Convegno SIDRA 2010
L'Aquila 13-15 Settembre 2010

Data-driven optimization through the scenario approach

Algo Carè
algo.care@ing.unibs.it

Simone Garatti
sgaratti@elet.polimi.it
Marco C. Campi
campi@ing.unibs.it

The scenario approach

Uncertain convex program

$$\min_{\alpha \in \mathbb{R}^d} c^T \alpha$$

$$\text{subject to: } f_\delta(\alpha) \leq 0, \quad \delta \in \Delta$$

convex

Infinite constraints parametrized by δ, an *uncertain parameter*

We observe a finite number of situations (i.e. measure a finite number of values taken by δ) and obtain an optimal solution satisfying them.

Scenario program

$$\min_{\alpha \in \mathbb{R}^d} c^T \alpha$$

$$\text{subject to: } f_{\delta^{(i)}}(\alpha) \leq 0, \quad \delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)} \text{ from } \Delta$$

N situations observed

→ Optimal solution: α_N^*

Under certain conditions, α_N^* generalizes well, i.e, it satisfies "the large majority" of the infinite unseen constraints…

Probabilistic guarantees through the scenario approach

δ distributed according to an unknown probability on Δ;

Scenario program

$$\min_{\alpha \in \mathbb{R}^d} c^T \alpha$$

subject to: $f_{\delta^{(i)}}(\alpha) \leq 0,\quad \delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}$ from Δ

N situations observed (i.i.d)

→ optimal solution: α_N^*

Then you can be sure that

$$\Pr[\, f_\delta(\alpha_N^*) > 0] \leq \varepsilon$$

It depends only on *N*

not on the probability distribution!

Mathematical details

Indeed, $\Pr[f_\delta(\alpha_N^*) > 0] \leq \varepsilon$ holds
with very high probability $1-\beta$, ($\beta \approx 0$, e.g. $\beta = 10^{-6}$) independently
of the unknown probability distribution on Δ,

on condition that: $$\beta \leq \sum_{i=0}^{d-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}.$$

In practice…

N → min ε such that $$\beta \leq \sum_{i=0}^{d-1} \binom{N}{i} \varepsilon^i (1-\varepsilon)^{N-i}$$ → ε

Fact 1: For general convex programs with *d* decision variables, you cannot find any better ε given N: this result is unimprovable!

Fact 2: Setting smaller β does not impact significantly on ε given *N*

Data-driven approach in short

Data acquisition procedures

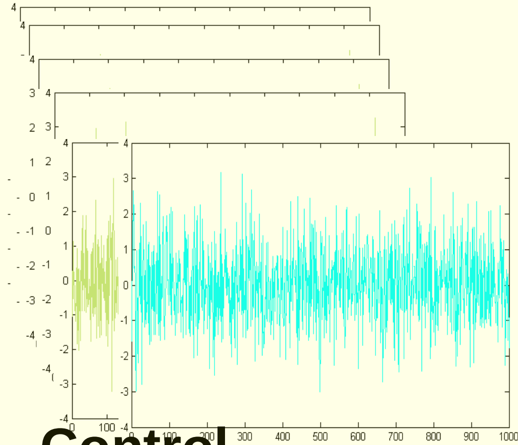
$\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}$

Scenario program

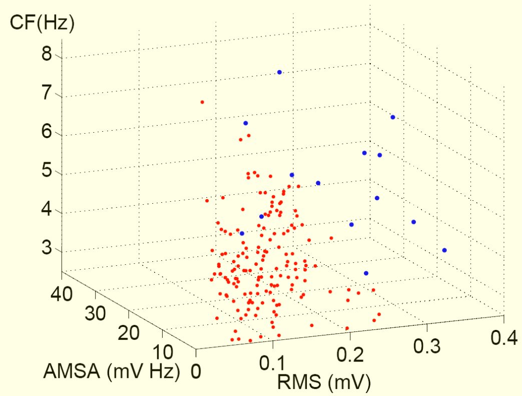
α_N^* solution

- *Historical series*
- *Performed experiments*
- *…*

Finance
(stock movements)



Control
(Input noise realizations)



Medical diagnosis
(after-CA resuscitations)

N → Exact formula → ε reliability level

α_N^* satisfies all the constraints $\delta \in \Delta$ except at most a proportion ε of them

Some issues about data-driven optimization

I. Shortage of samples

What if we have "too few" observations at our disposal to reach a desirable ε?

II. Bias-Variance trade-off

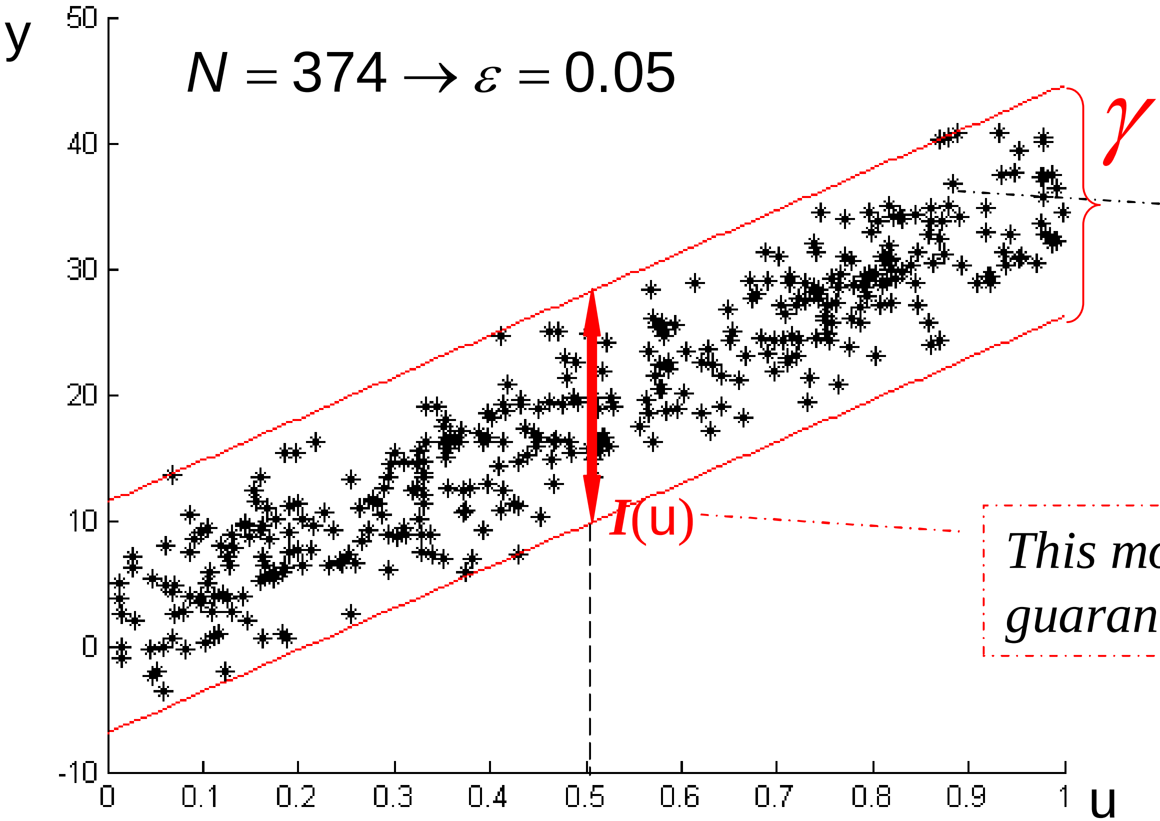
How to choose "the right model" given the available data?

III. Deep into a-priori knowledge

Is it possible to characterise more deeply the scenario solution α_N^* *before observing data?*

We'll face these issues using the example of **Interval Predictor Models** as a paradigm.

Example: Interval Predictor Models (I.P.M.)



374 *i.i.d observations of couples* $(u,y) \in \mathbb{R}^d \times \mathbb{R}$

This model (IPM of order 1) associates a guaranteed interval of y *to each value of* u

The **red strip** contains at least the **95%** of the probability mass, viz. $\Pr[y \notin I(u)] \leq 0.05$

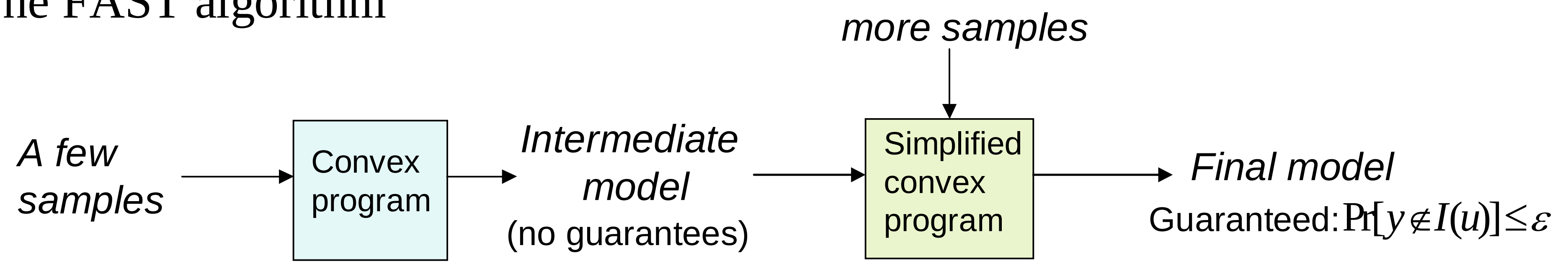
In general, to find the parameters of an IPM of order *n*, just solve the convex problem:

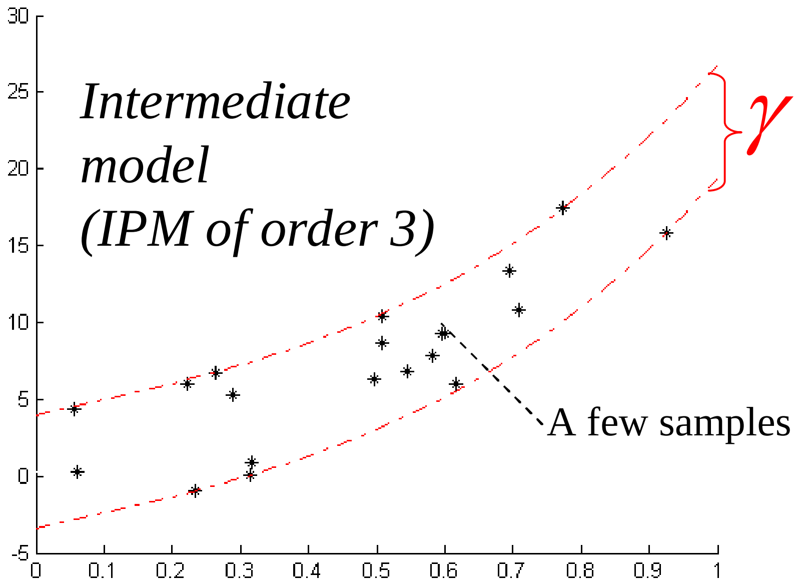
$$\min_{q_0, q_1, \ldots, q_n} \gamma$$

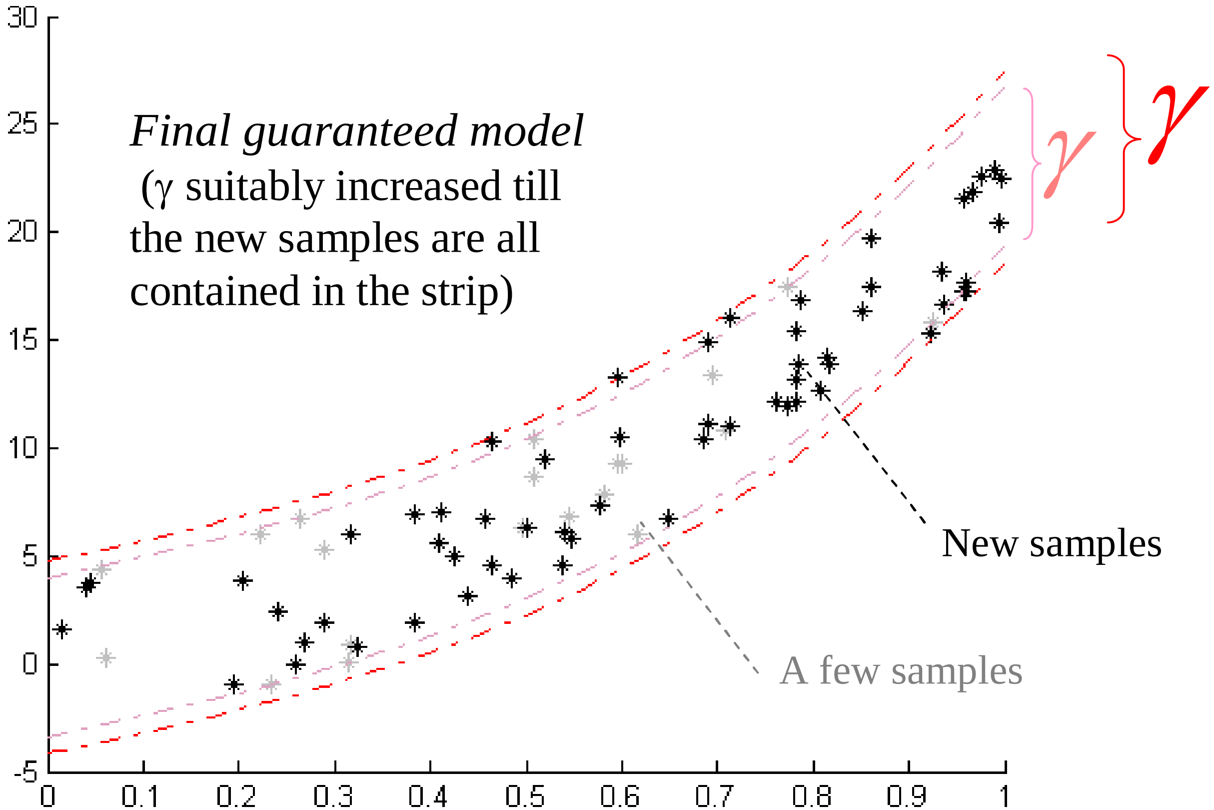
$$\text{subject to: } \left| y^{(i)} - \left[q_n \left(u^{(i)}\right)^n + q_{n-1} \left(u^{(i)}\right)^{n-1} + \ldots + q_1 \left(u^{(i)}\right) + q_0 \right] \right| \leq \gamma \quad i = 1, \ldots, N$$

(I) Same guarantees through a reduced number of samples

The FAST algorithm



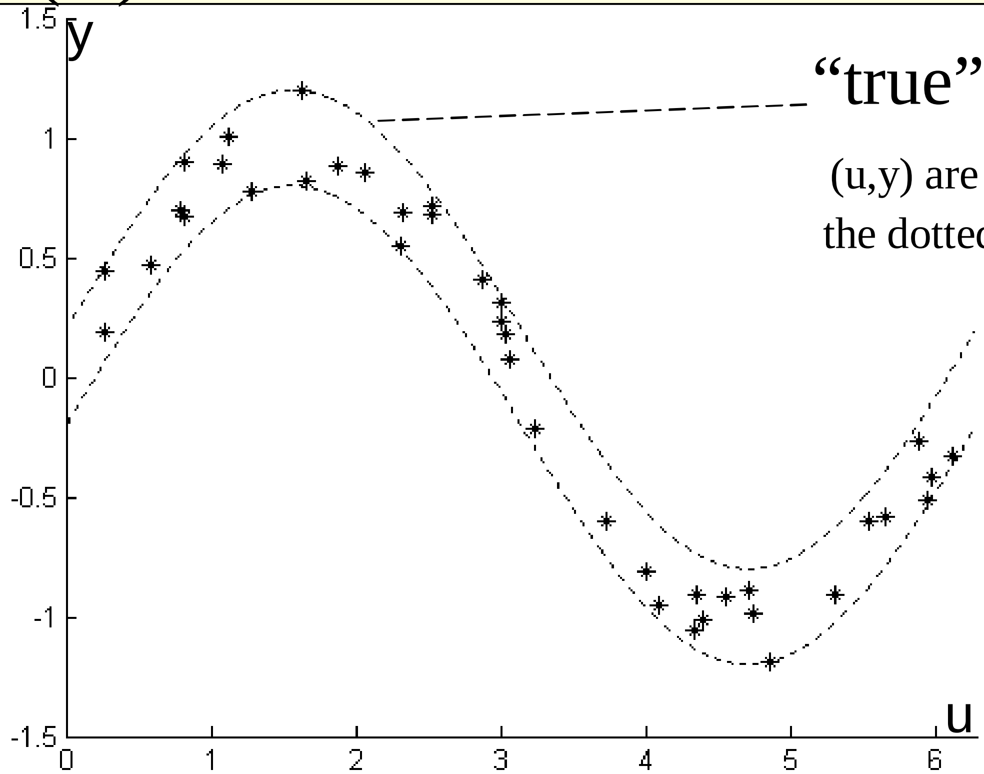




•**Ex.** If *order*=600, you may guarantee ε=0.1 with <u>one tenth</u> of the samples needed with the classic procedure.

•Computational advantages for medium and large scale problems.

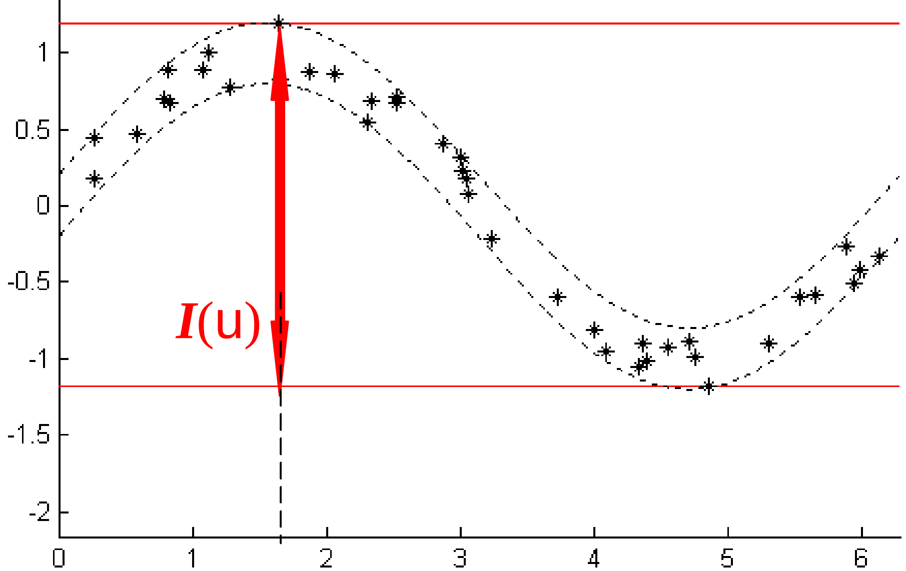
(II) Bias vs Variance trade-off



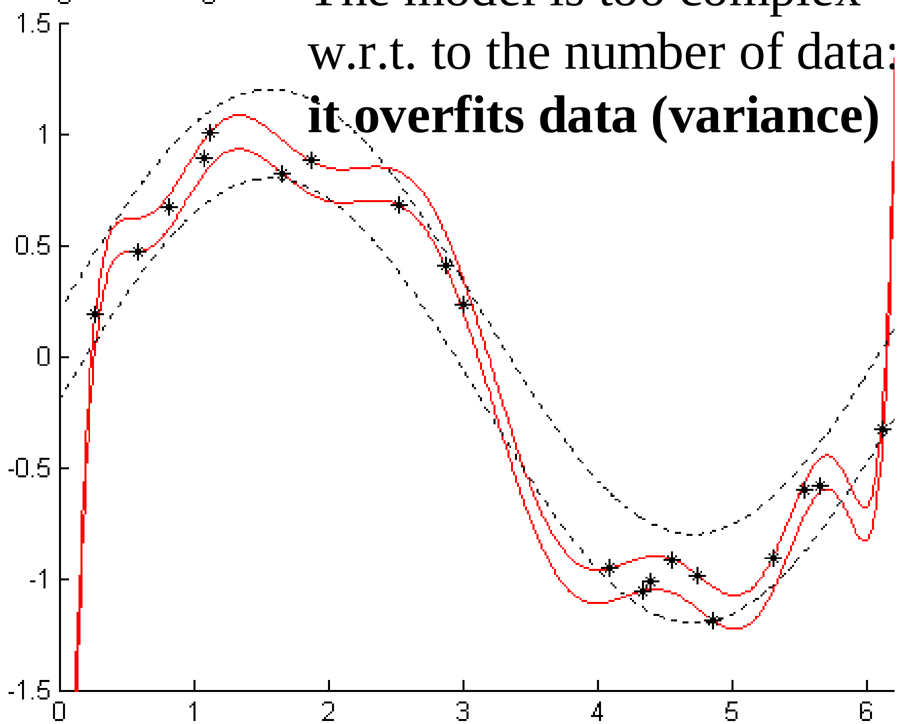
“true” model

(u,y) are uniformly distributed inside the dotted strip

The model is too simple:
it underfits data (bias)



The model is too complex w.r.t. to the number of data:
it overfits data (variance)



(II) How to trade bias and variance with the scenario approach

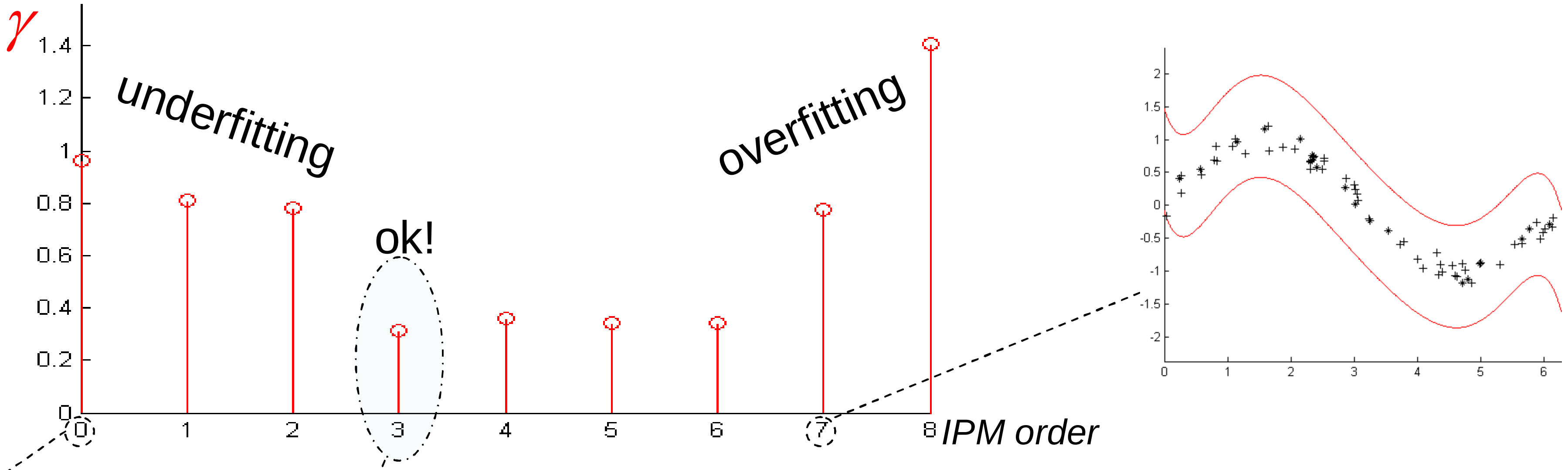
N samples → **Model builder** ← ε

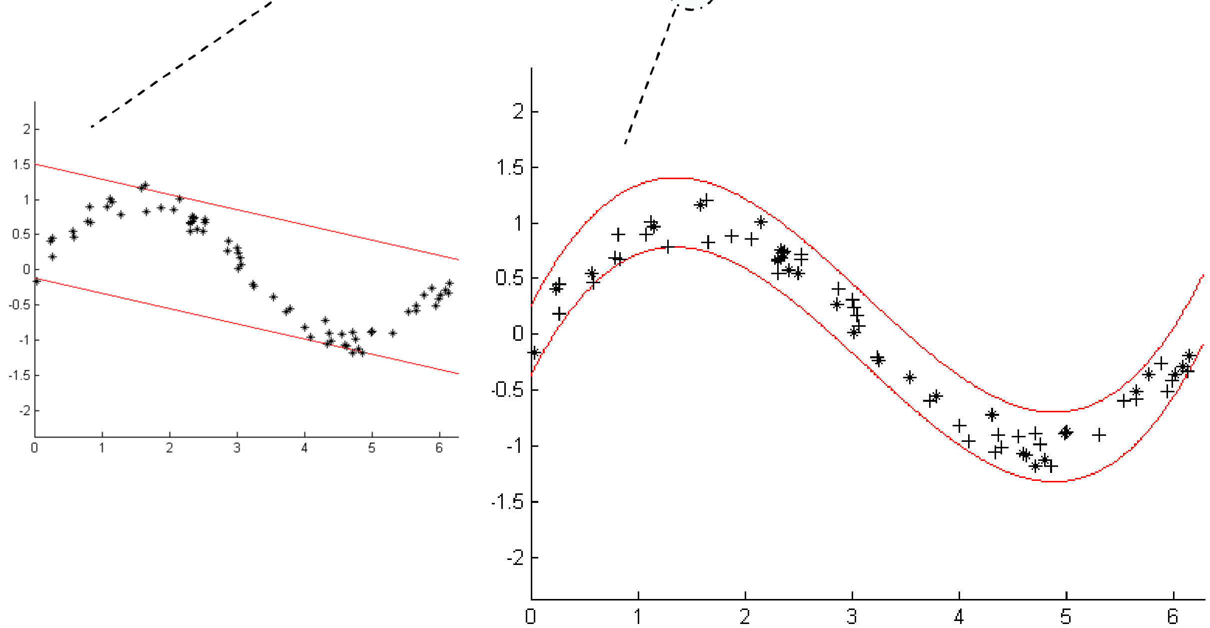
→

- ε-*guaranteed model of order 0 with perform.* γ_1
- ε-*guaranteed model of order 1 with perform.* γ_2

  …
- ε-*guaranteed model of order m with perform.* γ_m

Choose the best one





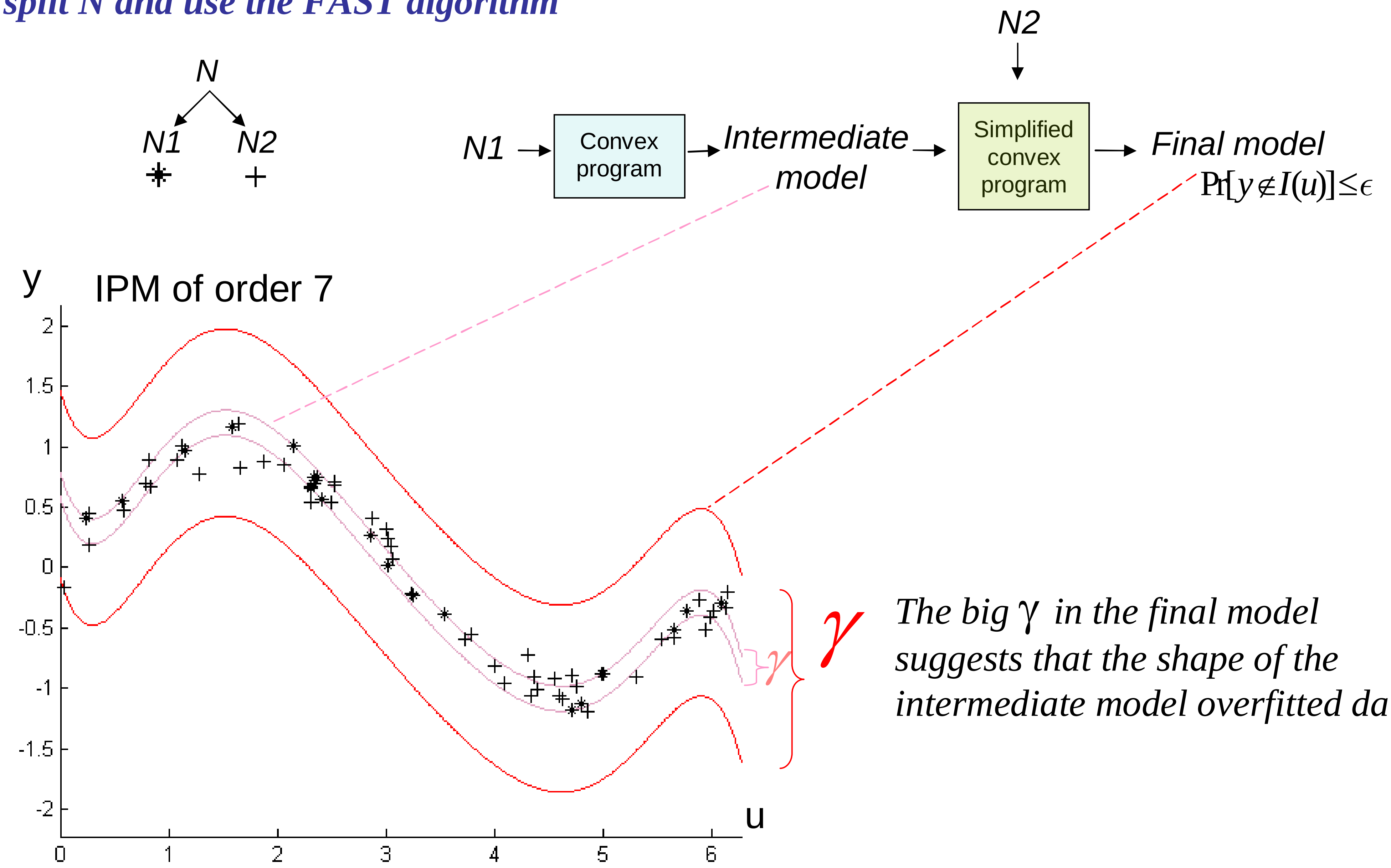
For all these models the same guarantee holds, with the same ε, i.e. $\Pr[y \notin I(u)] \leq \varepsilon$

The procedure allows to diclose overfitting or underfitting phenomena

just by looking at $\boldsymbol{\gamma}$

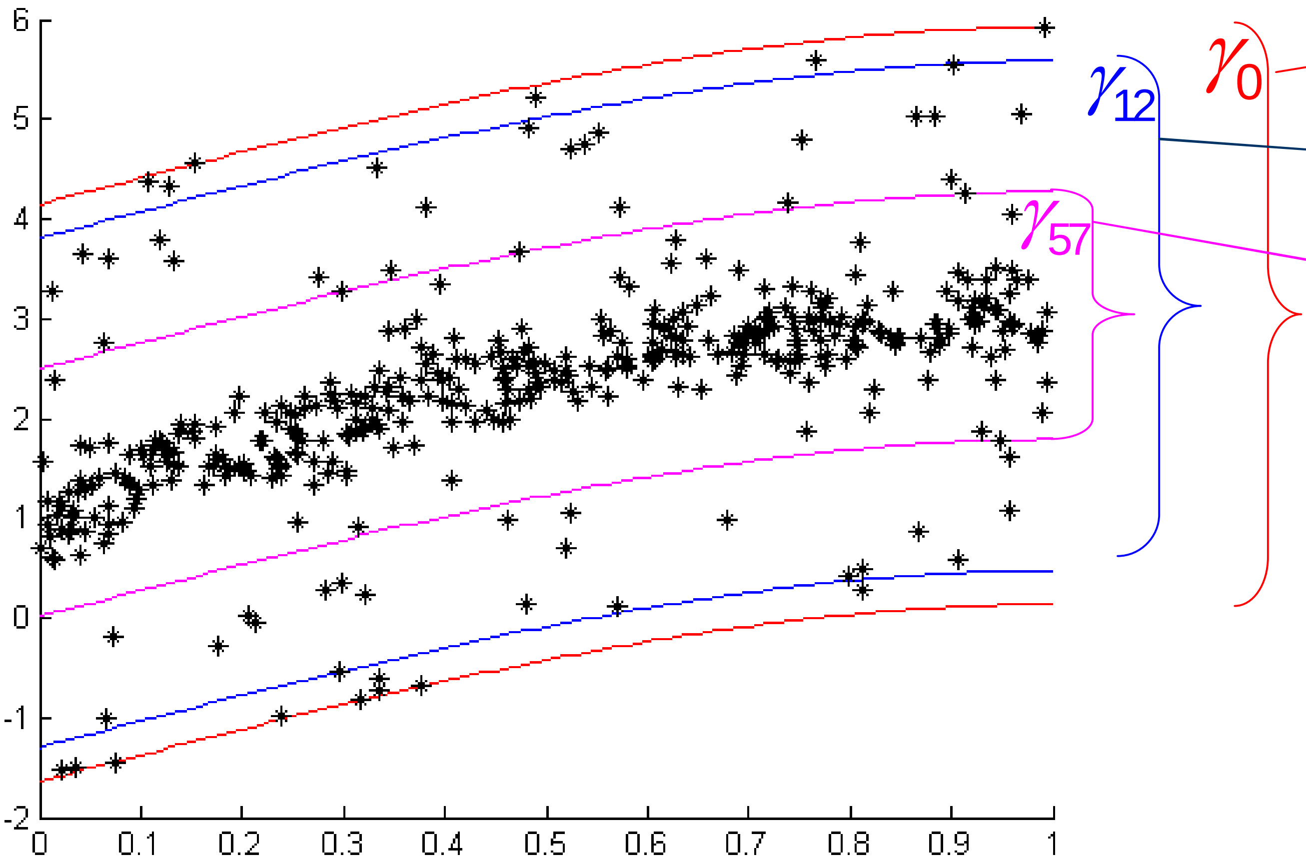
(II) Revealing overfitting through FAST

When the IPM order is too large to guarantee ε through the classic scenario approach, split N and use the FAST algorithm



The big γ in the final model suggests that the shape of the intermediate model overfitted data

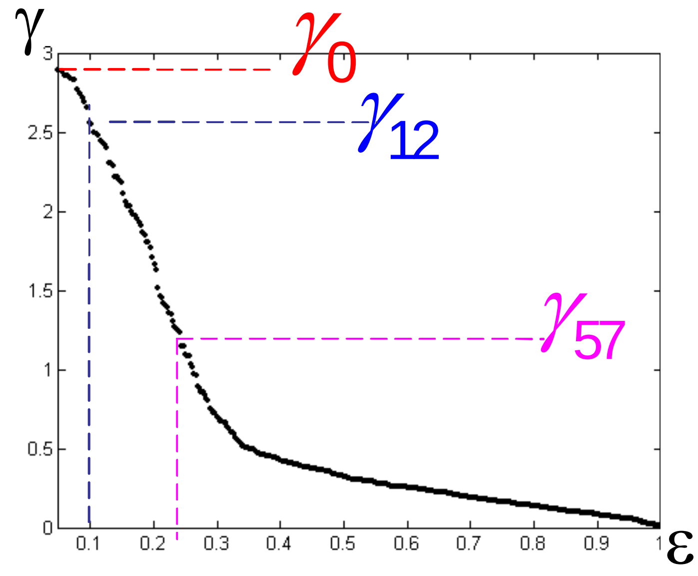
(III) Many a-priori reliability levels: $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{N-1}$ instead of just one ε



$\varepsilon_0 = 0.05 \Rightarrow$ No more than **95%** of mass probability outside this strip

$\varepsilon_{12} = 0.10 \Rightarrow$ No more than **90%** outside this

$\varepsilon_{57} = 0.24 \Rightarrow$ No more than **76%** outside this



N (e.g.=418)

↓

Exact Formula

↓

$\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{N-1}$ (e.g. **0.05**,...,**0.10**,…,**0.24**,…,1)

N samples → Convex program → IPM model

$(\gamma_0, \varepsilon_0), (\gamma_1, \varepsilon_1), \ldots, (\gamma_{N-1}, \varepsilon_{N-1})$