Smooth and flexible skew-symmetric distributions using B-splines and penalties

Patrizio Frederic

Dipartimento di Economia Politica,
Università di Modena e Reggio Emilia,
patrizio.frederic@unimore.it

ERCIM 2009
Data from a chemical distillation process: 3101 measures (21.53 days) of the temperature of the top of a distillation column. The process was supposed to be in control, then engineers would expect $Temp = 133.125 + \epsilon$, and $\epsilon \sim N(0, \sigma^2)$. 

Is it possible to model departures from symmetry of the distributions?
Data from a chemical distillation process: 3101 measures (21.53 days) of the temperature of the top of a distillation column. The process was supposed to be in control, then engineers would expect $Temp = 133.125 + \epsilon$, and $\epsilon \sim N(0, \sigma^2)$. 

---

**temerature of the top of the coulmn 67**

**Normal Q–Q Plot**

---

P. Frederic (univ. of Modena)  B-spline Skew Symmetric Distributions  ERCIM 2009 2 / 44
Data from a chemical distillation process: 3101 measures (21.53 days) of the temperature of the top of a distillation column. The process was supposed to be in control, then engineers would expect $Temp = 133.125 + \epsilon$, and $\epsilon \sim N(0, \sigma^2)$.

Is it possible to model departures from symmetry of the distributions?
Introduction to Skew-Symmetric models.
Introduction to Skew-Symmetric models.

B-spline Skew-Symmetric distributions.
Table of Contents

- Introduction to Skew-Symmetric models.
- B-spline Skew-Symmetric distributions.
- Numerical and real data examples.
Table of Contents

- Introduction to Skew-Symmetric models.
- B-spline Skew-Symmetric distributions.
- Numerical and real data examples.
- Discussion.
Proposition 1. Every $p$-dimension pdf $f$, $\forall \xi \in \mathbb{R}^p$, admits a unique SS around $\xi$ representation:

$$f(x) = 2f_\xi(x - \xi)\pi_\xi(x - \xi)$$  \hspace{1cm} (1)

where

- $f_\xi$ is a symmetric around 0 pdf:
  $$f_\xi(x) = f_\xi(-x)$$

- $\pi_\xi$ is the skewing function, thus:
  $$0 \leq \pi_\xi(x) = 1 - \pi_\xi(-x) \leq 1, \forall x$$
Proposition 1. Every $p$-dimension pdf $f$, $\forall \xi \in \mathbb{R}^p$, admits a unique SS around $\xi$ representation:

$$f(x) = 2f_\xi(x - \xi)\pi_\xi(x - \xi)$$

where

- $f_\xi$ is a symmetric around 0 pdf: $f_\xi(x) = f_\xi(-x)$

Proposition 2. Let $f_S$ be a symmetric around 0 pdf, and let $\pi$ be a skewing function then:

$$f(x) = 2f_S(x)\pi(x)$$

then $f_{SS}$ is a valid pdf.

Definition. Let $f_S(\cdot|\nu)$ be a parametric, symmetric around 0 pdf, with shape $\nu$, let $\xi$ be a location, and $\omega$ a scale matrix. We write $X \sim SS(\xi,\omega,\nu,\pi)$ iff $X$ has pdf:

$$f_{SS}(x|\xi,\omega,\nu,\pi) = 2 \det(\omega)^{-1}f_S(z|\nu)\pi(z), \quad z = \omega^{-1}(x - \xi)$$
Proposition 1. Every $p$-dimension pdf $f$, $\forall \xi \in \mathbb{R}^p$, admits a unique SS around $\xi$ representation:

$$f(x) = 2f_\xi(x - \xi)\pi_\xi(x - \xi)$$  \hspace{1cm} (1)

where

- $f_\xi$ is a symmetric around 0 pdf: $f_\xi(x) = f_\xi(-x)$
- $\pi_\xi$ is the skewing function, thus: $0 \leq \pi_\xi(-x) = 1 - \pi_\xi(x) \leq 1$, $\forall x$
**Proposition 1.** Every $p$-dimension pdf $f$, $\forall \xi \in \mathbb{R}^p$, admits a unique SS around $\xi$ representation:

$$f(x) = 2f_\xi(x - \xi)\pi_\xi(x - \xi)$$  \hspace{1cm} (1)$$

where

- $f_\xi$ is a symmetric around 0 pdf: $f_\xi(x) = f_\xi(-x)$
- $\pi_\xi$ is the skewing function, thus: $0 \leq \pi_\xi(-x) = 1 - \pi_\xi(x) \leq 1$, $\forall x$

**Proposition 2.** Let $f_S$ be a symmetric around 0 pdf, and let $\pi$ be a skewing function then:

$$f(x) = 2f_S(x)\pi(x),$$  \hspace{1cm} (2)$$

then $f_{SS}$ is a valid pdf.
Proposition 1. Every $p$-dimension pdf $f$, $\forall \xi \in \mathbb{R}^p$, admits a unique SS around $\xi$ representation:

$$f(x) = 2f_\xi(x - \xi)\pi_\xi(x - \xi)$$

(1)

where

- $f_\xi$ is a symmetric around 0 pdf: $f_\xi(x) = f_\xi(-x)$
- $\pi_\xi$ is the skewing function, thus: $0 \leq \pi_\xi(-x) = 1 - \pi_\xi(x) \leq 1$, $\forall x$

Proposition 2. Let $f_S$ be a symmetric around 0 pdf, and let $\pi$ be a skewing function then:

$$f(x) = 2f_S(x)\pi(x),$$

(2)

then $f_{SS}$ is a valid pdf.

Definition. Let $f_S(\cdot|\nu)$ be a parametric, symmetric around 0 pdf, with shape $\nu$, let $\xi$ be a location, and $\omega$ a scale matrix. We write $X \sim SS(\xi, \omega, \nu, \pi)$ iff $X$ has pdf:

$$f_{SS}(x|\xi, \omega, \nu, \pi) = 2 \det(\omega)^{-1}f_S(z|\nu)\pi(z), \quad z = \omega^{-1}(x - \xi)$$

(3)
An Example of SS family

Skewing function and symm. pdf

Resulting pdf

pnnorm(B %*% initb)
−6 0 6
0.0 0.2 0.4 0.6 0.8 1.0
0 0.08 0.16 0.24 0.32 0.4

Skewing function and symm. pdf

−5 0 5
0.0 0.1 0.2 0.3 0.4 0.5 0.6

x

dgsst(x, xi, omega, ni, odd)

Resulting pdf

x
SS examples for fixed \( f_S \) and \( \pi \)

Let \( H \) be any invertible, symmetric in 0 cdf. \( H \) is a one-to-one map from the space of odd functions to the space of skewing functions:

\[
\pi(x) = H(\eta(x)), \quad \eta \text{ odd}
\]

- **skew normal (Azzalini 1985)**
  - \( f_S \) standard normal, \( \pi(x) = \Phi(\alpha'x) \)
  - \( \alpha \in \mathbb{R}^p \), \( \Phi \) standard normal cdf
Let $H$ be any invertible, symmetric in 0 cdf. $H$ is a one-to-one map from the space of odd functions to the space of skewing functions:

$$\pi(x) = H(\eta(x)), \quad \eta \text{ odd}$$

- **skew normal (Azzalini 1985)**
  - $f_S$ standard normal, $\pi(x) = \Phi(\alpha'x)$
  - $\alpha \in \mathbb{R}^p$, $\Phi$ standard normal cdf
- **skew-t (Azzalini and Capitanio 2003)**
  - $f_S$ t-distribution pdf with $\nu$ df, $\pi(x) = T(\alpha'x)$
  - $\alpha \in \mathbb{R}^p$, $T$ t-distribution cdf
SS examples for fixed $f_S$ and $\pi$

Let $H$ be any invertible, symmetric in 0 cdf. $H$ is a one-to-one map from the space of odd functions to the space of skewing functions:

$$\pi(x) = H(\eta(x)), \quad \eta \text{ odd}$$

- **skew normal (Azzalini 1985)**
  
  $f_S$ standard normal, $\pi(x) = \Phi(\alpha'x)$
  
  $\alpha \in \mathbb{R}^p$, $\Phi$ standard normal cdf

- **skew-t (Azzalini and Capitanio 2003)**
  
  $f_S$ t-distribution pdf with $\nu$ df, $\pi(x) = T(\alpha'x)$
  
  $\alpha \in \mathbb{R}^p$, $T$ t-distribution cdf

- **Generalized Skew-Elliptical (Genton and Loperfido 2003)**
  
  $f_S$ any elliptical contoured in zero pdf, $\pi$ any skewing function
SS examples for fixed $f_S$ and $\pi$

Let $H$ be any invertible, symmetric in 0 cdf. $H$ is a one-to-one map from the space of odd functions to the space of skewing functions:

$$\pi(x) = H(\eta(x)), \quad \eta \text{ odd}$$

- **skew normal (Azzalini 1985)**
  $f_S$ standard normal, $\pi(x) = \Phi(\alpha'x)$
  $\alpha \in \mathbb{R}^p$, $\Phi$ standard normal cdf

- **skew-t (Azzalini and Capitanio 2003)**
  $f_S$ t-distribution pdf with $\nu$ df, $\pi(x) = T(\alpha'x)$
  $\alpha \in \mathbb{R}^p$, $T$ t-distribution cdf

- **Generalized Skew-Elliptical (Genten and Loperfido 2003)**
  $f_S$ any elliptical contoured in zero pdf, $\pi$ any skewing function

- **Flexible Skew-Symmetric (FSS, Ma and Genten 2005)**
  $f_S$ any symmetric in zero pdf, $\pi = H(P_k(x))$
  $P_K(x)$ $K$-degree, odd polynomial function.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$. 

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use of B-spline functions for modeling $\eta$. 

Setup a probabilistic framework.

Characterize odd B-splines, and define B-spline-based SS (BSS).

Proof the class of BSS is as large as the class of FSS.

Setup an inferential framework.

Propose an estimation method.

Model selection criteria.

Setup a computational framework.

Setup fast and efficient computations.

Provide a software.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use B-spline functions for modeling $\eta$. 
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use B-spline functions for modeling $\eta$.
- Setup a probabilistic framework.
- Proposition an estimation method.
- Model selection criteria.
- Setup a inferential framework.
- Setup fast and efficient computations.
- Provide a software.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use B-spline functions for modeling $\eta$.

- Setup a probabilistic framework.
  - Characterize odd B-splines, and define B-spline based SS (BSS)
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use B-spline functions for modeling $\eta$.

- Setup a probabilistic framework.
  - Characterize odd B-splines, and define B-spline based SS (BSS)
  - Proof the class of BSS is as large as the class of FSS.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use B-spline functions for modeling $\eta$.

- Setup a probabilistic framework.
  - Characterize odd B-splines, and define B-spline based SS (BSS)
  - Proof the class of BSS is as large as the class of FSS.
- Setup a inferential framework.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use B-spline functions for modeling $\eta$.

- Setup a probabilistic framework.
  - Characterize odd B-splines, and define B-spline based SS (BSS)
  - Proof the class of BSS is as large as the class of FSS.
- Setup a inferential framework.
  - Propose an estimations method.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

**In this work** we propose the use B-spline functions for modeling $\eta$.

- Setup a probabilistic framework.
  - Characterize odd B-splines, and define B-spline based SS (BSS).
  - Proof the class of BSS is as large as the class of FSS.
- Setup a inferential framework.
  - Propose an estimations method.
  - Model selection criteria.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use B-spline functions for modeling $\eta$.

- Setup a probabilistic framework.
  - Characterize odd B-splines, and define B-spline based SS (BSS)
  - Proof the class of BSS is as large as the class of FSS.
- Setup a inferential framework.
  - Propose an estimations method.
  - Model selection criteria.
- Setup a computational framework.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use B-spline functions for modeling $\eta$.

- Setup a probabilistic framework.
  - Characterize odd B-splines, and define B-spline based SS (BSS)
  - Proof the class of BSS is as large as the class of FSS.
- Setup a inferential framework.
  - Propose an estimations method.
  - Model selection criteria.
- Setup a computational framework.
  - Setup fast and efficient computations.
Ma and Genton (2005) prove that under mild regularity conditions FSS is dense in SS, under the $L_\infty$.

Polynomials are computationally unstable also for moderately small values of the degree.

In this work we propose the use B-spline functions for modeling $\eta$.

- Setup a probabilistic framework.
  - Characterize odd B-splines, and define B-spline based SS (BSS)
  - Proof the class of BSS is as large as the class of FSS.
- Setup a inferential framework.
  - Propose an estimations method.
  - Model selection criteria.
- Setup a computational framework.
  - Setup fast and efficient computations.
  - Provide a software.
Univariate splines, and B-spline functions

- A polynomial spline \( s \in \mathcal{C}_{d-1} \), is a piecewise polynomial function with fixed degree \( d \) defined on break points \( t_N = (a = t_1 < \ldots < t_N = b) \), \( a \in \mathbb{R}, b \in \mathbb{R} \).
A polynomial spline $s \in \mathcal{C}_{d-1}$, is a piecewise polynomial function with fixed degree $d$ defined on break points $t_N = (a = t_1 < ... < t_N = b)$, $a \in \mathbb{R}, b \in \mathbb{R}$.

Odd degree spline functions have excellent numerical proprieties.
Univariate splines, and B-spline functions

- A polynomial spline \( s \in C_{d-1} \), is a piecewise polynomial function with fixed degree \( d \) defined on break points \( t_N = (a = t_1 < \ldots < t_N = b) \), \( a \in \mathbb{R}, b \in \mathbb{R} \).
- Odd degree spline functions have excellent numerical properties.
- B-splines (De Boor 1974) provide a well-conditioned, numerically robust representation for splines.
A polynomial spline \( s \in \mathcal{C}_{d-1} \), is a piecewise polynomial function with fixed degree \( d \) defined on break points \( t_N = (a = t_1 < ... < t_N = b) \), \( a \in \mathbb{R}, b \in \mathbb{R} \).

Odd degree spline functions have excellent numerical proprieties.

B-splines (De Boor 1974) provide a well-conditioned, numerically robust representation for splines.

Given a set of knots \( t_N \), and a degree \( d \), B-spline are \( M = N - (d + 1) \), positive, hat-shaped, polynomial splines of degree \( d \)

\[
B_m : \mathbb{R} \rightarrow \mathbb{R}, \quad m = 1, ..., M
\]
Univariate splines, and B-spline functions

- A polynomial spline \( s \in \mathcal{C}_{d-1} \), is a piecewise polynomial function with fixed degree \( d \) defined on break points \( t_N = (a = t_1 < \ldots < t_N = b) \), \( a \in \mathbb{R}, b \in \mathbb{R} \).
- Odd degree spline functions have excellent numerical properties.
- B-splines (De Boor 1974) provide a well-conditioned, numerically robust representation for splines.
- Given a set of knots \( t_N \), and a degree \( d \), B-spline are \( M = N - (d + 1) \), positive, hat-shaped, polynomial splines of degree \( d \)

\[ B_m : \mathbb{R} \to \mathbb{R}, \quad m = 1, \ldots, M \]

- B-splines \( \mathbf{B}(x) = (B_m(x))_{m=1,\ldots,M} \) forms a basis for splines:

\[ \mathcal{B} = \{ h(x) = \mathbf{B}(x)\beta, \beta \in \mathbb{R}^M \} \]
B-Splines

P. Frederic (univ. of Modena)

B-spline Skew Symmetric Distributions

ERCIM 2009
B-Splines

\[ 2 + Bx \%*% be1 \]

P. Frederic (univ. of Modena)  B-spline Skew Symmetric Distributions  ERCIM 2009  10 / 44
Odd B-spline functions

- Let $d$ be an odd number (a common choice is $d = 3$). Let $N \in \mathbb{N}$, $a \in \mathbb{R}$, and let $t_{a,N}$ be the sequence

$$t_{a,N} = (-a = -t_N, -t_{N-1}, ..., -t_1, t_1, ..., t_{N-1}, t_N = a).$$
Let $d$ be an odd number (a common choice is $d = 3$). Let $N \in \mathbb{N}$, $a \in \mathbb{R}$, and let $t_{a,N}$ be the sequence

$$t_{a,N} = (-a = -t_N, -t_{N-1}, ..., -t_1, t_1, ..., t_{N-1}, t_N = a).$$

Resulting number of $d$-degree B-spline basis functions is

$$2M = 2(N - (d + 1)/2): M \text{ B-splines whose first knots is negative, and } M \text{ B-splines whose last knots if positive}.$$
Odd B-spline functions

- Let $d$ be an odd number (a common choice is $d = 3$). Let $N \in \mathbb{N}$, $a \in \mathbb{R}$, and let $t_{a,N}$ be the sequence

  $$t_{a,N} = (-a = -t_N, -t_{N-1}, ..., -t_1, t_1, ..., t_{N-1}, t_N = a).$$

- Resulting number of $d$-degree B-spline basis functions is $2M = 2(N - (d + 1)/2)$: $M$ B-splines whose first knots is negative, and $M$ B-splines whose last knots if positive.

- **NOTE** Given such a setting, every function $\eta_B \in \mathcal{B}$ can be written as

  $$\eta_B(x) = B(x)\beta = \sum_{m=-M}^{-1} B_m(x)\beta_m + \sum_{m=1}^{M} B_m(x)\beta_m$$
Odd B-spline functions

- Let $d$ be an odd number (a common choice is $d = 3$). Let $N \in \mathbb{N}$, $a \in \mathbb{R}$, and let $t_{a,N}$ be the sequence

  $$t_{a,N} = (-a = -t_N, -t_{N-1}, ..., -t_1, t_1, ..., t_{N-1}, t_N = a).$$

- Resulting number of $d$-degree B-spline basis functions is

  $$2M = 2(N - (d + 1)/2): \ M \text{ B-splines whose first knots is negative, and } M \text{ B-splines whose last knots if positive}$$

- **NOTE** Given such a setting, every function $\eta_B \in \mathcal{B}$ can be written as

  $$\eta_B(x) = B(x)\beta = \sum_{m=-M}^{-1} B_m(x)\beta_m + \sum_{m=1}^{M} B_m(x)\beta_m$$

- **Theorem 1** Let $\eta_B \in \mathcal{B}$ then $\eta_B$ is odd iff

  $$\beta_{-m} = -\beta_m, m = 1, ..., M$$
Odd B-splines

\[ b\text{approx}(x, \beta) \]

P. Frederic (univ. of Modena)
BSS Define the B-spline Skew-Symmetric Models by:

\[ f_{BSS}(x|\xi, \omega, \nu, \beta) = 2\omega^{-1} f_S(z|\nu) H(\eta_B(z)) \]  

where \( \eta_B \) is a member of \( \mathcal{B}_{\text{odd}} = \{ \eta(x) = B(x)\beta, \eta \text{ odd} \} \).
**BSS** Define the B-spline Skew-Symmetric Models by:

\[ f_{\text{BSS}}(x|\xi, \omega, \nu, \beta) = 2\omega^{-1} f_S(z|\nu) H(\eta_B(z)) \]  

(4)

where \( \eta_B \) is a member of \( \mathcal{B}_{\text{odd}} = \{ \eta(x) = B(x)\beta, \eta \text{ odd} \} \).

**Some special cases**
B-spline Skew-Symmetric Models

**BSS** Define the B-spline Skew-Symmetric Models by:

\[
f_{BSS}(x|\xi, \omega, \nu, \beta) = 2\omega^{-1} f_S(z|\nu) H(\eta_B(z))
\]

where \( \eta_B \) is a member of \( \mathcal{B}_{\text{odd}} = \{ \eta(x) = \mathbf{B}(x)\beta, \eta \text{ odd} \} \).

**Some special cases**

- B-SS normal distribution (\( M + 2 \) parameters)
  \[
  2\phi(z) H(\eta_B(z)), \quad \phi \text{ standard normal pdf}
  \]
B-spline Skew-Symmetric Models

- BSS Define the B-spline Skew-Symmetric Models by:
  \[
  f_{BSS}(x|\xi, \omega, \nu, \beta) = 2\omega^{-1} f_S(z|\nu) H(\eta_B(z))
  \] (4)
  
  where \( \eta_B \) is a member of \( B_{\text{odd}} = \{ \eta(x) = B(x)\beta, \eta \text{ odd} \} \).

- Some special cases
  - B-SS normal distribution (\( M + 2 \) parameters)
    \[
    2\phi(z) H(\eta_B(z)), \quad \phi \text{ standard normal pdf}
    \]
  - B-SS t distribution (\( M + 3 \) parameters)
    \[
    2f_\nu(z) H(\eta_B(z)), \quad f_\nu \text{ t pdf with } \nu \text{ df}
    \]
B-spline Skew-Symmetric Models

- **BSS** Define the B-spline Skew-Symmetric Models by:

  \[ f_{\text{BSS}}(x|\xi, \omega, \nu, \beta) = 2\omega^{-1} f_S(z|\nu) H(\eta_B(z)) \]  

  where \( \eta_B \) is a member of \( \mathcal{B}_{\text{odd}} = \{ \eta(x) = B(x)\beta, \eta \text{ odd} \} \).

- **Some special cases**
  - B-SS normal distribution (\( M + 2 \) parameters)

    \[ 2\phi(z) H(\eta_B(z)), \quad \phi \text{ standard normal pdf} \]

  - B-SS t distribution (\( M + 3 \) parameters)

    \[ 2f_\nu(z) H(\eta_B(z)), \quad f_\nu \text{ t pdf with } \nu \text{ df} \]

- **Theorem 2** For every precision \( \epsilon > 0 \), it exists a real value \( a \) and a sequence of knots \( t_{a,N} \) such that:

  \[ ||f_{\text{SS}} - f_{\text{BSS}}||_\infty < \epsilon \]

  (under mild regularity conditions).
Let $x_S$ be an $n$-dimensional sample from (4), $x_S = (x_1, \ldots, x_n)$. The log likelihood is:

$$\ell(\theta, \beta) = \text{constant} - n \log \omega + \sum_{i=1}^{n} \log f_S(z_i|\nu) + \sum_{i=1}^{n} \log H(\eta_B(z_i)),$$

where $\theta = (\xi, \omega, \nu)$, and $z_i = \omega^{-1}(x_i - \xi)$. 

In order to avoid over smoothing we proposed a penalized estimation criteria by minimizing:

$$\hat{\theta}, \hat{\beta} = \arg \min_{\theta, \beta} -\ell(\theta, \beta) + \lambda P(\beta),$$

where $\lambda > 0$ is a tuning parameter.
Let $x_S$ be an $n$-dimensional sample from (4), $x_S = (x_1, ..., x_n)$. The log likelihood is:

$$
\ell(\theta, \beta) = \text{constant} - n \log \omega + \sum_{i=1}^{n} \log f_S(z_i | \nu) + \sum_{i=1}^{n} \log H(\eta_B(z_i)),
$$

where $\theta = (\xi, \omega, \nu)$, and $z_i = \omega^{-1}(x_i - \xi)$.

In order to avoid over smoothing we proposed a penalized estimation criteria by minimizing of the following

$$(\hat{\theta}, \hat{\beta}) = \arg\min_{\theta, \beta} - \ell(\theta, \beta) + \lambda P(\beta)$$

where $\lambda > 0$ is a tuning parameter.
We consider the usual roughness penalty and its B-spline approximation (Eiler, and Marx (1996))

\[ P(\beta) = \int_{-a}^{a} \left( \frac{d^2 \eta_B(t)}{dt^2} \right)^2 dt \approx \beta' D_2' D_2 \beta \]

where \( D_2 \) is the second order difference matrix.
Penalization and Complexity of the model

- We consider the usual roughness penalty and its B-spline approximation (Eiler, and Marx (1996))

\[ P(\beta) = \int_{-a}^{a} \left( \frac{d^2 \eta_B(t)}{dt^2} \right)^2 dt \approx \beta' D'_2 D_2 \beta \]

where \( D_2 \) is the second order difference matrix.

- We approximate the dimension of the model by

\[ df \approx \text{tr}((B(z_S)'B(z_S) + \lambda D'_2 D_2)^{-1}B(z_S)'B(z_S)) + 3 \]

where \( B(z_S) \in \mathbb{R}^{n \times M} \), \( z_S = \omega^{-1}(x_S - \xi) \).
Penalization and Complexity of the model

- We consider the usual roughness penalty and its B-spline approximation (Eiler, and Marx (1996))

\[ P(\beta) = \int_{-a}^{a} \left( \frac{d^2 \eta_B(t)}{dt^2} \right)^2 dt \approx \beta' D_2' D_2 \beta \]

where \( D_2 \) is the second order difference matrix.

- We approximate the dimension of the model by

\[ df \approx \text{tr}((B(z_S)'B(z_S) + \lambda D_2' D_2)^{-1} B(z_S)'B(z_S)) + 3 \]

where \( B(z_S) \in \mathbb{R}^{n \times M} \), \( z_S = \omega^{-1}(x_S - \xi) \).

- Selection of \( \lambda \) is made by minimizing Information Criteria (IC) such as AIC, BIC, etc.

\[ AIC = -\ell(\hat{\theta}, \hat{\beta}) + 2 \cdot df \]

\[ BIC = -\ell(\hat{\theta}, \hat{\beta}) + \sqrt{n} \cdot df \]
Beside its appealing SS distributions have severe computational issues (Azzalini and Genton (2008)) such as:

- Non-log-concavity,
- Possible multi-modality,
- In general SS distributions not belong to the exponential family.
- In general there not exists closed forms for computing minimal sufficient statistics.
Beside its appealing SS distributions have severe computational issues (Azzalini and Genton (2008)) such as:

- non-log-concavity,
Beside its appealing SS distributions have severe computational issues (Azzalini and Genton (2008)) such as:
- non-log-concavity,
- possible multi-modality,
Computational Issues

• Beside its appealing SS distributions have severe computational issues (Azzalini and Genton (2008)) such as:
  • non-log-concavity,
  • possible multi-modality,
• In general SS distributions not below to the exponential family.
Beside its appealing SS distributions have severe computational issues (Azzalini and Genton (2008)) such as:

- non-log-concavity,
- possible multi-modality,

In general SS distributions not below to the exponential family.

In general there not exists closed forms for computing minimal sufficient statistics.
Let \( \hat{f} \) be a non-parametric density estimation of \( f \) (eg Kernel).

We use proposition (1) to decompose \( \hat{f} \) by

\[
\hat{f}(x) = 2 \hat{f}_\xi(x - \xi) \hat{\pi}_\xi(x - \xi),
\]

for any \( \xi \).

Numerically find the 3-dim vector:

\[
(\tilde{\xi}, \tilde{\omega}, \tilde{\nu}) = \text{argmin}_{\xi, \omega, \nu} ||\omega - 1 f_S(\omega - 1(\cdot - \xi))|\nu - \hat{f}_\xi(\cdot - \xi)||
\]

where \( ||\cdot|| \) is a suitable norm (eg the \( L_\infty \), or the \( L_2 \) norm).

Use linear interpolation to fit \( \pi_{\tilde{\xi}} \) with B-splines

\[
\tilde{\beta} = \text{argmin}_\beta ||H - 1(\pi_{\tilde{\xi}}) - \eta_B||^2_2
\]

Use the vector \( (\tilde{\xi}, \tilde{\omega}, \tilde{\nu}, \tilde{\beta}) \) as starting point for a newton-like optimization algorithm to find the maximum penalized likelihood,
Computational Strategy (for BSS-\(t\))

1. Let \(\hat{f}\) be a non-parametric density estimation of \(f\) (eg Kernel).
2. We use proposition (1) to decompose \(\hat{f}\) by

\[
\hat{f}(x) = 2\hat{f}_\xi(x - \xi)\hat{\pi}_\xi(x - \xi),
\]

for any \(\xi\).
1. Let \( \hat{f} \) be a non-parametric density estimation of \( f \) (eg Kernel).

2. We use proposition (1) to decompose \( \hat{f} \) by

\[
\hat{f}(x) = 2\hat{f}_\xi(x - \xi)\hat{\pi}_\xi(x - \xi),
\]

for any \( \xi \).

3. Numerically find the 3-dim vector:

\[
(\tilde{\xi}, \tilde{\omega}, \tilde{\nu}) = \arg\min_{\xi, \omega, \nu} ||\omega^{-1}f_S(\omega^{-1}(\cdot - \xi)\mid \nu) - \hat{f}_\xi(\cdot - \xi)||
\]

where \( || \cdot || \) is a suitable norm (eg the \( L_\infty \), or the \( L_2 \) norm).
Computational Strategy (for BSS-$t$)

1. Let $\hat{f}$ be a non-parametric density estimation of $f$ (eg Kernel).

2. We use proposition (1) to decompose $\hat{f}$ by

$$\hat{f}(x) = 2\hat{f}_\xi(x - \xi)\hat{\pi}_\xi(x - \xi),$$

for any $\xi$.

3. Numerically find the 3-dim vector:

$$(\hat{\xi}, \hat{\omega}, \hat{\nu}) = \arg\min_{\xi, \omega, \nu} ||\omega^{-1} f_S(\omega^{-1}(\cdot - \xi)|\nu) - \hat{f}_\xi(\cdot - \xi)||$$

where $||\cdot||$ is a suitable norm (eg the $L_\infty$, or the $L_2$ norm).

4. Use linear interpolation to fit $\pi_\xi$ with B-splines

$$\tilde{\beta} = \arg\min_\beta ||H^{-1}(\pi_\xi) - \eta_B||_2^2$$
Let $\hat{f}$ be a non-parametric density estimation of $f$ (eg Kernel).

We use proposition (1) to decompose $\hat{f}$ by

$$\hat{f}(x) = 2\hat{f}_\xi(x - \xi)\hat{\pi}_\xi(x - \xi),$$

for any $\xi$.

Numerically find the 3-dim vector:

$$(\tilde{\xi}, \tilde{\omega}, \tilde{\nu}) = \arg\min_{\xi, \omega, \nu} \|\omega^{-1}f_S(\omega^{-1}(\cdot - \xi)|\nu) - \hat{f}_\xi(\cdot - \xi)\|$$

where $\| \cdot \|$ is a suitable norm (eg the $L_\infty$, or the $L_2$ norm).

Use linear interpolation to fit $\pi_\xi$ with B-splines

$$\tilde{\beta} = \arg\min_{\beta} \|H^{-1}(\pi_\xi) - \eta_B\|^2$$

Use the vector $(\tilde{\xi}, \tilde{\omega}, \tilde{\nu}, \tilde{\beta})$ as starting point for a newton-like optimization algorithm to find the maximum penalized likelihood,
Ma-Hart (2007), $f_S = f_t(\cdot | \nu = 10)$, $\pi = \Phi(\sin(3x))$, $\xi = 3$, $\nu = 1,$

Skewing function and symm. pdf

Resulting pdf

\[ p\text{norm}(B \cdot \text{initb}) \]

\[ d\text{gsst}(x, \xi, \omega, n, \text{odd}) \]
1 sample (n=100) and SS detection

Skewing function: kernel based (wide, black), B-approximate (red), Symmetric density: kernel based (blue), t pdf (gray)

hist, kernel (blue), bss-guess (dashed green), bss-maxlik (red)
10 simulations and the resulting estimated pdf (n=100)
The graph shows the density function (true pdf) of a B-spline skew symmetric distribution. The quantiles at 0.05 and 0.95, 0.25 and 0.75, and the median are indicated on the distribution.

The function is defined as `dgsst(x, xi = xi, omega = omega, ni = ni, odd = odd)`.
Frederic (2009), \( f_S = f_t(\cdot | \nu = 10), \pi = \Phi(\sin(-20x \exp(-x^2))) \), \( \xi = 3, \nu = 1 \),

Skewing function and symm. pdf

Resulting pdf
1 sample (n=100) and SS detection

Kernel density estimation: kernel based (wide, black), B-approximate (red)

Symmetric density: kernel based (blue), t pdf (gray)

Hist, kernel (blue), bss-guess (red)
10 simulations and the resulting estimated pdf (n=100)
true pdf
.05 and .95 quantile
.25 and .75 quantile
Median

dgsst(x, xi = xi, omega = omega, ni = ni, odd = odd)

P. Frederic (univ. of Modena)
B-spline Skew Symmetric Distributions

ERCIM 2009 25 / 44
The BSS package

The BSS package is an R package (R Development Core Team (2009)) for computing SS, FSS, and BSS distributions and SS B-spline penalized likelihood maximization.

The package consists in series of tools such as

1. random generator, and pdf of generic SS-t distribution;

An alpha release of the software is available at the address:
http://www.economia.unimore.it/frederic_patrizio/BSS/
The BSS package is an R package (R Development Core Team (2009)) for computing SS, FSS, and BSS distributions and SS B-spline penalized likelihood maximization.

The package consists in series of tools such as

1. random generator, and pdf of generic SS-t distribution;
2. maximum penalized likelihood for BSS models;

An alpha release of the software is available at the address: http://www.economia.unimore.it/frederic_patrizio/BSS/
The BSS package

The BSS package is an R package (R Development Core Team (2009)) for computing SS, FSS, and BSS distributions and SS B-spline penalized likelihood maximization.

The package consists in series of tools such as

1. random generator, and pdf of generic SS-t distribution;
2. maximum penalized likelihood for BSS models;
3. maximum likelihood for FSS models;

An alpha release of the software is available at the address: http://www.economia.unimore.it/frederic_patrizio/BSS/
The BSS package

The BSS package is an R package (R Development Core Team (2009)) for computing SS, FSS, and BSS distributions and SS B-spline penalized likelihood maximization.

The package consists in series of tools such as

1. random generator, and pdf of generic SS-t distribution;
2. maximum penalized likelihood for BSS models;
3. maximum likelihood for FSS models;
4. generic plot methods for various diagnostic of the model;

An alpha release of the software is available at the address: http://www.economia.unimore.it/frederic_patrizio/BSS/
The BSS package

The BSS package is an R package (R Development Core Team (2009)) for computing SS, FSS, and BSS distributions and SS B-spline penalized likelihood maximization.

The package consists in series of tools such as:

1. random generator, and pdf of generic SS-t distribution;
2. maximum penalized likelihood for BSS models;
3. maximum likelihood for FSS models;
4. generic plot methods for various diagnostic of the model;
5. Skewing detection functions;

An alpha release of the software is available at the address:
http://www.economia.unimore.it/frederic_patrizio/BSS/
Application to Control Processes Data (Frederic (2003)), plan of the site.

1° Dist. column

Synth. wash C1

Residuals boiler

2° Dist. column

Recover

Reflux Recycle

Reflux boiler

Output

Distillation

C2

P. Frederic (univ. of Modena) B-spline Skew Symmetric Distributions ERCIM 2009 27 / 44
Application to Control Processes Data

Kernel function: kernel based (wide, black), B-approximate
Symmetric density: kernel based (blue), t pdf (gray)

Hist, kernel (blue), bss-guess (red)
Histogram of cg67
Tensor products allow a natural extension of one-dimensional B-spline smoothing to two or more dimensions.
Tensor products allow a natural extension of one-dimensional B-spline smoothing to two or more dimensions.

Let \( x \in \mathbb{R}^k \) be a vector with coordinates \( x = \{x_1, ..., x_k\} \). For all \( j = 1, ..., k \) let \( B_{M_j}^{(j)} = \{B_1^{(j)}, ..., B_{M_j}^{(j)}\} \) be a \( M_j \) dimensional B-spline basis.
Tensor products allow a natural extension of one-dimensional B-spline smoothing to two or more dimensions.

Let $\mathbf{x} \in \mathbb{R}^k$ be a vector with coordinates $\mathbf{x} = \{x_1, \ldots, x_k\}$. For all $j = 1, \ldots, k$ let $\mathbf{B}_{M_j}^{(j)} = \{B_1^{(j)}, \ldots, B_{M_j}^{(j)}\}$ be a $M_j$ dimensional B-spline basis.

Tensor product of $\mathbf{B}_{M_1}^{(1)}, \ldots, \mathbf{B}_{M_k}^{(k)}$ represent a basis for a $k$-dimensional function space, where the generic element is:

$$h(\mathbf{x}) = \sum_{j_1=1}^{M_1} \cdots \sum_{j_k=1}^{M_k} B_{j_1}^{(1)}(x_1) \cdots B_{j_k}^{(k)}(x_k) \beta_{j_1,\ldots,j_k}$$

the dimension of the basis is $M^{(k)} = M_1 \cdot \ldots \cdot M_k$. The parameter $\text{vec}((\beta_{j_1,\ldots,j_k}) \in \mathbb{R}^{M^{(k)}}$ is the coefficient of the linear combination.
2dim B-spline odd surfaces

1 dim B-spline basis function

2 dim tensor B-spline basis function

random odd surface: contour plot

random odd surface: persp plot
Measurements on 100 genuine and 100 forged old Swiss 1000 franc, $X_1 =$ distance from the inner frame to the lower border, $X_2$ length of the diagonal of the bills.
SS models which can capture skewness, heavy tails and multimodality systematically.
SS models which can capture skewness, heavy tails and multimodality systematically.

BSS provide an computationally efficient setting for approximating general SS models.
- SS models which can capture skewness, heavy tails and multimodality systematically.
- BSS provide an computationally efficient setting for approximating general SS models.
- Numerical studies give evidence that our computation strategy for defining the starting point leads quite often to the global maximum.
SS models which can capture skewness, heavy tails and multimodality systematically.

BSS provide an computationally efficient setting for approximating general SS models.

Numerical studies give evidence that our computation strategy for defining the starting point leads quite often to the global maximum.

Tensor product of B-splines are the natural extension from 1-dimensional to more complex \( p \)-dimensional BSS models.
THANK YOU!
Appendix 6: Wang et alt. SS propositions

Let $f : \mathbb{R}^p \rightarrow \mathbb{R}^+$ be a pdf and $\xi$ be any point in $\mathbb{R}^p$. Then

$$f(x) = 2f_\xi(x - \xi)\pi_\xi(x - \xi)$$

where $f_\xi$ is a pdf, symmetric around 0, and $\pi_\xi$ is a skewing function. This representation is unique for any $\xi$, and

$$f_\xi(s) = \frac{f(\xi + s) + f(\xi - s)}{2}$$

$$\pi_\xi(s) = \frac{f(\xi + s)}{f(\xi + s) + f(\xi - s)}$$
Algorithm 1 Let \( Y \) be a random vector with pdf \( f_S(\cdot|\nu) \);
Algorithm 1 Let \( Y \) be a random vector with pdf \( f_S(\cdot | \nu) \);
Let \( U \) be a \([0,1]\) uniform random variable, independent of \( Y \);

Algorithm 2 Let \( W \) be a RV with cdf \( H \), independent of \( Y \);
Algorithm 1 Let $Y$ be a random vector with pdf $f_S(\cdot|\nu)$; 
Let $U$ be a [0,1] uniform random variable, independent of $Y$; 
Let $X$ be the random vector

$$X = \begin{cases} 
\omega Y + \xi, & \text{if } U \leq \pi(Y) \\
-\omega Y + \xi, & \text{if } U > \pi(Y)
\end{cases}, \text{ then } X \sim SS(\xi, \omega, \nu, \pi)$$
Algorithm 1 Let $\mathbf{Y}$ be a random vector with pdf $f_S(\cdot|\nu)$; let $U$ be a [0,1] uniform random variable, independent of $\mathbf{Y}$; let $\mathbf{X}$ be the random vector

$$
\mathbf{X} = \begin{cases} 
\omega \mathbf{Y} + \xi, & \text{if } U \leq \pi(\mathbf{Y}) \\
-\omega \mathbf{Y} + \xi, & \text{if } U > \pi(\mathbf{Y})
\end{cases}
$$

then $\mathbf{X} \sim \text{SS}(\xi, \omega, \nu, \pi)$

Algorithm 2 Let $W$ be a RV with cdf $H$, independent of $\mathbf{Y}$.
Algorithm 1 Let $\mathbf{Y}$ be a random vector with pdf $f_S(\cdot | \nu)$; Let $U$ be a [0,1] uniform random variable, independent of $\mathbf{Y}$; Let $\mathbf{X}$ be the random vector

$$
\mathbf{X} = \begin{cases} 
\omega \mathbf{Y} + \xi, & \text{if } U \leq \pi(\mathbf{Y}) \\
-\omega \mathbf{Y} + \xi, & \text{if } U > \pi(\mathbf{Y}) 
\end{cases}, \text{ then } \mathbf{X} \sim \text{SS}(\xi, \omega, \nu, \pi)
$$

Algorithm 2 Let $W$ be a RV with cdf $H$, independent of $\mathbf{Y}$; Let $\mathbf{X}$ be the random vector

$$
\mathbf{X} = \begin{cases} 
\omega \mathbf{Y} + \xi, & \text{if } W \leq \eta(\mathbf{Y}) \\
-\omega \mathbf{Y} + \xi, & \text{if } W > \eta(\mathbf{Y}) 
\end{cases}, \text{ then } \mathbf{X} \sim \text{SS}(\xi, \omega, \nu, \pi)
Let $\eta : \mathbb{R} \rightarrow \mathbb{R}$, $\eta \in C^l[a, b]$, and $\eta^{(j)} \in L_q[a, b]$, $j \in \{1, ..., d\}$. Let $h_{t_N}$ be a $d$ degree spline on knots $t_N$. Then:

$$||h^{(j)}(x) - \eta^{(j)}(x)||_{L_p[a,b]} \leq c ||\Delta_N||^{1-1/p+1/q} \omega_{d-j}(h^{j+1}, ||\Delta_N||_q)$$

where $\omega_q(f, \delta)$ is the modulus of smoothing of order $q$:

$$\omega_q(f, \delta) = \sup_{0 < h \leq \delta} \left|\sum_{i=0}^{q} (-1)^{q-i} \binom{q}{i} f(q - 2i) \frac{h}{2} \right|,$$

and

$$||\Delta_N|| = \max_{1 \leq i \leq N-1} (t_{i+1} - t_i)$$
Theorem 2 For every precision $\epsilon > 0$, it exists a real value $a$ and a sequence of knots $t_{a,N}$ such that:

$$\|f_{SS} - f_{BSS}\|_\infty < \epsilon$$

(under mild regularity conditions).
Theorem 2 For every precision $\epsilon > 0$, it exists a real value $a$ and a sequence of knots $t_{a,N}$ such that:

$$\|f_{SS} - f_{BSS}\|_\infty < \epsilon$$

(under mild regularity conditions).

Sketch of the proof
Theorem 2 For every precision $\epsilon > 0$, it exists a real value $a$ and a sequence of knots $t_{a,N}$ such that:

$$||f_{SS} - f_{BSS}||_\infty < \epsilon$$

(under mild regularity conditions).

Sketch of the proof

Let $a_\epsilon$ such that: $f_S(t) < \epsilon/2, \forall |t| \geq a_\epsilon$. 
Appendix 1: Richness of BSS in SS

- **Theorem 2** For every precision $\epsilon > 0$, it exists a real value $a$ and a sequence of knots $t_{a,N}$ such that:

$$||f_{SS} - f_{BSS}||_{\infty} < \epsilon$$

(under mild regularity conditions).

- **Sketch of the proof**
  - Let $a_\epsilon$ such that: $f_S(t) < \epsilon/2, \forall |t| \geq a_\epsilon$.
  - Choose $t_{a_\epsilon,N} \exists \eta_B \in \mathcal{B}_{\text{odd}}$

$$||\eta - \eta_B||_{\infty} < \epsilon/(2f_0), \quad f_0 = \max f_S$$

Then

$$||f_{SS} - f_{BSS}||_{\infty} = \sup_{x \in \mathbb{R}} 2f_S(z|\nu)|H(\eta(z)) - H(\eta_B(z))|$$

$$< \begin{cases} 2f_S(z|\nu)\epsilon/2(f_0) < \epsilon, & \text{if } |x| \leq a_\epsilon \\ 2f_S(z|\nu) < \epsilon, & \text{if not} \end{cases}$$
Graphical Sketch of the Proof

Spline approximation (30 Knots in (0,6])

Resulting pdf

\[ \text{pnorm}(B \times \text{initb}) \]

\[ -6, 0, 6 \]

\[ 0.0, 0.2, 0.4, 0.6, 0.8, 1.0 \]

Spline approximation (30 Knots in (0,6])

\[ -5, 0, 5 \]

\[ 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 \]

\[ \text{dgsst}(x, \xi, \omega, n, \text{odd}) \]

True pdf

BSS-approx

Resulting pdf

P. Frederic (univ. of Modena)
By proposition (2) we have $f_{SS} \leq 2f_S$ a.e.
Reference

A class of distributions which includes the normal ones.

The skew-normal distribution and related multivariate families.
With discussion by Marc G. Genton and a rejoinder by the author.

Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution.

Robust likelihood methods based on the skew-t and related distributions.

The skew-normal distribution and related multivariate families.
With discussion by Marc G. Genton and a rejoinder by the author.

A class of distributions which includes the normal ones.

The skew-normal distribution and related multivariate families.
With discussion by Marc G. Genton and a rejoinder by the author.

Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution.

Robust likelihood methods based on the skew-t and related distributions.

Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution.

Robust likelihood methods based on the skew-t and related distributions.
*A practical guide to splines*, volume 27 of *Applied Mathematical Sciences*.  
Springer-Verlag, New York, revised edition.

*Moduli of smoothness*, volume 9 of *Springer Series in Computational Mathematics*.  
Springer-Verlag, New York.

Fast and compact smoothing on large multidimensional grids.  

Flexible smoothing with $B$-splines and penalties.  
With comments and a rejoinder by the authors.

Splines, knots, and penalties.  
Unpublished.

CRC Press LLC, Boca Raton, FL.


Spline approximation.

A skew-symmetric representation of multivariate distributions.