Maximum Likelihood vs. Least Squares for Estimating Mixtures of Truncated Exponentials

Helge Langseth[1] Thomas D. Nielsen[2] Rafael Rumí[3]
Antonio Salmerón[3]

[1]Department of Computer and Information Science
The Norwegian University of Science and Technology, Trondheim (Norway)

[2]Department of Computer Science
Aalborg University, Aalborg (Denmark)

[3]Department of Statistics and Applied Mathematics
University of Almería, Almería (Spain)

INFORMS, Seattle, November 2007

Outline

1. Motivation.
2. The MTE (Mixture of Truncated Exponentials) model.
3. Maximum Likelihood (ML) estimation for MTEs.
4. Least Squares (LS) estimation of MTEs.
5. Experimental analysis.
6. Conclusions.

Outline

1. Motivation.
2. The MTE (Mixture of Truncated Exponentials) model.
3. Maximum Likelihood (ML) estimation for MTEs.
4. Least Squares (LS) estimation of MTEs.
5. Experimental analysis.
6. Conclusions.

Outline

1. Motivation.
2. The MTE (Mixture of Truncated Exponentials) model.
3. Maximum Likelihood (ML) estimation for MTEs.
4. Least Squares (LS) estimation of MTEs.
5. Experimental analysis.
6. Conclusions.

Outline

1. Motivation.
2. The MTE (Mixture of Truncated Exponentials) model.
3. Maximum Likelihood (ML) estimation for MTEs.
4. Least Squares (LS) estimation of MTEs.
5. Experimental analysis.
6. Conclusions.

Outline

1. Motivation.
2. The MTE (Mixture of Truncated Exponentials) model.
3. Maximum Likelihood (ML) estimation for MTEs.
4. Least Squares (LS) estimation of MTEs.
5. Experimental analysis.
6. Conclusions.

Outline

1. Motivation.
2. The MTE (Mixture of Truncated Exponentials) model.
3. Maximum Likelihood (ML) estimation for MTEs.
4. Least Squares (LS) estimation of MTEs.
5. Experimental analysis.
6. Conclusions.

Motivation

- Graphical models are a common tool for decision analysis.
- Problems in which continuous and discrete variables interact are frequent.
- A very general solution is the use of MTE models.
- When learning from data, the only existing method in the literature is based on least squares.
- The feasibility of ML estimation seems to be worth studying.

Motivation

- Graphical models are a common tool for decision analysis.
- Problems in which continuous and discrete variables interact are frequent.
- A very general solution is the use of MTE models.
- When learning from data, the only existing method in the literature is based on least squares.
- The feasibility of ML estimation seems to be worth studying.

Motivation

- Graphical models are a common tool for decision analysis.
- Problems in which continuous and discrete variables interact are frequent.
- A very general solution is the use of MTE models.
- When learning from data, the only existing method in the literature is based on least squares.
- The feasibility of ML estimation seems to be worth studying.

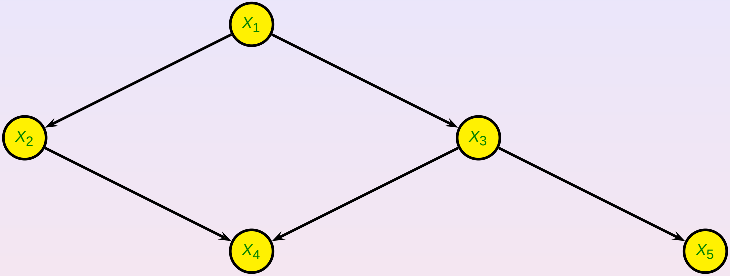
Motivation

- Graphical models are a common tool for decision analysis.
- Problems in which continuous and discrete variables interact are frequent.
- A very general solution is the use of MTE models.
- When learning from data, the only existing method in the literature is based on least squares.
- The feasibility of ML estimation seems to be worth studying.

Motivation

- Graphical models are a common tool for decision analysis.
- Problems in which continuous and discrete variables interact are frequent.
- A very general solution is the use of MTE models.
- When learning from data, the only existing method in the literature is based on least squares.
- The feasibility of ML estimation seems to be worth studying.

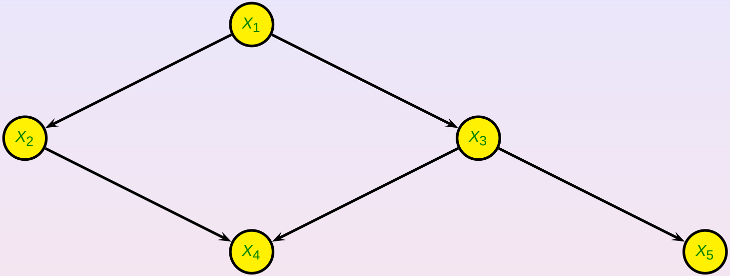
Bayesian networks



- D.A.G.
- The nodes represent random variables.
- Arc $\Rightarrow$ dependence.

$$p(\mathbf{x}) = \prod_{i=1}^{n} p(x_i|\pi_i) \quad \mathbf{x} \in \Omega_{\mathbf{X}}$$

Bayesian networks



- D.A.G.
- The nodes represent random variables.
- Arc $\Rightarrow$ dependence.

$$p(x_1, x_2, x_3, x_4, x_5) = p(x_1)p(x_2|x_1)p(x_3|x_1)p(x_5|x_3)p(x_4|x_2, x_3) \quad .$$

The MTE model (Moral et al. 2001)

Definition (MTE potential)

- $\mathbf{X}$: mixed n-dimensional random vector. $\mathbf{Y} = (Y_1, \ldots, Y_d)$, $\mathbf{Z} = (Z_1, \ldots, Z_c)$ its discrete and continuous parts. A function $f : \Omega_{\mathbf{X}} \mapsto \mathbb{R}_0^+$ is a Mixture of Truncated Exponentials potential (MTE potential) if for each fixed value $\mathbf{y} \in \Omega_{\mathbf{Y}}$ of the discrete variables $\mathbf{Y}$, the potential over the continuous variables $\mathbf{Z}$ is defined as:

$$f(\mathbf{z}) = a_0 + \sum_{i=1}^{m} a_i \exp\left\{ \sum_{j=1}^{c} b_i^{(j)} z_j \right\}$$

  for all $\mathbf{z} \in \Omega_{\mathbf{Z}}$, where a_i, $b_i^{(j)}$ are real numbers.
- Also, f is an MTE potential if there is a partition $D_1, \ldots, D_k$ of $\Omega_{\mathbf{Z}}$ into hypercubes and in each D_i, f is defined as above.

The MTE model (Moral et al. 2001)

Definition (MTE potential)

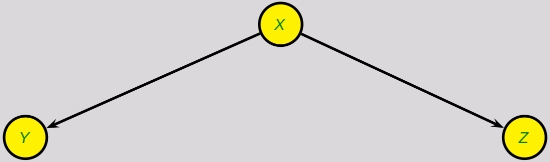
- $\mathbf{X}$: mixed n-dimensional random vector. $\mathbf{Y} = (Y_1, \ldots, Y_d)$, $\mathbf{Z} = (Z_1, \ldots, Z_c)$ its discrete and continuous parts. A function $f : \Omega_{\mathbf{X}} \mapsto \mathbb{R}_0^+$ is a Mixture of Truncated Exponentials potential (MTE potential) if for each fixed value $\mathbf{y} \in \Omega_{\mathbf{Y}}$ of the discrete variables $\mathbf{Y}$, the potential over the continuous variables $\mathbf{Z}$ is defined as:

$$f(\mathbf{z}) = a_0 + \sum_{i=1}^{m} a_i \exp\left\{ \sum_{j=1}^{c} b_i^{(j)} z_j \right\}$$

  for all $\mathbf{z} \in \Omega_{\mathbf{Z}}$, where a_i, $b_i^{(j)}$ are real numbers.
- Also, f is an MTE potential if there is a partition $D_1, \ldots, D_k$ of $\Omega_{\mathbf{Z}}$ into hypercubes and in each D_i, f is defined as above.

Example

Consider a model with continuous variables X and Y, and discrete variable Z.



The MTE model (Moral et al. 2001)

Example

One example of conditional densities for this model is given by the following expressions:

$$f(x) = \begin{cases} 1.16 - 1.12e^{-0.02x} & \text{if } 0.4 \leq x < 4 \;, \\ 0.9e^{-0.35x} & \text{if } 4 \leq x < 19 \;. \end{cases}$$

$$f(y|x) = \begin{cases} 1.26 - 1.15e^{0.006y} & \text{if } 0.4 \leq x < 5,\ 0 \leq y < 13 \;, \\ 1.18 - 1.16e^{0.0002y} & \text{if } 0.4 \leq x < 5,\ 13 \leq y < 43 \;, \\ 0.07 - 0.03e^{-0.4y} + 0.0001e^{0.0004y} & \text{if } 5 \leq x < 19,\ 0 \leq y < 5 \;, \\ -0.99 + 1.03e^{0.001y} & \text{if } 5 \leq x < 19,\ 5 \leq y < 43 \;. \end{cases}$$

$$f(z|x) = \begin{cases} 0.3 & \text{if } z = 0,\ 0.4 \leq x < 5 \;, \\ 0.7 & \text{if } z = 1,\ 0.4 \leq x < 5 \;, \\ 0.6 & \text{if } z = 0,\ 5 \leq x < 19 \;, \\ 0.4 & \text{if } z = 1,\ 5 \leq x < 19 \;. \end{cases}$$

Example

One example of conditional densities for this model is given by the following expressions:

$$f(x) = \begin{cases} 1.16 - 1.12e^{-0.02x} & \text{if } 0.4 \le x < 4 \ , \\ 0.9e^{-0.35x} & \text{if } 4 \le x < 19 \ . \end{cases}$$

$$f(y|x) = \begin{cases} 1.26 - 1.15e^{0.006y} & \text{if } 0.4 \le x < 5,\ 0 \le y < 13 \ , \\ 1.18 - 1.16e^{0.0002y} & \text{if } 0.4 \le x < 5,\ 13 \le y < 43 \ , \\ 0.07 - 0.03e^{-0.4y} + 0.0001e^{0.0004y} & \text{if } 5 \le x < 19,\ 0 \le y < 5 \ , \\ -0.99 + 1.03e^{0.001y} & \text{if } 5 \le x < 19,\ 5 \le y < 43 \ . \end{cases}$$

$$f(z|x) = \begin{cases} 0.3 & \text{if } z = 0,\ 0.4 \le x < 5 \ , \\ 0.7 & \text{if } z = 1,\ 0.4 \le x < 5 \ , \\ 0.6 & \text{if } z = 0,\ 5 \le x < 19 \ , \\ 0.4 & \text{if } z = 1,\ 5 \le x < 19 \ . \end{cases}$$

The MTE model (Moral et al. 2001)

Example

One example of conditional densities for this model is given by the following expressions:

$$f(x) = \begin{cases} 1.16 - 1.12e^{-0.02x} & \text{if } 0.4 \leq x < 4 \ , \\ 0.9e^{-0.35x} & \text{if } 4 \leq x < 19 \ . \end{cases}$$

$$f(y|x) = \begin{cases} 1.26 - 1.15e^{0.006y} & \text{if } {\scriptstyle 0.4 \leq x < 5,\ 0 \leq y < 13} \ , \\ 1.18 - 1.16e^{0.0002y} & \text{if } {\scriptstyle 0.4 \leq x < 5,\ 13 \leq y < 43} \ , \\ 0.07 - 0.03e^{-0.4y} + 0.0001e^{0.0004y} & \text{if } {\scriptstyle 5 \leq x < 19,\ 0 \leq y < 5} \ , \\ -0.99 + 1.03e^{0.001y} & \text{if } {\scriptstyle 5 \leq x < 19,\ 5 \leq y < 43} \ . \end{cases}$$

$$f(z|x) = \begin{cases} 0.3 & \text{if } z = 0,\ 0.4 \leq x < 5 \ , \\ 0.7 & \text{if } z = 1,\ 0.4 \leq x < 5 \ , \\ 0.6 & \text{if } z = 0,\ 5 \leq x < 19 \ , \\ 0.4 & \text{if } z = 1,\ 5 \leq x < 19 \ . \end{cases}$$

In this work we are concerned with the univariate case.

The learning task involves three basic steps:

- Determination of the splits into which Ω_X will be partitioned.
- Determination of the number of exponential terms in the mixture for each split.
- Estimation of the parameters.

Learning MTEs from data

In this work we are concerned with the univariate case.

The learning task involves three basic steps:

- Determination of the splits into which Ω_X will be partitioned.
- Determination of the number of exponential terms in the mixture for each split.
- Estimation of the parameters.

Learning MTEs from data

In this work we are concerned with the univariate case.

The learning task involves three basic steps:

- Determination of the splits into which Ω_X will be partitioned.
- Determination of the number of exponential terms in the mixture for each split.
- Estimation of the parameters.

Learning MTEs from data

In this work we are concerned with the univariate case.

The learning task involves three basic steps:

- Determination of the splits into which Ω_X will be partitioned.
- Determination of the number of exponential terms in the mixture for each split.
- Estimation of the parameters.

Learning MTEs from data by ML

Why ML?

- Well developed core theory.
- Good asymptotic properties under regularity conditions.
- Several procedures connected to Bayesian networks rely on ML estimations.

Problems for applying ML to MTEs

- The likelihood equations cannot be solved for the MTE model.
- Numerical methods are slow and potentially unstable.

Question

Can the LS estimates be used as approximations to the ML ones?

Learning MTEs from data by ML

Why ML?

- Well developed core theory.
- Good asymptotic properties under regularity conditions.
- Several procedures connected to Bayesian networks rely on ML estimations.

Problems for applying ML to MTEs

- The likelihood equations cannot be solved for the MTE model.
- Numerical methods are slow and potentially unstable.

Question

Can the LS estimates be used as approximations to the ML ones?

Learning MTEs from data by ML

Why ML?

- Well developed core theory.
- Good asymptotic properties under regularity conditions.
- Several procedures connected to Bayesian networks rely on ML estimations.

Problems for applying ML to MTEs

- The likelihood equations cannot be solved for the MTE model.
- Numerical methods are slow and potentially unstable.

Question

Can the LS estimates be used as approximations to the ML ones?

Learning MTEs from data by ML

Why ML?

- Well developed core theory.
- Good asymptotic properties under regularity conditions.
- Several procedures connected to Bayesian networks rely on ML estimations.

Problems for applying ML to MTEs

- The likelihood equations cannot be solved for the MTE model.
- Numerical methods are slow and potentially unstable.

Question

Can the LS estimates be used as approximations to the ML ones?

Learning MTEs from data by ML

Why ML?

- Well developed core theory.
- Good asymptotic properties under regularity conditions.
- Several procedures connected to Bayesian networks rely on ML estimations.

Problems for applying ML to MTEs

- The likelihood equations cannot be solved for the MTE model.
- Numerical methods are slow and potentially unstable.

Question

Can the LS estimates be used as approximations to the ML ones?

Learning MTEs from data by ML

Why ML?

- Well developed core theory.
- Good asymptotic properties under regularity conditions.
- Several procedures connected to Bayesian networks rely on ML estimations.

Problems for applying ML to MTEs

- The likelihood equations cannot be solved for the MTE model.
- Numerical methods are slow and potentially unstable.

Question

Can the LS estimates be used as approximations to the ML ones?

Learning MTEs from data

Starting point

R. Rumí, A. Salmerón, S. Moral (2006) Estimating mixtures of truncated exponentials in hybrid Bayesian networks. *Test* **15**:397–421.

- Estimation based on least squares (LS).
- Empiric density approximated by a histogram.

Improved version

V. Romero, R. Rumí, A. Salmerón (2006) Learning hybrid Bayesian networks using mixtures of truncated exponentials. *International Journal of Approximate Reasoning* **42**:54-68.

- Empiric density approximated by a kernel.

Starting point

R. Rumí, A. Salmerón, S. Moral (2006) Estimating mixtures of truncated exponentials in hybrid Bayesian networks. *Test* **15**:397–421.

- Estimation based on least squares (LS).
- Empiric density approximated by a histogram.

Improved version

V. Romero, R. Rumí, A. Salmerón (2006) Learning hybrid Bayesian networks using mixtures of truncated exponentials. *International Journal of Approximate Reasoning* **42**:54-68.

- Empiric density approximated by a kernel.

Learning MTEs from data

Starting point

R. Rumí, A. Salmerón, S. Moral (2006) Estimating mixtures of truncated exponentials in hybrid Bayesian networks. *Test* **15**:397–421.

- Estimation based on least squares (LS).
- Empiric density approximated by a histogram.

Improved version

V. Romero, R. Rumí, A. Salmerón (2006) Learning hybrid Bayesian networks using mixtures of truncated exponentials. *International Journal of Approximate Reasoning* **42**:54-68.

- Empiric density approximated by a kernel.

Learning MTEs from data

Starting point

R. Rumí, A. Salmerón, S. Moral (2006) Estimating mixtures of truncated exponentials in hybrid Bayesian networks. *Test* **15**:397–421.

- Estimation based on least squares (LS).
- Empiric density approximated by a histogram.

Improved version

V. Romero, R. Rumí, A. Salmerón (2006) Learning hybrid Bayesian networks using mixtures of truncated exponentials. *International Journal of Approximate Reasoning* **42**:54-68.

- Empiric density approximated by a kernel.

Learning MTEs from data

- The split points are determined observing the extreme and inflexion points.
  - The number of points (N) to locate is established beforehand.
  - The N higher changes from concavity/convexity or increase/decrease are selected.

- The number of exponential terms can be determined beforehand or decided during the parameter estimation procedure.

Learning MTEs from data

- The split points are determined observing the extreme and inflexion points.
  - The number of points (N) to locate is established beforehand.
  - The N higher changes from concavity/convexity or increase/decrease are selected.
- The number of exponential terms can be determined beforehand or decided during the parameter estimation procedure.

Learning MTEs from data

- The split points are determined observing the extreme and inflexion points.
  - The number of points (N) to locate is established beforehand.
  - The N higher changes from concavity/convexity or increase/decrease are selected.
- The number of exponential terms can be determined beforehand or decided during the parameter estimation procedure.

Target density

$$f(x) = k + ae^{bx} + ce^{dx}$$

- Assume we have initial estimates a_0, b_0 and k_0.
- c and d are estimated by fitting to points $(\mathbf{x}, \mathbf{w})$, where

$$\mathbf{w} = \mathbf{y} - a_0 \exp\{b_0 \mathbf{x}\} - k_0 \quad ,$$

a function

$$w = c \exp\{dx\} \quad ,$$

minimising the weighted mean squared error.

Learning MTEs: estimating the parameters by LS

Target density

$$f(x) = k + ae^{bx} + ce^{dx}$$

- Assume we have initial estimates a_0, b_0 and k_0.
- c and d are estimated by fitting to points $(\mathbf{x}, \mathbf{w})$, where

$$\mathbf{w} = \mathbf{y} - a_0 \exp\{b_0 \mathbf{x}\} - k_0 \quad ,$$

a function

$$w = c \exp\{dx\} \quad ,$$

minimising the weighted mean squared error.

Target density

$$f(x) = k + ae^{bx} + ce^{dx}$$

- Assume we have initial estimates a_0, b_0 and k_0.
- c and d are estimated by fitting to points $(\mathbf{x}, \mathbf{w})$, where

$$\mathbf{w} = \mathbf{y} - a_0 \exp\{b_0 \mathbf{x}\} - k_0 \quad ,$$

a function

$$w = c \exp\{dx\} \quad ,$$

minimising the weighted mean squared error.

Learning MTEs: estimating the parameters by LS

- Taking logarithms, the problem reduces to linear regression:

$$\ln\{w\} = \ln\{c\}\exp\{dx\} = \ln\{c\} + dx \quad ,$$

that can be written as

$$w^* = c^* + dx \quad ,$$

where $c^* = \ln\{c\}$ and $w^* = \ln\{w\}$.

- The solution is

$$(c^*, d) = \arg\min_{c^*,d} \sum_{i=1}^{n} (w_i^* - c^* - dx_i)^2 f(x_i) \quad .$$

Learning MTEs: estimating the parameters by LS

- Taking logarithms, the problem reduces to linear regression:

$$\ln\{w\} = \ln\{c\}\exp\{dx\} = \ln\{c\} + dx \quad ,$$

that can be written as

$$w^* = c^* + dx \quad ,$$

where $c^* = \ln\{c\}$ and $w^* = \ln\{w\}$.

- The solution is

$$(c^*, d) = \arg\min_{c^*, d} \sum_{i=1}^{n} (w_i^* - c^* - dx_i)^2 f(x_i) \quad .$$

Learning MTEs: estimating the parameters by LS

- The solution can be obtained by analytical means:

$$
\begin{aligned}
c^* &= \frac{\left(\sum_{i=1}^{n} w_i x_i f(x_i)\right) - d\left(\sum_{i=1}^{n} x_i f(x_i)\right)^2}{\left(\sum_{i=1}^{n} x_i f(x_i)\right)} \\
d &= \frac{\left(\sum_{i=1}^{n} w_i f(x_i)\right)\left(\sum_{i=1}^{n} x_i f(x_i)\right) - \left(\sum_{i=1}^{n} f(x_i)\right)\left(\sum_{i=1}^{n} w_i x_i f(x_i)\right)}{\left(\sum_{i=1}^{n} x_i f(x_i)\right)^2 - \left(\sum_{i=1}^{n} f(x_i)\right)\left(\sum_{i=1}^{n} x_i^2 f(x_i)\right)}
\end{aligned}
$$

Learning MTEs: estimating the parameters by LS

- The solution can be obtained by analytical means:

$$c^* = \frac{\left(\sum_{i=1}^n w_i x_i f(x_i)\right) - d\left(\sum_{i=1}^n x_i f(x_i)\right)^2}{\left(\sum_{i=1}^n x_i f(x_i)\right)}$$

$$d = \frac{\left(\sum_{i=1}^n w_i f(x_i)\right)\left(\sum_{i=1}^n x_i f(x_i)\right) - \left(\sum_{i=1}^n f(x_i)\right)\left(\sum_{i=1}^n w_i x_i f(x_i)\right)}{\left(\sum_{i=1}^n x_i f(x_i)\right)^2 - \left(\sum_{i=1}^n f(x_i)\right)\left(\sum_{i=1}^n x_i^2 f(x_i)\right)}$$

- Once a, b, c and d are known, we go for k:

$$f^*(x) = k + ae^{\{bx\}} + ce^{\{dx\}} \quad ,$$

where $k \in \mathbb{R}$ should be such that minimises the error

$$E(k) = \sum_{i=1}^{n} \frac{(f(x_i) - ae^{bx_i} - ce^{dx_i} - k)^2 f(x_i)}{n} \quad ,$$

- This is optimised for

$$\hat{k} = \frac{\sum_{i=1}^{n}(f(x_i) - ae^{bx_i} - ce^{dx_i})f(x_i)}{\sum_{i=1}^{n} f(x_i)} \quad .$$

- Once a, b, c and d are known, we go for k:

$$f^*(x) = k + ae^{\{bx\}} + ce^{\{dx\}} \quad ,$$

where $k \in \mathbb{R}$ should be such that minimises the error

$$E(k) = \sum_{i=1}^{n} \frac{(f(x_i) - ae^{bx_i} - ce^{dx_i} - k)^2 f(x_i)}{n} \quad ,$$

- This is optimised for

$$\hat{k} = \frac{\sum_{i=1}^{n}(f(x_i) - ae^{bx_i} - ce^{dx_i})f(x_i)}{\sum_{i=1}^{n} f(x_i)} \quad .$$

Learning MTEs: estimating the parameters by LS

- The contribution of each exponential term can be refined. Assume that we have an estimated model

$$\hat{f}(x) = \hat{a}_0 e^{\hat{b}_0 x} + \hat{c}_0 e^{\hat{d}_0 x} + \hat{k}_0 \ .$$

- The impact of the second exponential term can be determined by introducing a factor h in the regression equation, for a sample $(\mathbf{x}, \mathbf{y})$, given by

$$\mathbf{y} = \hat{a}_0 e^{\hat{b}_0 \mathbf{x}} + h \hat{c}_0 e^{\hat{d}_0 \mathbf{x}} + \hat{k}_0 \ ,$$

and the value of h is computed by least squares, obtaining

$$h = \frac{\sum_{i=1}^{n} (y_i - \hat{a}_0 e^{\hat{b}_0 x_i} - \hat{k}_0)(e^{\hat{d}_0 x_i}) f(x_i)}{\sum_{i=1}^{n} \hat{c}_0 e^{2\hat{d}_0 x_i} f(x_i)} \ .$$

- The contribution of each exponential term can be refined. Assume that we have an estimated model

$$\hat{f}(x) = \hat{a}_0 e^{\hat{b}_0 x} + \hat{c}_0 e^{\hat{d}_0 x} + \hat{k}_0 \quad .$$

- The impact of the second exponential term can be determined by introducing a factor h in the regression equation, for a sample $(\mathbf{x}, \mathbf{y})$, given by

$$\mathbf{y} = \hat{a}_0 e^{\hat{b}_0 \mathbf{x}} + h \hat{c}_0 e^{\hat{d}_0 \mathbf{x}} + \hat{k}_0 \quad ,$$

and the value of h is computed by least squares, obtaining

$$h = \frac{\sum_{i=1}^{n} (y_i - \hat{a}_0 e^{\hat{b}_0 x_i} - \hat{k}_0)(e^{\hat{d}_0 x_i}) f(x_i)}{\sum_{i=1}^{n} \hat{c}_0 e^{2\hat{d}_0 x_i} f(x_i)} \quad .$$

Initialising a, b and k

- The initial values of a, b and k can be arbitrary, but a good selection of them can speed up the convergence of the method.
- These values can be initialised fitting a curve $\mathbf{y} = a \exp\{b\mathbf{x}\}$ to the modified sample by exponential regression, and computing k as before.
- Another alternative is to force the empiric density and the initial model to have the same derivative.

Initialising a, b and k

- The initial values of a, b and k can be arbitrary, but a good selection of them can speed up the convergence of the method.
- These values can be initialised fitting a curve $\mathbf{y} = a \exp\{b\mathbf{x}\}$ to the modified sample by exponential regression, and computing k as before.
- Another alternative is to force the empiric density and the initial model to have the same derivative.

Initialising a, b and k

- The initial values of a, b and k can be arbitrary, but a good selection of them can speed up the convergence of the method.
- These values can be initialised fitting a curve $\mathbf{y} = a \exp\{b\mathbf{x}\}$ to the modified sample by exponential regression, and computing k as before.
- Another alternative is to force the empiric density and the initial model to have the same derivative.

Experimental setting

- Tested distributions:
  - Normal.
  - Log-normal.
  - χ^2.
  - Beta.
  - MTE.
- Sample size: 1000.
- Split points detection:
  - ML: Manually determined.
  - LS: Automatic detection.

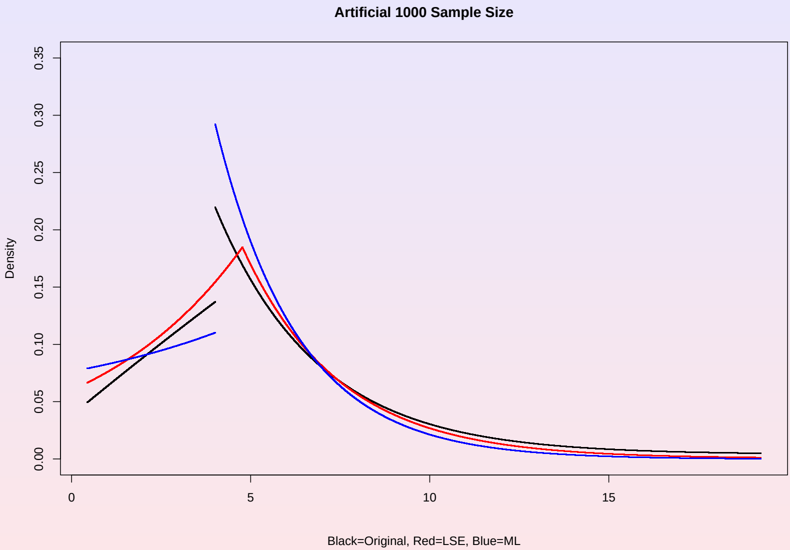
Experimental setting

- Tested distributions:
  - Normal.
  - Log-normal.
  - χ^2.
  - Beta.
  - MTE.
- Sample size: 1000.
- Split points detection:
  - ML: Manually determined.
  - LS: Automatic detection.

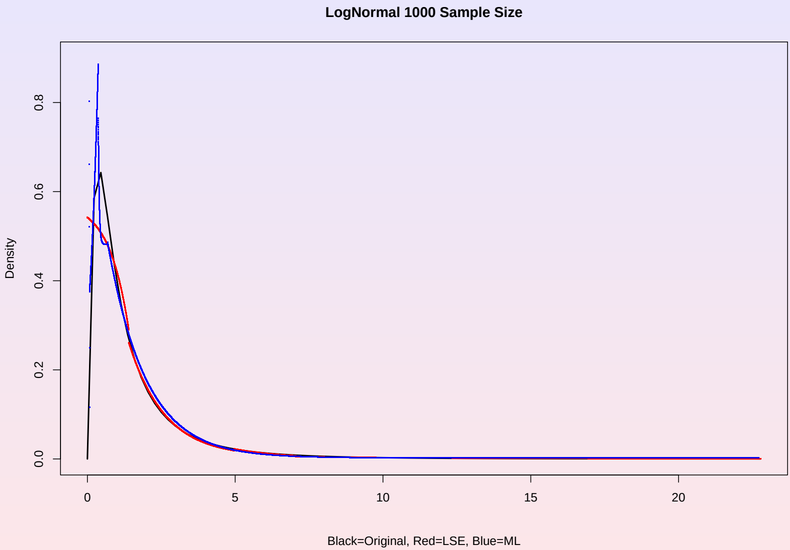
Experimental setting

- Tested distributions:
  - Normal.
  - Log-normal.
  - χ^2.
  - Beta.
  - MTE.
- Sample size: 1000.
- Split points detection:
  - ML: Manually determined.
  - LS: Automatic detection.

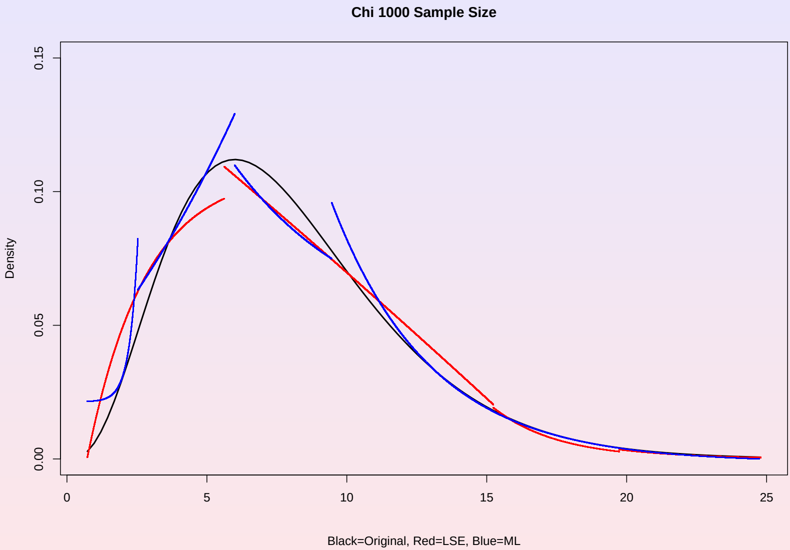
Graphical comparison: MTE



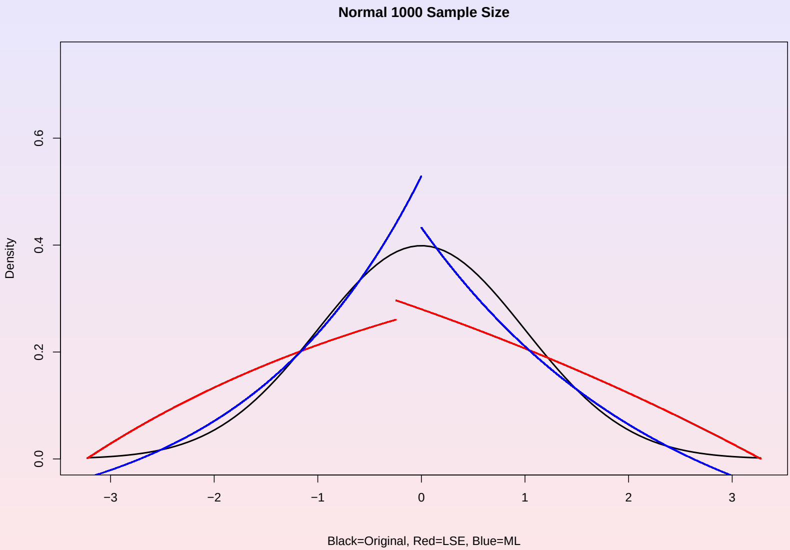
Graphical comparison: Log-normal



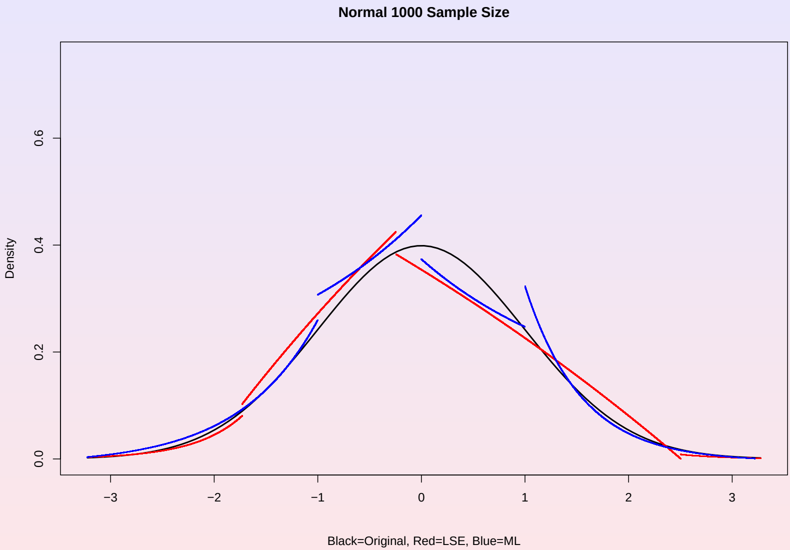
Graphical comparison: χ^2



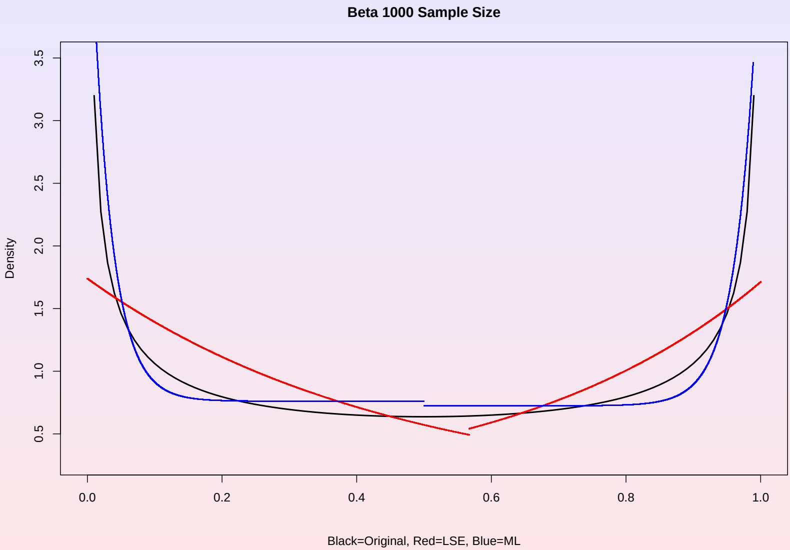
Graphical comparison: Normal (2 splits)



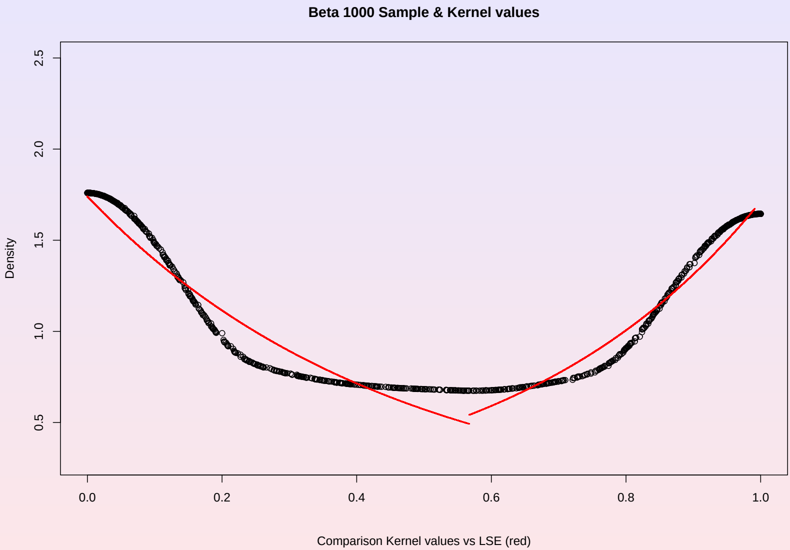
Graphical comparison: Normal (4 splits)



Graphical comparison: Beta



Graphical comparison: Beta, kernel fitting



Comparison in terms of likelihood

	Artificial	χ^2	Beta	Normal-2	Normal-4	Log-normal
ML	-2263.132	160.687	-2685.765	-1373.499	-1364.643	-1398.300
LS	-2293.473	90.234	-2726.950	-1533.889	-1404.747	-1451.568

Comparison through a two-sided paired *t*-test

- Including all the data: $p = 0.02039$.
- Excluding the Beta: $p = 0.05418$.

Comparison in terms of likelihood

	Artificial	χ^2	Beta	Normal-2	Normal-4	Log-normal
ML	-2263.132	160.687	-2685.765	-1373.499	-1364.643	-1398.300
LS	-2293.473	90.234	-2726.950	-1533.889	-1404.747	-1451.568

Comparison through a two-sided paired *t*-test

- Including all the data: $p = 0.02039$.
- Excluding the Beta: $p = 0.05418$.

Conclusions

- LS highly dependent on the empirical density (kernel or histogram).
- LS and ML very close except for the Beta case.
- ML reaches higher likelihood values.
- LS efficient from a computational point of view.

Conclusions

- LS highly dependent on the empirical density (kernel or histogram).
- LS and ML very close except for the Beta case.
- ML reaches higher likelihood values.
- LS efficient from a computational point of view.

Conclusions

- LS highly dependent on the empirical density (kernel or histogram).
- LS and ML very close except for the Beta case.
- ML reaches higher likelihood values.
- LS efficient from a computational point of view.

Conclusions

- LS highly dependent on the empirical density (kernel or histogram).
- LS and ML very close except for the Beta case.
- ML reaches higher likelihood values.
- LS efficient from a computational point of view.

Ongoing work

- Refining LS estimation.
- More exhaustive comparison with ML.
- Extension to the conditional case.
- Possible mixed approach ML and LS.

Ongoing work

- Refining LS estimation.
- More exhaustive comparison with ML.
- Extension to the conditional case.
- Possible mixed approach ML and LS.

Ongoing work

- Refining LS estimation.
- More exhaustive comparison with ML.
- Extension to the conditional case.
- Possible mixed approach ML and LS.

Ongoing work

- Refining LS estimation.
- More exhaustive comparison with ML.
- Extension to the conditional case.
- Possible mixed approach ML and LS.