Modeling and Estimation of Multivariate Densities

Modellierung und Schätzung von multivariaten Dichten

Eckhard Liebscher

University of Applied Sciences Merseburg, Germany

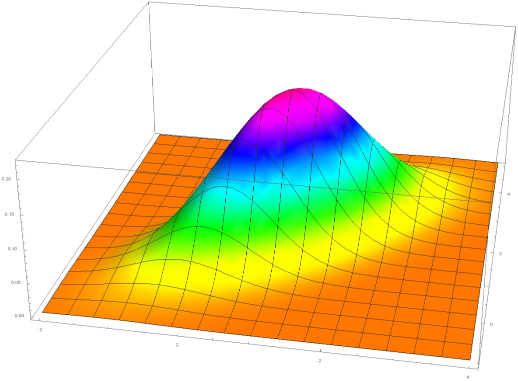
28.05.2024

Outline

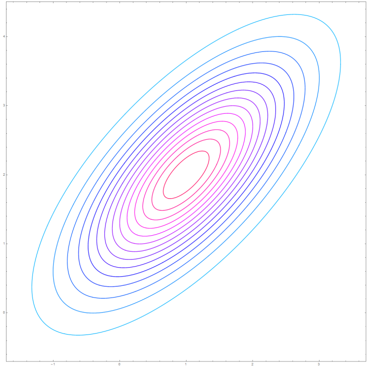
Multivariate Normal Distribution

(X, Y) normally distributed with $\mathbb{E}(X) = 1, \mathbb{E}(Y) = 2$, $\text{Var}(X) = 1$, $\text{Var}(Y) = 1$, $\text{corr}(X, Y) = 0.7$

bivariate density



contour plot



Modeling Multivariate Distributions

- elliptical distributions
- skewed normal distributions, mixture of normal distributions
- copulas
- star-shaped distributions

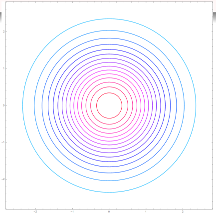
Fields of applications:

- methods of applied statistics: cluster analysis,...
- finance
- survival analysis - reliability
- time series analysis - multivariate time series
- Statistical Process Control

Spherical and Elliptical Densities

Random vector $X \in \mathbb{R}^p$ has a **continuous spherical distribution** $SC_p(g_p)$, if the density exist and can be written as

$$g_p(t^T t) = g_p(t_1^2 + \ldots + t_p^2) \text{ for } t \in \mathbb{R}^p$$



with a function $g_p : [0, \infty) \to [0, \infty)$, $X \sim SC_p(g_p)$.

$g_p \ldots$ **density generator** for the dimension p if $g_p \geq 0$, and $\int_0^\infty r^{p/2-1} g_p(r) dr = \frac{\Gamma(p/2)}{\pi^{p/2}}$ for $p \geq 1$.

Random vector $Y \in \mathbb{R}^p$ has a **continuous elliptical distribution** with parameter $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p,p}$, $rank(\Sigma) = p$, $\Sigma = AA^T$, in symbols $Y \sim EC_p(\mu; \Sigma; g_p)$ if

$$Y \stackrel{d}{=} \mu + AZ, \quad Z \sim SC_p(g_p)$$

Elliptical Densities

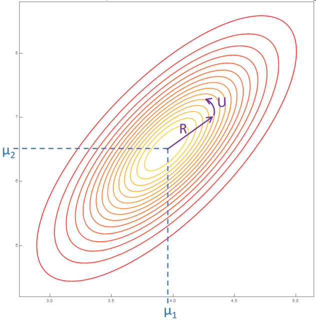
- **density of the random vector** $X = (X^{(1)}, \dots, X^{(p)})^T$:

$$f_X(x) = \det(\Sigma)^{-1/2} g\left((x-\mu)^T \Sigma^{-1}(x-\mu)\right),\ x \in \mathbb{R}^p,$$

generator fcn. $g : [0, +\infty) \to [0, +\infty)$,

location $\mu \in \mathbb{R}^d$, scaling matrix $\Sigma = AA^T$

contour plot of the density



- Representation:

$$X \stackrel{d}{=} \mu + R\ AU$$

random vector $U \in \mathcal{S}_{p-1}$ uniform distribution

random variable $R \in [0, +\infty)$,

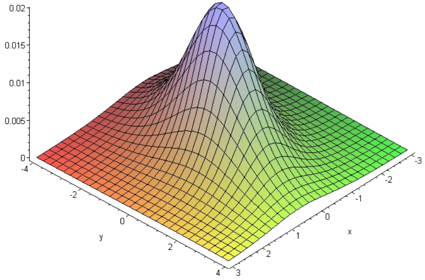
R and U independent

Elliptical Densities

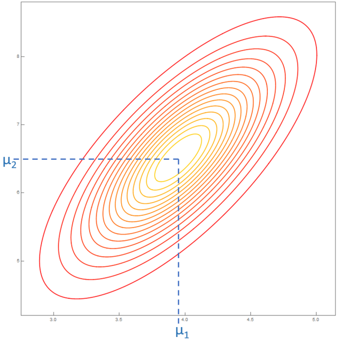
- **density of the random vector** $X = (X^{(1)}, \ldots, X^{(p)})^T \sim EC_p(\mu; \Sigma; g_p)$:

$$f_X(x) = \det(\Sigma)^{-1/2} g_p\left((x-\mu)^T \Sigma^{-1}(x-\mu)\right), \; x \in \mathbb{R}^d,$$

location $\mu \in \mathbb{R}^d$, scaling matrix Σ



contour plot of the density



Fang/Zhang 1990, Fang/Kotz/Ng 1990, Bilodeau/Brenner 1999

Consistent Families of Generators

$$f_X(x) = \det(\Sigma)^{-1/2} g_p\left((x-\mu)^T \Sigma^{-1}(x-\mu)\right),\ x \in \mathbb{R}^d,$$

$\{g_p\}_{p=1,2\ldots}$ family of density generators

This family is **consistent** $:\Longleftrightarrow$

$$g_p(r) = 2\int_0^\infty g_{p+1}(r+t^2)dt = \int_r^\infty \frac{1}{\sqrt{y-r}} g_{p+1}(y)\,dy$$

for $r \geq 0, 1 \leq p < p_0$. Then: $z \longmapsto g_q(z_1^2 + \ldots + z_q^2)$ $(z = (z_1, \ldots, z_q)^T \in \mathbb{R}^q)$ is the q-dimensional marginal density of X for $q < p$.

Consistent Families of Generators

$\{g_p\}_{p=1,2...}$ family of density generators

Theorem.

Assume that $g_1 : [0,\infty) \to [0,\infty)$ *completely monotonic on* $[0,\infty)$,

$\lim_{r\to\infty} g_1(r) = 0$,

for $p \geq 1, r \geq 0$ *and*

$$g_{p+2}(r) = -\frac{1}{\pi} g_p'(r),$$

$$g_2(r) = 2\int_0^\infty g_3\left(r + y^2\right) dy \qquad (r \geq 0).$$

$$\int_0^\infty r^{-1/2} g_1(r) dr = 1$$

$\Longrightarrow \{g_p\}_{p\geq 1}$ *is a consistent family of density generators.*

Algorithm for the construction of consistent families of density generators

Consistent Families of Generators

$\{g_p\}_{p=1,2\dots}$ consistent family of density generators

$\Longrightarrow$

$$(g_p(-u))^{(1/2)} = \sqrt{\pi} g_{p+1}(-u) \quad (u \leq 0, p \geq 1),$$

Let $a \in [-\infty, \infty)$, $\nu > 0, \nu \notin \mathbb{N}$ and $n = [\nu] + 1$, where $[\alpha]$ is the integer part of α. the fractional derivative of order ν in the sense of Caputo of function $\varphi : \mathbb{R} \to \mathbb{R}$ is defined by

$$_C D_a^\nu \varphi(x) = \frac{1}{\Gamma(n-\nu)} \int_a^x \frac{\varphi^{(n)}(t)}{(x-t)^{\nu-n+1}} dt \quad \text{for } x > a.$$

Examples of Generator Families

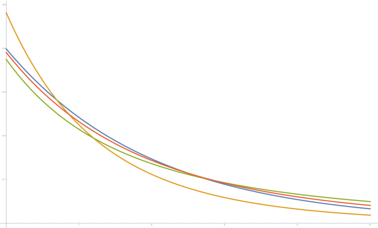
- **normal distribution**

$$g_p(r) = \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{r}{2}\right)$$

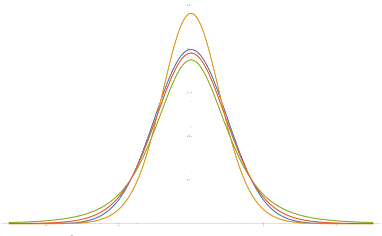
- **normal mixture generator**

$$g_p(r) = \frac{1}{(2\pi)^{p/2}} \left(q a^{p/2} \exp(-\frac{ar}{2}) + (1-q) \exp(-\frac{r}{2})\right)$$

generator fcn. g_p



one-dimensional density



- ***t*-distribution (Pearson VII)**, $m > 0$

$$g_p(r) = \frac{\Gamma(p/2+m)}{(2m)^{p/2}\pi^{p/2}\Gamma(m)} \left(1 + \frac{r}{2m}\right)^{-p/2-m}$$

Examples of Generator Families

- **logarithmic generator** $a > 1, b > 0, b \leq a - 1$

$$g_1(r) = \frac{\sqrt{b}\ln(r+a)}{2\pi\ln(\sqrt{a}+\sqrt{b})\ (r+b)} \to g_3, g_2$$

- **fractional-exponential generator**

$$\begin{aligned} g_1(r) &= \frac{e^{-a-r}\sqrt{a}}{(r+a)\mathrm{erfc}(\sqrt{a})\ \pi} \to \\ g_p(r) &= \frac{\sqrt{a}\bar{\Gamma}(\frac{p+1}{2}, a+r)}{\mathrm{erfc}(\sqrt{a})\ \pi^{(p+1)/2}(a+r)^{(p+1)/2}}, \end{aligned}$$

- **mixtures**
- **Kotz-type family**

$$g_p(r) = \frac{t}{\Gamma(\frac{s+d}{t})} r^s e^{-r^t} \text{ for } r > 0.$$

problem: no explicit formulas for $g_q, q \neq p$

Moments and Identifiability

$$f_X(x) = \det(\Sigma)^{-1/2} g\left((x-\mu)^T \Sigma^{-1}(x-\mu)\right),\ x \in \mathbb{R}^p,$$

$$\mathbb{E}(X) = \mu,\ \mathrm{cov}(X) = \frac{2\pi^{p/2}}{p\Gamma(p/2)} \int_0^\infty r^{p+1} g_p(r^2) dr\ \Sigma$$

Theorem.

$X \sim EC_p(\mu; \Sigma; g_p), X \sim EC_p(\bar{\mu}; \bar{\Sigma}; \bar{g}_p) \Longrightarrow$

$$\bar{\mu} = \mu,\ \bar{\Sigma} = c\Sigma,\ \bar{g}_p(z) = c^{p/2} g_p(cz)\ \forall z \geq 0$$

Choose c such that

$$\frac{2\pi^{p/2}}{p\Gamma(p/2)} \int_0^\infty r^{p+1} g_p(r^2) dr = 1 \Longleftrightarrow \mathrm{cov}(X) = \Sigma$$

Estimation problems

$$\mathbb{E}(X) = \mu, \ \mathrm{cov}(X) = \Sigma$$

$X_1, X_2, \ldots X_n$ sample, $X \sim EC_p(\mu; \Sigma; g_p)$

1) Nonparametric estimation of μ and Σ

$$\hat{\mu} = \bar{X}, \quad \hat{\Sigma} = \text{empirical covariance matrix}$$

alternatively: M-estimator Maronna 1976

2) Parametric estimation of the density

Maximum-Likelihood estimator

3) Semiparametric estimation of the density: Kernel density estimator for g_p

Maximum likelihood estimation

model for the density $g_p(\cdot \mid \tau)$ with parameter $\tau \in \Theta_1 \subset \mathbb{R}^{q_1}$

▷ Sample $X_1, \ldots, X_n,\ X \sim EC_p(\mu; \Sigma; g_p)$

parameter $\theta = (\mu, \underbrace{\sigma_{11}, \sigma_{12}, \ldots, \sigma_{pp}}_{\text{without repetition of identical entries}}, \tau)^T \in \Theta$

• Likelihood function

$$L(\theta) = \prod_{i=1}^{n} g_p\left((X_i - \mu)^T \Sigma^{-1} (X_i - \mu) \mid \tau\right)$$

Maximum-likelihood estimator

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

The estimator $\hat{\theta}$ is asymptotically normally distributed.

Maximum likelihood estimation - Example

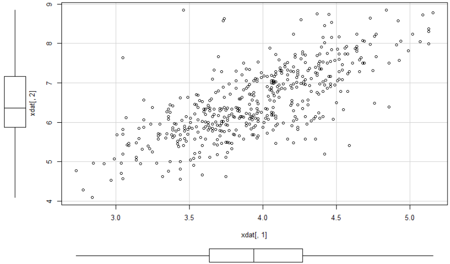
dataset 5 of Andrews and Herzberg (1985)

variables: yield of grain and straw

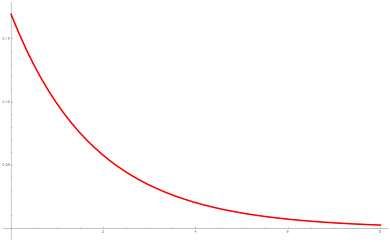
model function $g_p(r) = (qa^{p/2}\exp(-ra/2) + (1-q)\exp(-r/2))/(2\pi)^{p/2}$

ml-method $\rightarrow \hat{q} = 0.09476697, \hat{a} = 1.68244509$

scatter plot of the data



plot of g_p,

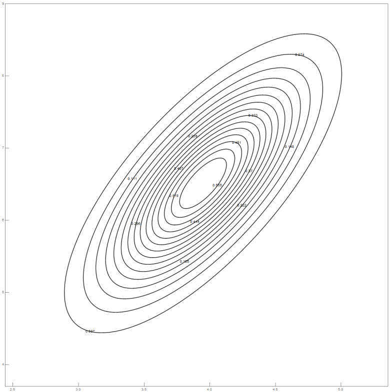


Maximum likelihood estimation - Example

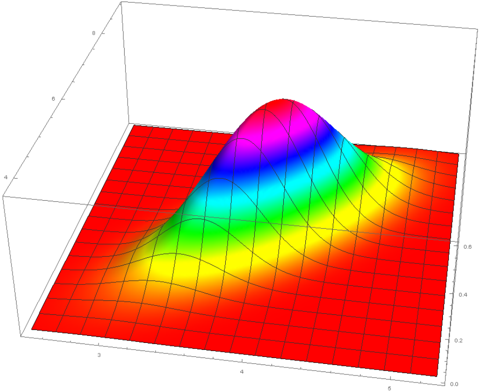
dataset 5 of Andrews and Herzberg (1985) - yield of grain and straw

ml-method $\rightarrow \hat{q} = 0.09476697, \hat{a} = 1.68244509$

contour plot of the density



density



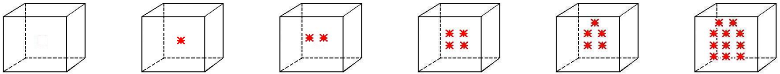
Curse of dimensionality

Data example: Liver-disorders (UCI Machine Learning Repository)
Variable $X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}$,
The range is divided into 3 subintervals:

81 cuboids are formed $D_{1i_1} \times D_{2i_2} \times D_{3i_3} \times D_{4i_4}$.

$n = 200$ daten vectors are distributed as follows:

more than 5 vect.



54 cuboids	8 cuboids	5 cuboids	2 cuboids	2 cuboids	10 cuboids
67%	10%	6%	2.5%	2.5%	12%

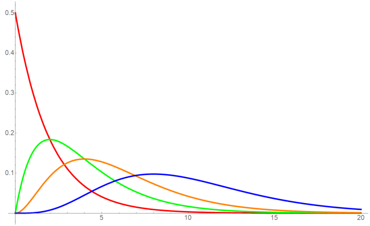
$\Longrightarrow$ Dimension reduction techniques

Preliminaries - Volcano effect

We consider: $Y = (X-\mu)^T\Sigma^{-1}(X-\mu)$ with density

$$f_Y(y) = s_p \cdot y^{p/2-1} g(y) \quad (y \in \mathbb{R}^+), \quad s_p = \frac{\pi^{p/2}}{\Gamma(p/2)}$$

density of Y in the multinormal case



- **Idea 1:** $\hat{f}_Y$ estimator for $f_Y \Longrightarrow$

$\hat{g}(y) = s_p^{-1} y^{-p/2+1} \hat{f}_Y(y)$

estimator for g.

Problem: $\hat{g}(y) \to \infty$ for $y \to 0$.

the mass moves away from zero as $p \to \infty$

Construction of the estimator

$$f(x) = \det(\Sigma)^{-1/2} g\left((x-\mu)^T \Sigma^{-1}(x-\mu)\right) \qquad (x \in \mathbb{R}^p)$$

transformation $\tilde{Y} = \psi(Y)$, $Y = (X-\mu)^T \Sigma^{-1}(X-\mu)$

Assume that $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ has a derivative ψ' with $\psi'(y) > 0$ for $y \geq 0$, and $\psi(0) = 0$. Ψ inverse function of ψ.

$\tilde{Y}$ has density h:

$$h(t) = \Psi'(t)\, f_Y(\Psi(t)) = s_p \cdot \Psi'(t)\, \Psi(t)^{p/2-1} g(\Psi(t)),$$

$\Longrightarrow$

$$g(y) = s_p^{-1} y^{-p/2+1} \psi'(y)\, h(\psi(y))$$

Transformation function ψ

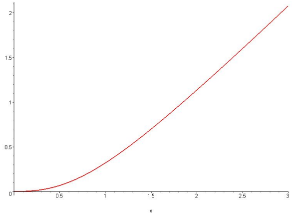
Assumption $\mathcal{T}(p)$. $\Psi = \psi^{-1}$, $\Psi^{(p+1)}$ exists and is continuous on $(0, \infty)$, $\psi' > 0$, ψ' is supposed to be bounded on $(0, +\infty)$, and ψ'' is assumed to be bounded on $(0, +\infty)$. Function $x \mapsto x^{p/2-1}\psi'(x)^{-1}$ has a bounded derivative on $[0, M_1]$ with $M_1 > 0$,

$$\lim_{x \downarrow 0} x^{-p/2+1}\psi'(x) = C_1 > 0.$$

There exist constants $\alpha \in (0, 1]$, $C_2, M_2 > 0$, such that

$$|\Psi(t)| \leq C_2 |t|^{\alpha} \quad \text{for } t \in [0, M_2].$$

Example:
$\psi(x) = -a + (a^{p/2} + x^{p/2})^{2/p}$
image for $p = 5, a = 1$:



Construction of the estimator

$$f(x) = \det(\Sigma)^{-1/2} g\left((x-\mu)^T \Sigma^{-1}(x-\mu)\right) \qquad (x \in \mathbb{R}^p)$$

Transformation $\tilde{Y} = \psi(Y)$ mit Dichte h

$$g(y) = s_p^{-1} y^{-p/2+1} \psi'(y)\, h(\psi(y))$$

- **kernel estimators for g, f, h:**

$$\hat{g}_n(z) = s_p^{-1} z^{-p/2+1} \psi'(z)\, \hat{h}_n(\psi(z)) \quad (z \geq 0),$$

$$\hat{f}_n(x) = \det(\hat{\Sigma}_n)^{-1/2} \hat{g}_n\left((x-\hat{\mu}_n)^T \hat{\Sigma}_n^{-1}(x-\hat{\mu}_n)\right) \quad (x \geq 0).$$

Construction of the estimator

$X_1, X_2, \dots X_n$ sample, $X \sim EC_p(\mu; \Sigma; g_p)$

- transformed sample: $Y_{n1}, \dots, Y_{nn}$,

$$Y_{ni} = \psi\left((X_i - \hat{\mu}_n)^T \hat{\Sigma}_n^{-1}(X_i - \hat{\mu}_n)\right), \quad i = 1, \dots, n.$$

- Assumption $$\limsup_{n\to\infty} \sqrt{\frac{n}{\ln\ln(n)}} \left\|\hat{\theta}_n - \theta\right\| = C_0 \quad f.s.$$

with a constant $C_0 > 0$.

- **Kernel estimator for h:** mirror method

$$\hat{h}_n(y) = \frac{1}{nb}\sum_{i=1}^{n}\left(K\left((y - Y_{in})b^{-1}\right) + \underbrace{K\left((y + Y_{in})b^{-1}\right)}_{\textbf{additional term}}\right) \qquad (y \geq 0)$$

Construction of the estimator

- **Assumptions on the kernel function:** **Condition $\mathcal{K}(p)$**

$K : \mathbb{R} \to \mathbb{R}$ bounded, $K(-t) = K(t)$,

$$\begin{aligned} \int_{-1}^{1} K(t)\,\mathrm{d}t &= 1, \quad K(t) = 0 \text{ for } |t| > 1, \\ \int_{-1}^{1} u^j K(u)\,\mathrm{d}u &= 1 \quad (j \le p-1 \text{ even}) \end{aligned}$$

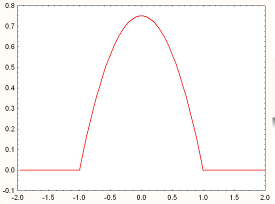
- **Assumptions on the bandwidth** $b = b(n)$:

$$\begin{aligned} \lim_{n\to\infty} b(n) \ln\ln n &= 0 \quad \text{and} \\ b(n) &\ge C_5 \cdot n^{-1/5}. \end{aligned}$$

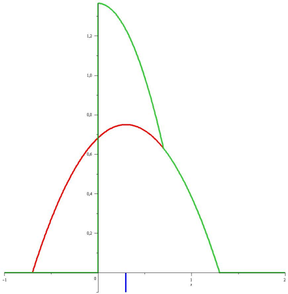
Construction of the estimator

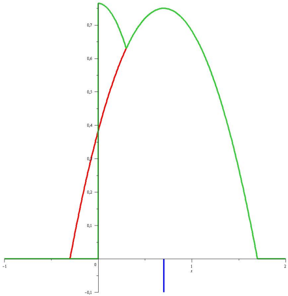
- **Epanechnikov kernel**

$$K(x) = 0.75(1 - x^2) \text{ for } -1 \leq x \leq 1$$

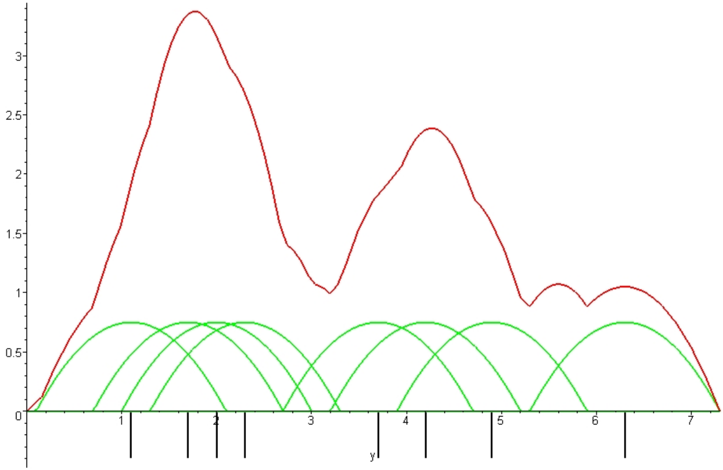


Kernel function at boundary for $b = 1$ and data items at 0.3 and 0.7





Kernel estimator



Covergence rate of the estimator

Theorem.

Assume that $g^{(q)}$ exists and is bounded on $\mathbb{R}^+$ for even $q \geq 2$, Assumptions $\mathcal{K}(q)$, $\mathcal{T}(q)$ are satisfied. Then for a compact set D with $\mu \notin D$:

$$\sup_{x \in D} \left| \hat{f}_n(x) - f(x) \right| = O\left(\sqrt{\ln n} (nb(n))^{-1/2} + b^q(n) \right) \quad a.s.$$

For a compact set D with $\mu \in D$ we have

$$\sup_{x \in D} \left| \hat{f}_n(x) - f(x) \right| = O\left(\sqrt{\ln n} (nb(n))^{-1/2} + b^\gamma(n) \right) \quad a.s.$$

with $\gamma > 0$.

- optimized convergence rate: $O\left(\left(\frac{n}{\ln n}\right)^{-q/(2q+1)}\right)$

improves rates in Cui and He (1995)

Asymptotic distribution of the estimator

Theorem.

$b(n) \sim C_1 \; n^{-1/(2p+1)}$ with a constant $C_1 > 0$,

$x \neq \mu$ such that $g^{(p)}$ is continuous in $u := (x-\mu)^T \Sigma^{-1} (x-\mu) \Longrightarrow$

$$\sqrt{nb(n)} \left(\hat{f}_n(x) - f(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(\bar{\mu}, \bar{\sigma}^2), \text{ where}$$

$$\begin{aligned} \bar{\sigma}^2 &= \det(\Sigma)^{-1} s_d^{-1} u^{-d/2+1} \psi'(u) g(u) \int_{-1}^{1} K^2(t)\, dt \\ \bar{\mu} &= \det(\Sigma)^{-1/2} s_d^{-1} u^{-d/2+1} \psi'(u) C_1^{(2p+1)/2} \frac{1}{p!} h^{(p)}(\psi(u)) \int_{-1}^{1} t^p K(t)\, dt. \end{aligned}$$

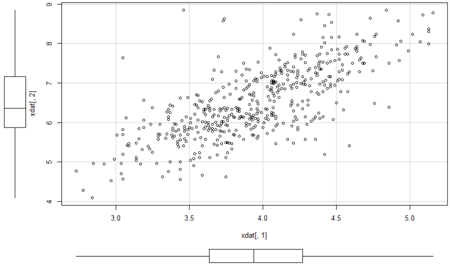
Stute/Werner (1991): known μ and $\psi(x) \equiv x$

Semiparametric estimation - Example

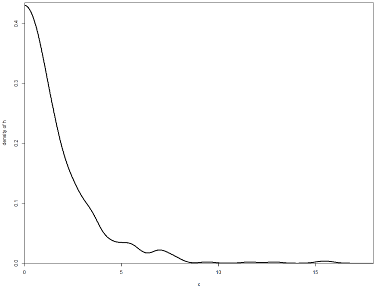
dataset 5 of Andrews and Herzberg (1985)

variables: yield of grain and straw

scatter plot of the data



plot of g_p,



Semiparametric estimation - High-dimensional case

📖 Liebscher/Okhrin (2023)

dimension p may tend to ∞, $p \leq n^{1/2}$

density h of R^2/p

$$g(r) = s_p^{-1} p^{-1} r^{-p/2+1} h(r/p)$$

h bounded

- **kernel estimators for g, f, h:** $\hat{h}_n$ as before

$$\hat{g}_n(r) = s_p^{-1} r^{-p/2+1}\, \hat{h}_n(r) \quad (r \geq 0),$$

$$\hat{f}_n(x) = \det(\hat{\Sigma}_n)^{-1/2} \hat{g}_n\left((x - \hat{\mu}_n)^T \hat{\Sigma}_n^{-1} (x - \hat{\mu}_n)\right) \quad (x \geq 0).$$

Semiparametric estimation - High-dimensional case

$p \to \infty$

Assumption: $C_1 \leq \lambda_{min} \leq \|\Sigma\|_2 = \lambda_{max} \leq C_2$

$$\hat{f}_n(x) = \det(\hat{\Sigma}_n)^{-1/2} \hat{g}_n \left((x - \hat{\mu}_n)^T \hat{\Sigma}_n^{-1} (x - \hat{\mu}_n) \right) \quad (x \geq 0).$$

Theorem.

Assume that $g^{(q)}$ exists and is bounded on $\mathbb{R}^+$ for even $q \geq 2$, Assumption $\mathcal{K}(q)$ is satisfied, $b(n) \geq C_3 n^{-1/5}$. Then for $0 < m < M$:

$$\sup_{x: x^T \Sigma x \in [mp, Mp]} \left| \hat{f}_n(x) - f(x) \right| = O\left(\sqrt{\ln n} (nb(n))^{-1/2} + b^q(n) \right)$$
$$+ O(pn^{-1/2} \sqrt{\ln n} + p^2 n^{-1} \ln n b^{-2}) \quad a.s.$$

- optimized convergence rate: $O((\frac{n}{\ln n})^{-q/(2q+1)})$

Contour function

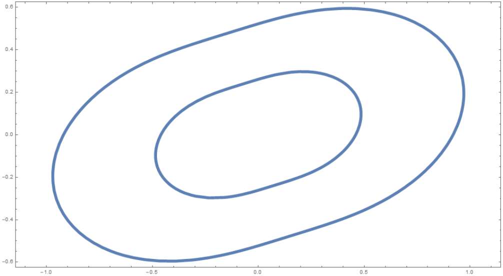
$h : \mathbb{R}^p \to [0, +\infty)$ positively homogeneous function:

$h(\lambda x) = \lambda h(x)\ \forall x \in \mathbb{R}^p, \lambda > 0$

Let $x_0 \in \mathbb{R}^p$ belong to the star-shaped sphere $\{x : h(x) = 1\}$

$\to h(\lambda x_0) = \lambda$

star-shaped sphere and set $\{x : h(x) = 0.5\}$



Example for h

$h(x) = \|x\|_q$

Star-shaped density

- **Density of the random vector**

$X = (X^{(1)}, \dots, X^{(p)})^T \sim CSS(g, h, \mu, \Sigma)$:

$$f_X(x) = \left(C_0 \det(\Sigma)\right)^{-1} g\left(h(\Sigma^{-1}(x-\mu))\right),\ x \in \mathbb{R}^p,$$

generator fcn. $g : [0, +\infty) \to [0, +\infty)$, contour fcn. $h : \mathbb{R}^p \to [0, +\infty)$,

location $\mu \in \mathbb{R}^d, \sigma = (\sigma_1, \dots, \sigma_p)$, scaling $\Sigma =$diag$(\sigma_1, \dots, \sigma_p)$

- Representation: Richter 2014

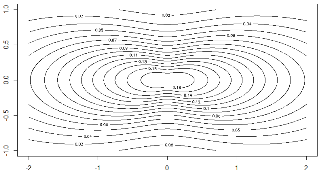
$$X \stackrel{d}{=} \mu + R\,\Sigma U$$

$R = h(\Sigma^{-1}(X-\mu)) \in [0, +\infty)$ and

$U = \frac{1}{R}\left(\Sigma^{-1}(X-\mu)\right) \in \mathcal{S}_{p-1}$ independ.

$g(r) = r^{1-p} f_R(r)$, f_R density of R

density contour plot - example



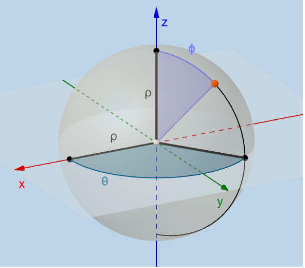
volcano effect for bounded g

Models for h

- $h(x) = \|x\|_q \to l_q$-elliptical distribution, Gupta/Song 1997
- p-dimensional spherical coordinates $\to$

transformation T

$$
\begin{aligned}
x_1 &= \cos\alpha_1, \ x_2 = \cos\alpha_2 \sin\alpha_1, \\
x_j &= \cos\alpha_j \prod_{k=1}^{j-1} \sin\alpha_k, \quad x_p = \prod_{k=1}^{p-1} \sin\alpha_k
\end{aligned}
$$



source: geogebra

$T : \{x \in \mathbb{R}^p : \|x\| = 1\} \to [0,\pi]^{p-2} \times [0, 2\pi)$,

$x \mapsto (\alpha_1, \ldots, \alpha_{p-1})$,

- **Model function** $H_\eta : [0,\pi]^{p-2} \times [0,2\pi) \to (0,\infty)$,

parameter $\eta \in \Theta_2 \subset \mathbb{R}^{q_2}$

$$h_\eta(x) = \|x\|_2 \, H_\eta(T(\|x\|_2^{-1} \, x))$$

Models for h

- p-dimensional polar coordinate transformation

$T : \{x \in \mathbb{R}^d : \|x\| = 1\} \to [0,\pi)^{p-2} \times [0,2\pi]$ with $x \rightsquigarrow (\alpha_1, \ldots, \alpha_{p-1})$,

$$h_\eta(x) = \|x\|_2 \, H_\eta(T(\|x\|_2^{-1}\, x))$$

$$H_\eta(\alpha_1, \ldots, \alpha_{p-1}) = h_\eta(T^{-1}(\alpha_1, \ldots, \alpha_{p-1}))$$

Model function H_η, $\eta \in \Theta_2 \subset \mathbb{R}^{q_2}$

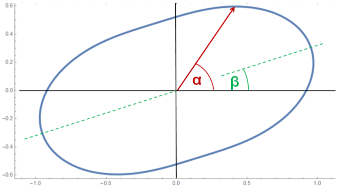
- Example for $d = 2$, $\alpha = \alpha_1$:

model with parameters $\eta = (a, \beta)^T$

$H_\eta(\alpha) = 1 + a - a\cos(2\alpha - 2\beta)$

length of the red arrow: $1/H(\alpha)$

contour of the density $h = 1$



Parametric Models for h - case $d = 2$

$$h_\eta(x) = \|x\|_2 \, H_\eta(\alpha)$$

α... angle between point and x-axis, $H_\eta : [0, 2\pi) \to (0, \infty)$

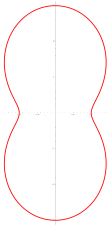
- general properties: H_η is a bounded, continuously differentiable function,

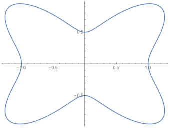
$$\lim_{t \to 2\pi - 0} H_\eta(t) = H_\eta(0), \quad H_\eta(\pi + \alpha) = H_\eta(\alpha), \quad \inf_{\alpha \in [0, 2\pi)} H_\eta(\alpha) > 0.$$

Model class 1:

$\eta = (a, b, c, \beta)^T$, $a, b, c > 0, \beta \in [0, \pi)$

$$H(\alpha) = \left(\frac{1 + a \tan^2(\alpha - \beta)}{1 + b \tan^2(\alpha - \beta)} \right)^c \quad \text{for } \alpha \in [0, 2\pi)$$





Model class 2:

$$H(\alpha) = \left(1 + a \sin^2(\alpha - \beta) + b \sin^4(\alpha - \beta) \right)^c$$

for $\alpha \in [0, 2\pi]$

Models for higher-dimensional h

Assumption: $H : [0,\pi]^{d-2} \times [0, 2\pi) \to (0,\infty)$ is contin. differentiable,

$$\lim_{t \to 2\pi - 0} H(\alpha_1, \dots, \alpha_{d-2}, t) = H(\alpha_1, \dots, \alpha_{d-2}, 0),$$
$$\lim_{\alpha_j \to 0+0} H(\alpha) = \lim_{\alpha_j \to \pi - 0} H(\alpha) = \tilde{H}_j(\alpha_1, \dots, \alpha_{j-1}) \text{ for } j = 1, \dots, d-1$$

with a constant $\tilde{H}_1$ and appropriate functions $\tilde{H}_j : [0,\pi]^{j-1} \to (0,\infty), j \ge 2$

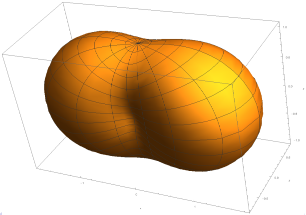
- Model function H_η with parameter $\eta \in \Theta_2 \subset \mathbb{R}^{q_2}$

Q_η orthogonal matrix describing the rotation

$$h_\eta(x) = \|x\|_2 \, H_\eta(T(\|Q_\eta x\|_2^{-1} \, Q_\eta x))$$

Example: contour plot of the density $\to$

$$H(\alpha) = \left(1 + \sum_{j=1}^{d-1} a_j \prod_{k=1}^{j} \sin^2 \alpha_k\right)^{-1}$$



Estimation for star-shaped distributions

$X_1, \ldots, X_n$ sample of random vectors having density

$$f_Y(x) = \left(C_0 \det(\Sigma)\right)^{-1} g\left(h(\Sigma^{-1}(x-\mu))\right), \forall x \in \mathbb{R}^p,$$

$\Sigma =$diag$(\sigma_1, \ldots, \sigma_p)$

- for the sake of simplicity, assumption $h(-x) = h(x) \; \forall x \in \mathbb{R}^p \rightarrow \mathbb{E}X = \mu$

- **Estimation problems**

1) Estimation of μ, σ

2) Parametric estimation of h

3) Maximum-likelihood estimation of f

4) Semiparametric estimation of f: kernel density estimator for g

Estimators for μ, σ

$X_1, \ldots, X_n$ sample, $X_i = (X_i^{(1)}, \ldots X_i^{(d)})^T$ has density f,

$\mathbb{E}\,\|X\|_2 < +\infty$.

$\Longrightarrow \mathbb{E}X = \mu$, w.l.o.g. we can assume $\mathbb{V}(X_i) = \sigma_i^2$ for $i = 1, \ldots, n$.

- Estimator

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \hat{\sigma}_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^{(j)} - \hat{\mu}_j\right)^2}$$

where $\hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_d)^T, \hat{\Sigma} =$diag$(\hat{\sigma}_1, \ldots, \hat{\sigma}_d)$

standardisized sample items

$$\tilde{X}_i^{(j)} = \frac{1}{\hat{\sigma}_j}\left(X_i^{(j)} - \hat{\mu}_j\right), \ \tilde{X}_i = (\tilde{X}_i^{(1)}, \ldots, \tilde{X}_i^{(d)})^T.$$

Models for h - Parametric estimation

Model function $H = H_\eta$ with parameter $\eta \in \Theta_2 \subset \mathbb{R}^{q_2}$

- Density of spherical angle vect. $\Gamma = T\left(\left\|\Sigma^{-1}(X_i-\mu)\right\|_2^{-1}\Sigma^{-1}(X_i-\mu)\right)$:

$$f_\Gamma(\alpha) = C_0(\eta)^{-1} H_\eta(\alpha)^{-d} \prod_{k=1}^{d-2} \sin^{d-k-1}\alpha_k, \quad \alpha \in [0,\pi)^{d-1} \times [0,2\pi]$$

standardized sample $\tilde{X}_1, \ldots, \tilde{X}_n$,

spherical angle vectors $\Gamma_i = T\left(\left\|\tilde{X}_i\right\|_2^{-1}\tilde{X}_i\right)$

Maximizing the likelihood function by using the pseudo sample $\Gamma_1, \ldots, \Gamma_n$

$\rightarrow$ estimator $\hat{\eta}$

- Alternative: moment estimator for η

Maximum likelihood estimation

model functions g_τ, h_η with parameter $\tau \in \Theta_1 \subset \mathbb{R}^{q_1}, \eta \in \Theta_2 \subset \mathbb{R}^{q_2}$

▷ Sample $X_1, \ldots, X_n, X_i \sim CSS(g, h, \mu, \Sigma)$,

parameter $\theta = (\mu^T, \sigma^T, \eta^T, \theta^T)^T \in \Theta = \mathbb{R}^p \times SPD(\mathbb{R}^p) \times \Theta_1 \times \Theta_2$

- Loglikelihood function

$$\ln L(\theta) = \sum_{i=1}^{n} \ln g_\tau(h_\eta(\Sigma^{-1}(X_i - \mu))) - n \sum_{j=1}^{d} \ln \sigma_j - n \ln(C_0(\eta)).$$

Maximum-likelihood estimator

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \ln L(\theta)$$

Theorem.

Under regularity assumptions,

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathcal{N}(0, I(\delta)^{-1}) \text{ for } n \to \infty, I \ldots \text{information matrix}$$

Semiparametric estimator - transformation and sample

- **Transformation** $\psi : [0, \infty) \to [0, +\infty)$ as above

density χ of $Y = \psi(R), R = h(\Sigma^{-1}(X - \mu)), \Psi = \psi^{-1}$:

$$\chi(y) = \Psi(y)^{d-1} g(\Psi(y)) \cdot \Psi'(y) \text{ for } y \geq 0.$$

$$\Longrightarrow \qquad g(z) = z^{1-d}\ \psi'(z)\ \chi(\psi(z)) \text{ for } z \geq 0.$$

$0 < \lim_{z \to 0+} z^{1-d}\ \psi'(z) < +\infty$

- Sample $X_1, \ldots, X_n$, $\hat{h} = h_{\hat{\eta}}$

transformed sample $Y_{1n}, \ldots, Y_{nn}$,

$$Y_{in} = \psi(\hat{h}(\hat{\Sigma}^{-1}(X_i - \hat{\mu})))$$

Density estimator

$$Y_{in} = \psi(\hat{h}(\hat{\Sigma}^{-1}(X_i - \hat{\mu})))$$

- kernel density estimator for density χ of $Y = \psi(R)$:

$$\hat{\chi}_n(y) = n^{-1} b^{-1} \sum_{i=1}^{n} \left(k\left((y - Y_{in}) b^{-1}\right) + k\left((y + Y_{in}) b^{-1}\right)\right) \quad \text{for } y \geq 0,$$

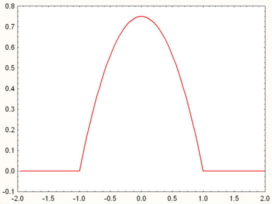
$b = b(n) \longrightarrow 0$ bandwidth

Assumptions: $k : \mathbb{R} \longrightarrow \mathbb{R}$ continuous kernel function, k' Lipschitz on $[-1, 1]$, $k(t) = 0$ for $t \in \mathbb{R} \backslash [-1, 1]$, $k(-t) = k(t)$ for $t \in [-1, 1]$,

$$\int_{-1}^{1} k(t)\,\mathrm{d}t = 1 \text{ and } \int_{-1}^{1} t^j k(t)\,\mathrm{d}t = 0$$

for $j : 0 < j < p \in \mathbb{Z}$, even $p \geq 2$

Epanechnikov kernel



Semiparametric density estimator

Assumptions:

$$\lim_{n\to\infty} b \ln \ln n = 0 \quad \text{and } b_0 \geq b \geq C_1 \cdot n^{-1/5}$$

with constants $b_0, C_1 > 0$.

The partial derivatives $G_1, \ldots, G_d$ of h exist and are bounded. Moreover, $x \rightsquigarrow \psi'(h(x))G_j(x)$ is Hölder continuous of order $\alpha > 0.2$ for each j.

- **Estimator**

$$\hat{g}_n(z) = z^{1-d} \; \psi'(z) \; \hat{\chi}_n(\psi(z)),$$

$$\hat{f}_n(x) = \left(\hat{C}_0 \det(\hat{\Sigma})\right)^{-1} \hat{g}_n \left(\hat{h}(\hat{\Sigma}^{-1}(x - \hat{\mu}))\right)$$

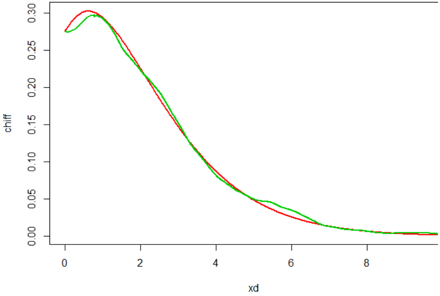
Density estimator - Simulated example

Simulation of the generalized q-norm-disribution, $q = 3, d = 2$,

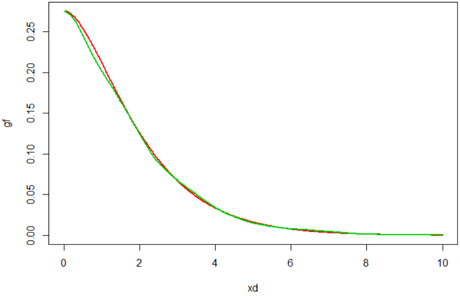
f_R modified exponential distribution, $n = 1000$ data points

green: estimator, red: true function

density χ of the transformed data



generator function g



Convergence of the density estimator

Suppose that $g^{(p)}$ exists, is bounded on $[0,\infty)$ for some even integer $p \geq 2$.

Theorem.

$\Longrightarrow$ *for any compact set D with $\mu \notin D$ and $n \to \infty$,*

$$\sup_{x \in D} \left| \hat{f}_n(x) - f(x) \right| = O\left(\sqrt{\ln n}(nb)^{-1/2} + b^p \right) \quad a.s.$$

case $\mu \in D$: b^p is replaced by $b^{1/d}$

- $\mu \notin D$: the same rate as for one-dimensional kernel density estimators, does not depend on d

LR 2016

Asymptotic normality of the density estimator

Assume that $g^{(p)}$ is continuous at $\tilde{x} := h(\Sigma^{-1}(x-\mu)), x \neq \mu$

Theorem.

$\lim_{n\to\infty} n^{1/(2p+1)} b = C_2$ *with a constant* $C_2 \geq 0$, $\Longrightarrow$ *for* $n \to \infty$

$$\sqrt{nb}\left(\hat{f}_n(x) - f(x)\right) \xrightarrow{d} \mathcal{N}(C_2^{(2p+1)/2}\Lambda(\tilde{x}), \bar{\sigma}^2(\tilde{x})).$$

where

$$\begin{aligned}
\bar{\sigma}^2(\tilde{x}) &= C_0^{-2}\tilde{x}^{1-d}\psi'(\tilde{x})g(\tilde{x})\int_{-1}^{1} k^2(t)\,dt, \\
\Lambda(\tilde{x}) &= C_0^{-1}\tilde{x}^{1-d}\psi'(\tilde{x})\frac{1}{p!}\chi^{(p)}(\psi(\tilde{x}))\int_{-1}^{1} t^p k(t)\,dt.
\end{aligned}$$

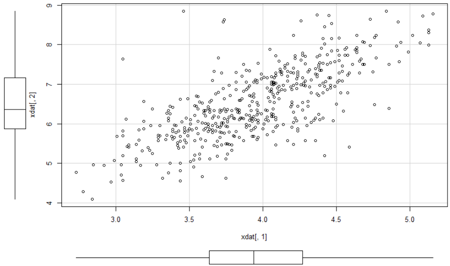
LR 2016

Data example

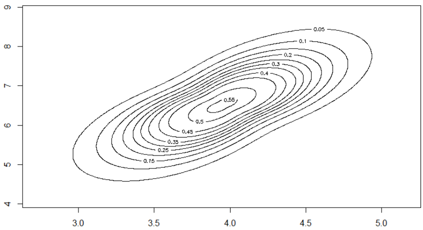
dataset 5 of Andrews and Herzberg (1985)

variables: yield of grain and straw

scatter plot of the data



contour plot of $\hat{\varphi}_n$, $b = 0.5$



moment estimator for a_1: $\hat{a}_1 = 0.83641$

Data example

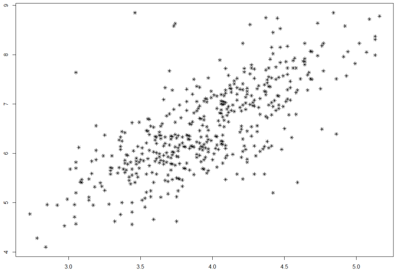
dataset 5 📖 Andrews/Herzberg 1985, variables: yield of grain and straw

model g: Pearson VII,

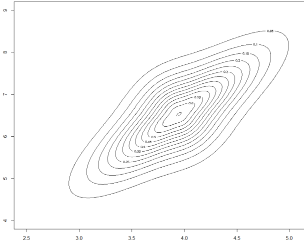
H class 1 with $c = 1$

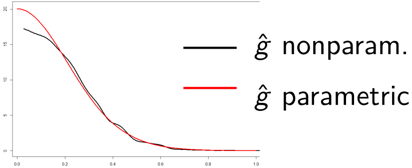
goodness-of-fit $\rho = 0.9846$

scatter plot of the data



contour plot of $\varphi_{\hat{g},\hat{h},\hat{\mu},\hat{\sigma}}$





Simulating the distribution

Theorem.

$R = h(\Sigma^{-1}(X-\mu))$ and $\Psi = T(\|\Sigma^{-1}(X-\mu)\|^{-1}\Sigma^{-1}(X-\mu))$ are independent with densities

$$f(r) = g(r)r^{d-1},\ \varphi_\Psi(\alpha) = \kappa^{-1}H(\alpha)^{-d}\prod_{k=1}^{d-2}\sin^{d-k-1}\alpha_k$$

- **algorithm** to simulate X:

1) Generate R with density f.

2) Generate Ψ (independent to R) with density φ_Ψ.

   Formulas for marginal densities of every dimension are available.

3) Evaluate $U = H(\Psi)^{-1}T^{-1}(\Psi)$ and $X = \mu + \Sigma R U$

Summary

- very general class of star-shaped distributions including distributions with non-convex contours

- parametric density estimators

- flexible semiparametric density estimators on the basis of a one-dimensional kernel estimator

- good performance, no dimensionality problems

• Tasks for future investigations:

- model validation

- skewed distributions

Literature

Liebscher, E. (2005). A semiparametric density estimator based on elliptical distributions, *J. Multivariate Analysis* **92** (2005), 205-225

Liebscher, E. (2023). Constructing models for spherical and elliptical densities. *Dependence Modeling, Special Issue: 10 years of Dependence Modeling*, 0111.

Liebscher, E.; W.-D. Richter (2016). Estimation of star-shaped distributions. *Risks*, special issue "Selected Papers from the 10th Tartu Conference on Multivariate Statistics" **4**(44), 1-37.

Liebscher, E.; W.-D. Richter (2020). Modelling with star-shaped distributions. *Dependence Modeling* **8**, 45-69.

Liebscher, E.; O. Okhrin (2023) Semiparametric estimation of the high-dimensional elliptical distribution. *J. Multivariate Analysis* **195** Issue C

Richter, W.-D. (2014). Geometric disintegration and star-shaped distributions. *Journal of Statistical Distributions and Applications* 2014, 1-20.

Richter, W.-D. (2015). Convex and Radially Concave Contoured Distributions. *Journal of Probability and Statistics*, Article ID 165468