NHH

Multivariate density estimation using the local Gaussian correlation

Building Bridges at Bislett

Håkon Otneim and Dag Tjøstheim

22 May 2017

The Pont Notre-Dame, 1756

The basic idea

NHH

- Let $X_1, \ldots, X_n$ be (possibly multivariate) stochastic variables having density $f(x)$.
- For some parametric family $\Psi = \{\psi(x, \theta) : \theta \in \Theta\}$ we may conclude that
  - $f(x) \in \Psi \Rightarrow$ estimate f by $\widehat{f}(x) = \psi(x, \widehat{\theta})$, or
  - $f(x) \notin \Psi \Rightarrow$ obtain $\widehat{f}(x)$ non-parametrically.

The basic idea

NHH

- Let $X_1, \ldots, X_n$ be (possibly multivariate) stochastic variables having density $f(x)$.
- For some parametric family $\Psi = \{\psi(x, \theta) : \theta \in \Theta\}$ we may conclude that
  - $f(x) \in \Psi \Rightarrow$ estimate f by $\widehat{f}(x) = \psi(x, \widehat{\theta})$, or
  - $f(x) \notin \Psi \Rightarrow$ obtain $\widehat{f}(x)$ non-parametrically.

One possible **bridge** between the parametric and non-parametric worlds is provided by Hjort & Jones (1996):

- Estimate $f(x)$ by fitting $\psi(x, \theta)$ *locally*, obtaining $\widehat{\theta} = \widehat{\theta}(x)$, and

$$\widehat{f}(x) = \psi\left(x, \widehat{\theta}(x)\right).$$

Local likelihood density estimation

The parameter is estimated for each x by maximizing the local likelihood function:

$$\widehat{\theta}(x) = \arg\max_{\theta} n^{-1} \sum_{i=1}^{n} K_h(X_i - x) \log \psi(X_i, \theta) - \int K_h(x - y) \psi(y, \theta) \, \mathrm{d}y,$$

where the smoothing parameter h is our location on the **bridge**:

- If $h \to 0$ (as $n \to \infty$), the estimator compares with classical non-parametric methods.
- If $h \to \infty$, the estimator appraches a global ML-fit $\psi(x, \widehat{\theta})$.

Local likelihood density estimation

NHH

The parameter is estimated for each x by maximizing the local likelihood function:

$$\widehat{\theta}(x) = \arg\max_{\theta} n^{-1} \sum_{i=1}^{n} K_h(X_i - x) \log \psi(X_i, \theta) - \int K_h(x - y) \psi(y, \theta) \, \mathrm{d}y,$$

where the smoothing parameter h is our location on the **bridge**:

- If $h \to 0$ (as $n \to \infty$), the estimator compares with classical non-parametric methods.
- If $h \to \infty$, the estimator appraches a global ML-fit $\psi(x, \widehat{\theta})$.

Choosing a bandwidth in between can thus be regarded as a *trade-off* between the two classical mindset - this is a semi-parametric estimator.

A problem

Suppose that we wish to estimate the p-variate probability density function $f(\boldsymbol{x})$ based on n samples $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.

The two traditional and straightforward methods both have serious problems in the multivariate case:

Parametric model: can we summarize the entire dependence structure between all p variables in just a few parameters?

Non-parametric kernel estimator: poor performance because of the *Curse of Dimensionality*.

A problem

It is not so easy to scale the local likelihood directly up to higher dimensions.

For example, consider a 3-dimensional local Gaussian fit. Instead of directly estimating

$$f(x_1, x_2, x_3),$$

we must rather estimate

$$\boldsymbol{\mu}(\boldsymbol{x}) = \begin{pmatrix} \mu_1(x_1, x_2, x_3) \\ \mu_2(x_1, x_2, x_3) \\ \mu_3(x_1, x_2, x_3) \end{pmatrix}, \quad \boldsymbol{\sigma}(\boldsymbol{x}) = \begin{pmatrix} \sigma_1(x_1, x_2, x_3) \\ \sigma_2(x_1, x_2, x_3) \\ \sigma_3(x_1, x_2, x_3) \end{pmatrix}$$

and

$$\boldsymbol{R}(\boldsymbol{x}) = \begin{pmatrix} 1 & \rho_{12}(x_1, x_2, x_3) & \rho_{13}(x_1, x_2, x_3) \\ \rho_{21}(x_1, x_2, x_3) & 1 & \rho_{23}(x_1, x_2, x_3) \\ \rho_{31}(x_1, x_2, x_3) & \rho_{32}(x_1, x_2, x_3) & 1 \end{pmatrix}$$

The simplification

Clearly, we need a simplification!

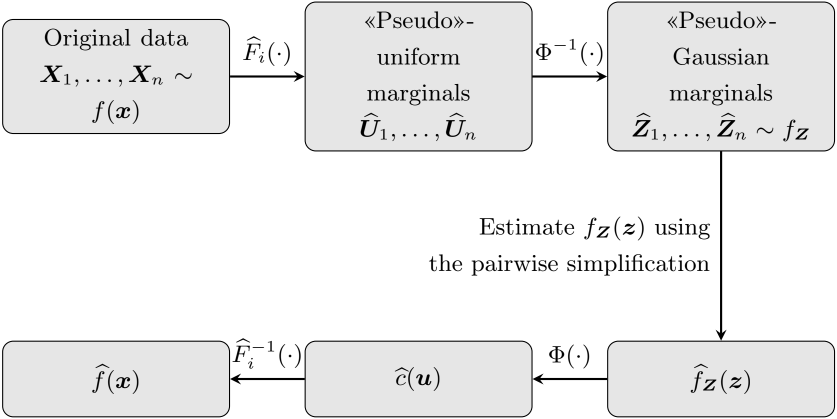
Assume that $f(\boldsymbol{x})$ has standard normal marginals.

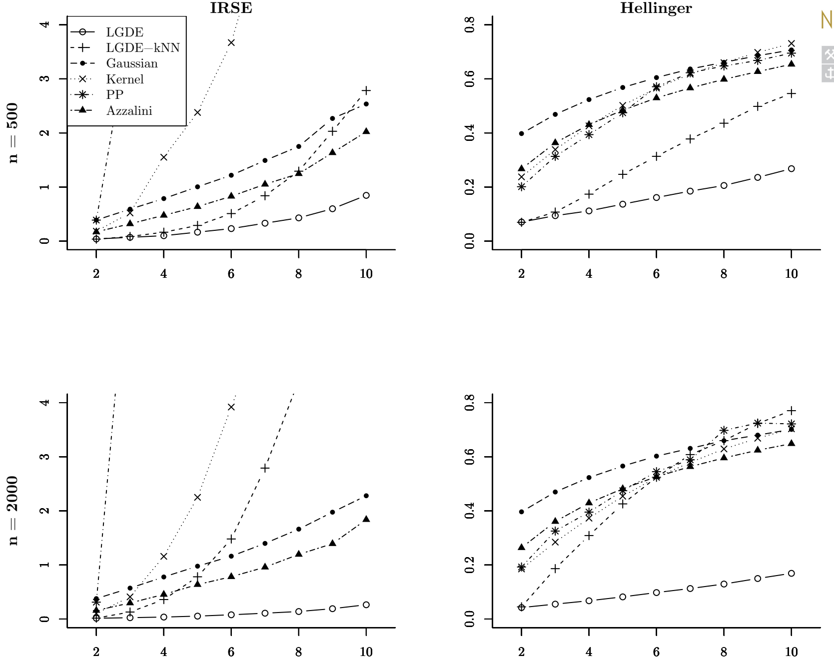
We can then fit the multivariate normal under the following restrictions:

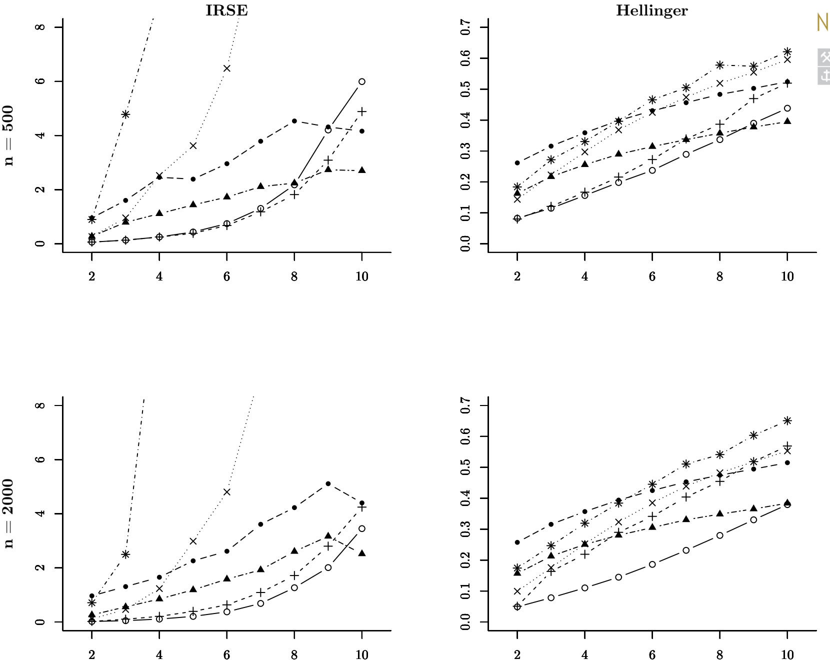
- Keep all local means and standard deviations fixed equal to 0 and 1 respectively.
- Estimate the local correlations *pairwise*:

$$\widehat{R}(\boldsymbol{z}) = \begin{pmatrix} 1 & \widehat{\rho}_{12}(z_1, z_2) & \cdots & \widehat{\rho}_{1p}(z_1, z_p) \\ \widehat{\rho}_{21}(z_1, z_2) & 1 & \cdots & \widehat{\rho}_{2p}(z_2, z_p) \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\rho}_{p1}(z_1, z_p) & \widehat{\rho}_{p2}(z_2, z_p) & \cdots & 1 \end{pmatrix}$$

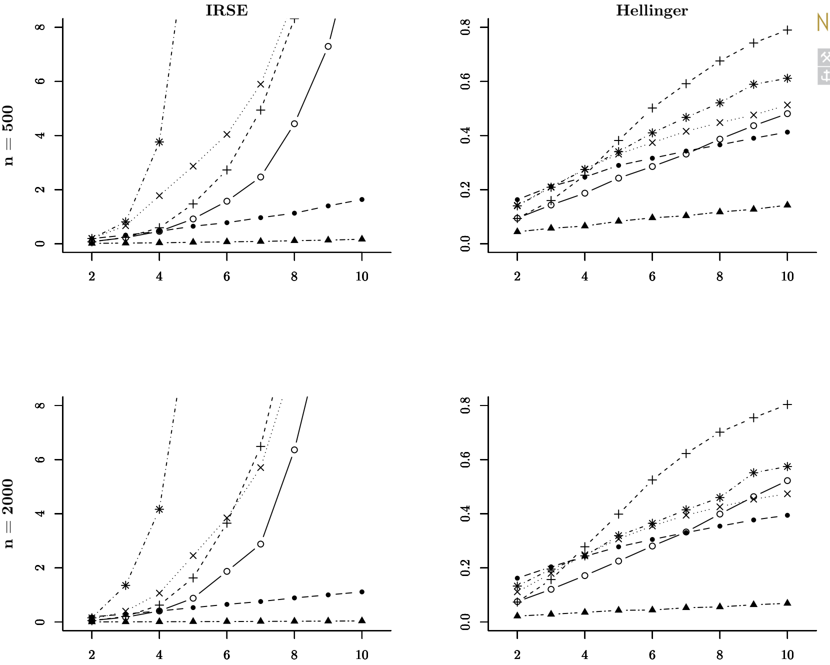
Algorithm for data that are not marginally st. normal




IRSE
Hellinger
n = 500
n = 2000
LGDE
LGDE−kNN
Gaussian
Kernel
PP
Azzalini
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
2
4
6
8
10



IRSE
Hellinger
n = 500
n = 2000



IRSE
Hellinger
n = 500
n = 2000

The conditional density

Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ be a stochastic vector with joint density $f(\boldsymbol{x}) = f(\boldsymbol{x}_1, \boldsymbol{x}_2)$. The conditional density of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is

$$f_{\boldsymbol{X}_1|\boldsymbol{X}_2=\boldsymbol{x}_2}(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f(\boldsymbol{x}_2)}.$$

The conditional density

Let $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2)$ be a stochastic vector with joint density $f(\boldsymbol{x}) = f(\boldsymbol{x}_1, \boldsymbol{x}_2)$. The conditional density of $\boldsymbol{X}_1$ given that $\boldsymbol{X}_2 = \boldsymbol{x}_2$ is

$$f_{\boldsymbol{X}_1|\boldsymbol{X}_2=\boldsymbol{x}_2}(\boldsymbol{x}_1|\boldsymbol{x}_2) = \frac{f(\boldsymbol{x}_1, \boldsymbol{x}_2)}{f(\boldsymbol{x}_2)}.$$

Existing estimation methods can be roughly classified as (not a complete list!)

- **Kernel based:** Rosenblatt (1969), Bashtannyk & Hyndman (2001), Li & Racine (2007), Faugeras (2009).
- **Semi-parametric:** Hyndman et. al. (1996), Fan et. al. (1996), Hyndman & Yao (2002), Fan & Yim (2004).

Very few methods are implemented in available software, and even fewer allow both explanatory and response variables to be vectors.

The conditional density

Let $\boldsymbol{Z} = (\boldsymbol{Z}_1, \boldsymbol{Z}_2)$ be multivariate normally distributed with expectation vector and covariance matrix

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \Sigma_{22} \end{pmatrix}.$$

Then $\boldsymbol{Z}_1 | \boldsymbol{Z}_2 = \boldsymbol{z}_2$ is multivariate normally distributed with

$$\begin{aligned} \boldsymbol{\mu}^* &= \boldsymbol{\mu_1} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{z}_2 - \boldsymbol{\mu}_2), \\ \boldsymbol{\Sigma}^* &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}. \end{aligned}$$

The conditional density

Because of our pairwise fits, the same is true for our *locally* Gausian fits:

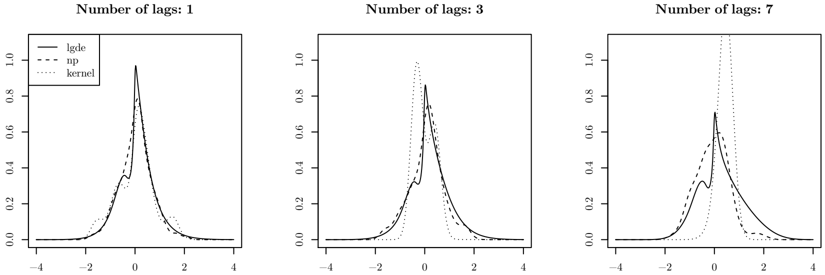
$$\frac{\psi\left(\boldsymbol{z}, \widehat{R}(\boldsymbol{z})\right)}{\psi\left(\boldsymbol{z}_2, \widehat{R}_{22}(\boldsymbol{z}_2)\right)} = \psi\left(\boldsymbol{z}_1, \widehat{\boldsymbol{\mu}}^*(\boldsymbol{z}), \widehat{\boldsymbol{\Sigma}}^*(\boldsymbol{z})\right),$$

where

$$\widehat{\boldsymbol{\mu}}^*(\boldsymbol{z}) = \widehat{\boldsymbol{R}}_{12}(\boldsymbol{z})\widehat{\boldsymbol{R}}_{22}^{-1}(\boldsymbol{z})\boldsymbol{z}_2,$$
$$\widehat{\boldsymbol{\Sigma}}^*(\boldsymbol{z}) = \widehat{\boldsymbol{R}}_{11}(\boldsymbol{z}) - \widehat{\boldsymbol{R}}_{12}(\boldsymbol{z})\widehat{\boldsymbol{R}}_{22}^{-1}(\boldsymbol{z})\widehat{\boldsymbol{R}}_{21}(\boldsymbol{z}).$$

An example

- Observe 500 daily log-returns from the S&P500 index.
  - Estimate $f(x_t|x_{t-1})$, $f(x_t|x_{t-1}, x_{t-2}, x_{t-3})$ and $f(x_t|x_{t-1}, x_{t-2}, x_{t-3}, x_{t-4}, x_{t-5}, x_{t-6}, x_{t-7})$



A quick overview

- For **jointly Gaussian variables** we can easily estimate their joint density by calculating the pairwise correlation coefficients.
- With the LGDE, we use the same strategy for general stochastic variables using the **local** correlation.

A quick overview

NHH

- For **jointly Gaussian variables** we can easily estimate their joint density by calculating the pairwise correlation coefficients.
- With the LGDE, we use the same strategy for general stochastic variables using the **local** correlation.

- For **jointly Gaussian variables** we can easily estimate conditional densities using the formulas from an earlier slide.
- We simply replace the correlation coefficient with the **local** correlation in the same expressions, and obtain an estimator for general conditional densities that performs very well.

Also in the same spirit

The same way of thinking has produced another useful result:

- **Jointly Gaussian variables** are independent *if and only if* the correlation between them is zero.
- In general, stochastic variables are independent *if and only if* the local Gaussian correlation is zero everywhere. (On the margnally standard normal scale)

Also in the same spirit

NHH

The same way of thinking has produced another useful result:

- **Jointly Gaussian variables** are independent *if and only if* the correlation between them is zero.
- In general, stochastic variables are independent *if and only if* the local Gaussian correlation is zero everywhere. (On the margnally standard normal scale)

So there seems to be another **bridge** here:

- Using the LGC we can formulate statistical properties in terms of the Gaussian distribution.

Current status

NHH

Another property that is true for **jointly Gaussian variables**: X and Y are *conditionally independent* given Z if and only if the *partial* correlation between X and Y is zero.

- Our version: X and Y are *conditionally independent* given Z if and only if the local partial correlation is zero everywhere.
- This can, for example, be used to test for conditional independence.