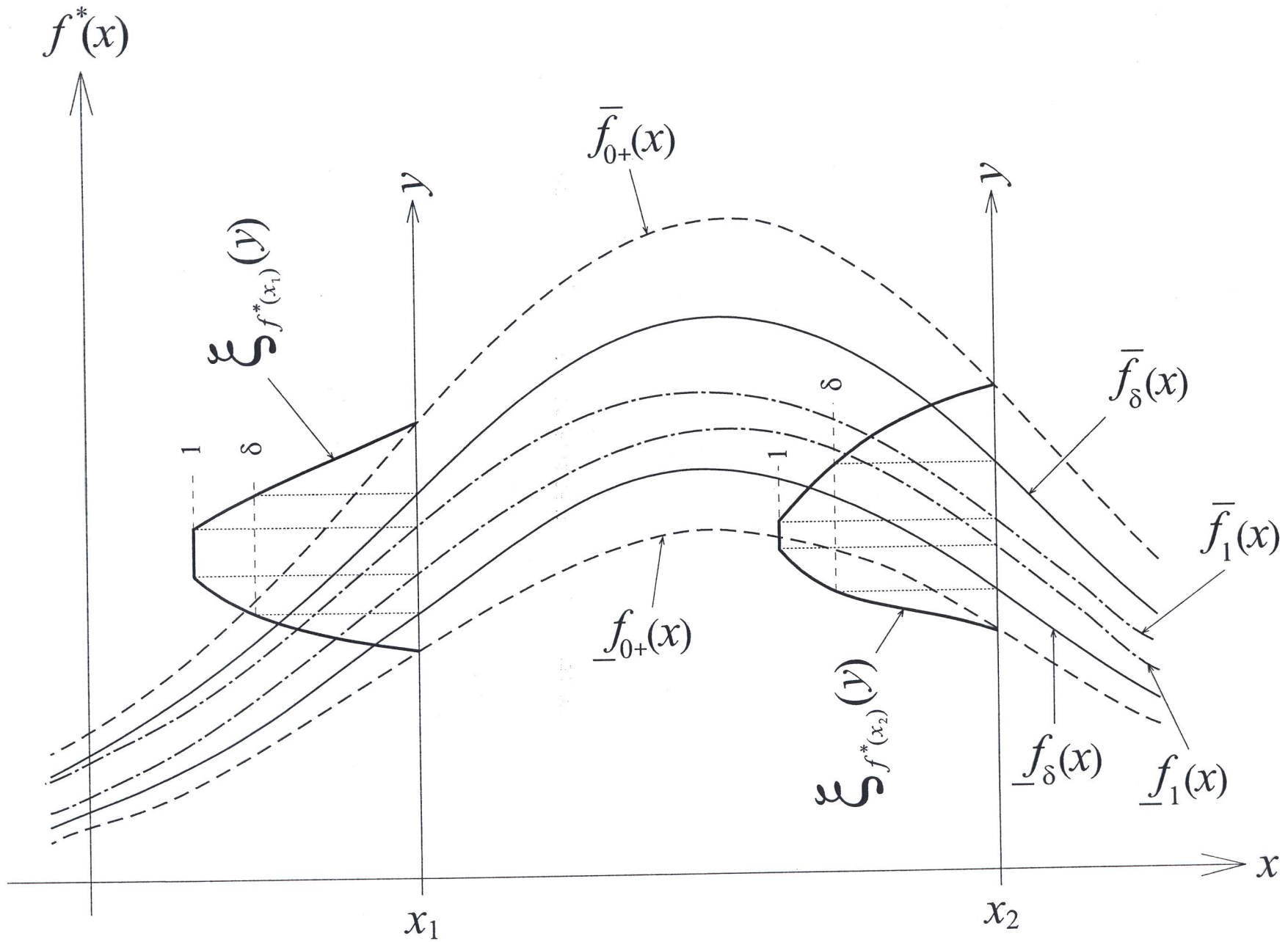
$$f^*(x) = y^* \hat{=} \xi_x(\cdot) \quad \forall x \in M$$

δ -level functions $\underline{f}_\delta(\cdot)$ and $\bar{f}_\delta(\cdot)$

$$\text{defined by } C_\delta[f^*(x)] = [\underline{f}_\delta(x), \bar{f}_\delta(x)] \quad \forall x \in M$$
$$\forall \delta \in (0, 1]$$

For $M = \mathbb{R}$ δ -level curves (real functions)





FIRST IDEA

Extension Principle $\pi^*(\cdot | \underline{x}^*)$

$$\pi^*(\theta | \underline{x}^*) \hat{=} \psi_{\theta}(\cdot)$$

$\pi^*(\cdot | \underline{x}^*)$ fuzzy density on Θ

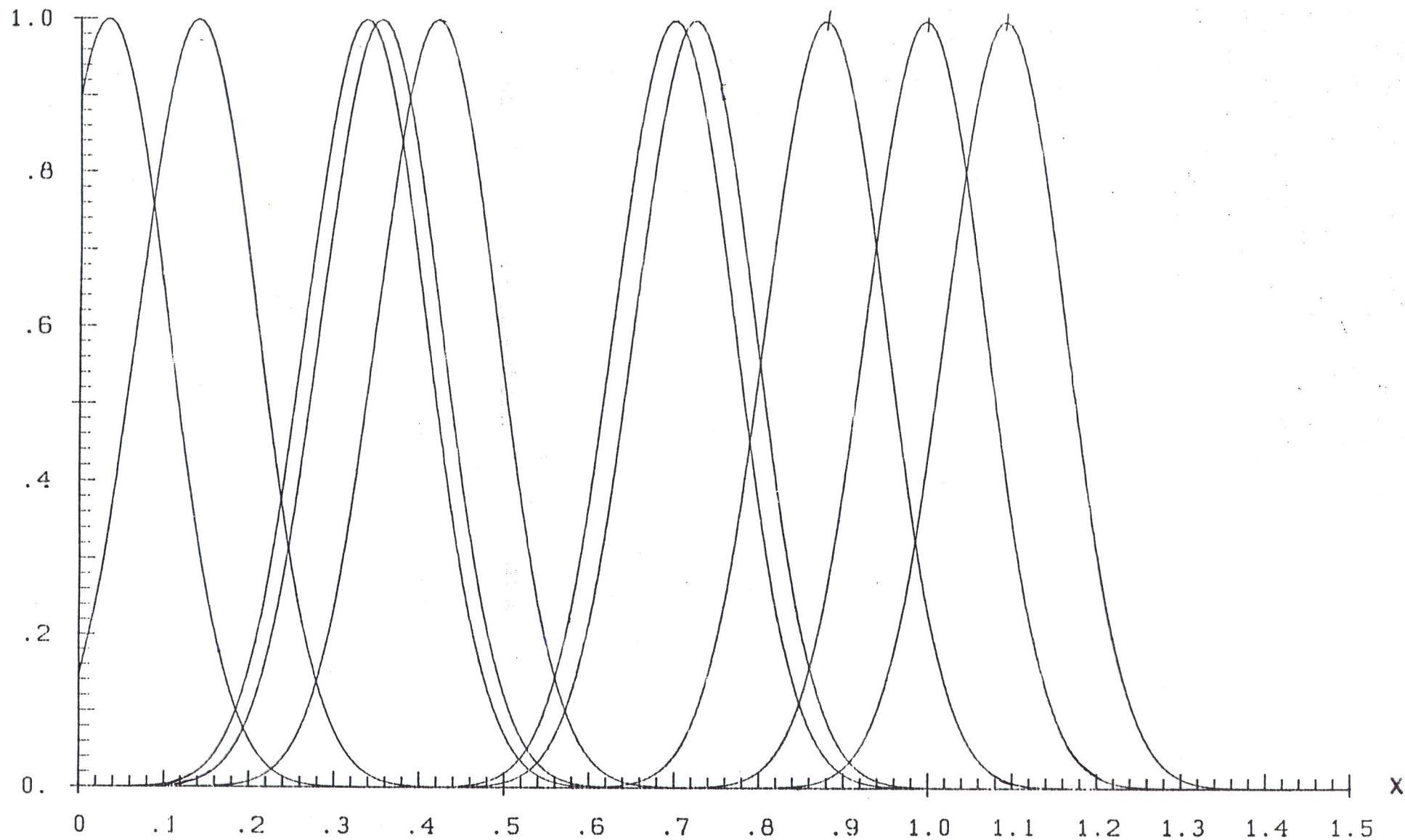
Example: $X \sim \text{Ex}_{\theta}$, $\Theta = (0, \infty)$

a-priori - distr. $\text{Gam}(1, 1)$

fuzzy sample x_1^*, \dots, x_{10}^*

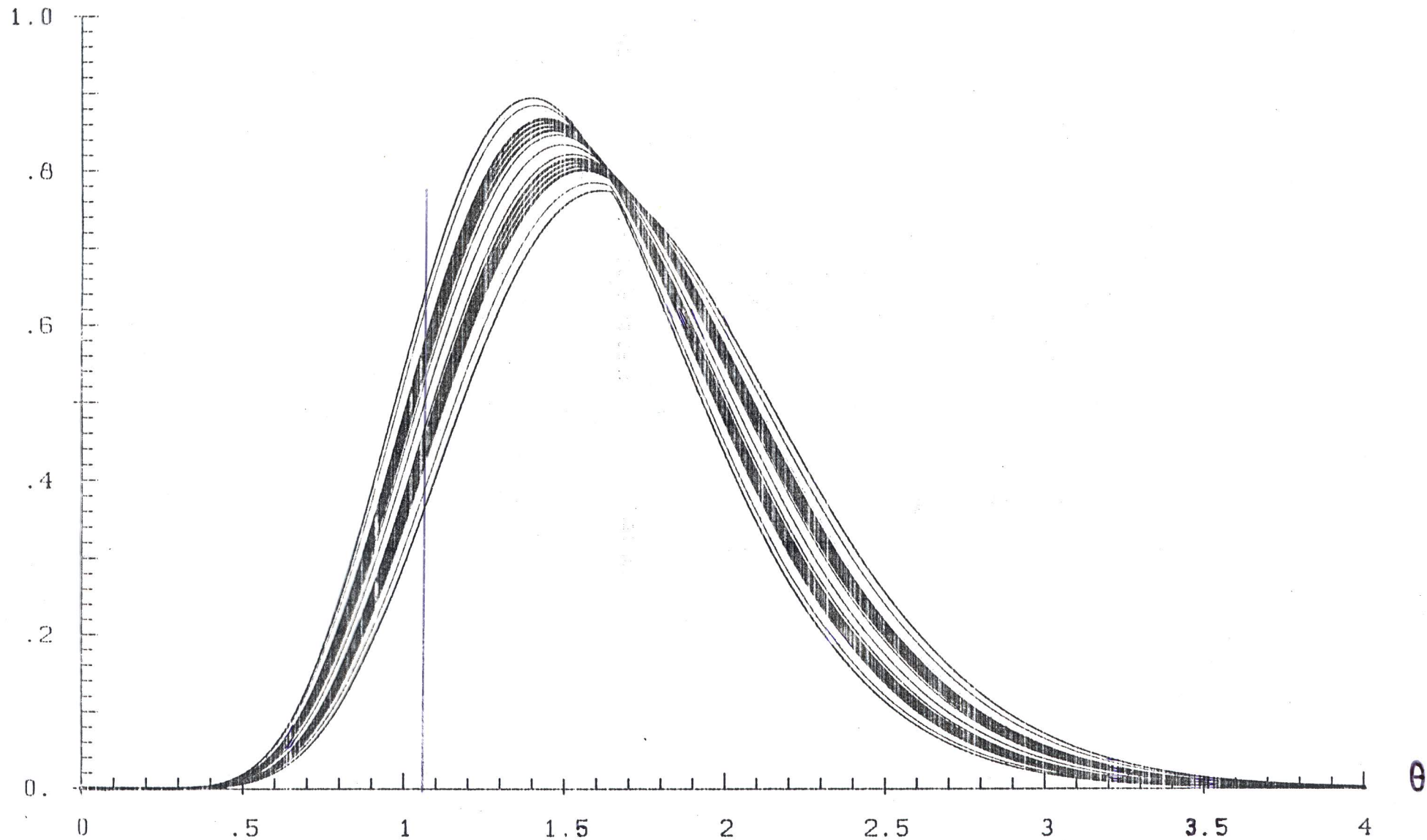
$$X \sim E_{x_0}$$

$n = 10$



$$\pi^*(\theta | \underline{x}^*)$$

$$\pi(\theta) \hat{=} \text{Gam}(1, 1)$$



FUZZY PROBABILITY DENSITY

Generalized densities $f^*(\cdot)$ on \mathbb{R} :

$f^*(\cdot)$ fuzzy function with δ -level functions
 $\underline{f}_\delta(\cdot)$ and $\bar{f}_\delta(\cdot)$ integrable with

$$\int_{\mathbb{R}} \bar{f}_\delta(x) dx < \infty \quad \forall \delta \in (0, 1]$$

and \exists classical density $f(\cdot)$ on \mathbb{R} with

$$\underline{f}_1(x) \leq f(x) \leq \bar{f}_1(x) \quad \forall x \in \mathbb{R}$$

The fuzzy probability $P^*(B)$ of $B \in \mathcal{B}$
is a fuzzy interval.

PROBLEMS

Sequential updating

Precise a-priori density

ALTERNATIVE SOLUTION

Based on δ -level functions

$$\bar{\pi}_\delta(\cdot), \quad \bar{l}_\delta(\cdot; \underline{x}^*), \quad \bar{\pi}_\delta(\cdot | \underline{x}^*)$$

$$\underline{\pi}_\delta(\cdot), \quad \underline{l}_\delta(\cdot; \underline{x}^*), \quad \underline{\pi}_\delta(\cdot | \underline{x}^*)$$

LIKELIHOOD FOR FUZZY DATA

\underline{x}^* combined fuzzy sample with v.c.f. $\xi(\cdot)$

$l^*(\theta; \underline{x}^*)$ fuzzy value of the likelihood $l(\theta; \underline{x})$
with c.f. $\eta_\theta(\cdot)$ defined by

$$\eta_\theta(y) = \left\{ \begin{array}{ll} \sup \{ \xi(\underline{x}) : l(\theta; \underline{x}) = y \} & \text{if } l^{-1}(\{y\}) \neq \emptyset \\ 0 & \text{if } l^{-1}(\{y\}) = \emptyset \end{array} \right\} \quad \forall y \in \mathbb{R}$$

Remark: For precise data \underline{x} the indicator function of $l(\theta; \underline{x})$ is obtained

GENERALIZED BAYES' THEOREM

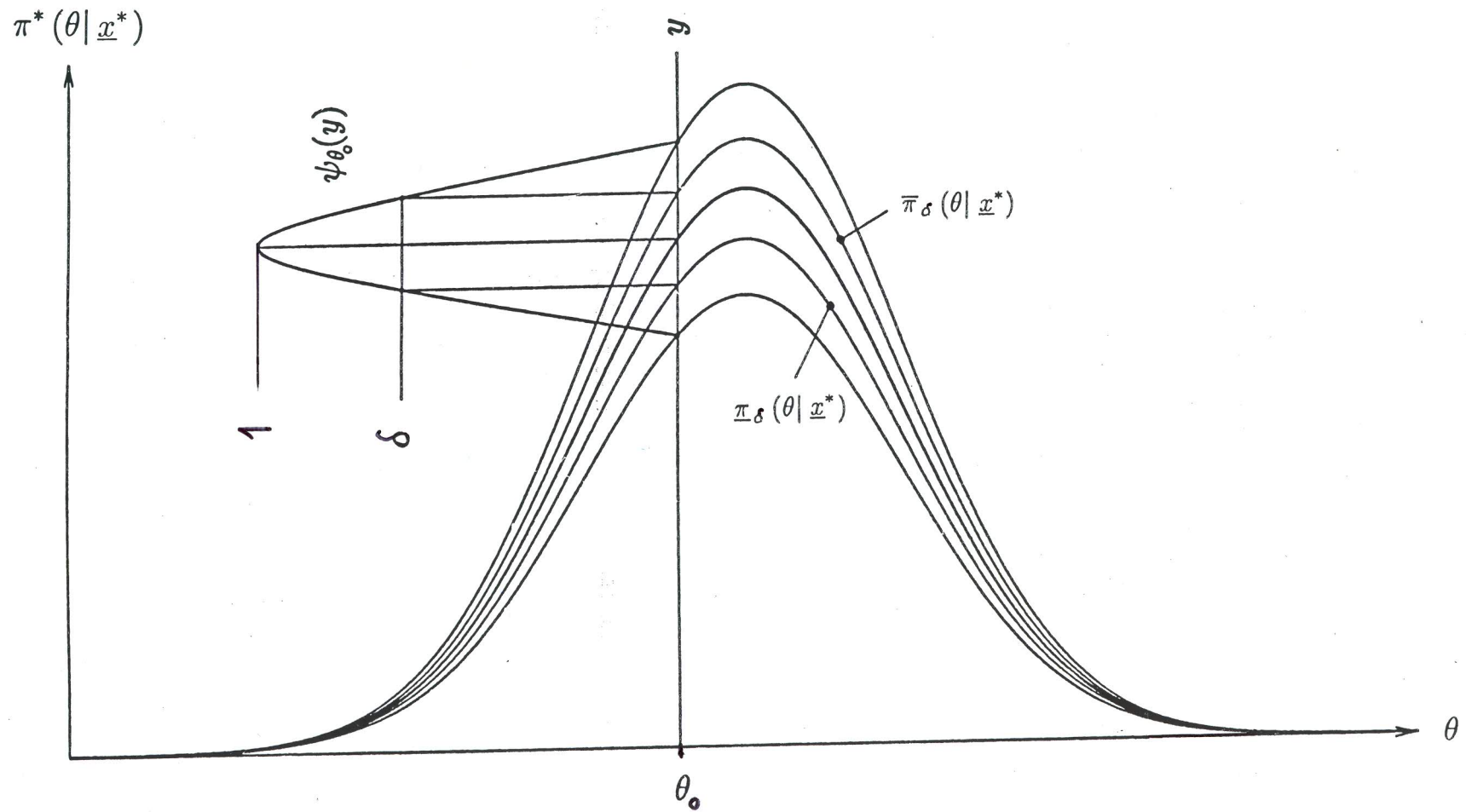
δ -level curves of the fuzzy a-posteriori density

$$\bar{\pi}_{\delta}(\theta | \underline{x}^*) = \frac{\bar{\pi}_{\delta}(\theta) \bar{l}_{\delta}(\theta; \underline{x}^*)}{\int_{\Theta} \frac{1}{2} [\underline{\pi}_{\delta}(\theta) \underline{l}_{\delta}(\theta; \underline{x}^*) + \bar{\pi}_{\delta}(\theta) \bar{l}_{\delta}(\theta; \underline{x}^*)] d\theta}$$

$$\underline{\pi}_{\delta}(\theta | \underline{x}^*) = \frac{\underline{\pi}_{\delta}(\theta) \underline{l}_{\delta}(\theta; \underline{x}^*)}{\int_{\Theta} \frac{1}{2} [\underline{\pi}_{\delta}(\theta) \underline{l}_{\delta}(\theta; \underline{x}^*) + \bar{\pi}_{\delta}(\theta) \bar{l}_{\delta}(\theta; \underline{x}^*)] d\theta}$$

$$\forall \theta \in \Theta$$

Figure Fuzzy a-posteriori density



EXAMPLE

$$X \sim \text{Exp}_\theta, \quad \theta \in \Theta = (0, \infty)$$

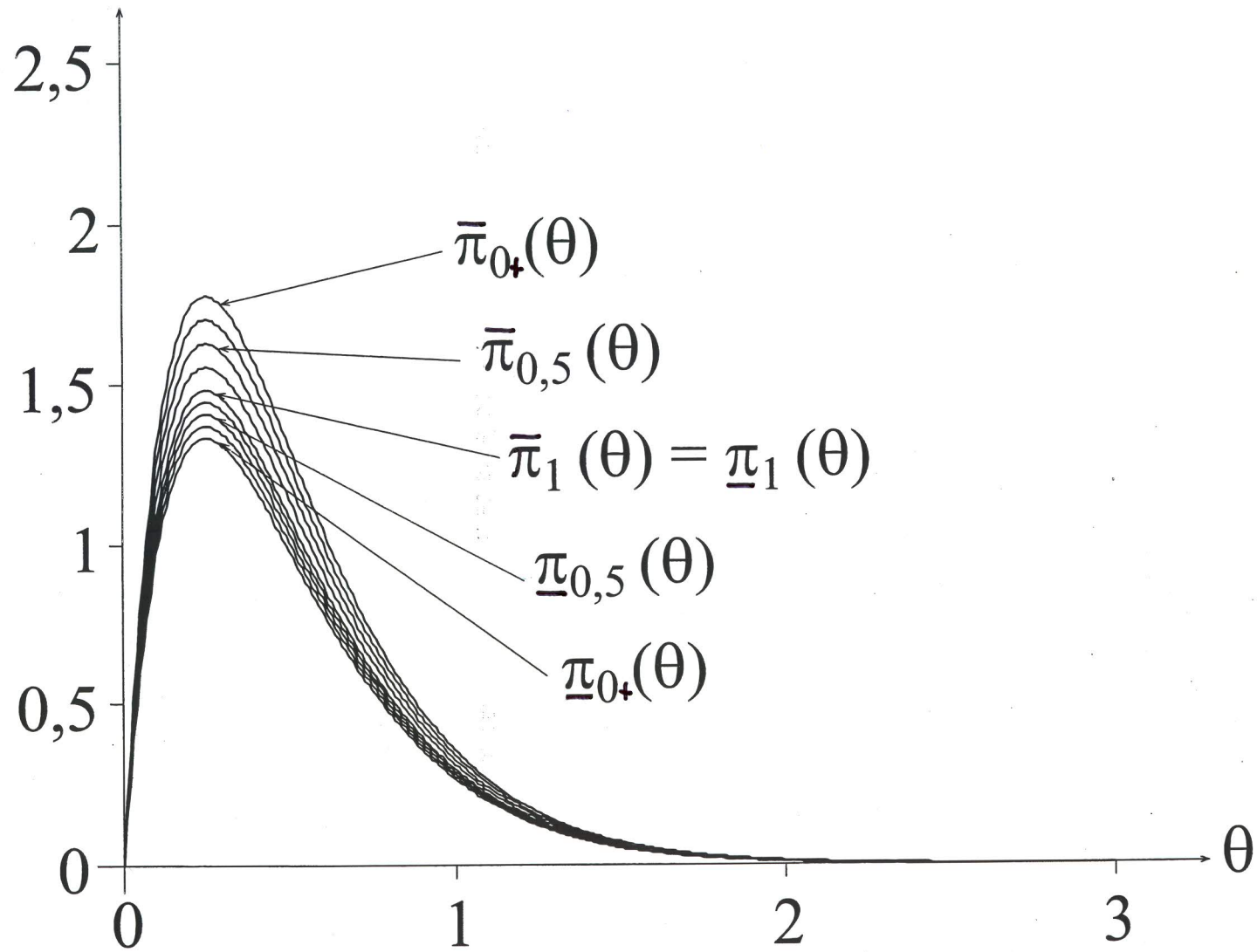
$$f(x|\theta) = \frac{1}{\theta} \exp\left\{-\frac{x}{\theta}\right\} \cdot I_{(0, \infty)}(x)$$

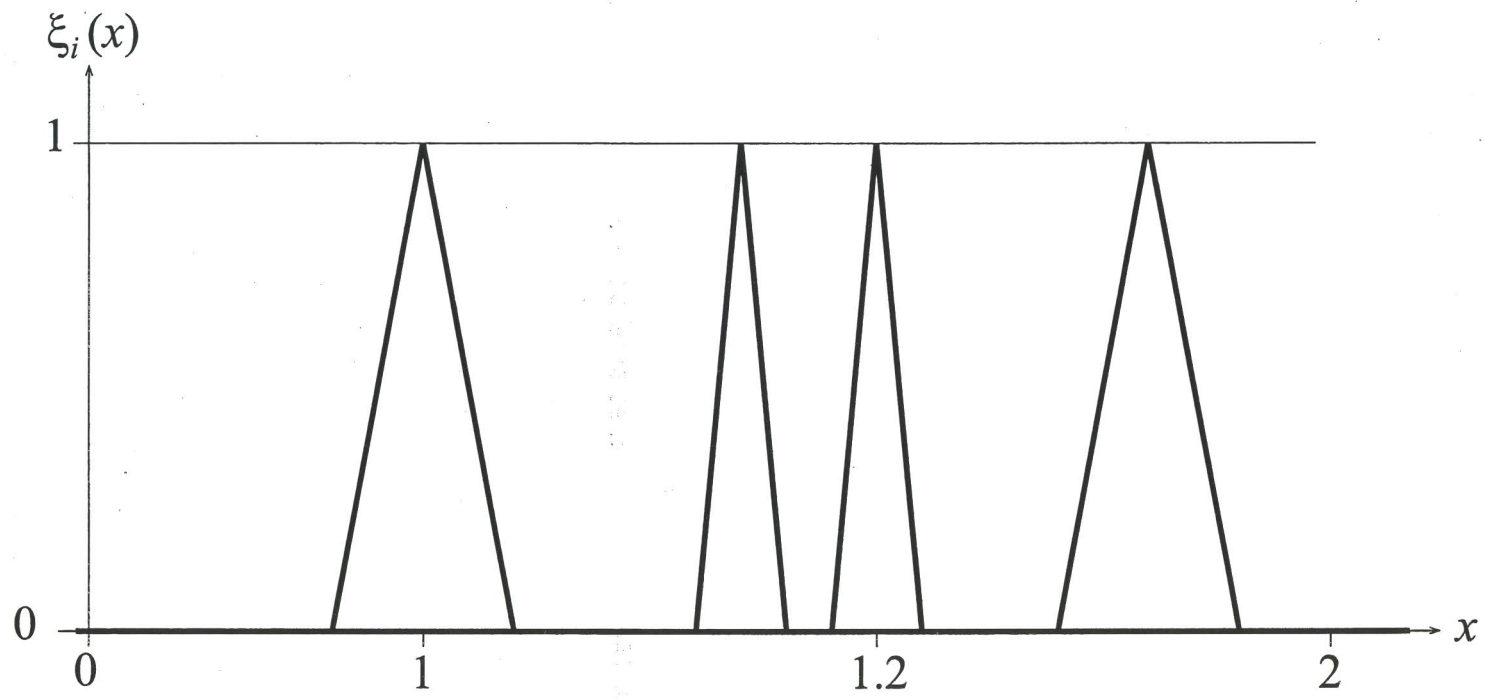
Fuzzy a-priori distribution

$\pi^*(\cdot)$ fuzzy gamma density

$\bar{\pi}_\delta(\cdot)$ upper } δ -level curves
 $\underline{\pi}_\delta(\cdot)$ lower }

$\bar{\pi}_\delta(\theta), \underline{\pi}_\delta(\theta)$





COMBINED FUZZY SAMPLE

$$\underline{x}^* = (x_1, x_2, x_3, x_4)^*$$

vector char. function $\int(\cdot, \cdot, \cdot, \cdot)$

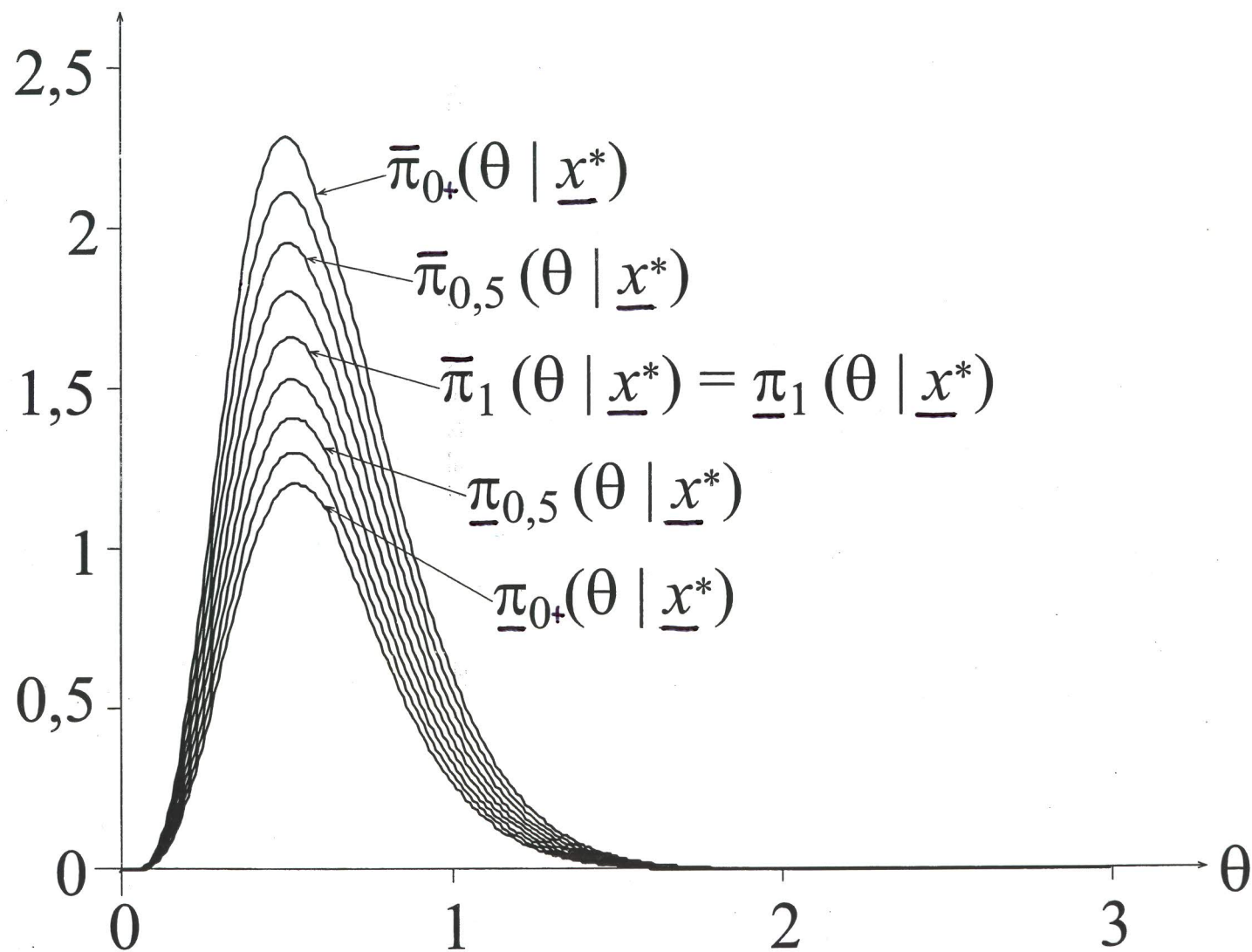
$$\int(x_1, x_2, x_3, x_4) = \min\{\int_1(x_1), \int_2(x_2), \int_3(x_3), \int_4(x_4)\}$$

$$\bar{\pi}_\delta(\cdot | \underline{x}^*)$$

by gen. Bayes' theorem

$$\underline{\pi}_\delta(\cdot | \underline{x}^*)$$

$\bar{\pi}_\delta(\theta | \underline{x}^*), \underline{\pi}_\delta(\theta | \underline{x}^*)$



HPD - Regions

$\pi(\cdot|D)$ a-posteriori Density

$1-\alpha$ Confidence level

$\Theta_{1-\alpha} \subseteq \Theta$ obeying:

$$1) \int_{\Theta_{1-\alpha}} \pi(\theta|D) d\theta = 1-\alpha$$

$$2) \pi(\theta|D) \text{ max. on } \Theta_{1-\alpha}$$

GENERALIZED HPD-Regions

$\pi^*(\cdot | D^*)$ Fuzzy a-posteriori Density

$$\mathcal{D}_\delta := \{g : g \text{ density with } \underline{\pi}_\delta(\theta) \leq g(\theta) \leq \overline{\pi}_\delta(\theta) \quad \forall \theta \in \Theta\}$$

${}^\delta \text{HPD}_{1-\alpha}(g)$ HPD-Region based on g

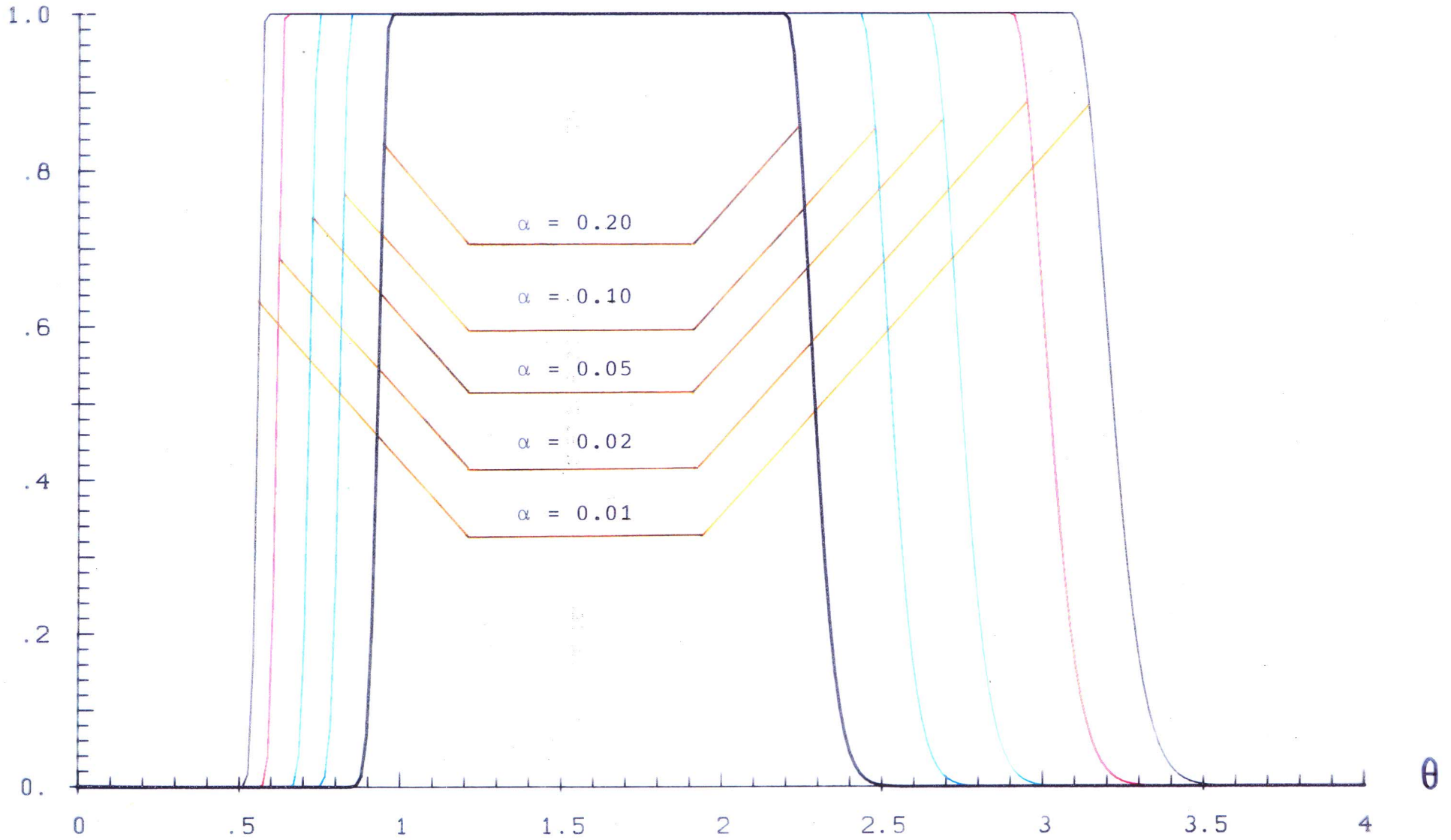
$$A_\delta := \bigcup_{g \in \mathcal{D}_\delta} {}^\delta \text{HPD}_{1-\alpha}(g) \quad \forall \delta \in (0, 1]$$

$\Rightarrow (A_\delta; \delta \in (0, 1])$ nested family of subsets of Θ

Construction Lemma for Membership Functions:

$$\varphi(\theta) := \sup \{ \delta \cdot \mathbf{1}_{A_\delta}(\theta) : \delta \in [0, 1] \} \quad \forall \theta \in \Theta$$

$E_{x\theta}$



PREDICTIVE DENSITIES

$X \sim f(\cdot|\theta), \theta \in \Theta$ Stochastic Model

$\pi(\cdot)$ a-priori density

$(x_1, \dots, x_n) = D$ data

$\Rightarrow \pi(\cdot|D)$ a-posteriori density

$p(\cdot|D)$ predictive density

$$p(x|D) = \int_{\Theta} f(x|\theta) \cdot \pi(\theta|D) d\theta \quad \forall x \in M_x$$

FUZZY PREDICTIVE DENSITY

$$p^*(\cdot | D^*)$$

$$p^*(x | D^*) = \int_{\Theta} f(x|\theta) \odot \pi^*(\theta | D^*) d\theta \quad \forall x \in M_x$$

$$\mathcal{D}_\delta := \{ g(\cdot) \text{ density on } \Theta : \underline{\pi}_\delta(\theta) \leq g(\theta) \leq \bar{\pi}_\delta(\theta) \forall \theta \in \Theta \}$$

$$a_\delta := \inf \left\{ \int_{\Theta} f(x|\theta) g(\theta) d\theta : g(\cdot) \in \mathcal{D}_\delta \right\}$$

$$\forall \delta \in (0, 1]$$

$$b_\delta := \sup \left\{ \int_{\Theta} f(x|\theta) g(\theta) d\theta : g(\cdot) \in \mathcal{D}_\delta \right\}$$

The nested family of intervals $[a_\delta; b_\delta]$ define a fuzzy number by the construction lemma:

$$\psi_x(y) = \sup \{ \delta \cdot \mathbf{1}_{[a_\delta; b_\delta]}(y) : \delta \in [0, 1] \} \quad \forall y \in \mathbb{R}$$

$$p^*(x|D^*) \hat{=} \psi_x(\cdot)$$

For variable x this is a fuzzy density

SOFTWARE

- Some Programs

C++, R

- Under Development:

AFD, SAFD

CONCLUSIONS

- Fuzziness can be described quantitatively
- Statistics based on fuzzy information is possible: Two different uncertainties
- Kolmogorov's probability concept has to be generalized
- Hybrid approach: Fuzzy and Stochastics

SOME REFERENCES

H. Bandemer: Mathematics of Uncertainty,
Springer, Berlin, 2006

T. Ross et al. (Eds.): Fuzzy Logic and Probability
Applications - Bridging the Gap, ASA and
SIAM, Philadelphia, 2002

C. Borgelt et al. (Eds.): Combining Soft Computing
and Statistical Methods in Data Analysis,
Springer, Berlin, 2010

D. Dubois et al. (Eds.): Soft Methods for Handling Variability and Imprecision, Springer-Verlag, Berlin, 2008

R. Viertl: Foundations of Fuzzy Bayesian Inference, Journal of Uncertain Systems Vol. 2, No. 3 (2008)

R. Viertl: On reliability estimation based on fuzzy lifetime data, Journal of Statistical Planning and Inference 139 (2009)

R. Viertl: Statistical Methods for Fuzzy Data, Wiley, Chichester, 2011