Computational Information Geometry:
Model Sensitivity and Approximate Cuts

Karim Anaya-Izquierdo (The Open University)

joint work with Frank Critchley (Open University),
Paul Marriott (Waterloo) and Paul Vos (East Carolina)

WOGAS 2, April 2010

Introduction

- Overall objective:
  provide tools to help understand sensitivity to model choice

Introduction

- Overall objective:
  provide tools to help understand sensitivity to model choice
- Target:
  applications of Generalised Linear Models

Introduction

- Overall objective:
  provide tools to help understand sensitivity to model choice
- Target:
  applications of Generalised Linear Models
- Delivered via ...
  Computational Information Geometry

Introduction

- Overall objective:
  provide tools to help understand sensitivity to model choice
- Target:
  applications of Generalised Linear Models
- Delivered via ...
  Computational Information Geometry

Model sensitivity

- Key ingredients

Model sensitivity

- Key ingredients
  - θ quantity/parameter of interest

Model sensitivity

- Key ingredients
  - θ quantity/parameter of interest
  - x data

Model sensitivity

- Key ingredients
  - θ quantity/parameter of interest
  - x data
  - $f(x;\theta)$ base/working model

Model sensitivity

- Key ingredients
  - θ quantity/parameter of interest
  - x data
  - $f(x; \theta)$ base/working model

$\Rightarrow$

Model sensitivity

- Key ingredients
  - θ quantity/parameter of interest
  - x data
  - $f(x;\theta)$ base/working model

$\Rightarrow$ inference about θ

Model sensitivity

- Key ingredients
  - θ quantity/parameter of interest
  - x data
  - $f(x;\theta)$ base/working model

  $\Rightarrow$ inference about θ
- What is the effect on inference about θ when considering models apart from $f(x;\theta)$?

Model sensitivity

- Key ingredients
  - θ quantity/parameter of interest
  - x data
  - $f(x;\theta)$ base/working model

  $\Rightarrow$ inference about θ
- What is the effect on inference about θ when considering models apart from $f(x;\theta)$?

Model sensitivity

- Perturb base model $f(x;\theta)$

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x;\theta) \longrightarrow g(x;\theta,\eta)$$

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x;\theta) \longrightarrow g(x;\theta,\eta)$$

so that

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x;\theta) \longrightarrow g(x;\theta,\eta)$$

so that

1. θ does not lose its meaning

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x;\theta) \longrightarrow g(x;\theta,\eta)$$

so that

1. θ does not lose its meaning
2. $g(x;\theta,\eta_0) = f(x;\theta)$ for some η_0

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb
$$f(x;\theta) \longrightarrow g(x;\theta,\eta)$$
    so that
  1. θ does not lose its meaning
  2. $g(x;\theta,\eta_0) = f(x;\theta)$ for some η_0
- Examples of 'directions' η

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x;\theta) \longrightarrow g(x;\theta,\eta)$$

    so that

  1. θ does not lose its meaning
  2. $g(x;\theta,\eta_0) = f(x;\theta)$ for some η_0

- Examples of 'directions' η
  - introduce extra variability (mixture model)

Model sensitivity

- Perturb base model $f(x; \theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x; \theta) \rightarrow g(x; \theta, \eta)$$

  so that

  1. θ does not lose its meaning
  2. $g(x; \theta, \eta_0) = f(x; \theta)$ for some η_0

- Examples of 'directions' η
  - introduce extra variability (mixture model)
  - introduce skewness

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x;\theta) \to g(x;\theta,\eta)$$

  so that

  1. θ does not lose its meaning
  2. $g(x;\theta,\eta_0) = f(x;\theta)$ for some η_0

- Examples of 'directions' η
  - introduce extra variability (mixture model)
  - introduce skewness
  - introduce heavier tails

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x;\theta) \rightarrow g(x;\theta,\eta)$$

so that

1. θ does not lose its meaning
2. $g(x;\theta,\eta_0) = f(x;\theta)$ for some η_0

- Examples of 'directions' η
  - introduce extra variability (mixture model)
  - introduce skewness
  - introduce heavier tails
  - introduce dependence

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x;\theta) \to g(x;\theta,\eta)$$

so that

1. θ does not lose its meaning
2. $g(x;\theta,\eta_0) = f(x;\theta)$ for some η_0

- Examples of 'directions' η
  - introduce extra variability (mixture model)
  - introduce skewness
  - introduce heavier tails
  - introduce dependence
- What is the effect on inference about θ when perturbing along the direction η?

Model sensitivity

- Perturb base model $f(x;\theta)$
  - 'Pick a direction' η
  - Perturb

$$f(x;\theta) \to g(x;\theta,\eta)$$

  so that

  1. θ does not lose its meaning
  2. $g(x;\theta,\eta_0) = f(x;\theta)$ for some η_0

- Examples of 'directions' η
  - introduce extra variability (mixture model)
  - introduce skewness
  - introduce heavier tails
  - introduce dependence

- What is the effect on inference about θ when perturbing along the direction η?

Cuts

- Model $g(x;\theta,\eta)$

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance

Cuts

- Model $g(x;\theta,\eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η
- Key idea is factorisation of the likelihood

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η
- Key idea is factorisation of the likelihood

if $(T(x), S(x)) \Leftrightarrow x$,

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η
- Key idea is factorisation of the likelihood

if $(T(x), S(x)) \Leftrightarrow x$, $(\theta, \eta) \in \Theta \times \Lambda$

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η
- Key idea is factorisation of the likelihood

if $(T(x), S(x)) \Leftrightarrow x$, $(\theta, \eta) \in \Theta \times \Lambda$

$$L(\theta, \eta; x) =$$

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η
- Key idea is factorisation of the likelihood

if $(T(x), S(x)) \Leftrightarrow x$, $(\theta, \eta) \in \Theta \times \Lambda$

$$L(\theta, \eta; x) = L_i(\theta; T(x)) \times$$

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η
- Key idea is factorisation of the likelihood

if $(T(x), S(x)) \Leftrightarrow x$, $(\theta, \eta) \in \Theta \times \Lambda$

$$L(\theta, \eta; x) = L_i(\theta; T(x)) \times L_n(\eta; S(x)) \quad \forall\, \theta, \eta, x$$

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η
- Key idea is factorisation of the likelihood

if $(T(x), S(x)) \Leftrightarrow x$, $(\theta, \eta) \in \Theta \times \Lambda$

$$L(\theta, \eta; x) = L_i(\theta; T(x)) \times L_n(\eta; S(x)) \quad \forall\, \theta, \eta, x$$

then T is a cut

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η
- Key idea is factorisation of the likelihood

if $(T(x), S(x)) \Leftrightarrow x$, $(\theta, \eta) \in \Theta \times \Lambda$

$$L(\theta, \eta; x) = L_i(\theta; T(x)) \times L_n(\eta; S(x)) \quad \forall\, \theta, \eta, x$$

then T is a cut

- Very tight definition valid in very special cases:

Cuts

- Model $g(x; \theta, \eta)$
  - θ is of interest
  - η is nuisance
- A cut is a statistic which allow inference about θ independently of the values of η
- Key idea is factorisation of the likelihood

if $(T(x), S(x)) \Leftrightarrow x$, $(\theta, \eta) \in \Theta \times \Lambda$

$$L(\theta, \eta; x) = L_i(\theta; T(x)) \times L_n(\eta; S(x)) \quad \forall \theta, \eta, x$$

then T is a cut

- Very tight definition valid in very special cases: Exponential Families with so-called mixed parametrisation

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:
  - there exists an enlarged model $g(x;\theta,\eta)$ with cut $(T(x), S(x))$

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:
  - there exists an enlarged model $g(x; \theta, \eta)$ with cut $(T(x), S(x))$
    $\Rightarrow$ perturbation $\eta_0 \to \eta$ does not affect inferences about θ

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:
  - there exists an enlarged model $g(x;\theta,\eta)$ with cut $(T(x), S(x))$
    $\Rightarrow$ perturbation $\eta_0 \rightarrow \eta$ does not affect inferences about θ
- In practice relax conditions to define an approximate cut:

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:
  - there exists an enlarged model $g(x;\theta,\eta)$ with cut $(T(x), S(x))$
    $\Rightarrow$ perturbation $\eta_0 \rightarrow \eta$ does not affect inferences about θ
- In practice relax conditions to define an approximate cut:
  - Inferences about θ based on $T(x)$ 'do not change much' when perturbing in the direction η

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:
  - there exists an enlarged model $g(x; \theta, \eta)$ with cut $(T(x), S(x))$
    $\Rightarrow$ perturbation $\eta_0 \to \eta$ does not affect inferences about θ
- In practice relax conditions to define an approximate cut:
  - Inferences about θ based on $T(x)$ 'do not change much' when perturbing in the direction η
  - data x is fixed

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:
  - there exists an enlarged model $g(x;\theta,\eta)$ with cut $(T(x), S(x))$
    $\Rightarrow$ perturbation $\eta_0 \rightarrow \eta$ does not affect inferences about θ
- In practice relax conditions to define an approximate cut:
  - Inferences about θ based on $T(x)$ 'do not change much' when perturbing in the direction η
  - data x is fixed
  - For (θ, η) values where data x supports the use of $g(x;\theta,\eta)$

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:
  - there exists an enlarged model $g(x;\theta,\eta)$ with cut $(T(x), S(x))$
    $\Rightarrow$ perturbation $\eta_0 \rightarrow \eta$ does not affect inferences about θ
- In practice relax conditions to define an approximate cut:
  - Inferences about θ based on $T(x)$ 'do not change much' when perturbing in the direction η
  - data x is fixed
  - For (θ,η) values where data x supports the use of $g(x;\theta,\eta)$
- By complement we can find directions η with large effect on inference:

Cuts for model sensitivity

- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:
  - there exists an enlarged model $g(x;\theta,\eta)$ with cut $(T(x), S(x))$
    $\Rightarrow$ perturbation $\eta_0 \rightarrow \eta$ does not affect inferences about θ
- In practice relax conditions to define an approximate cut:
  - Inferences about θ based on $T(x)$ 'do not change much' when perturbing in the direction η
  - data x is fixed
  - For (θ,η) values where data x supports the use of $g(x;\theta,\eta)$
- By complement we can find directions η with large effect on inference:
  - Confidence interval length substantially increased

Cuts for model sensitivity

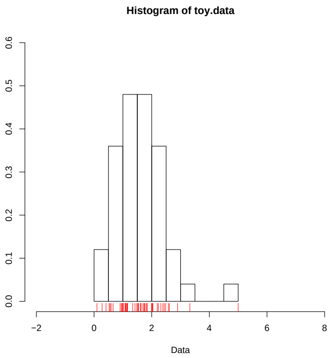
- Inferences on θ are made using $T(x)$, for example
  - an interval estimate for θ
  - the rejection region of a hypothesis test
- Ideal case:
  - there exists an enlarged model $g(x;\theta,\eta)$ with cut $(T(x), S(x))$
    $\Rightarrow$ perturbation $\eta_0 \rightarrow \eta$ does not affect inferences about θ
- In practice relax conditions to define an approximate cut:
  - Inferences about θ based on $T(x)$ 'do not change much' when perturbing in the direction η
  - data x is fixed
  - For (θ,η) values where data x supports the use of $g(x;\theta,\eta)$
- By complement we can find directions η with large effect on inference:
  - Confidence interval length substantially increased
  - Do not reject instead of rejecting

Our approach

- Use Structured Extended Multinomial models (SEM's)
- Extended multinomials are multinomials but allow cell probabilities to be zero
- Discretising continuous data gives multinomials with structure on the cells
- SEM proxy for universal space of all distributions

Example

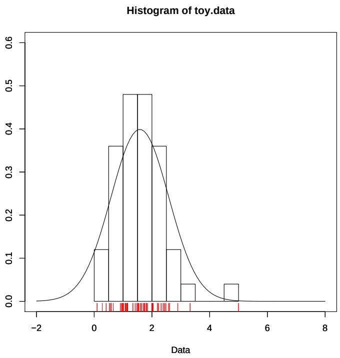
Question: what is the population mean θ?



- Data
- Model $N(\theta, 1)$
- Model $N(\theta, \sigma^2)$
- Model $N(\theta, 1)$ remove outlier (Base)
- Model Logistic$(\mu(\theta), \sigma)$
- Model log-Normal$(\mu'(\theta), \sigma^2)$

Example

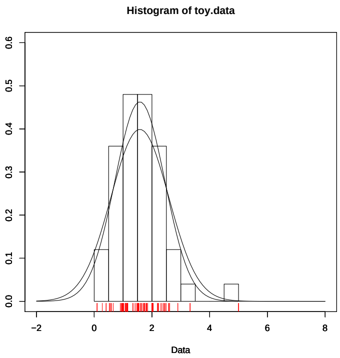
Question: what is the population mean θ?



- Data
- Model $N(\theta, 1)$
- Model $N(\theta, \sigma^2)$
- Model $N(\theta, 1)$ remove outlier (Base)
- Model Logistic$(\mu(\theta), \sigma)$
- Model log-Normal$(\mu'(\theta), \sigma^2)$

Example

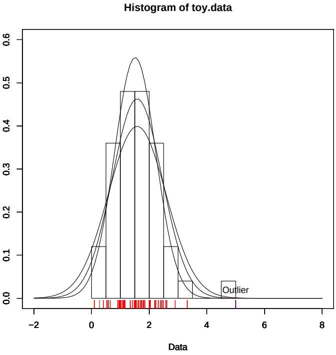
Question: what is the population mean θ?



- Data
- Model $N(\theta, 1)$
- Model $N(\theta, \sigma^2)$
- Model $N(\theta, 1)$ remove outlier (Base)
- Model Logistic$(\mu(\theta), \sigma)$
- Model log-Normal$(\mu'(\theta), \sigma^2)$

Example

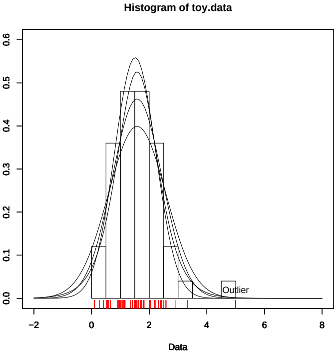
Question: what is the population mean θ?



- Data
- Model $N(\theta, 1)$
- Model $N(\theta, \sigma^2)$
- Model $N(\theta, 1)$ remove outlier (Base)
- Model Logistic$(\mu(\theta), \sigma)$
- Model log-Normal$(\mu'(\theta), \sigma^2)$

Example

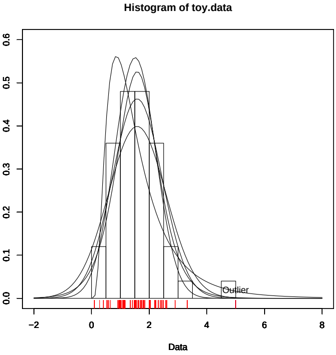
Question: what is the population mean θ?



- Data
- Model $N(\theta, 1)$
- Model $N(\theta, \sigma^2)$
- Model $N(\theta, 1)$ remove outlier (Base)
- Model Logistic$(\mu(\theta), \sigma)$
- Model log-Normal$(\mu'(\theta), \sigma^2)$

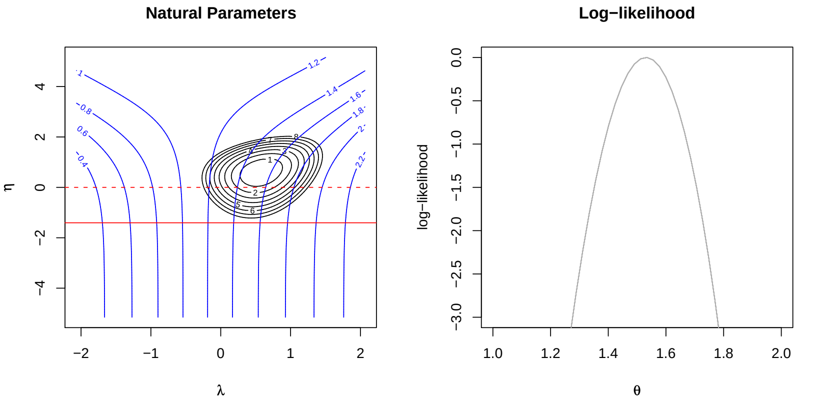
Example

Question: what is the population mean θ?



- Data
- Model $N(\theta, 1)$
- Model $N(\theta, \sigma^2)$
- Model $N(\theta, 1)$ remove outlier (Base)
- Model Logistic$(\mu(\theta), \sigma)$
- Model log-Normal$(\mu'(\theta), \sigma^2)$

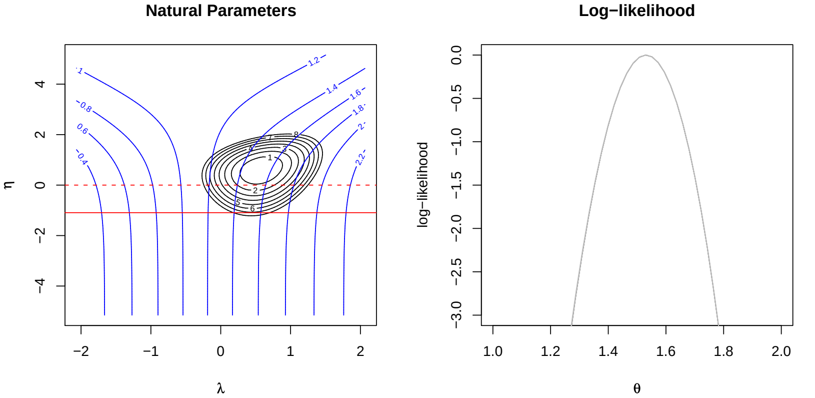
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

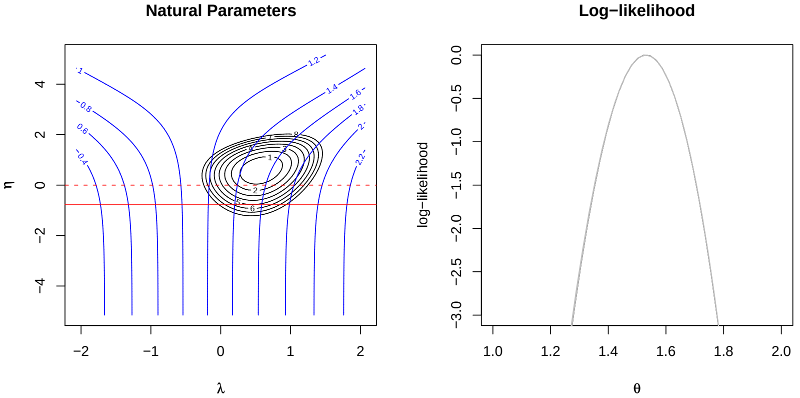
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

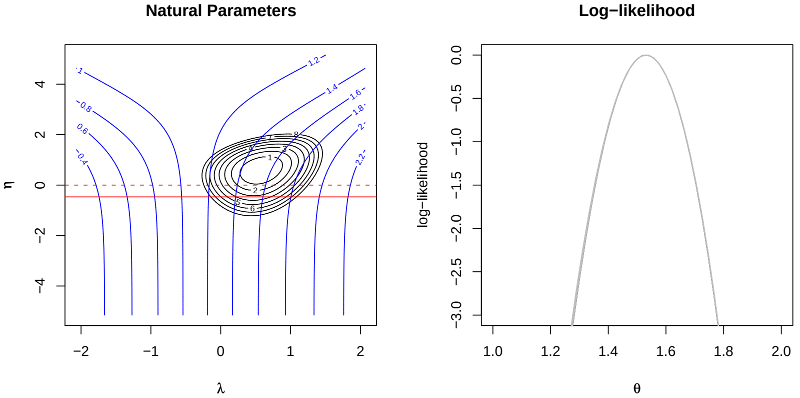
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

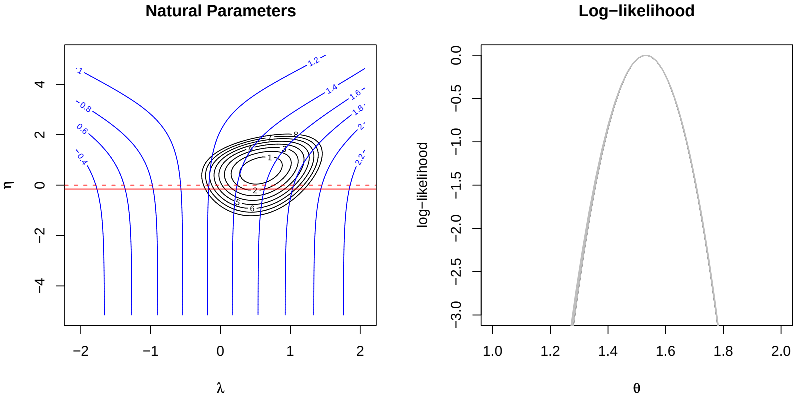
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

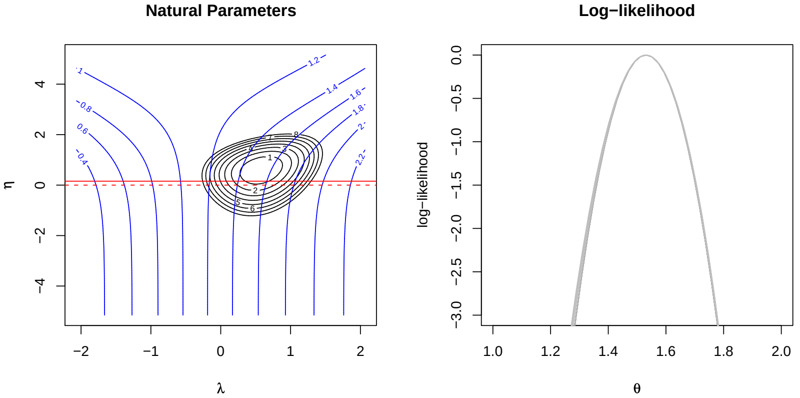
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

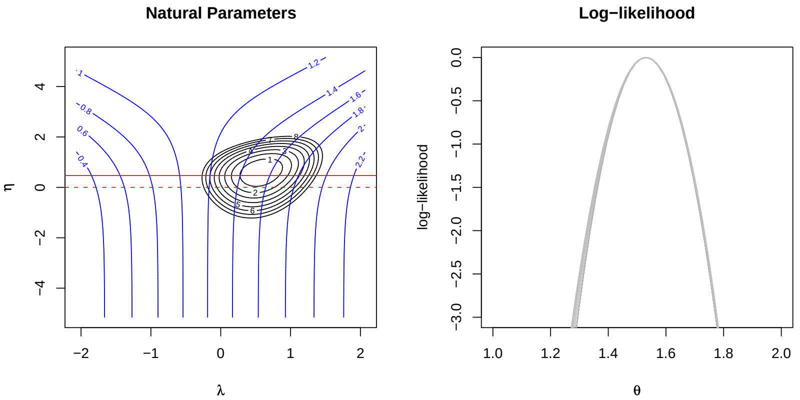
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

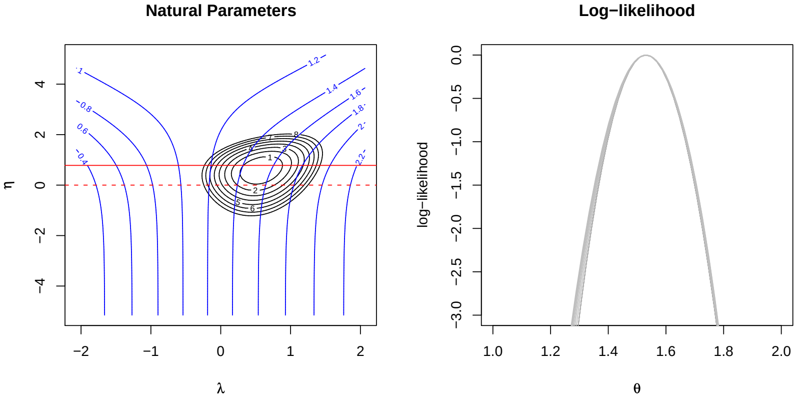
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

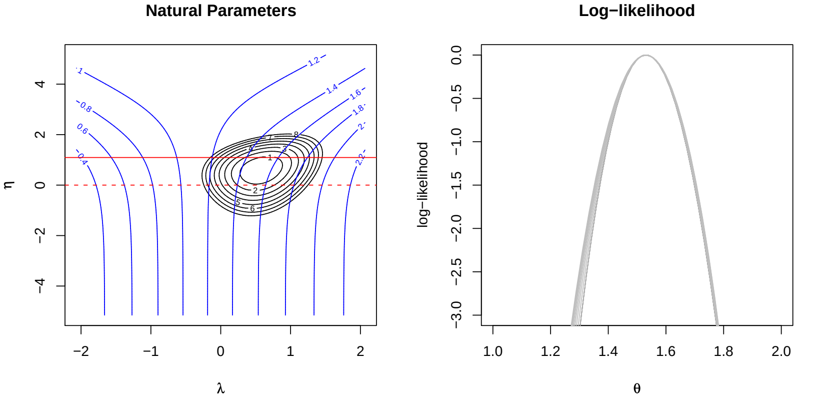
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

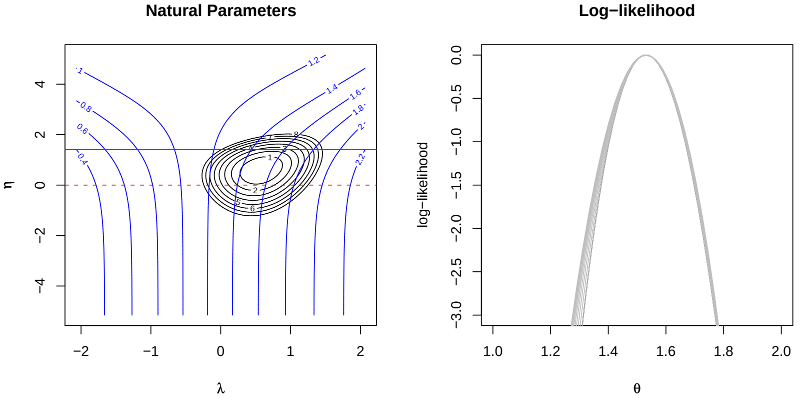
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

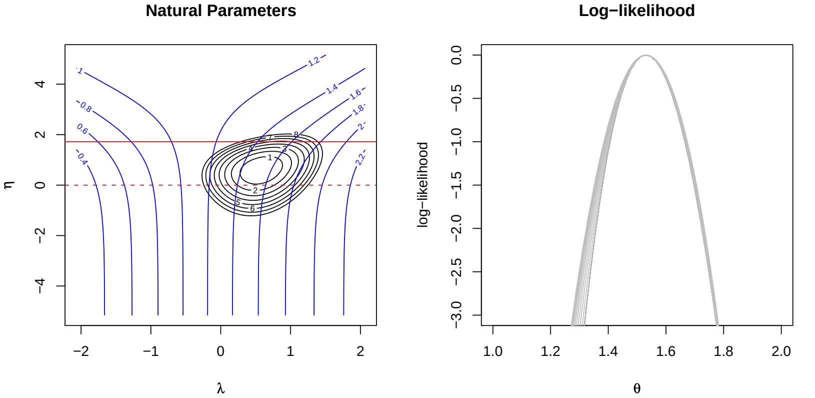
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

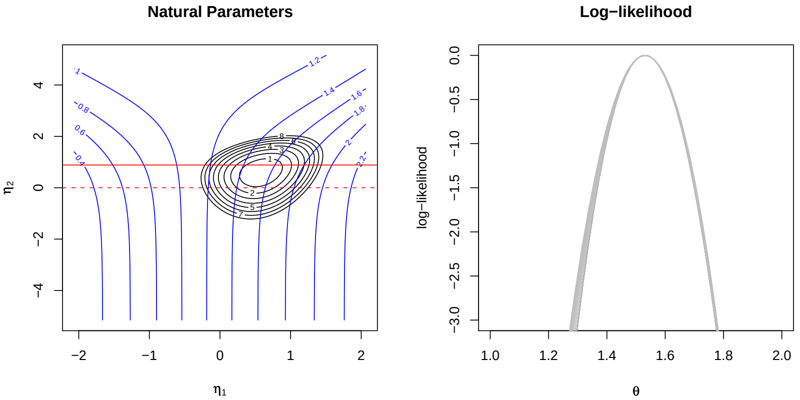
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

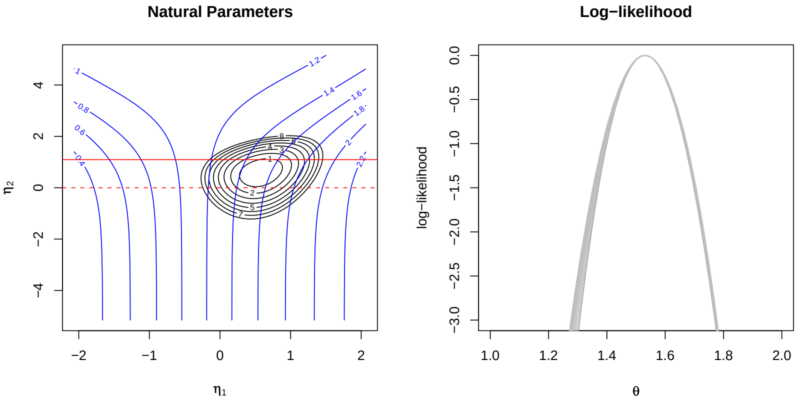
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

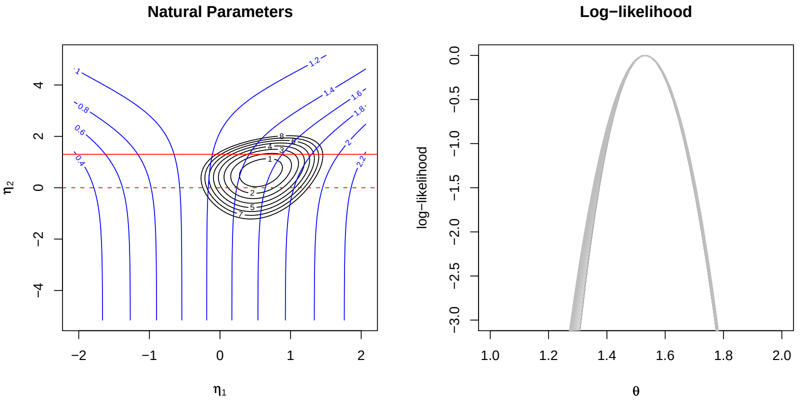
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

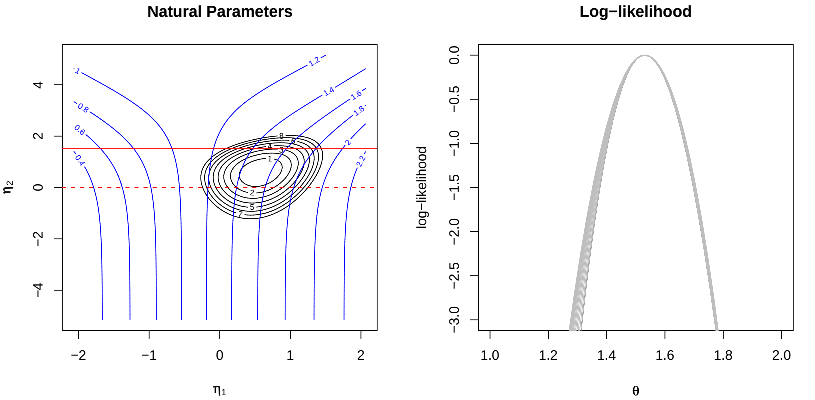
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

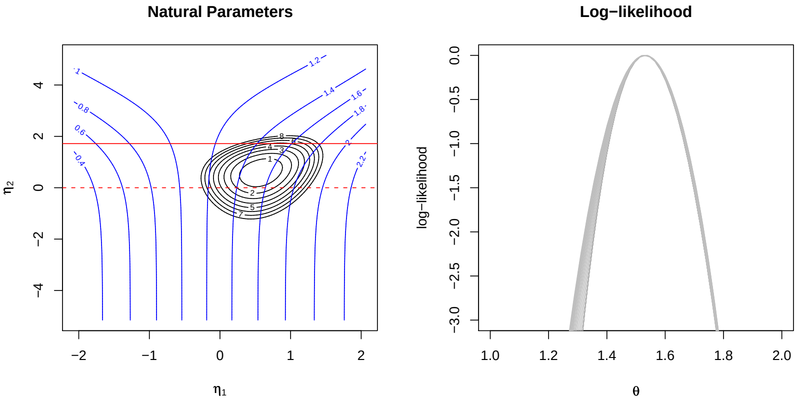
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

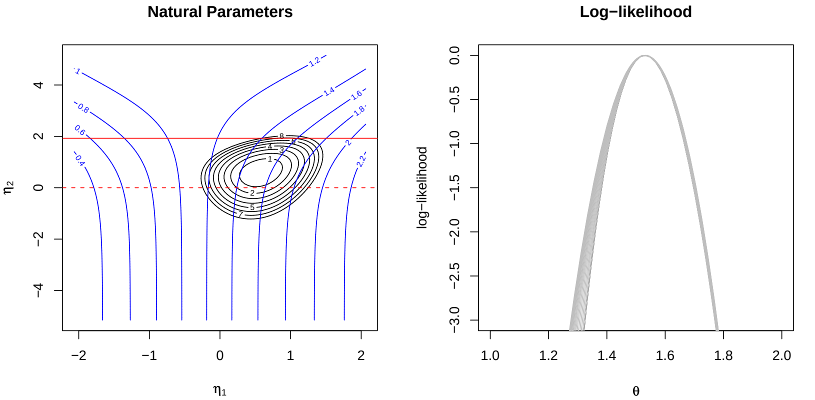
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

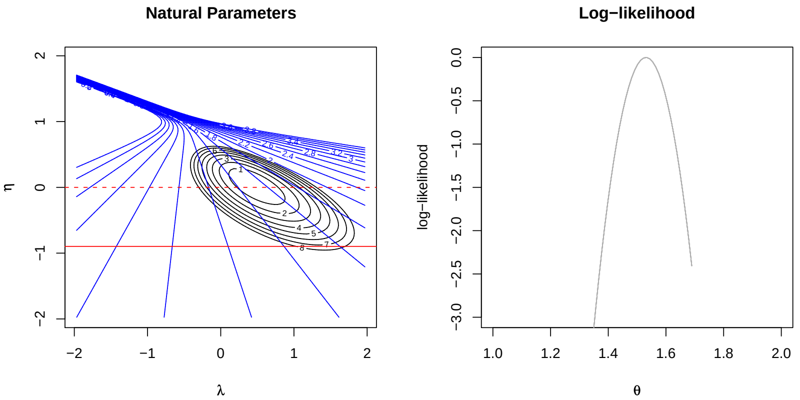
Insensitive direction: CUT



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

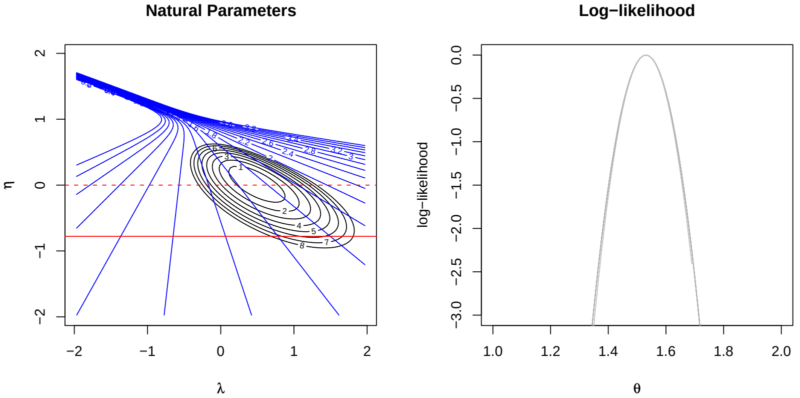
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

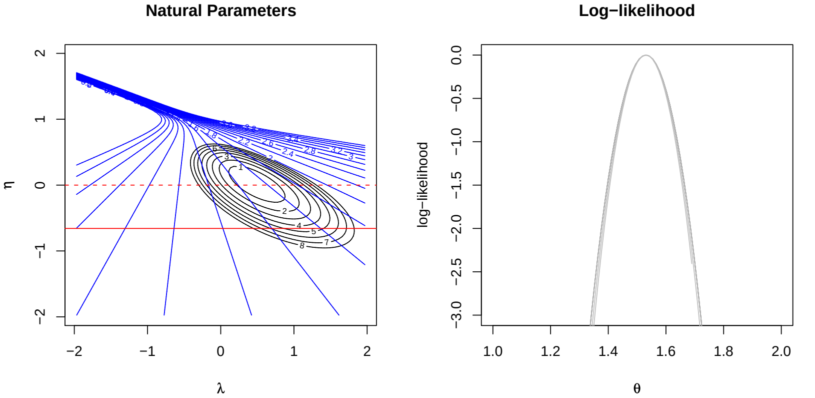
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

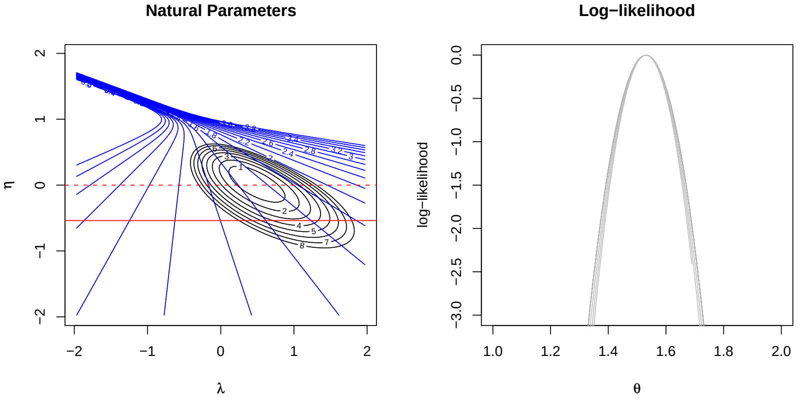
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

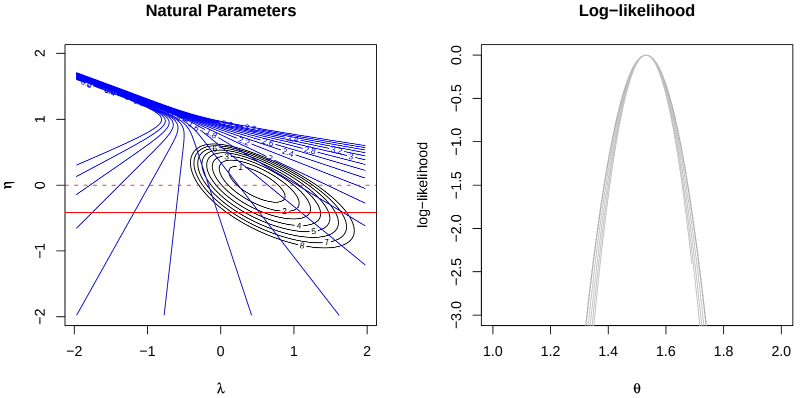
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

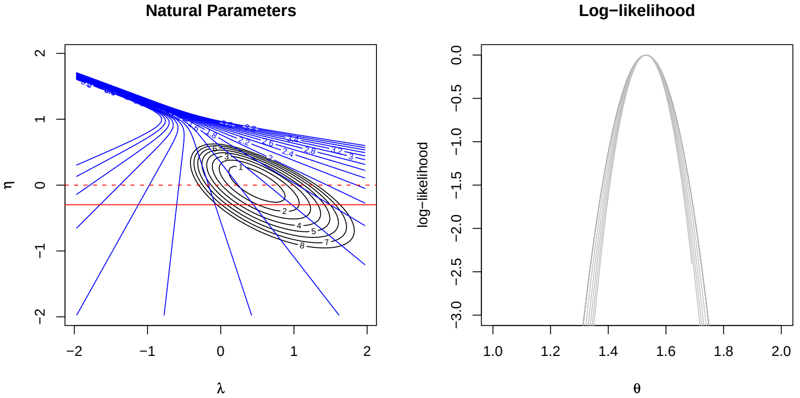
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

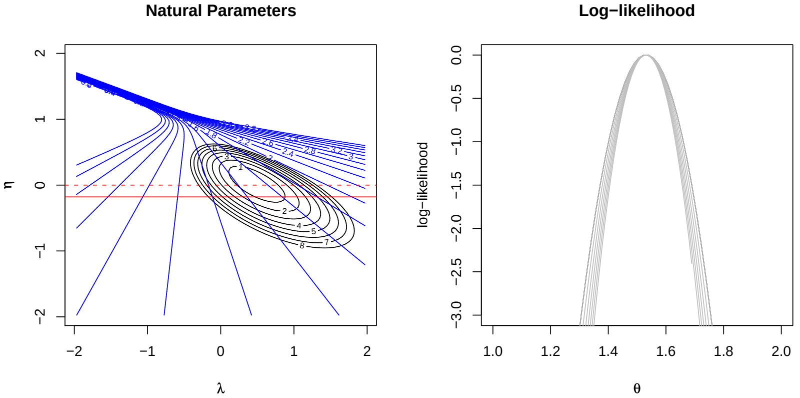
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

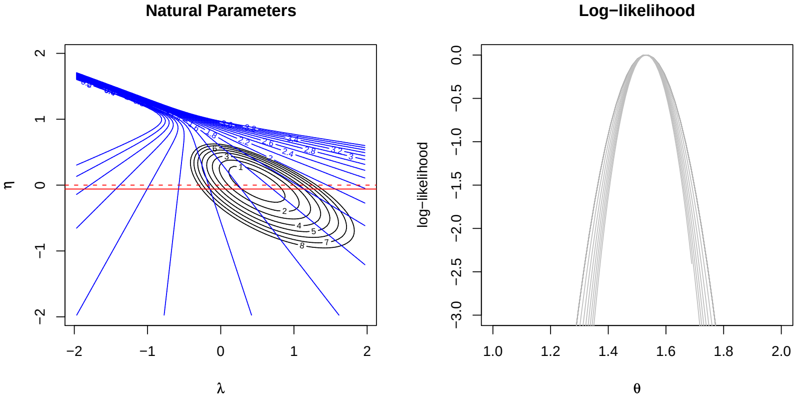
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

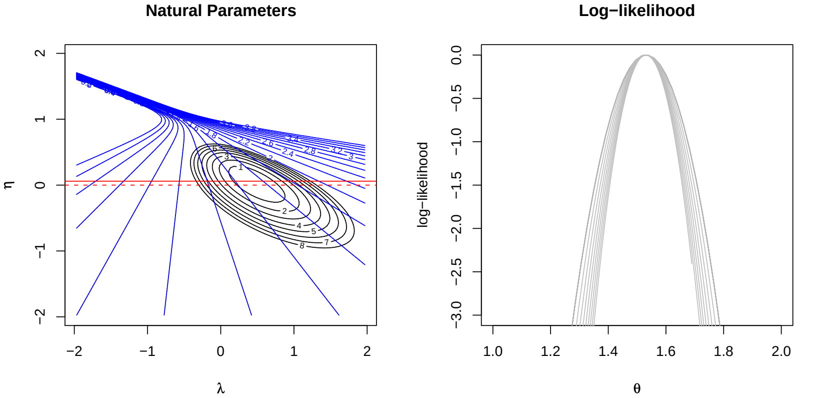
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

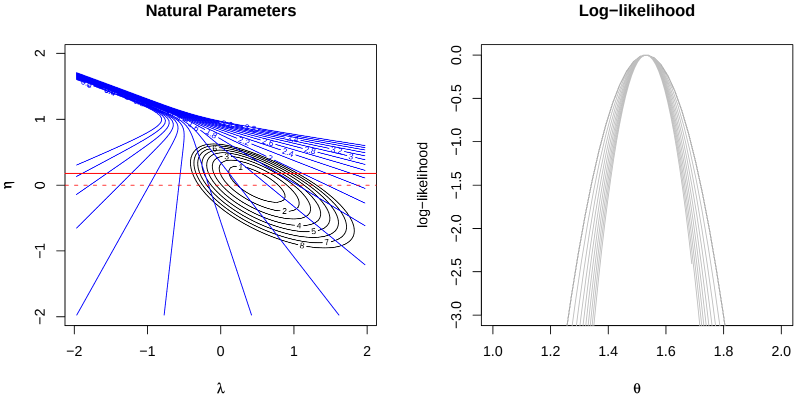
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

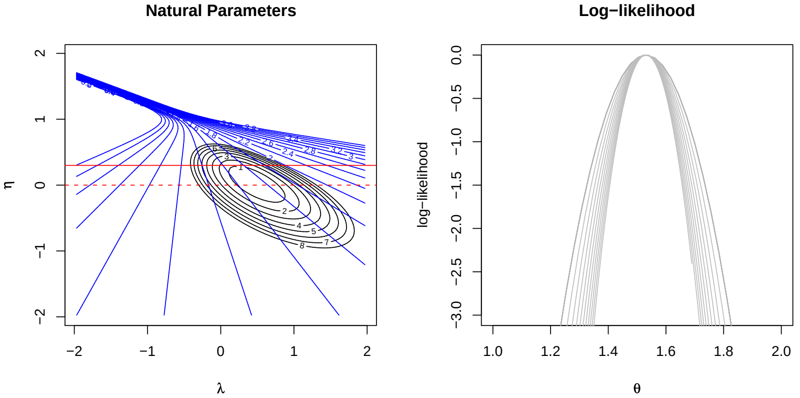
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

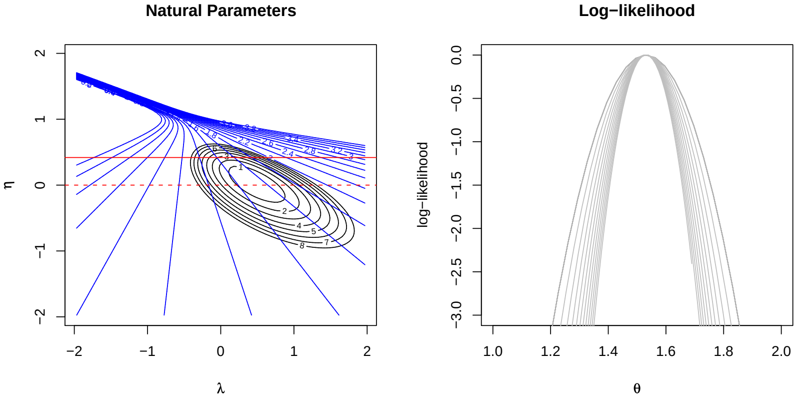
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

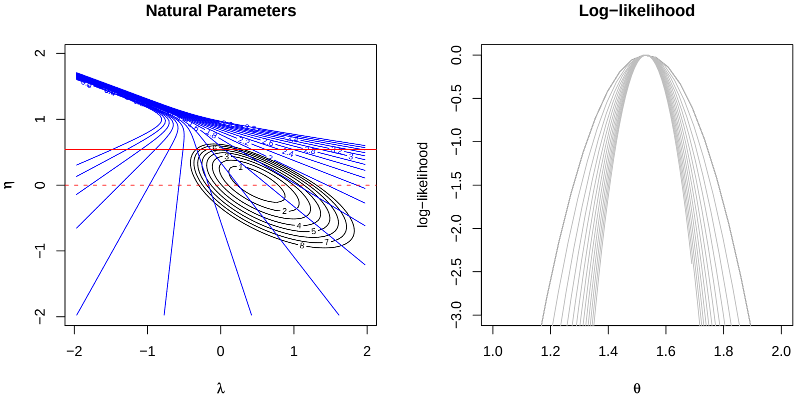
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

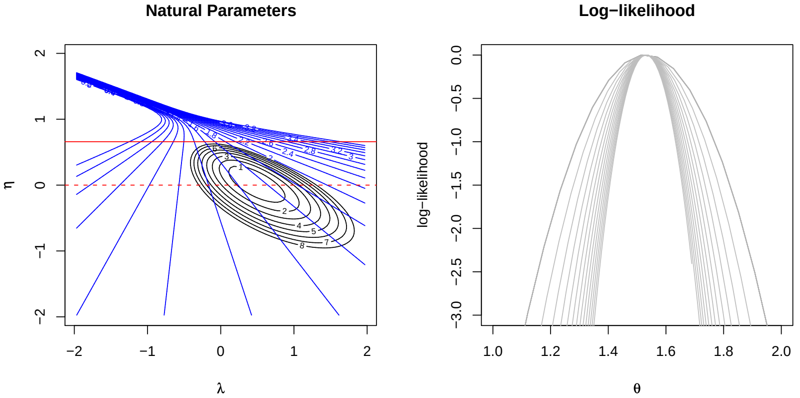
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

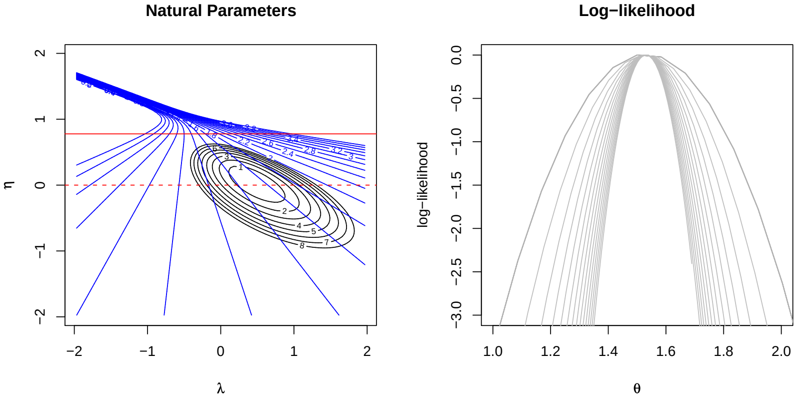
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

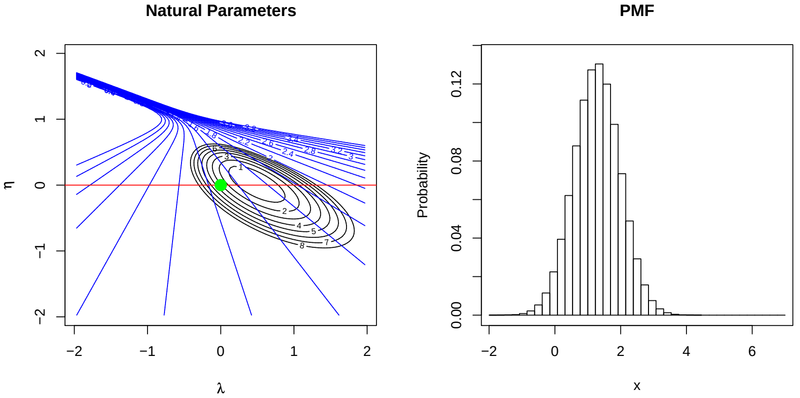
Highly sensitive direction



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

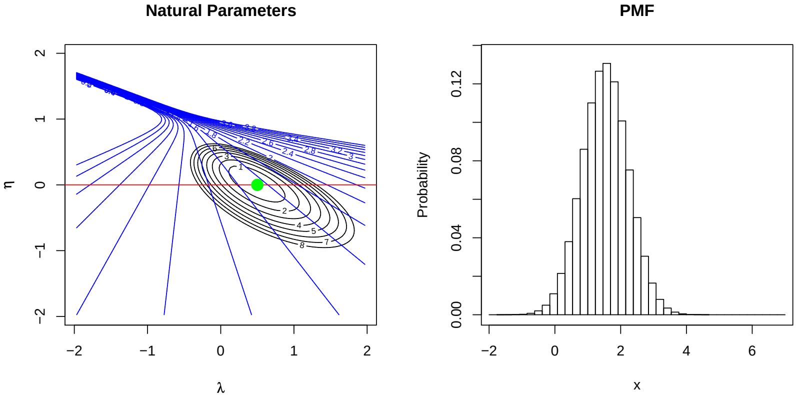
Highly sensitive direction: Interpretation



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

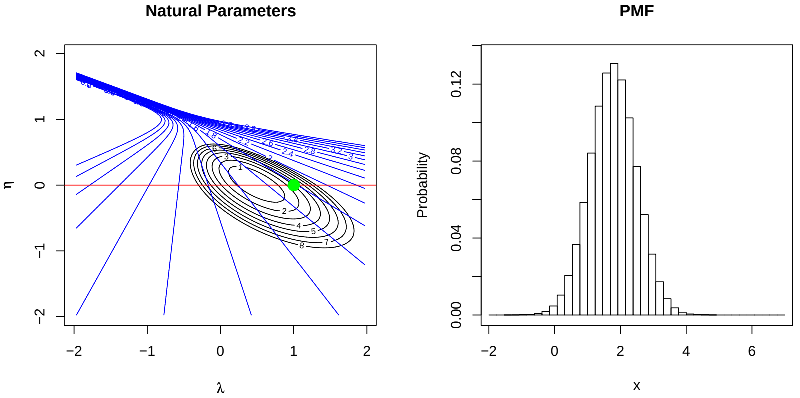
Highly sensitive direction: Interpretation



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

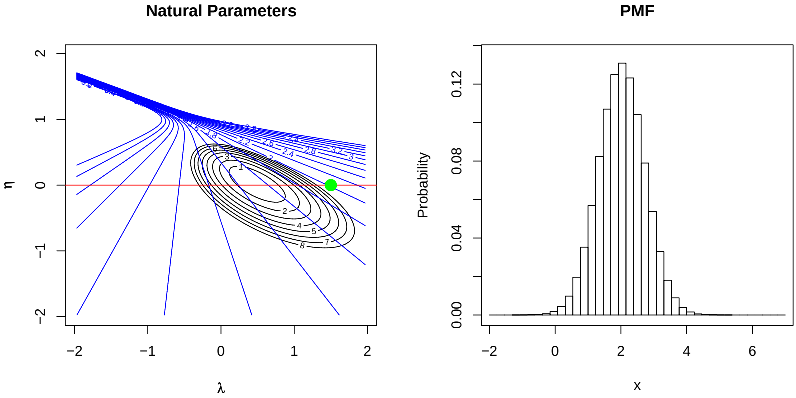
Highly sensitive direction: Interpretation



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

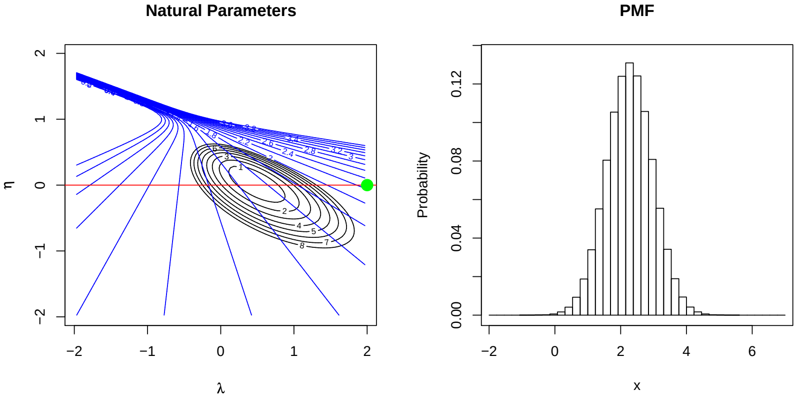
Highly sensitive direction: Interpretation



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

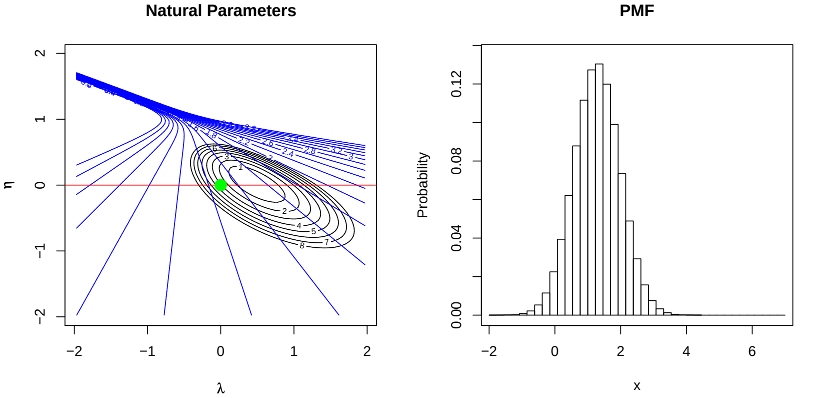
Highly sensitive direction: Interpretation



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

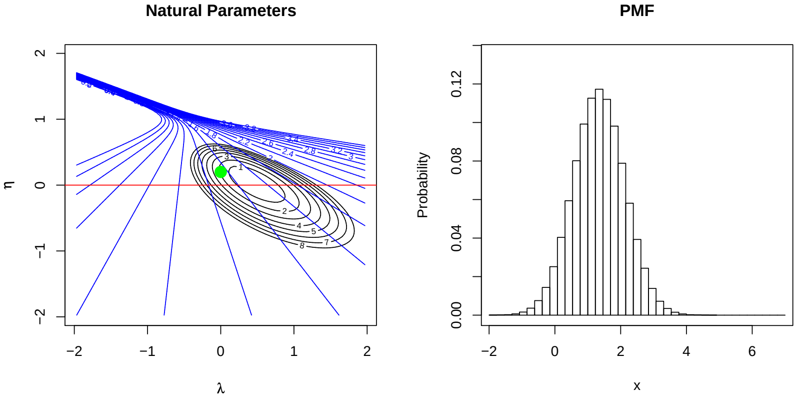
Highly sensitive direction: Interpretation



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

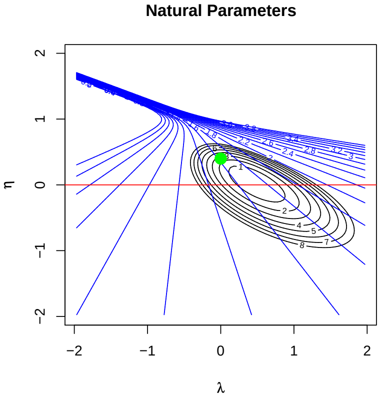
Highly sensitive direction: Interpretation

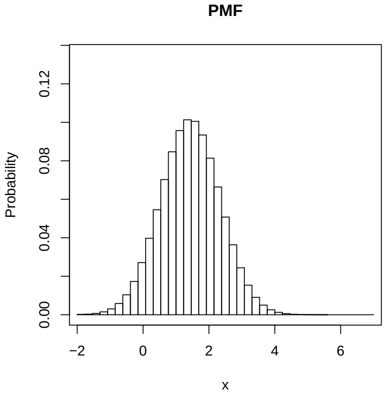


Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

Highly sensitive direction: Interpretation

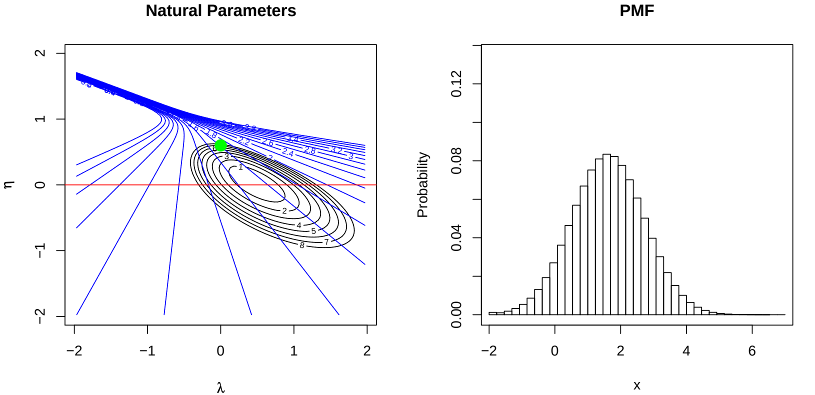




Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

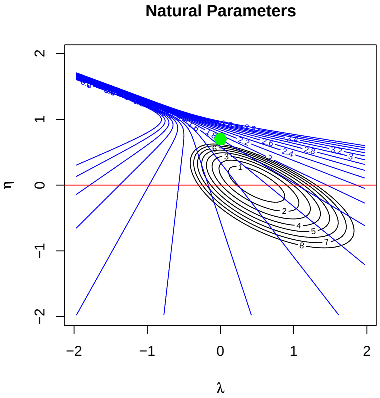
Highly sensitive direction: Interpretation

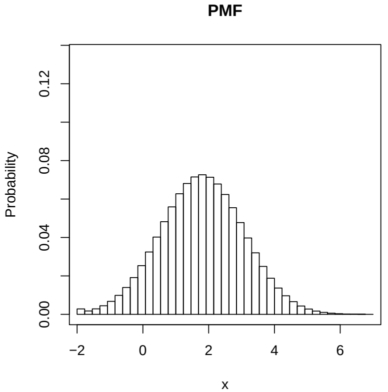


Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

Highly sensitive direction: Interpretation

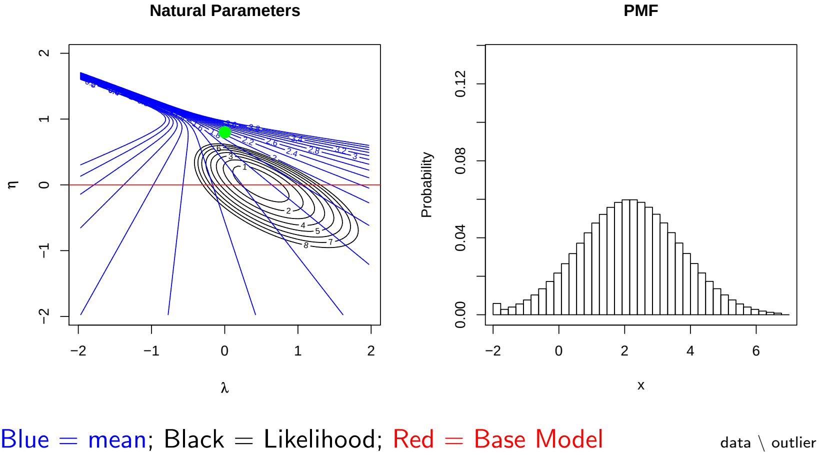




Blue = mean; Black = Likelihood; Red = Base Model

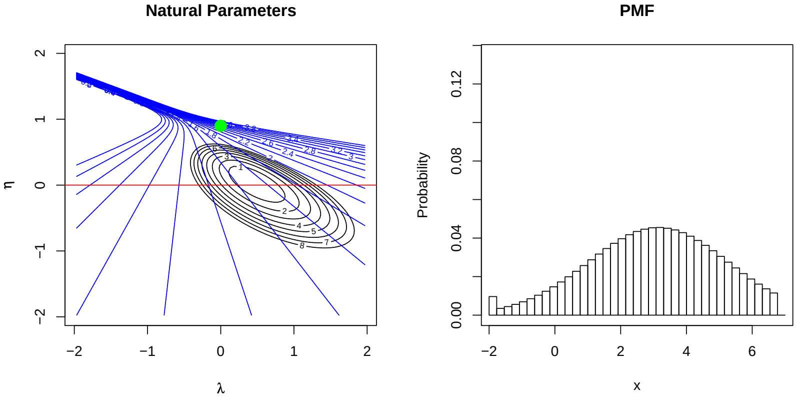
data \ outlier

Highly sensitive direction: Interpretation



Blue = mean; Black = Likelihood; Red = Base Model

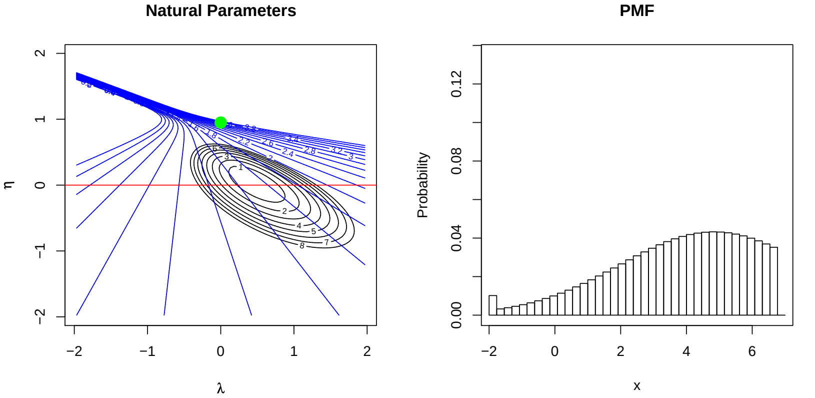
Highly sensitive direction: Interpretation



Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

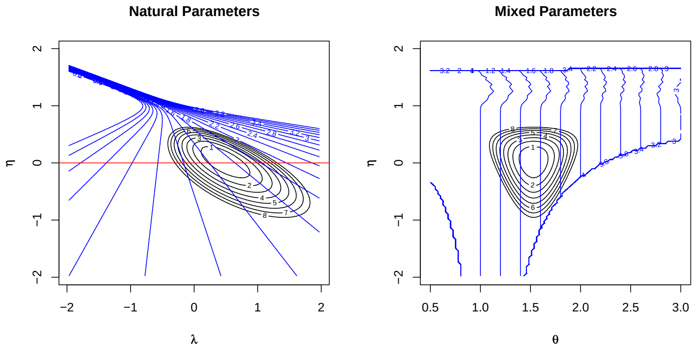
Highly sensitive direction: Interpretation



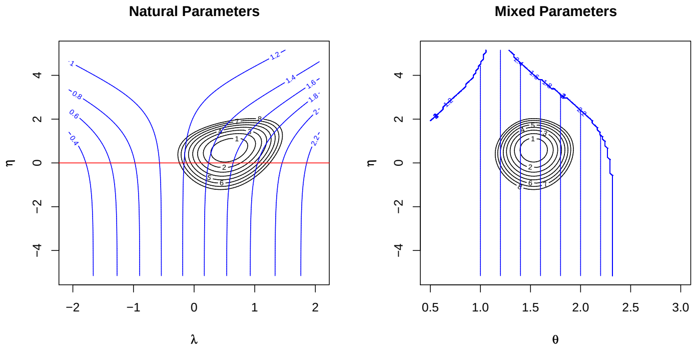
Blue = mean; Black = Likelihood; Red = Base Model

data \ outlier

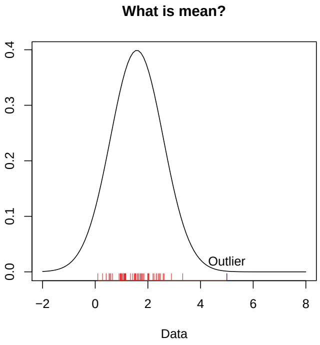
Sensitive direction: Mixed Parameters (Duality)



Insensitive direction: Mixed Parameters (Duality) CUT

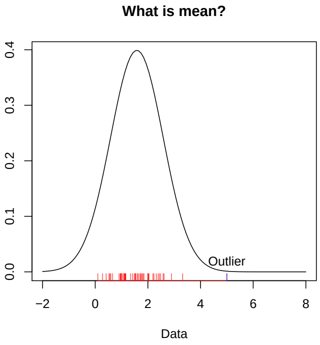


Perturbation Space



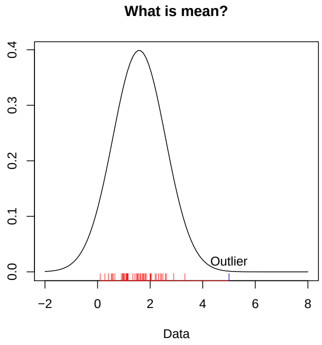
- Data $\setminus$ outlier

Perturbation Space



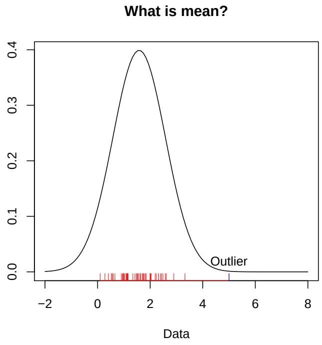
- Data $\setminus$ outlier

Perturbation Space



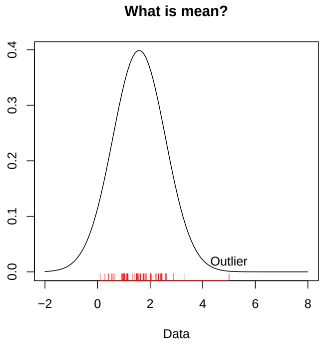
- Data $\setminus$ outlier and base model $N(\mu, 1)$:

Perturbation Space



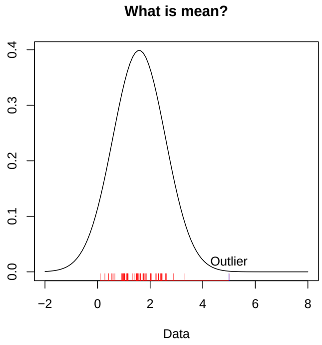
- Data $\setminus$ outlier and base model $N(\mu, 1)$: only one direction is highly sensitive

Perturbation Space



- Data $\setminus$ outlier and base model $N(\mu, 1)$: only one direction is highly sensitive
- This corresponds to changing variance giving model $N(\mu, \sigma^2)$

Perturbation Space



- Data $\setminus$ outlier and base model $N(\mu, 1)$: only one direction is highly sensitive
- This corresponds to changing variance giving model $N(\mu, \sigma^2)$
- If include outlier one more parameter needed

Other types of perturbations

We can perturb by:

Other types of perturbations

We can perturb by:

- Translations $\cdots$ Least informative directions

Other types of perturbations

We can perturb by:

- Translations $\cdots$ Least informative directions
- Rotations another story $\cdots$ bias, variance trade-off $\cdots$

Other types of perturbations

We can perturb by:

- Translations $\cdots$ Least informative directions
- Rotations another story $\cdots$ bias, variance trade-off $\cdots$
- Adding nuisance parameters $\cdots$

Other types of perturbations

We can perturb by:

- Translations $\cdots$ Least informative directions
- Rotations another story $\cdots$ bias, variance trade-off $\cdots$
- Adding nuisance parameters $\cdots$
- Down-weight/delete outliers $\cdots$

Other types of perturbations

We can perturb by:

- Translations $\cdots$ Least informative directions
- Rotations another story $\cdots$ bias, variance trade-off $\cdots$
- Adding nuisance parameters $\cdots$
- Down-weight/delete outliers $\cdots$

Computation

- Discretise and work in SEM

Computation

- Discretise and work in SEM
- Extension g is an exponential family

Computation

- Discretise and work in SEM
- Extension g is an exponential family
- Construct mixed parametrisation

Computation

- Discretise and work in SEM
- Extension g is an exponential family
- Construct mixed parametrisation
- Fisher Information (variance) is block diagonal

Computation

- Discretise and work in SEM
- Extension g is an exponential family
- Construct mixed parametrisation
- Fisher Information (variance) is block diagonal
- Choose directions where θ block changes most with η

Summary

- Overall objective:
  provide tools to help understand sensitivity to model choice
- Target:
  applications of Generalised Linear Models
- Delivered via ...
  Computational Information Geometry (hidden from the user)

Work supported by EPSRC grant EP/E017878/1