Generalizing Hodges-Lehmann: Nonparametric inference for location mixtures

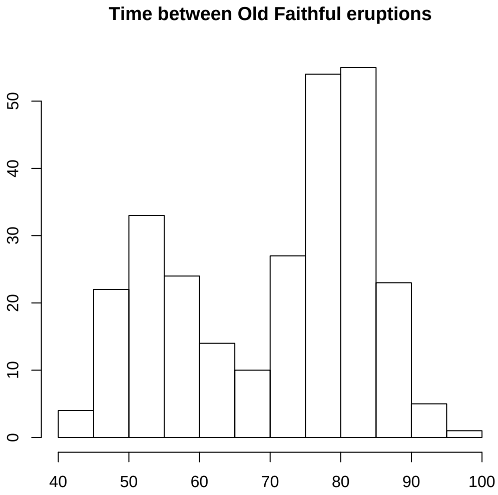
David Hunter[1]
Shaoli Wang[2]
Tom Hettmansperger[1]

[1]Department of Statistics
Penn State University

[2]Department of Epidemiology and Public Health
Yale University

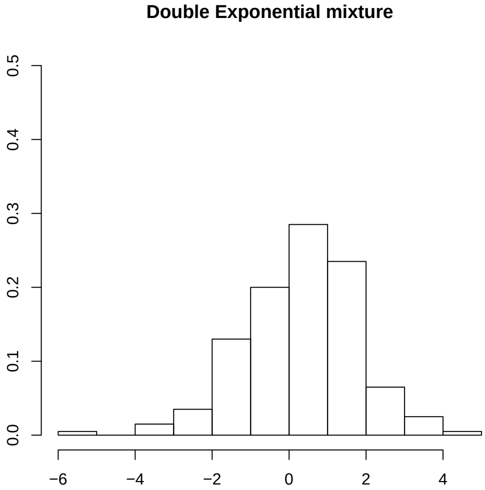
ENAR, March 27, 2006

Old Faithful dataset



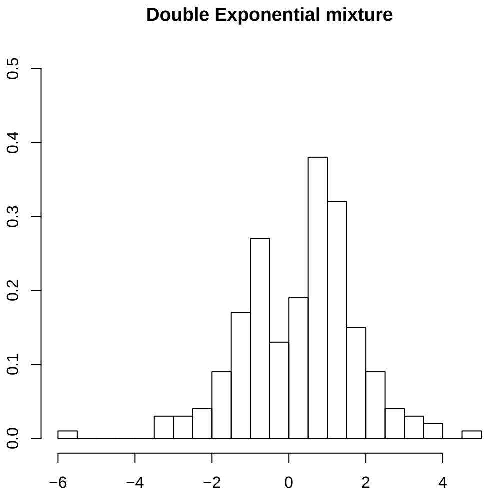
- Pretty clear two-component mixture structure
- Normal-looking components
- No problem!

Simulated dataset



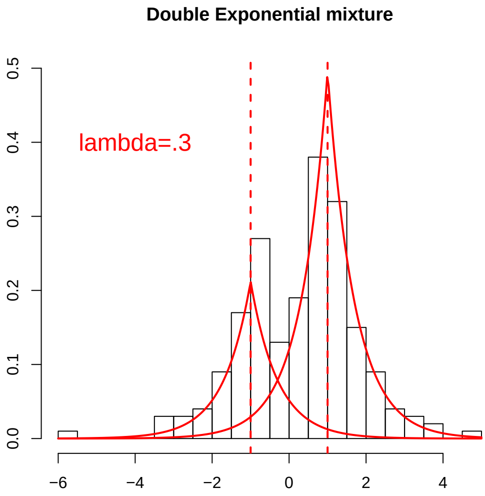
- Mixture structure not so clear

Simulated dataset



- Mixture structure not so clear
- Changing the bin widths helps, but. . .

Simulated dataset



- Mixture structure not so clear
- Changing the bin widths helps, but. . .
- . . . normal assumption still fails due to heavy tails

Finite mixture estimation problem

Goal

Estimate λ_j and F_j given a random sample from

$$F(x) = \sum_{j=1}^{K} \lambda_j F_j(x)$$

Finite mixture estimation problem

Goal ($K = 2$)

Estimate λ, F_1, F_2 given a random sample from

$$F(x) = \lambda F_1(x) + (1 - \lambda) F_2(x)$$

Finite mixture estimation problem

Goal ($K = 2$)

Estimate λ, F_1, F_2 given a random sample from

$$F(x) = \lambda F_1(x) + (1 - \lambda)F_2(x)$$

- However, we don't want to assume a particular parametric form for F_1 and F_2.

Finite mixture estimation problem

Goal ($K = 2$)

Estimate λ, F_1, F_2 given a random sample from

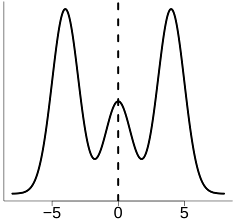
$$F(x) = \lambda F_1(x) + (1 - \lambda)F_2(x)$$

- However, we don't want to assume a particular parametric form for F_1 and F_2.
- Motivating question: How weak can our assumptions about the F_j be?

Outline

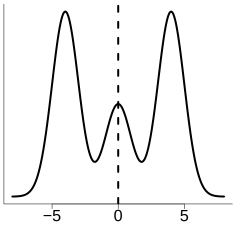
1. One-component "mixtures"
   - Location mixture as convolution
   - Re-convolution to achieve zero-symmetry
   - Basic result involving L_2-distance
2. Identifiability
   - Definition
   - A simple counterexample
   - Results on identifiability
3. Estimation
   - Formulas for $\hat{\lambda}$, $\hat{\theta}$, $\hat{G}$
   - Examples

One-component symmetric location “mixture”



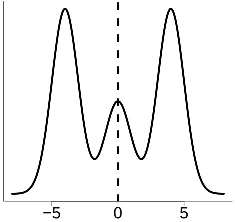
- Suppose $G(x)$ is a zero-symmetric distribution

One-component symmetric location “mixture”



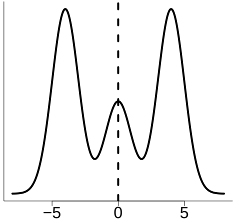
- Suppose $G(x)$ is a zero-symmetric distribution
- Let δ_θ denote a point mass distribution at θ.

One-component symmetric location "mixture"



- Suppose $G(x)$ is a zero-symmetric distribution
- Let δ_θ denote a point mass distribution at θ.
- Suppose we have a sample from $F = G \star \delta_{\theta_0}$, where θ_0 is unknown.

One-component symmetric location "mixture"



- Suppose $G(x)$ is a zero-symmetric distribution
- Let δ_θ denote a point mass distribution at θ.
- Suppose we have a sample from $F = G \star \delta_{\theta_0}$, where θ_0 is unknown.
- θ_0 is the center of symmetry to we'd like to estimate.

Re-convolution to achieve zero-symmetry

$F = G \star \delta_{\theta_0}$ (i.e., F is just G shifted by θ_0)

Re-convolution to achieve zero-symmetry

$F = G \star \delta_{\theta_0}$ (i.e., F is just G shifted by θ_0)

- Note that θ_0 is the unique θ such that $F \star \delta_{-\theta}$ is zero-symmetric.

Re-convolution to achieve zero-symmetry

$F = G \star \delta_{\theta_0}$ (i.e., F is just G shifted by θ_0)

- Note that θ_0 is the unique θ such that $F \star \delta_{-\theta}$ is zero-symmetric.
- A random variable W is zero-symmetric iff $W \stackrel{\mathcal{D}}{=} -W$.

Re-convolution to achieve zero-symmetry

$F = G \star \delta_{\theta_0}$ (i.e., F is just G shifted by θ_0)

- Note that θ_0 is the unique θ such that $F \star \delta_{-\theta}$ is zero-symmetric.
- A random variable W is zero-symmetric iff $W \stackrel{\mathcal{D}}{=} -W$.

These facts lead to…

Idea

Choose $\hat{\theta}$ so that $\hat{F}_n \star \delta_{-\hat{\theta}}$ is nearly zero-symmetric

…where $\hat{F}_n$ denotes the empirical distribution function.

Re-convolution to achieve zero-symmetry

Idea

Choose $\hat{\theta}$ so that $\hat{F}_n \star \delta_{-\hat{\theta}}$ is nearly zero-symmetric.

Re-convolution to achieve zero-symmetry

Idea

Choose $\hat{\theta}$ so that $\hat{F}_n \star \delta_{-\hat{\theta}}$ is nearly zero-symmetric.

In other words, letting $X^* \sim \hat{F}_n$, choose $\hat{\theta}$ so that

$$X^* - \hat{\theta} \quad \overset{\mathcal{D}}{\approx} \quad -X^* + \hat{\theta}.$$

Basic result involving L_2-distance

Notation: For any random variable W,

$$H_W(t) = P(W \leq t)$$

is the cdf of W.

Basic result involving L_2-distance

Notation: For any random variable W,

$$H_W(t) = P(W \leq t)$$

is the cdf of W.

Result

$$\int_{-\infty}^{\infty} [H_W(t) - H_{-W}(t)]^2 = \mathrm{E}\,|W_1 + W_2| - \mathrm{E}\,|W_1 - W_2|,$$

where W_1, W_2 are independent copies of W.

Estimating θ_0 using the basic result

Recall that $X^* \sim \hat{F}_n$, the empirical distribution.

Estimating θ_0 using the basic result

Recall that $X^* \sim \hat{F}_n$, the empirical distribution.

$$\text{Let } \hat{\theta} = \arg\min_{\theta} \int_{-\infty}^{\infty} [H_{X^*-\theta}(t) - H_{-X^*+\theta}(t)]^2$$

Estimating θ_0 using the basic result

Recall that $X^* \sim \hat{F}_n$, the empirical distribution.

$$\begin{aligned} \text{Let } \hat{\theta} &= \arg\min_{\theta} \int_{-\infty}^{\infty} [H_{X^*-\theta}(t) - H_{-X^*+\theta}(t)]^2 \\ &= \arg\min_{\theta} \{\mathrm{E}\,|X_1^* - \theta + X_2^* - \theta| - \mathrm{E}\,|X_1^* + X_2^*|\} \end{aligned}$$

Estimating θ_0 using the basic result

Recall that $X^* \sim \hat{F}_n$, the empirical distribution.

$$
\begin{aligned}
\text{Let } \hat{\theta} &= \arg\min_{\theta} \int_{-\infty}^{\infty} [H_{X^*-\theta}(t) - H_{-X^*+\theta}(t)]^2 \\
&= \arg\min_{\theta} \{\mathrm{E}\, |X_1^* - \theta + X_2^* - \theta| - \mathrm{E}\, |X_1^* + X_2^*|\} \\
&= \arg\min_{\theta} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i + x_j - 2\theta|
\end{aligned}
$$

Estimating θ_0 using the basic result

Recall that $X^* \sim \hat{F}_n$, the empirical distribution.

$$
\begin{aligned}
\text{Let } \hat{\theta} &= \arg\min_{\theta} \int_{-\infty}^{\infty} [H_{X^*-\theta}(t) - H_{-X^*+\theta}(t)]^2 \\
&= \arg\min_{\theta} \{\mathrm{E}\,|X_1^* - \theta + X_2^* - \theta| - \mathrm{E}\,|X_1^* + X_2^*|\} \\
&= \arg\min_{\theta} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i + x_j - 2\theta| \\
&= \text{median}_{i,j} \frac{x_i + x_j}{2}
\end{aligned}
$$

Estimating θ_0 using the basic result

Recall that $X^* \sim \hat{F}_n$, the empirical distribution.

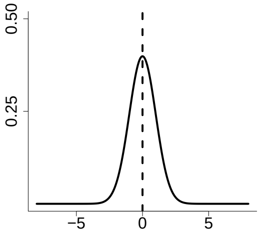
$$
\begin{aligned}
\text{Let } \hat{\theta} &= \arg\min_{\theta} \int_{-\infty}^{\infty} [H_{X^*-\theta}(t) - H_{-X^*+\theta}(t)]^2 \\
&= \arg\min_{\theta} \{\mathrm{E}\, |X_1^* - \theta + X_2^* - \theta| - \mathrm{E}\, |X_1^* + X_2^*|\} \\
&= \arg\min_{\theta} \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |x_i + x_j - 2\theta| \\
&= \text{median}_{i,j} \frac{x_i + x_j}{2}
\end{aligned}
$$

- It's (almost) the Hodges-Lehmann estimator!

K-component location mixture

Suppose that

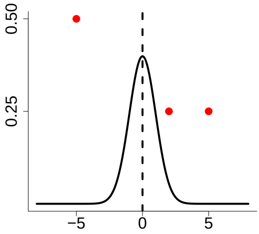
- $G(x)$ is some distribution



K-component location mixture

Suppose that

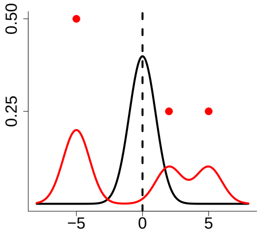
- $G(x)$ is some distribution
- θ is a K-vector w $\theta_1 < \cdots < \theta_K$
- λ is a K-vector with $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$
- $\delta_\theta(\lambda)$ is the discrete K-point distribution supported on θ with weights λ



K-component location mixture

Suppose that

- $G(x)$ is some distribution
- θ is a K-vector w $\theta_1 < \cdots < \theta_K$
- λ is a K-vector with $\lambda_i \geq 0$, $\sum_i \lambda_i = 1$
- $\delta_\theta(\lambda)$ is the discrete K-point distribution supported on θ with weights λ



Then $G \star \delta_\theta(\lambda)$ is a K-component location mixture whose jth component is $G(x - \theta_j)$ with weight λ_j.

Definition of identifiability

Let $F = G \star \delta_{\theta_0}(\lambda_0)$ be a K-component location mixture.

- Then F is identifiable (technically, K-identifiable) if

$$F = G^* \star \delta_{\theta^*}(\lambda^*)$$

implies that $G^* \overset{\mathcal{D}}{=} G$ and $\theta^* = \theta_0$ and $\lambda^* = \lambda_0$.

Definition of identifiability

Let $F = G \star \delta_{\theta_0}(\lambda_0)$ be a K-component location mixture.

- Then F is identifiable (technically, K-identifiable) if

$$F = G^* \star \delta_{\theta^*}(\lambda^*)$$

implies that $G^* \stackrel{\mathcal{D}}{=} G$ and $\theta^* = \theta_0$ and $\lambda^* = \lambda_0$.

Motivating Question (revised)

If we insist that G and G^* be zero-symmetric, is F always identifiable?

Definition of identifiability

Let $F = G \star \delta_{\theta_0}(\lambda_0)$ be a K-component location mixture.

- Then F is identifiable (technically, K-identifiable) if

$$F = G^* \star \delta_{\theta^*}(\lambda^*)$$

implies that $G^* \stackrel{\mathcal{D}}{=} G$ and $\theta^* = \theta_0$ and $\lambda^* = \lambda_0$.

Motivating Question (revised)

If we insist that G and G^* be zero-symmetric, is F always identifiable?

Answer: NO.

Definition of identifiability

Let $F = G \star \delta_{\theta_0}(\lambda_0)$ be a K-component location mixture.

- Then F is identifiable (technically, K-identifiable) if

$$F = G^* \star \delta_{\theta^*}(\lambda^*)$$

implies that $G^* \stackrel{\mathcal{D}}{=} G$ and $\theta^* = \theta_0$ and $\lambda^* = \lambda_0$.

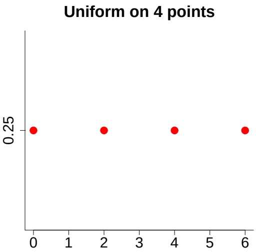
Motivating Question (revised)

If we insist that G and G^* be zero-symmetric, is F always identifiable?

Answer: NO. Not even if $\theta_1 < \cdots < \theta_K$ and all $\lambda_j > 0$.

A non-identifiable mixture ($K = 2$)

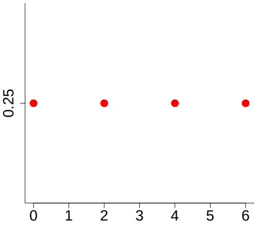
Zero-symmetric distribution G	Two-point distribution δ	Convolution $G \star \delta$
		uniform on $\{0, 2, 4, 6\}$
		uniform on $\{0, 2, 4, 6\}$



A non-identifiable mixture ($K = 2$)

Zero-symmetric distribution G	Two-point distribution δ	Convolution $G \star \delta$
uniform on $\{-1, 1\}$	uniform on $\{1, 5\}$	uniform on $\{0, 2, 4, 6\}$
		uniform on $\{0, 2, 4, 6\}$

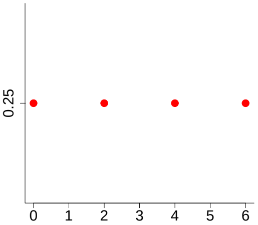
Uniform on 4 points



A non-identifiable mixture ($K = 2$)

Zero-symmetric distribution G	Two-point distribution δ	Convolution $G \star \delta$
uniform on $\{-1, 1\}$	uniform on $\{1, 5\}$	uniform on $\{0, 2, 4, 6\}$
uniform on $\{-2, 2\}$	uniform on $\{2, 4\}$	uniform on $\{0, 2, 4, 6\}$

Uniform on 4 points



Results on identifiability

Theorem

If G is assumed zero-symmetric and $\delta_{-\theta_0}(\lambda_0)$ is the unique K-point distribution such that

$$\delta_{\theta_0}(\lambda_0) \star \delta_{-\theta_0}(\lambda_0)$$

is zero-symmetric, then

$$F = G \star \delta_{\theta_0}(\lambda_0)$$

is K-identifiable.

Results on identifiability

Theorem

Suppose $K = 2$. If $\lambda_1 \notin \{0, \frac{1}{2}, 1\}$, then $F = G \star \delta_{\theta_0}(\lambda_0)$ is 2-identifiable. The converse is also true.

Theorem

Suppose $K = 3$. If $\lambda_1 \lambda_2 \lambda_3 \neq 0$ and

$$\frac{\theta_2 - \theta_1}{\theta_3 - \theta_2} \notin \left\{ \frac{1}{3}, \frac{1}{2}, 1, 2, 3 \right\},$$

then $F = G \star \delta_{\theta_0}(\lambda_0)$ is 3-identifiable. The converse is not true.

Estimating λ_0 and θ_0

Note that

$$F \star \delta_{-\theta_0}(\lambda_0) \quad = \quad G \star \delta_{\theta_0}(\lambda_0) \star \delta_{-\theta_0}(\lambda_0)$$

is zero-symmetric.

Estimating λ_0 and θ_0

Note that

$$F \star \delta_{-\theta_0}(\lambda_0) \quad = \quad G \star \delta_{\theta_0}(\lambda_0) \star \delta_{-\theta_0}(\lambda_0)$$

is zero-symmetric.

So take $W \sim \hat{F}_n \star \delta_{-\theta}(\lambda)$ and minimize

$$\int_{-\infty}^{\infty} [H_W(t) - H_{-W}(t)]^2 = \mathrm{E}\,|W_1 + W_2| - \mathrm{E}\,|W_1 - W_2|.$$

Estimating λ_0 and θ_0

Letting W_1 and W_2 be independent copies of $\hat{F}_n \star \delta_{-\theta}(\lambda)$, we take $(\hat{\lambda}, \hat{\theta})$ to minimize

$$
\begin{aligned}
& \mathrm{E}\left|W_1 + W_2\right| - \mathrm{E}\left|W_1 - W_2\right| \\
= \quad & \frac{1}{n^2} \sum_{a=1}^{K} \sum_{b=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_a \lambda_b \left| x_i + x_j - (\theta_a + \theta_b) \right| \\
& - \frac{1}{n^2} \sum_{a=1}^{K} \sum_{b=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_a \lambda_b \left| x_i - x_j - (\theta_a - \theta_b) \right| .
\end{aligned}
$$

Estimating $G(x)$ when $K = 2$

If $X \sim F$, then

$$\begin{aligned} H_X(t) &= \lambda_0 G(t-\theta_{01}) + (1-\lambda_0) G(t-\theta_{02}) \\ H_{-X}(t-\theta_{01}-\theta_{02}) &= \lambda_0 G(t-\theta_{02}) + (1-\lambda_0) G(t-\theta_{01}) \end{aligned}$$

Estimating $G(x)$ when $K = 2$

If $X \sim F$, then

$$\begin{pmatrix} H_X(t) \\ H_{-X}(t - \theta_{01} - \theta_{02}) \end{pmatrix} = \begin{pmatrix} \lambda_0 & 1 - \lambda_0 \\ 1 - \lambda_0 & \lambda_0 \end{pmatrix} \begin{pmatrix} G(t - \theta_{01}) \\ G(t - \theta_{02}) \end{pmatrix}$$

Invert (remember, $\lambda_0 \neq \frac{1}{2}$) to give two expressions for $G(t)$. Average them.

Estimating $G(x)$ when $K = 2$

If $X \sim F$, then

$$\frac{1}{2\lambda_0 - 1}\begin{pmatrix} \lambda_0 & \lambda_0 - 1 \\ \lambda_0 - 1 & \lambda_0 \end{pmatrix}\begin{pmatrix} H_X(t) \\ H_{-X}(t - \theta_{01} - \theta_{02}) \end{pmatrix} = \begin{pmatrix} G(t - \theta_{01}) \\ G(t - \theta_{02}) \end{pmatrix}$$

Invert (remember, $\lambda_0 \neq \frac{1}{2}$) to give two expressions for $G(t)$. Average them.

Then replace parameters (including H_X) by estimates.

Estimating $G(x)$ when $K = 2$

The resulting $\hat{G}(t)$ satisfies:

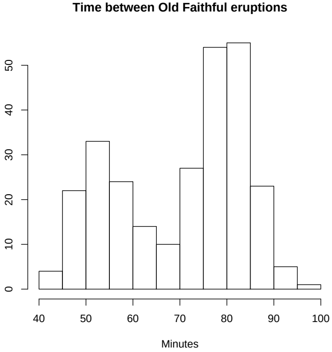
- $\lim_{t\to\infty} \hat{G}(t) = 1$
- $\lim_{t\to-\infty} \hat{G}(t) = 0$
- $\hat{G}(t) = 1 - \hat{G}(-t)$ at points of continuity (zero-symmetry)
- $\sup_t |\hat{G}(t) - G(t)| \to 0$ as $n \to \infty$

…but $\hat{G}(t)$ is not monotone.

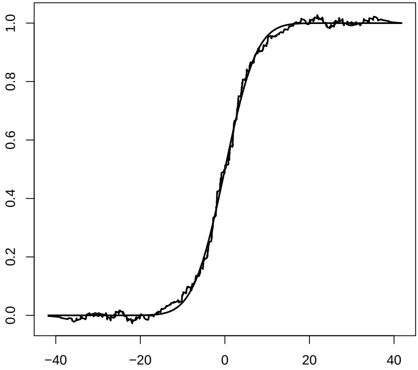
Old Faithful data: Two estimates of (λ_0, θ_0)

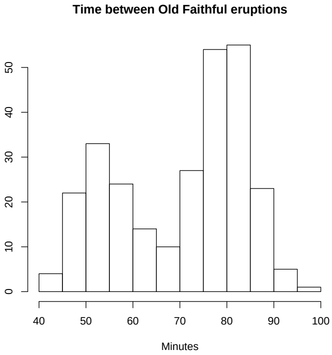
	NP	NML
$\hat{\theta}_1$ (SE)	54.0 (.76)	54.6 (.67)
$\hat{\theta}_2$ (SE)	80.0 (.50)	80.1 (.45)
$\hat{\lambda}$ (SE)	.352 (.032)	.361 (.032)
$\hat{\sigma}^2$ (SE)	30.7 (7.9)	34.5 (3.4)

NP: Nonparametric method
NML: Normal mixture assumed
SE: Bootstrapped, based on 200 resamples

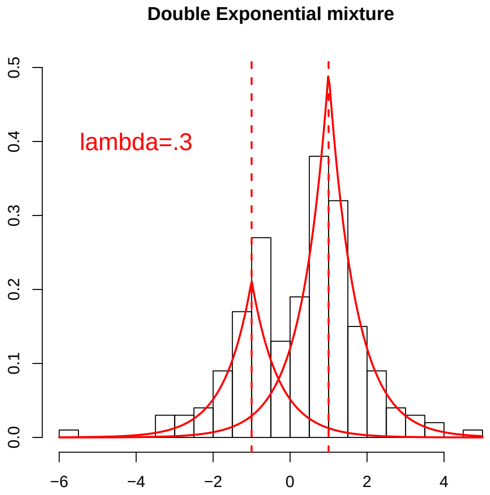


Old Faithful data: Two estimates of $G(t)$





Double exponential data: Two estimates of (λ_0, θ_0)



	NP	NML
$\hat{\theta}_1$	-1.04	-5.48
$\hat{\theta}_2$	0.97	0.33
$\hat{\lambda}$	0.33	0.006

NP: Nonparametric
NML: Normal MLE

Same problem, different approach

- L. Bourdes, S. Mottelet, P. Vandekerkhove (2006) "Semiparametric estimation of a two-component mixture", *Annals of Statistics*, to appear.

Same problem, different approach

- L. Bourdes, S. Mottelet, P. Vandekerkhove (2006) "Semiparametric estimation of a two-component mixture", *Annals of Statistics*, to appear.

Open problem

How can one improve the numerical optimization routine?

- L. Bourdes, D. Chaveau, P. Vandekerkhove (2006) "An EM algorithm for a semiparametric mixture model", unpublished manuscript.