Active Set Methods for Log-Concave Densities and Nonparametric Tail Inflation

Lutz Duembgen (Bern)

Alexandre Moesching and Christof Straehl (Bern)

Peter McCullagh and Nicholas G. Polson (Chicago)

January 2018

The general setting

Data. Summarized as a distribution

$$\widehat{P} = \sum_{i=1}^{n} w_i \delta_{x_i}$$

with

- weights $w_1, w_2, \ldots, w_n > 0$
- support points $x_1 < x_2 < \cdots < x_n$ in an open interval $\mathcal{X} \subset \mathbb{R}$

The general setting

Data. Summarized as a distribution

$$\widehat{P} \;=\; \sum_{i=1}^{n} w_i \delta_{x_i}$$

with

- weights $w_1, w_2, \ldots, w_n > 0$
- support points $x_1 < x_2 < \cdots < x_n$ in an open interval $\mathcal{X} \subset \mathbb{R}$

Model. $\widehat{P}$ estimates distribution

$$P(dx) \;=\; e^{\theta(x)}\, M(dx)$$

with

- given measure M on $\mathcal{X}$
- unknown function θ in given family Θ_1

Goal. Estimate $\theta \in \Theta_1$ via MLE

$$\widehat{\theta} \in \underset{\theta \in \Theta_1}{\arg\max} \int \theta \, d\widehat{P}$$

Goal. Estimate $\theta \in \Theta_1$ via MLE

$$\widehat{\theta} \;\in\; \arg\max_{\theta \in \Theta_1} \int \theta \, d\widehat{P}$$

Modification. Suppose that for a larger function family Θ

$$\Theta_1 \;=\; \left\{ \theta \in \Theta : \int e^{\theta} \, dM = 1 \right\}$$

$$\theta + c \;\in\; \Theta \quad \text{for all } \theta \in \Theta, c \in \mathbb{R}$$

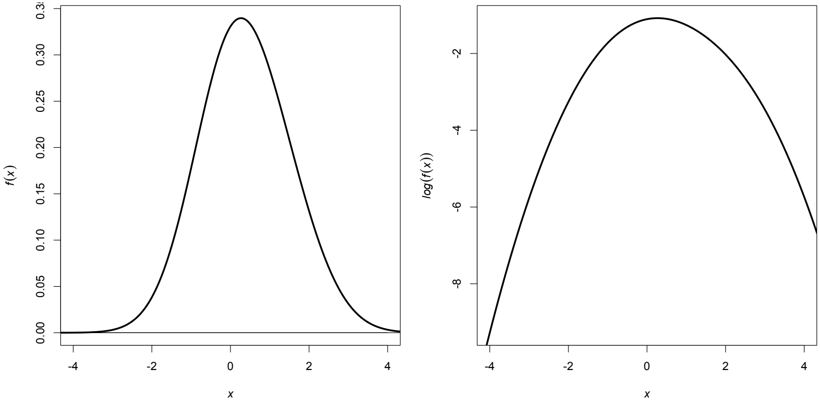
Then

$$\widehat{\theta} \;\in\; \arg\max_{\theta \in \Theta} \left(\int \theta \, d\widehat{P} - \int e^{\theta} \, dM \right)$$

Setting 1: Log-concave densities

P has log-concave density on $\mathcal{X}$, i.e.

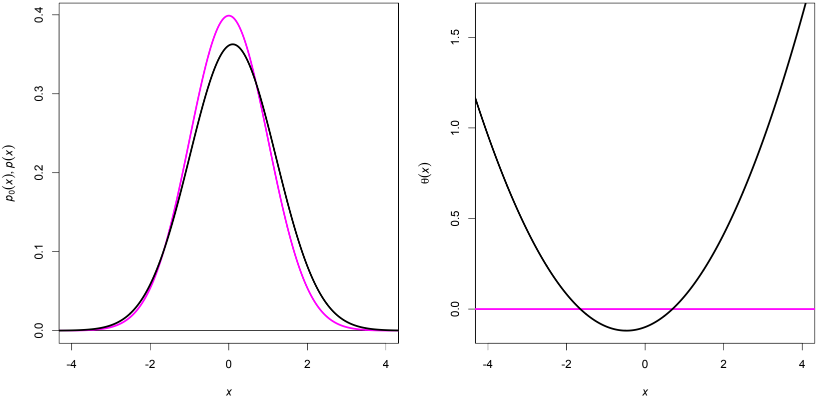
- $M =$ Lebesgue measure on $\mathcal{X}$
- $\Theta = \{\theta : \mathcal{X} \to [-\infty, \infty)$ concave and u.s.c.$\}$



Setting 2A: Tail inflation

P has log-convex density w.r.t. given distribution P_0 on $\mathcal{X}$, i.e.

- $M = P_0$
- $\Theta = \{\theta : \mathcal{X} \to \mathbb{R} \text{ convex}\}$



Example. Observe

$$X_i \;=\; \mu_i + \sigma_i \varepsilon_i, \quad 1 \le i \le n,$$

with unknown parameters $\mu_i \in \mathbb{R}$, $\sigma_i \ge 1$ and independent r.v.s

$$\varepsilon_i \;\sim\; P_0 := \mathcal{N}(0,1)$$

Example. Observe

$$X_i \;=\; \mu_i + \sigma_i \varepsilon_i, \quad 1 \le i \le n,$$

with unknown parameters $\mu_i \in \mathbb{R}$, $\sigma_i \ge 1$ and independent r.v.s

$$\varepsilon_i \;\sim\; P_0 := \mathcal{N}(0,1)$$

Marginal dist. $P := n^{-1} \sum_{i=1} \mathcal{L}(X_i)$:

$$\theta(x) \;:=\; \log \frac{dP}{dP_0}(x) \;=\; \log\Big(\frac{1}{n}\sum_{i=1}^{n} e^{\theta_i(x)}\Big)$$

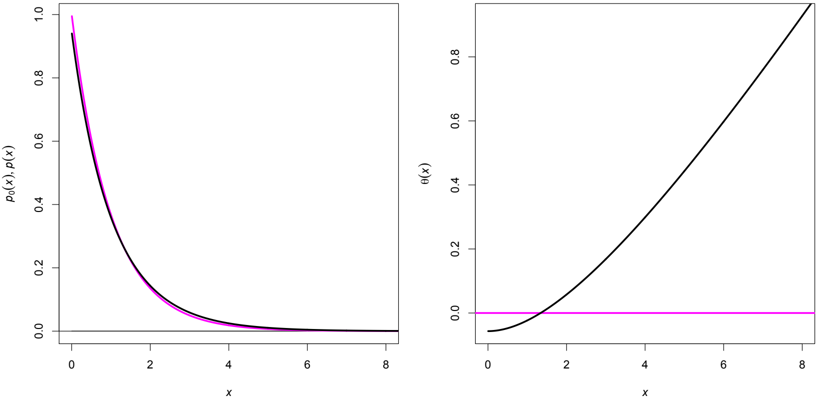
$$\theta_i(x) \;:=\; -\log(\sigma_i) + \frac{x^2}{2} - \frac{(x-\mu_i)^2}{2\sigma_i^2}$$

$$\theta_i'', \theta'' \;\ge\; 0 \qquad \text{(Artin's theorem)}$$

Setting 2B: Tail inflation (McCullagh and Polson 2012)

Assume that P has log-convex and isotonic density w.r.t. given distribution P_0 on $\mathcal{X}$, i.e.

- $M = P_0$
- $\Theta = \{\theta : \mathcal{X} \to \mathbb{R} \text{ convex and isotonic}\}$.



Example. Observe

$$X_i = S_i \varepsilon_i, \quad 1 \le i \le n,$$

with independent r.v.s

$$\varepsilon_i \sim \mathcal{N}(0,1), \quad S_i \ge 1.$$

Example. Observe

$$X_i = S_i \varepsilon_i, \quad 1 \le i \le n,$$

with independent r.v.s

$$\varepsilon_i \sim \mathcal{N}(0,1), \quad S_i \ge 1.$$

Then

$$Y_i := X_i^2 \sim \begin{cases} P_0 = \chi_1^2 = \mathrm{Gamma}(1/2, 2) & \text{if } S_i \equiv 1 \\ \mathbb{E}\,\mathrm{Gamma}(1/2, 2S_i^2) & \text{in general} \end{cases}$$

Example. Observe

$$X_i = S_i \varepsilon_i, \quad 1 \le i \le n,$$

with independent r.v.s

$$\varepsilon_i \sim \mathcal{N}(0,1), \quad S_i \ge 1.$$

Then

$$Y_i := X_i^2 \sim \begin{cases} P_0 = \chi_1^2 = \mathrm{Gamma}(1/2, 2) & \text{if } S_i \equiv 1 \\ \mathbb{E}\,\mathrm{Gamma}(1/2, 2S_i^2) & \text{in general} \end{cases}$$

Marginal dist. $P := n^{-1} \sum_{i=1} \mathcal{L}(Y_i)$:

$$\theta := \log \frac{dP}{dP_0} \quad \text{is convex and isotonic}$$

Existence and uniqueness of $\widehat{\theta}$

Lemma 1 (Log-concavity) In Setting 1,

$$\exists! \quad \widehat{\theta} \in \operatorname*{arg\,max}_{\theta \in \Theta} \left(\int \theta \, d\widehat{P} - \int_{\mathcal{X}} e^{\theta(x)} \, dx \right).$$

Existence and uniqueness of $\widehat{\theta}$

Lemma 1 (Log-concavity) In Setting 1,

$$\exists! \quad \widehat{\theta} \in \operatorname*{arg\,max}_{\theta \in \Theta} \left(\int \theta \, d\widehat{P} - \int_{\mathcal{X}} e^{\theta(x)} \, dx \right).$$

Additional properties:

- $\widehat{\theta}$ piecewise linear on $[x_1, x_n]$
- changes of slope only in $\{x_2, \ldots, x_{n-1}\}$
- $\widehat{\theta} \equiv -\infty$ on $\mathbb{R} \setminus [x_1, x_n]$

Lemma 2. In Settings 2A and 2B,

$$\exists! \quad \widehat{\theta} \;\in\; \underset{\theta \in \Theta}{\arg\max} \left(\int \theta \, d\widehat{P} - \int e^{\theta} \, dP_0 \right)$$

provided that $\mathrm{supp}(P_0) = \mathcal{X}$.

Lemma 2. In Settings 2A and 2B,

$$\exists! \quad \widehat{\theta} \in \operatorname*{arg\,max}_{\theta \in \Theta} \left(\int \theta \, d\widehat{P} - \int e^{\theta} \, dP_0 \right)$$

provided that $\mathrm{supp}(P_0) = \mathcal{X}$.

Additional properties:

- $\widehat{\theta}$ piecewise linear on $\mathcal{X}$
- changes of slope only in $\bigcup_{i=1}^{n-1} (x_i, x_{i+1})$
- at most one change of slope in (x_i, x_{i+1}), $1 \leq i < n$

Active set algorithm

In Settings 1 and 2A,

$$\widehat{\theta} \in \mathbb{V} \cap \Theta$$

with

$$\mathbb{V} := \{\text{linear splines on } \mathcal{X}_o \text{ with kinks on } \mathcal{D}\}$$

$$\mathcal{X}_o := \begin{cases} [x_1, x_n] & \text{in Setting 1} \\ \mathcal{X} & \text{in Setting 2A} \end{cases}$$

$$\mathcal{D} := \begin{cases} \{x_2, \ldots, x_{n-1}\} & \text{in Setting 1} \\ \mathcal{X} & \text{in Setting 2A} \end{cases}$$

For such a spline $v \in \mathbb{V}$ set

$$D(v) \ := \{\tau \in \mathcal{D} : v'(\tau-) \neq v'(\tau+)\}$$

(deactivated (equality) constraints)

For such a spline $v \in \mathbb{V}$ set

$$D(v) \;:=\; \{\tau \in \mathcal{D} : v'(\tau-) \neq v'(\tau+)\}$$

(deactivated (equality) constraints)

For finite set $D \subset \mathcal{D}$ define

$$\mathbb{V}_D \;:=\; \{v \in \mathbb{V} : D(v) \subset D\}$$

a linear space with

$$\dim(\mathbb{V}_D) \;=\; 2 + \#D$$

Target functional

$$L(\theta) \;:=\; \int \theta \, d\widehat{P} - \int e^{\theta} \, dM$$

with directional derivatives

$$DL(\theta, v) \;:=\; \lim_{t \to 0+} \frac{L(\theta + tv) - L(\theta)}{t}$$

Target functional

$$L(\theta) \;:=\; \int \theta \, d\widehat{P} - \int e^{\theta} \, dM$$

with directional derivatives

$$DL(\theta, v) \;:=\; \lim_{t \to 0+} \frac{L(\theta + tv) - L(\theta)}{t}$$

Characterization of $\widehat{\theta}$. $\theta \in \mathbb{V} \cap \Theta$ equals $\widehat{\theta}$ if, and only if,

$$DL(\theta, v) \;\le\; 0 \quad \text{whenever } \theta + tv \in \Theta \text{ for some } t > 0.$$

Local Search (Shape-constrained Newton). Let $\theta \in \mathbb{V} \cap \Theta$.

$\theta_{\text{new}} \leftarrow \mathsf{Newton}(\theta, \mathbb{V}_{D(\theta)})$
$\delta \leftarrow DL(\theta, \theta_{\text{new}} - \theta)$

while $\delta > \delta_o$ do

while $L(\theta_{\text{new}}) < L(\theta) + \delta/3$ do
$\theta_{\text{new}} \leftarrow (\theta + \theta_{\text{new}})/2$
$\delta \leftarrow \delta/2$
end while

if $\theta_{\text{new}} \notin \Theta$ do
$t_o \leftarrow \max\left\{t \in (0,1] : (1-t)\theta + t\theta_{\text{new}} \in \Theta\right\}$
$\theta_{\text{new}} \leftarrow (1-t_o)\theta + t_o\theta_{\text{new}}$
end if

$\theta \leftarrow \theta_{\text{new}}$
$\theta_{\text{new}} \leftarrow \text{Newton}(\theta, \mathbb{V}_{D(\theta)})$
$\delta \leftarrow DL(\theta, \theta_{\text{new}} - \theta)$

end while

Essential properties of local search:

Essential properties of local search:

- $L(\theta)$ increases

Essential properties of local search:

- $L(\theta)$ increases
- $D(\theta)$ decreases

Essential properties of local search:

- $L(\theta)$ increases
- $D(\theta)$ decreases
- Eventually, θ is locally optimal (approx.):

$$\theta = \arg\max_{\eta \in \mathbb{V}_{D(\theta)}} L(\eta)$$

Checking optimality. Let $\theta \in \mathbb{V} \cap \Theta$ be locally optimal (approx.). Then

$$\theta = \widehat{\theta}$$

if, and only if,

$$DL(\theta, V_\tau) \le 0 \quad \text{for all } \tau \in \mathcal{D} \setminus D(\theta)$$

where

$$V_\tau(x) := \begin{cases} -(x-\tau)^+ & \text{in Setting 1} \\ +(x-\tau)^+ & \text{in Setting 2A} \end{cases}$$

Checking optimality. Let $\theta \in \mathbb{V} \cap \Theta$ be locally optimal (approx.). Then

$$\theta \;=\; \widehat{\theta}$$

if, and only if,

$$DL(\theta, V_\tau) \;\leq\; 0 \quad \text{for all } \tau \in \mathcal{D} \setminus D(\theta)$$

where

$$V_\tau(x) \;:=\; \begin{cases} -(x-\tau)^+ & \text{in Setting 1} \\ +(x-\tau)^+ & \text{in Setting 2A} \end{cases}$$

If not, determine $\tau(\theta) \in \mathcal{D} \setminus D(\theta)$ such that

$$0 \;<\; DL(\theta, V_{\tau(\theta)}) \;\approx\; \max_{\tau \in \mathcal{D} \setminus D(\theta)} DL(\theta, V_\tau)$$

and run a modified local search.

Modified local search.

$\theta_{\text{new}} \leftarrow \text{Newton}(\theta, \mathbb{V}_{\boldsymbol{D(\theta)\cup\{\tau(\theta)\}}})$
$\delta \leftarrow DL(\theta, \theta_{\text{new}} - \theta)$

while $\delta > \delta_o$ do

    while $L(\theta_{\text{new}}) < L(\theta) + \delta/3$ do
        $\theta_{\text{new}} \leftarrow (\theta + \theta_{\text{new}})/2$
        $\delta \leftarrow \delta/2$
    end while

    if $\theta_{\text{new}} \notin \Theta$ do
        $t_o \ \leftarrow \ \max\{t \in (0,1] : (1-t)\theta + t\theta_{\text{new}} \in \Theta\}$
        $\theta_{\text{new}} \leftarrow (1 - t_o)\theta + t_o\theta_{\text{new}}$
    end if

    $\theta \leftarrow \theta_{\text{new}}$
    $\theta_{\text{new}} \leftarrow \text{Newton}(\theta, \mathbb{V}_{D(\theta)})$
    $\delta \leftarrow DL(\theta, \theta_{\text{new}} - \theta)$

end while

Remarks.

- After finitely many (modified) local searches algorithm will stop at $\theta = \widehat{\theta}$ (approx.)

Remarks.

- After finitely many (modified) local searches algorithm will stop at $\theta = \widehat{\theta}$ (approx.)
- Replace simple kink functions $V_\tau = \pm(\cdot - \tau)^+$ with 'localized versions' to gain numerical precision

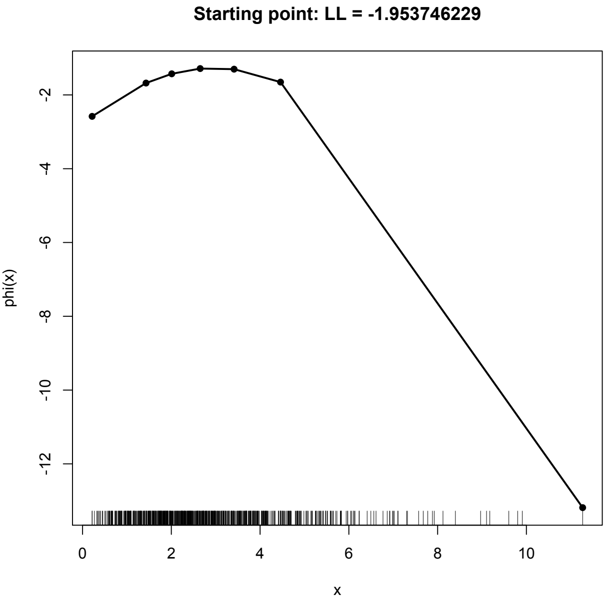
Remarks.

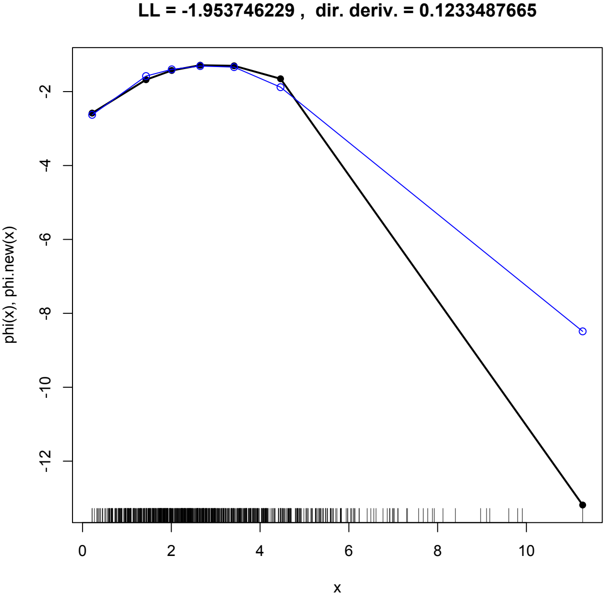
- After finitely many (modified) local searches algorithm will stop at $\theta = \widehat{\theta}$ (approx.)
- Replace simple kink functions $V_\tau = \pm(\cdot - \tau)^+$ with 'localized versions' to gain numerical precision
- In Settings 2A and 2B, the function

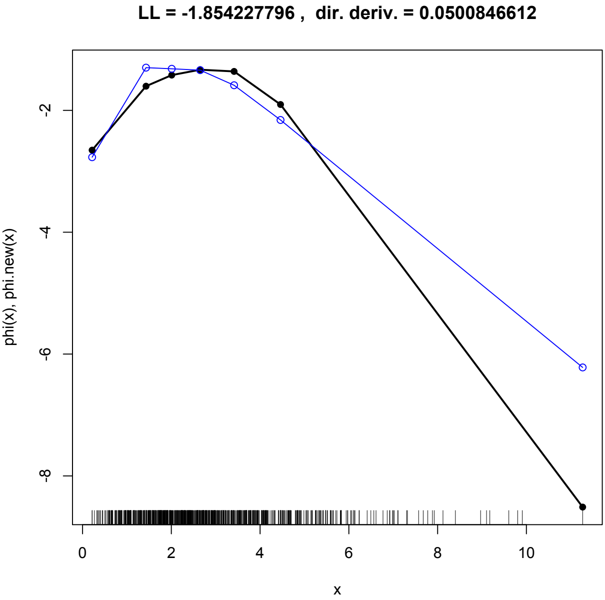
$$\tau \ \mapsto \ DL(\theta, V_\tau)$$

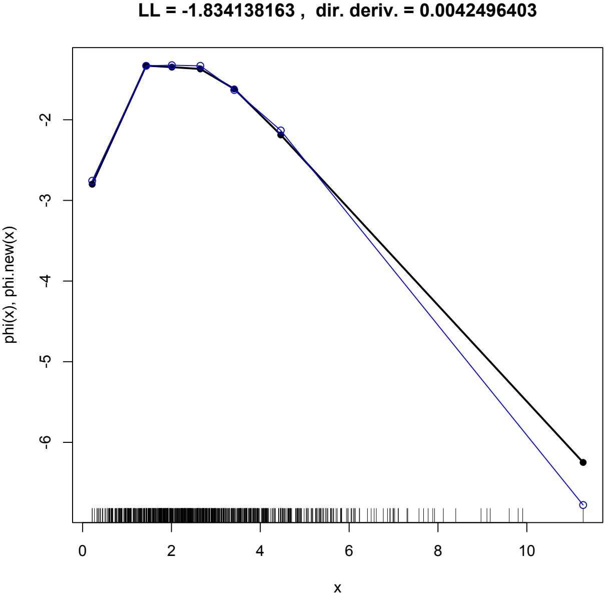
is strictly concave on any interval (x_i, x_{i+1}) ...

Example for Setting 1. $n = 800$ observations from $\mathrm{Gamma}(3, 1)$

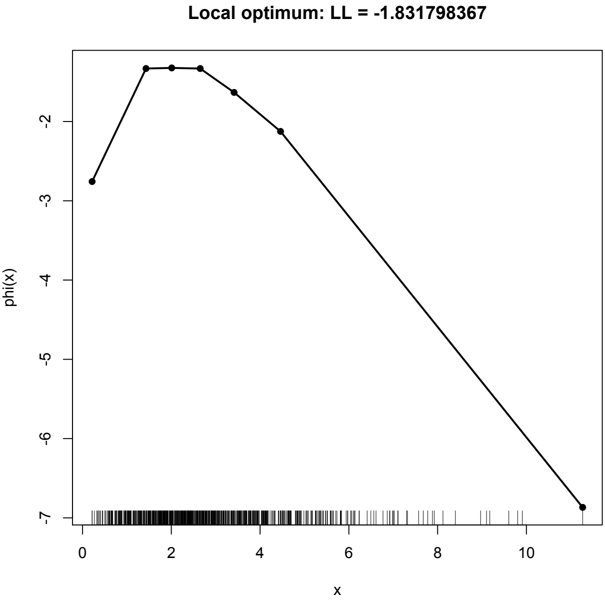



LL = -1.953746229 , dir. deriv. = 0.1233487665
phi(x), phi.new(x)
-2
-4
-6
-8
-10
-12
0
2
4
6
8
10
x

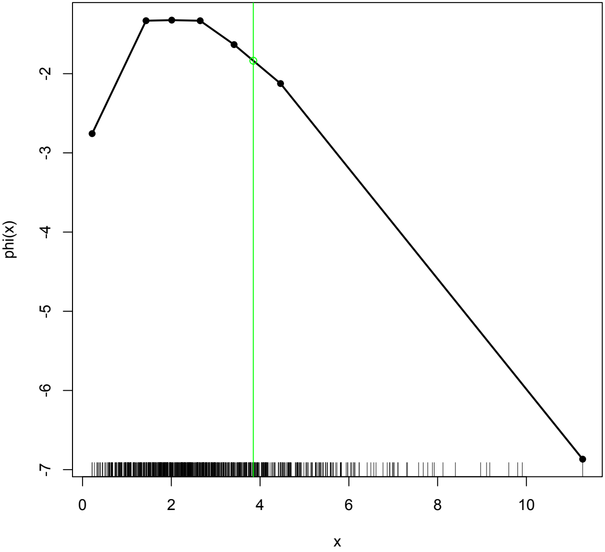

LL = -1.854227796 , dir. deriv. = 0.0500846612
phi(x), phi.new(x)
-2
-4
-6
-8
0
2
4
6
8
10
x


LL = -1.834138163 , dir. deriv. = 0.0042496403
phi(x), phi.new(x)
-2
-3
-4
-5
-6
0
2
4
6
8
10
x

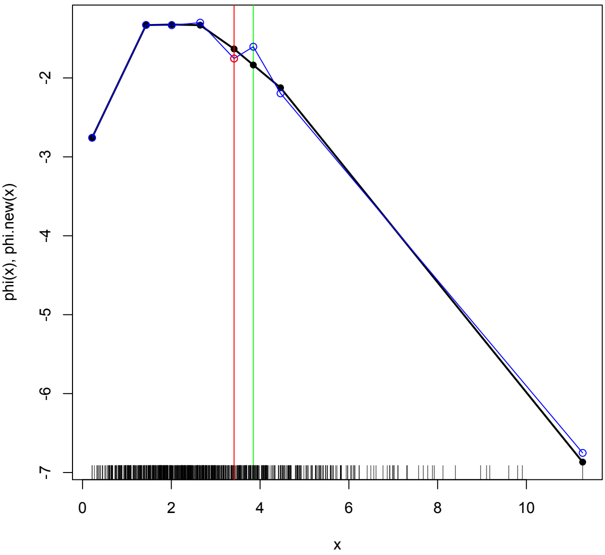
Newton proposals in dimensions $7, 7, 7, 7, 7, 7$

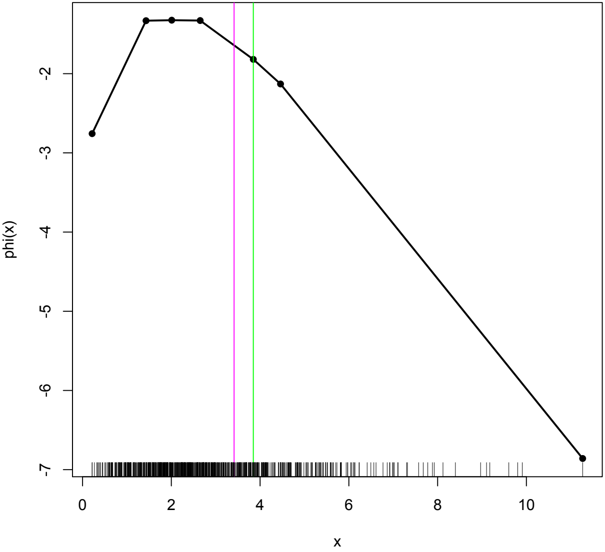


LL = -1.831798367 , dir. deriv. = 0.0025654828

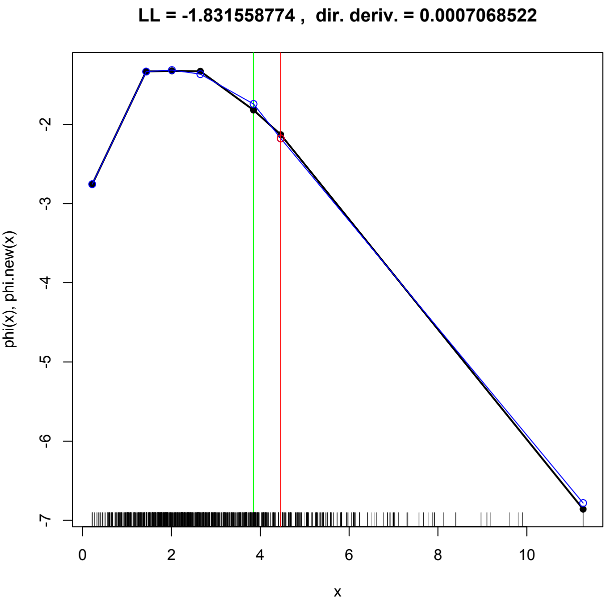


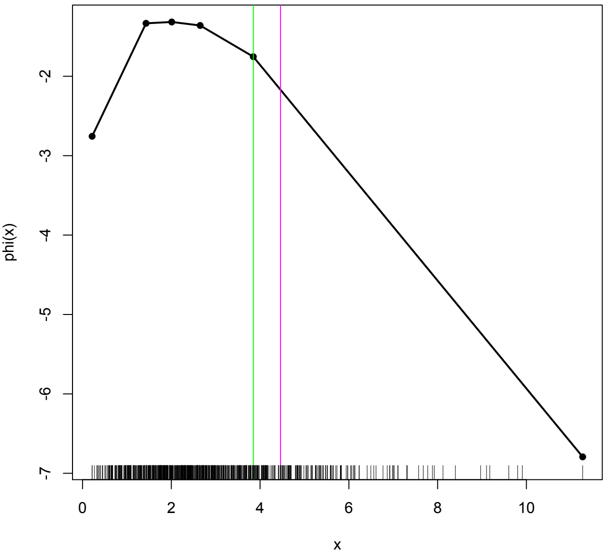
LL = -1.831798367 , dir. deriv. = 0.0033585062



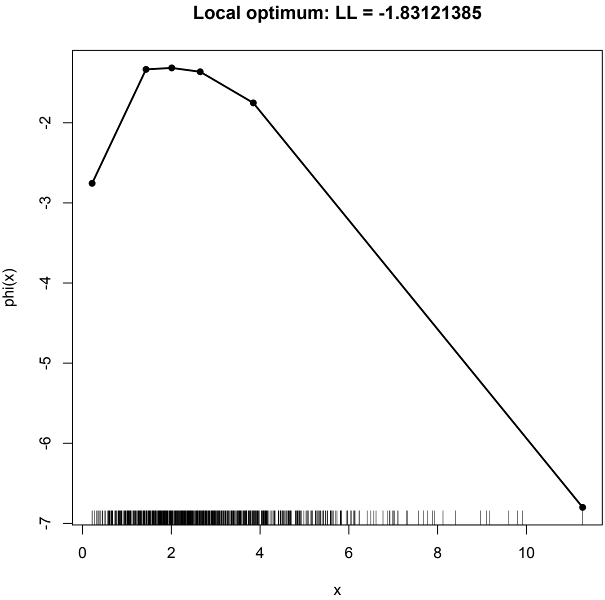


phi(x)
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x

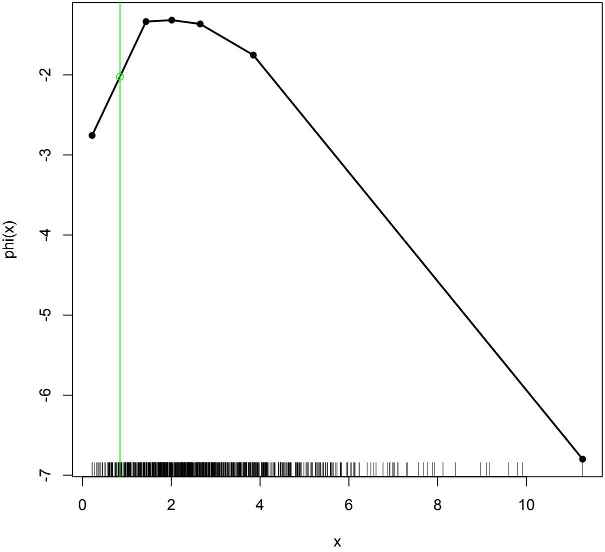

LL = -1.831558774 , dir. deriv. = 0.0007068522
phi(x), phi.new(x)
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x

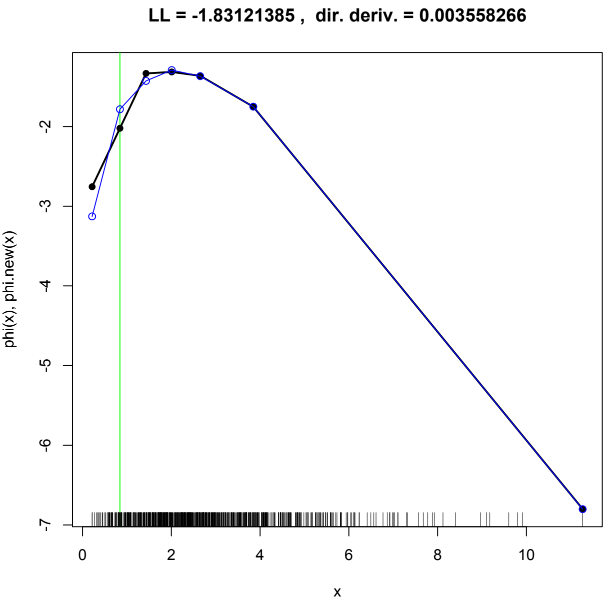

phi(x)
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x

Newton proposals in dimensions $8, 7, 6, 6$



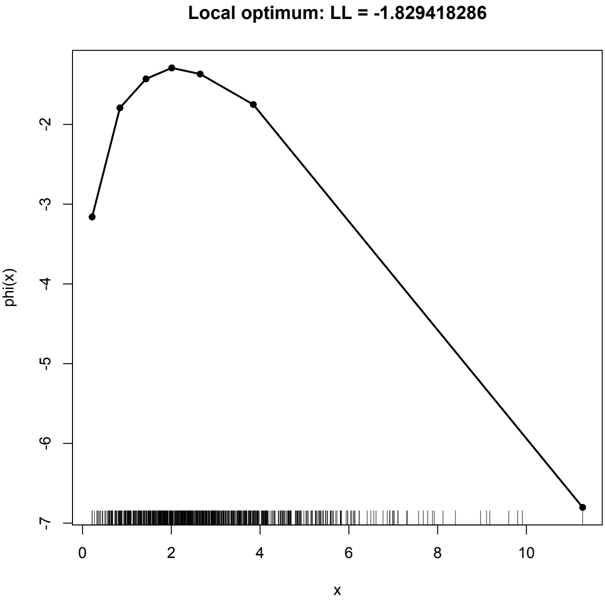
LL = -1.83121385 , dir. deriv. = 0.002308197

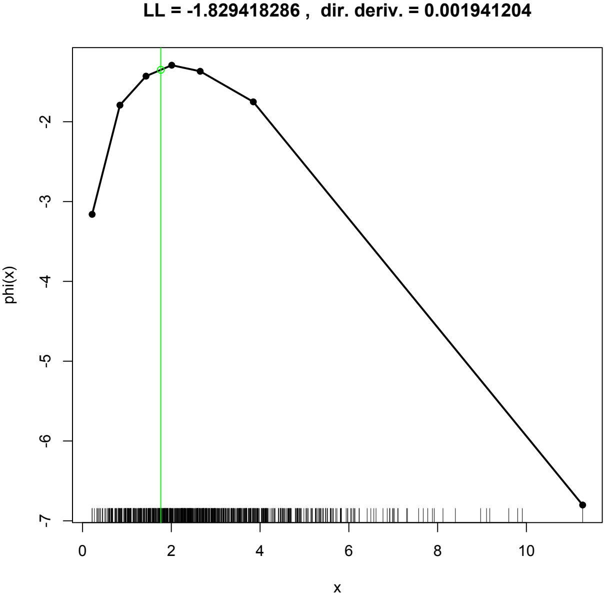




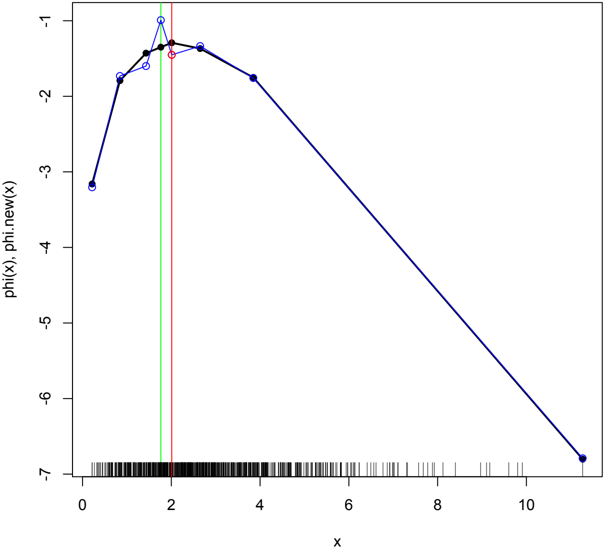
LL = -1.83121385 , dir. deriv. = 0.003558266
phi(x), phi.new(x)
x
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10

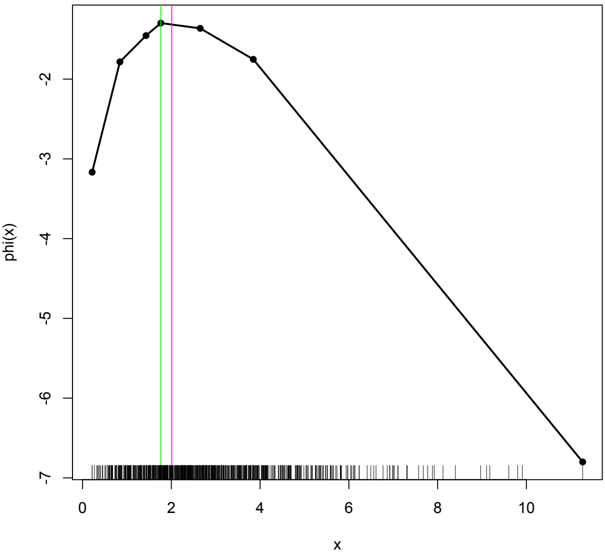
Newton proposals in dimensions $7, 7, 7, 7$



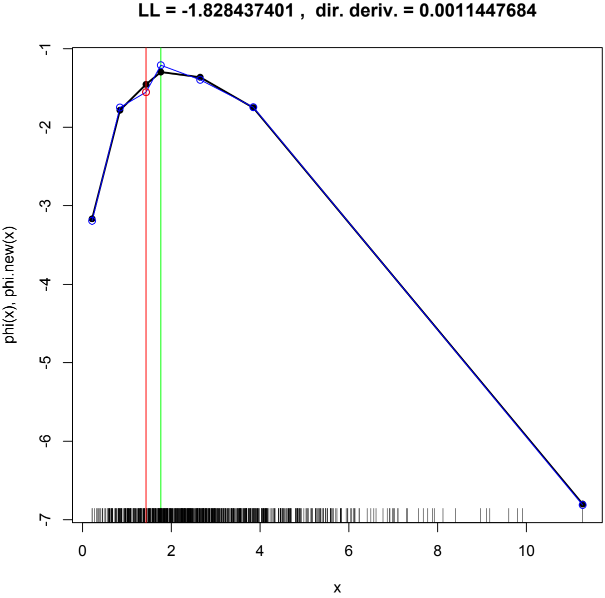

LL = -1.829418286 , dir. deriv. = 0.001941204
phi(x)
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x

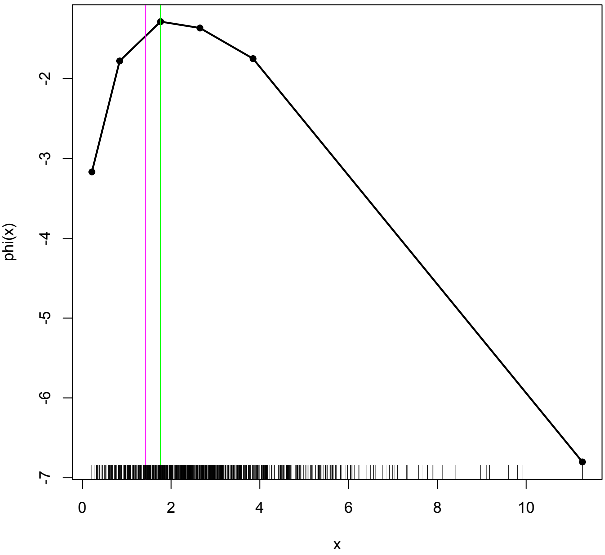
LL = -1.829418286 , dir. deriv. = 0.0071650095





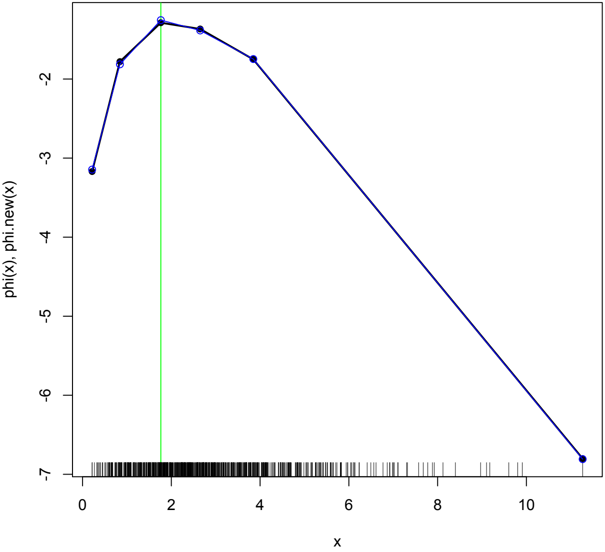
phi(x)
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x


LL = -1.828437401 , dir. deriv. = 0.0011447684
phi(x), phi.new(x)
-1
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x

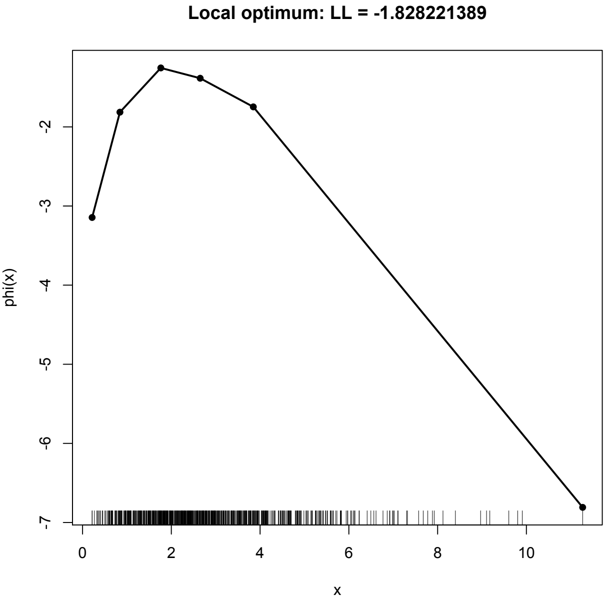


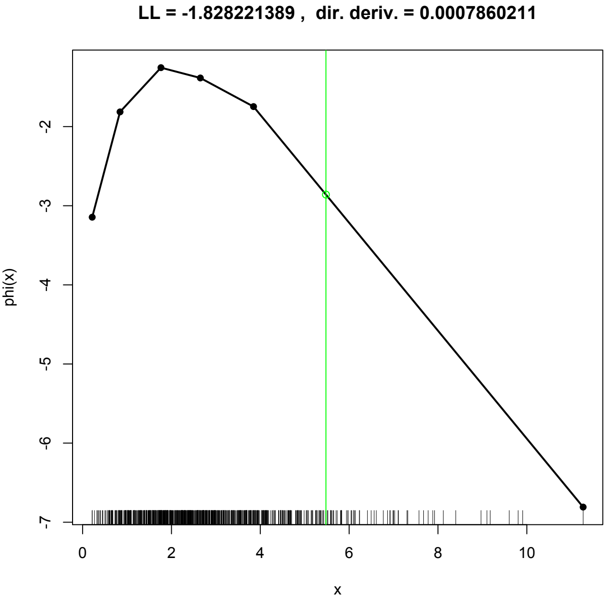
phi(x)
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x

LL = -1.828316099 , dir. deriv. = 0.0001895999

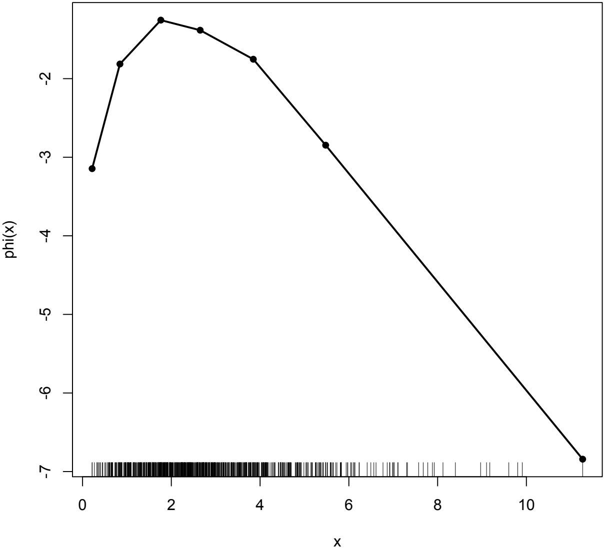


Newton proposals in dimensions $8, 7, 6, 6, 6$

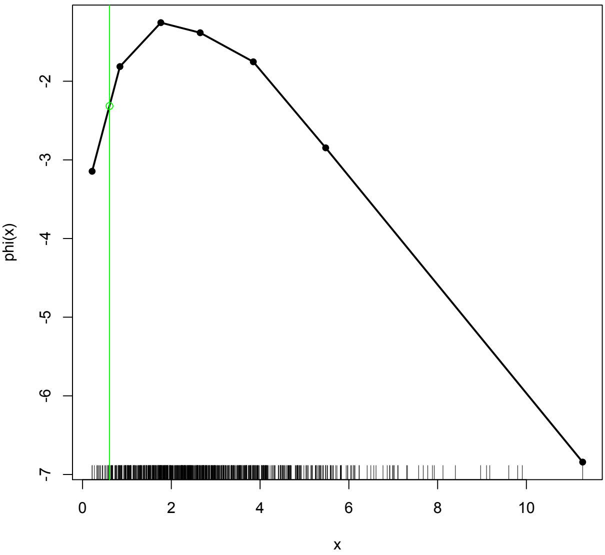


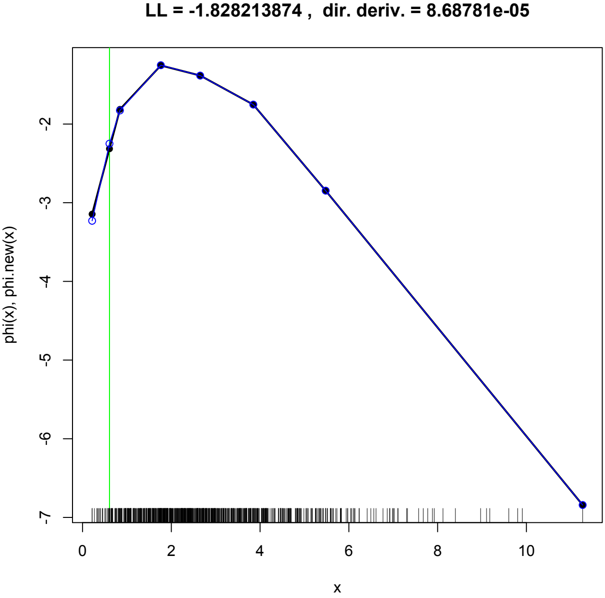

LL = -1.828221389 , dir. deriv. = 0.0007860211
phi(x)
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x

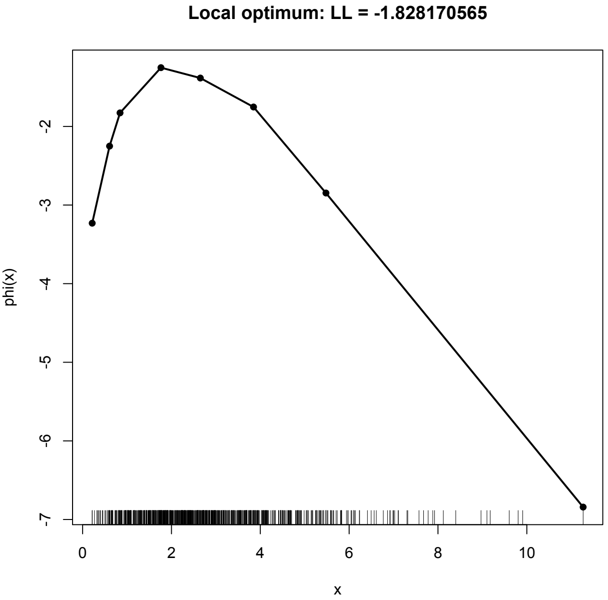
Local optimum: LL = -1.828213874



LL = -1.828213874 , dir. deriv. = 0.0001214971

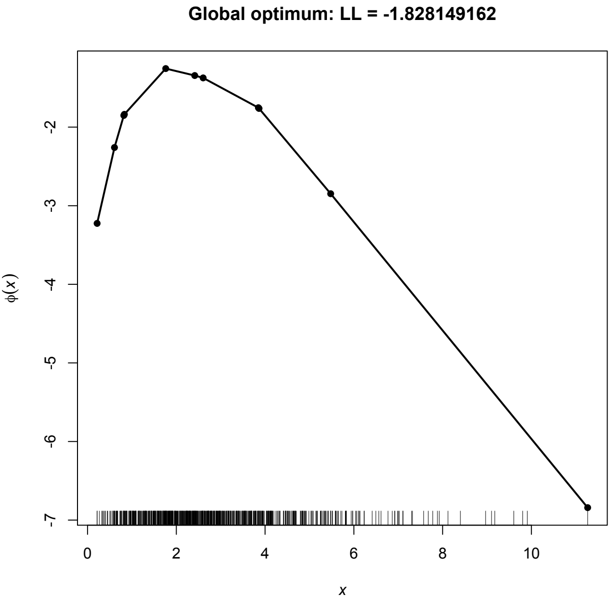



LL = -1.828213874 , dir. deriv. = 8.68781e-05
phi(x), phi.new(x)
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x

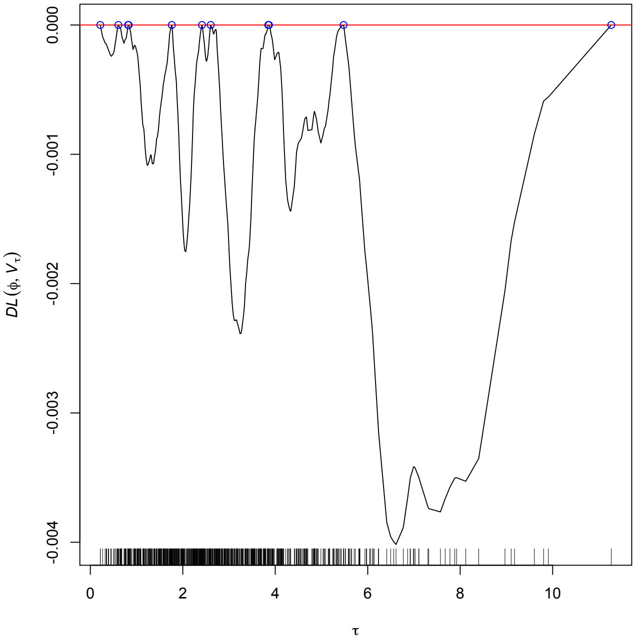

Local optimum: LL = -1.828170565
phi(x)
-2
-3
-4
-5
-6
-7
0
2
4
6
8
10
x

After 14 (modified) local searches, 45 Newton proposals:

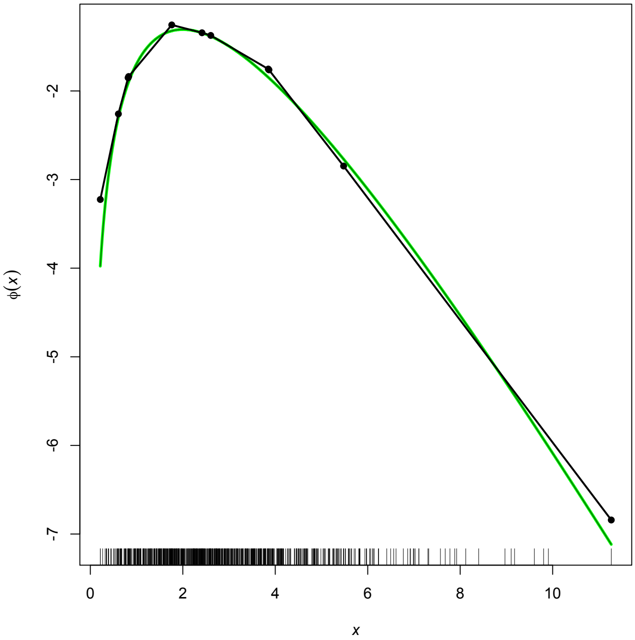
After 14 (modified) local searches, 45 Newton proposals:



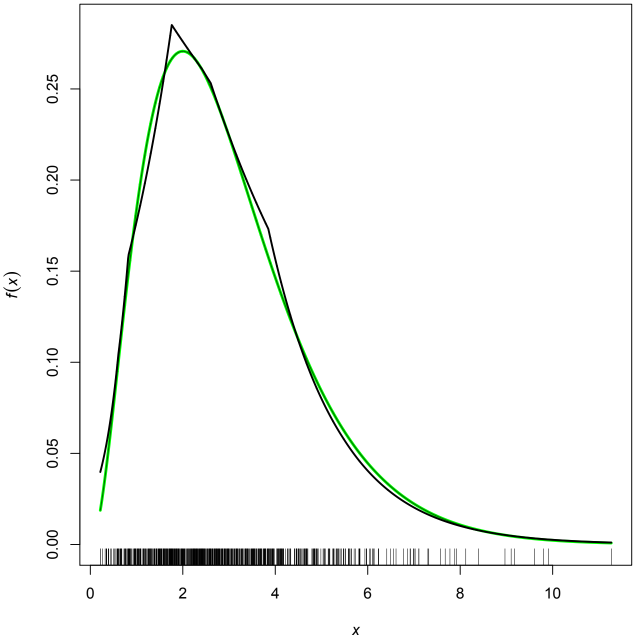
Directional derivatives for extra knot at τ:



Log-densities:

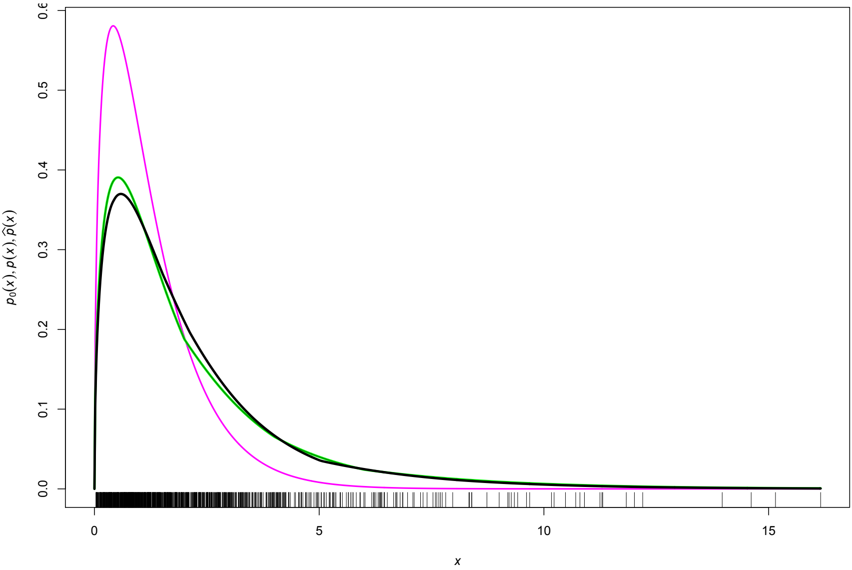


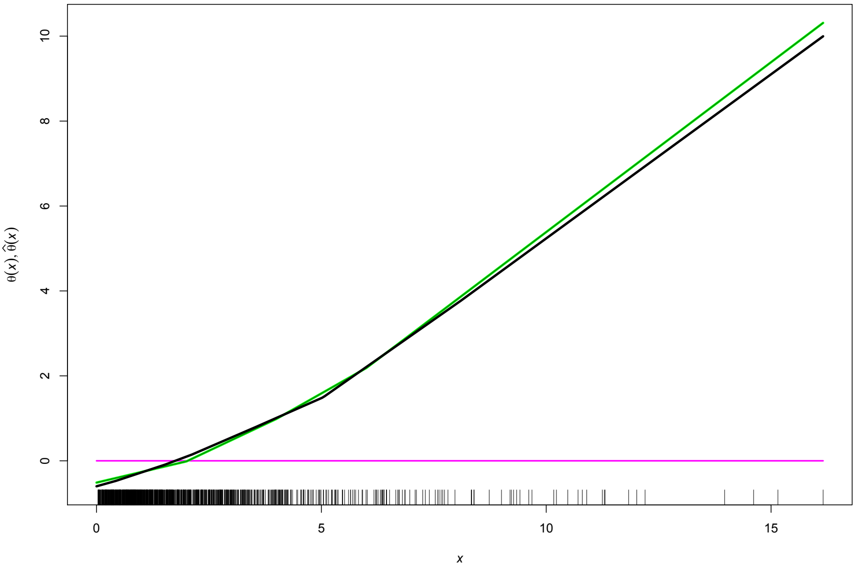
Densities:



Example for Setting 2B. $n = 1000$ observations from P

$P_0 = \mathrm{Gamma}(1.5, 1.2)$





$\theta(x), \hat{\theta}(x)$
0
2
4
6
8
10
0
5
10
15
x

Open questions for Settings 2A-B

Suppose we replace discrete $\widehat{P}$ with arbitrary distribution.

Open questions for Settings 2A-B

Suppose we replace discrete $\widehat{P}$ with arbitrary distribution.

Under which conditions on $\widehat{P}$ and P_0 exists a unique

$$\widehat{\theta} \;\in\; \underset{\theta \in \Theta}{\arg\max} \left(\int \theta \, d\widehat{P} - \int e^{\theta} \, dP_0 \right) ?$$

Open questions for Settings 2A-B

Suppose we replace discrete $\widehat{P}$ with arbitrary distribution.

Under which conditions on $\widehat{P}$ and P_0 exists a unique

$$\widehat{\theta} \;\in\; \underset{\theta \in \Theta}{\arg\max} \left(\int \theta \, d\widehat{P} - \int e^{\theta} \, dP_0 \right) ?$$

What continuity properties has the mapping

$$\widehat{P} \;\mapsto\; \widehat{\theta}\,?$$

Open questions for Settings 2A-B

Suppose we replace discrete $\widehat{P}$ with arbitrary distribution.

Under which conditions on $\widehat{P}$ and P_0 exists a unique

$$\widehat{\theta} \ \in \ \underset{\theta \in \Theta}{\arg\max} \left(\int \theta \, d\widehat{P} - \int e^{\theta} \, dP_0 \right) ?$$

What continuity properties has the mapping

$$\widehat{P} \ \mapsto \ \widehat{\theta}\, ?$$

(Analogous questions for Setting 1 well understood)

References

- Groeneboom, Jongbloed, Wellner (2008).
  The support reduction algorithm for computing nonparametric function estimates in mixture models. *Scand. J. Statist. 35*
- D., Hüsler, Rufibach (2007/2011).
  Active set and EM algorithms for log-concave densities based on complete and censored data.
  Technical report, IMSV, University of Bern (arxiv:0707.4643)
- D., Mösching, Strähl (2018).
  Active set algorithms for density estimators under shape-constraints.
  In preparation
- McCullagh, Polson (2012).
  Tail inflation. *Biometrika 99*
- D., Samworth, Schuhmacher (2011).
  Approximation by log-concave distributions, with applications to regression. *Ann. Statist. 39*