MS330: **Data-Driven Optimization: Algorithms and Theoretical Guarantees**

DATA-DRIVEN OPTIMIZATION VIA THE SCENARIO APPROACH: A THEORY FOR THE USER BASED ON CONDITIONAL RISK ASSESSMENTS

Speaker: **Simone Garatti**

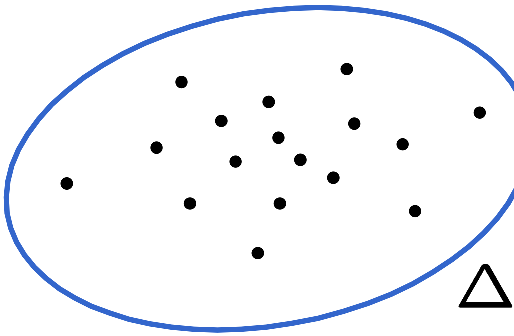
(Politecnico di Milano, Italy – email: simone.garatti@polimi.it)

in collaboration with: **Marco C. Campi**

(University of Brescia, Italy – email: marco.campi@unibs.it)

What is data-driven optimization?

δ = uncertain element ⇨ exercise caution

Data 

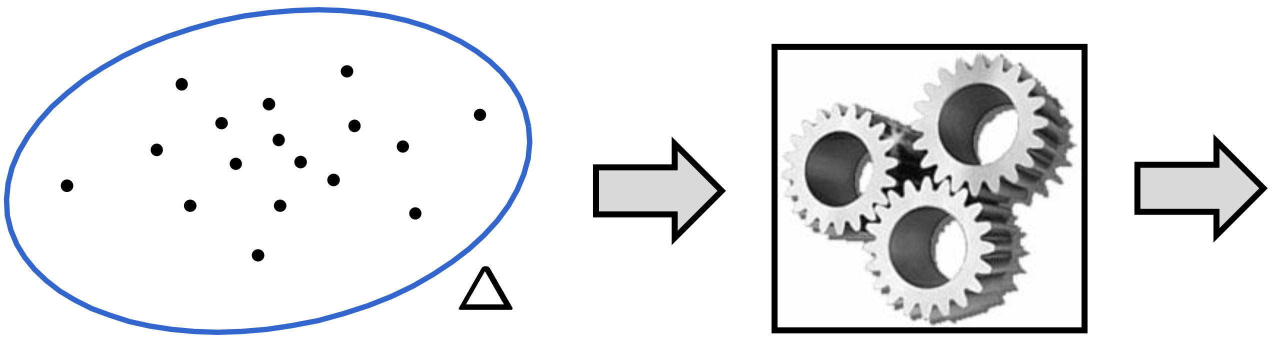
⇨  ⇨ **Decision** x^*

$\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}$ (i.i.d. sample, scenarios)

What is data-driven optimization?

δ = uncertain element ⇒ exercise caution

Data



Decision x^*

$\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}$ (i.i.d. sample, scenarios)

Desiderata

Effective ⇒ no over-conservatism

Dependable ⇒ certified decisions

Data-driven decision making via the Scenario Aproach

- Enforce design goals heuristically
- Solid theory to assess the quality of the solution

Data-driven decision making via the Scenario Approach

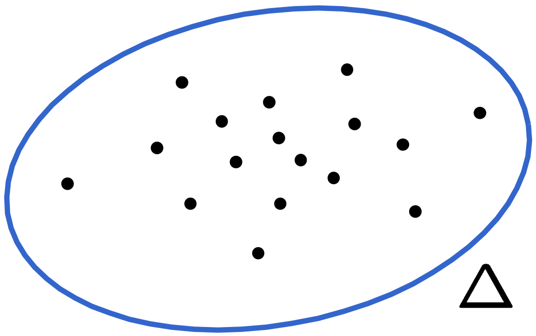
- Enforce design goals heuristically
- Solid theory to assess the quality of the solution

- Precise non-conditional distribution-free characterization of the risk (certification guaranteed to hold almost certainly with respect to the variability of data realizations)
- End-user perspective requires conditional evaluations, given what has been seen
- Are distribution-free certifications still possible? What is the extent of the needed information?

THIS TALK

A prototypical paradigm: Robust Scenario Optimization

Objectives: minimize $f(x)$ – satisfy $x \in \mathcal{X}_\delta$ (uncertain)



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

(scenarios)

Robust scenario program

$$\min_x \quad f(x)$$

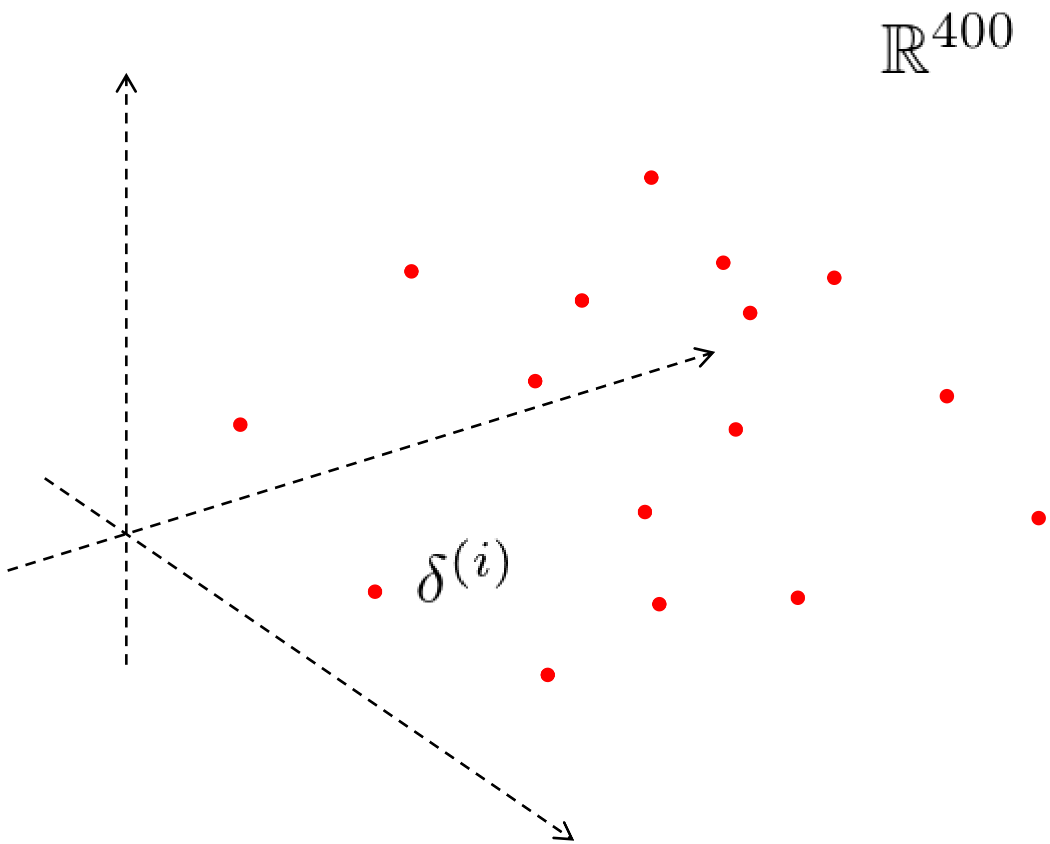
$$\text{s.t.} \quad x \in \bigcap_{i=1}^{N} \mathcal{X}_{\delta^{(i)}}$$

x^* = scenario solution

NOTE r.s.o. $\Longleftrightarrow$ safeguarding against the worst…

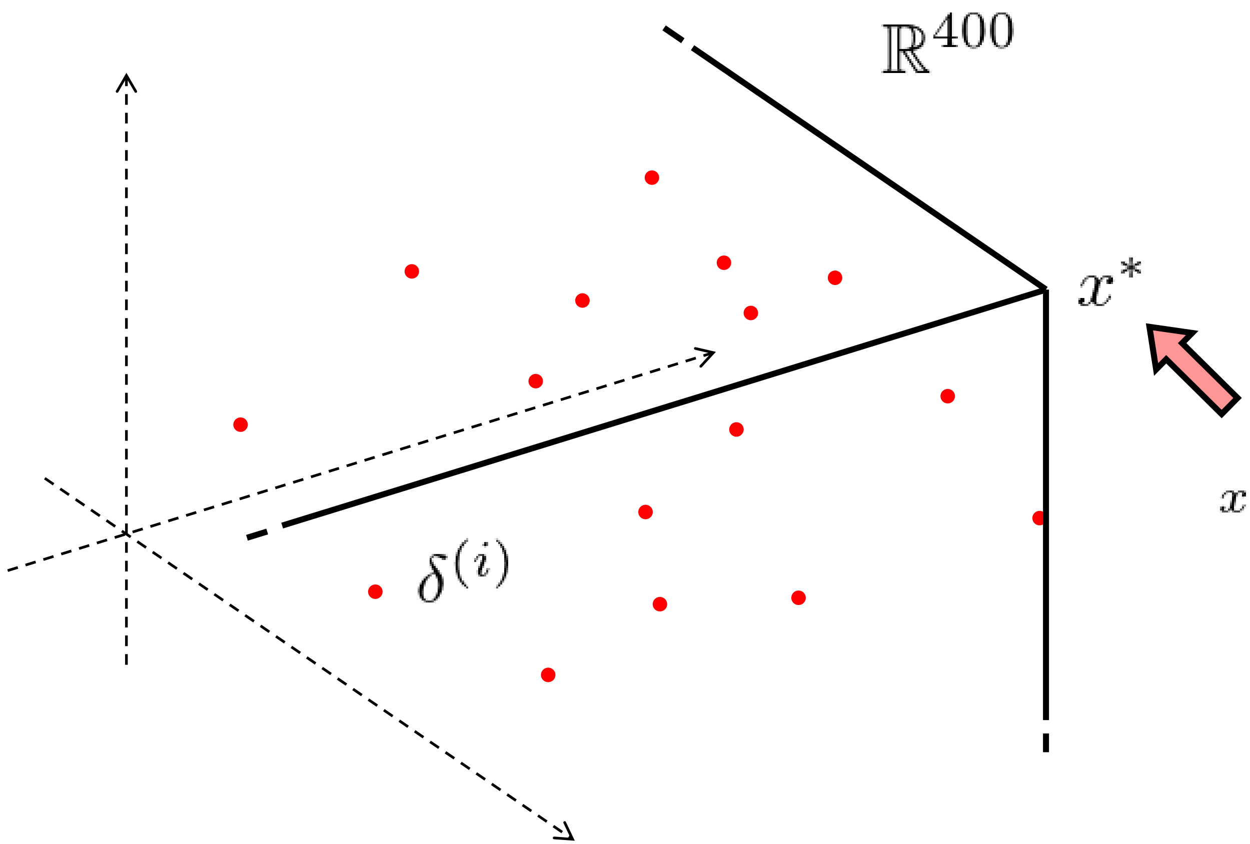
many other scenario paradigms exists!

An example



$$N = 1000$$

An example



$$\min_{x \in \mathbb{R}^{400}} \quad \sum_{k=1}^{400} x_k$$
$$\text{s.t.} \quad \delta^{(i)} \leq x,$$
$$i = 1, \ldots, 1000$$

$N = 1000$

Solution certification

$f(x)$ vs. $V(x) = \mathbb{P}\{\delta \in \Delta : \; x \notin \mathcal{X}_\delta\}$

cost

out-of-sample constraint violation (probabilistic mass outside a given orthant)

$\mathbb{P}$ = mechanism with which the δ are generated

scenario solution certification

- $f(x^*)$
- $V(x^*)$

Solution certification

$f(x)$ vs. $V(x) = \mathbb{P}\{\delta \in \Delta : \; x \notin \mathcal{X}_\delta\}$

cost

out-of-sample constraint violation (probabilistic mass outside a given orthant)

$\mathbb{P}$ = mechanism with which the δ are generated

scenario solution certification

- $f(x^*)$ accessible (once x^* is computed)
- $V(x^*)$

Solution certification

$f(x)$ vs. $V(x) = \mathbb{P}\{\delta \in \Delta : \; x \notin \mathcal{X}_\delta\}$

cost

out-of-sample constraint violation (probabilistic mass outside a given orthant)

$\mathbb{P}$ = mechanism with which the δ are generated

??????????

scenario solution certification

- $f(x^*)$ accessible (once x^* is computed)
- $V(x^*)$ **not accessible**

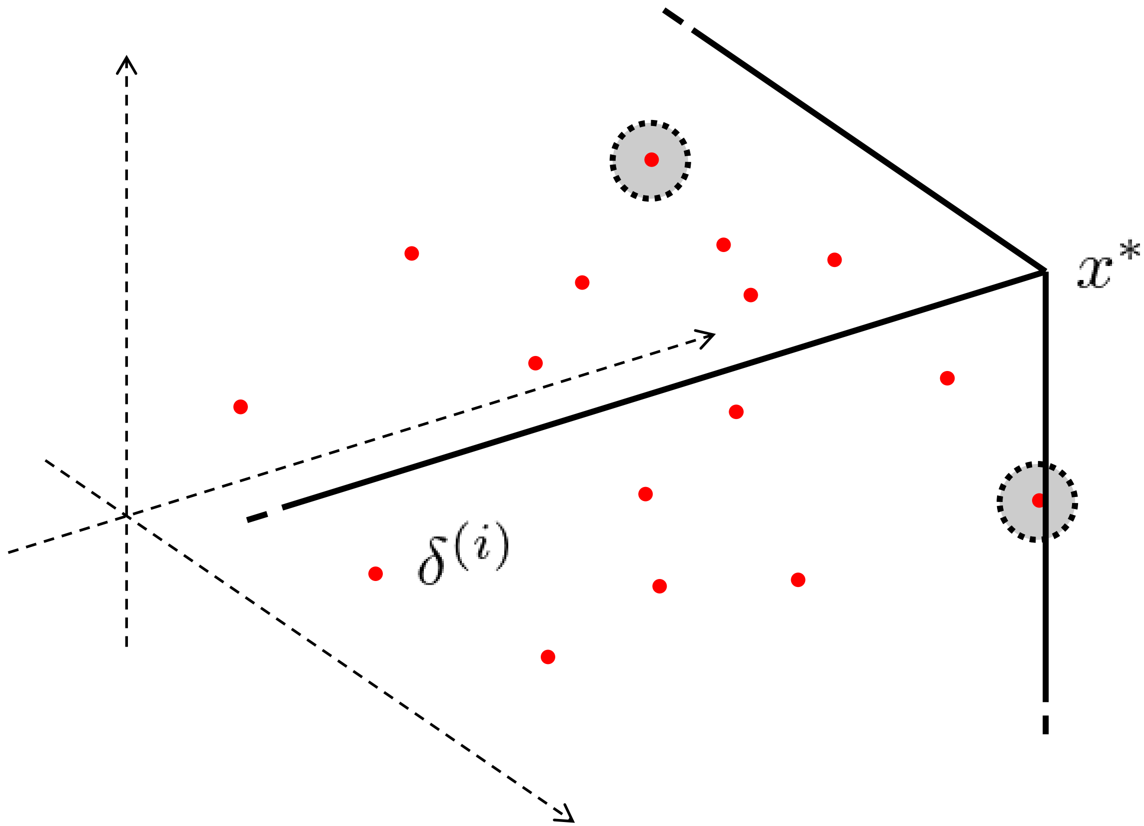
Risk and complexity

$$V(x^*) = V(x^*(\delta^{(1)}, \delta^{(2)}, \dots, \delta^{(N)})) \quad \left\{ \begin{array}{l} \text{random variable,} \\ \textbf{not accessible} \end{array} \right.$$

Risk and complexity

$$V(x^*) = V(x^*(\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}))$$ random variable, **not accessible**

s^* = least no. of $\delta^{(i)}$ that are needed to reconstruct x^*



Risk and complexity: an example

$$V(x^*) = V(x^*(\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}))$$

random variable,
not accessible

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

random variable,
observable

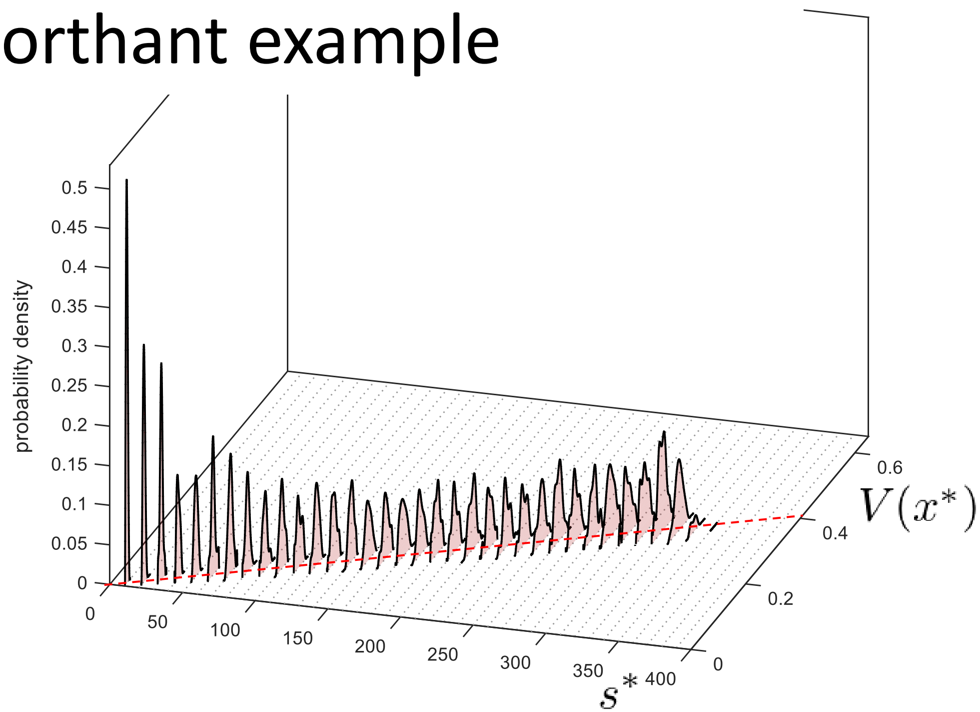
joint distribution for the orthant example

Risk and complexity: an example

$$V(x^*) = V(x^*(\delta^{(1)}, \delta^{(2)}, \dots, \delta^{(N)}))$$ $\begin{cases} \text{random variable,} \\ \text{not accessible} \end{cases}$

$$s^* = s^*(\delta^{(1)}, \delta^{(2)}, \dots, \delta^{(N)})$$ $\begin{cases} \text{random variable,} \\ \text{observable} \end{cases}$

joint distribution for the orthant example



high correlation

⇓

$V(x^*)$ can be accurately estimated from s^*

Main result

Choose $\beta \in (0,1)$ (confidence parameter)

$\epsilon(k)$ = unique root in (0,1) of polynomial

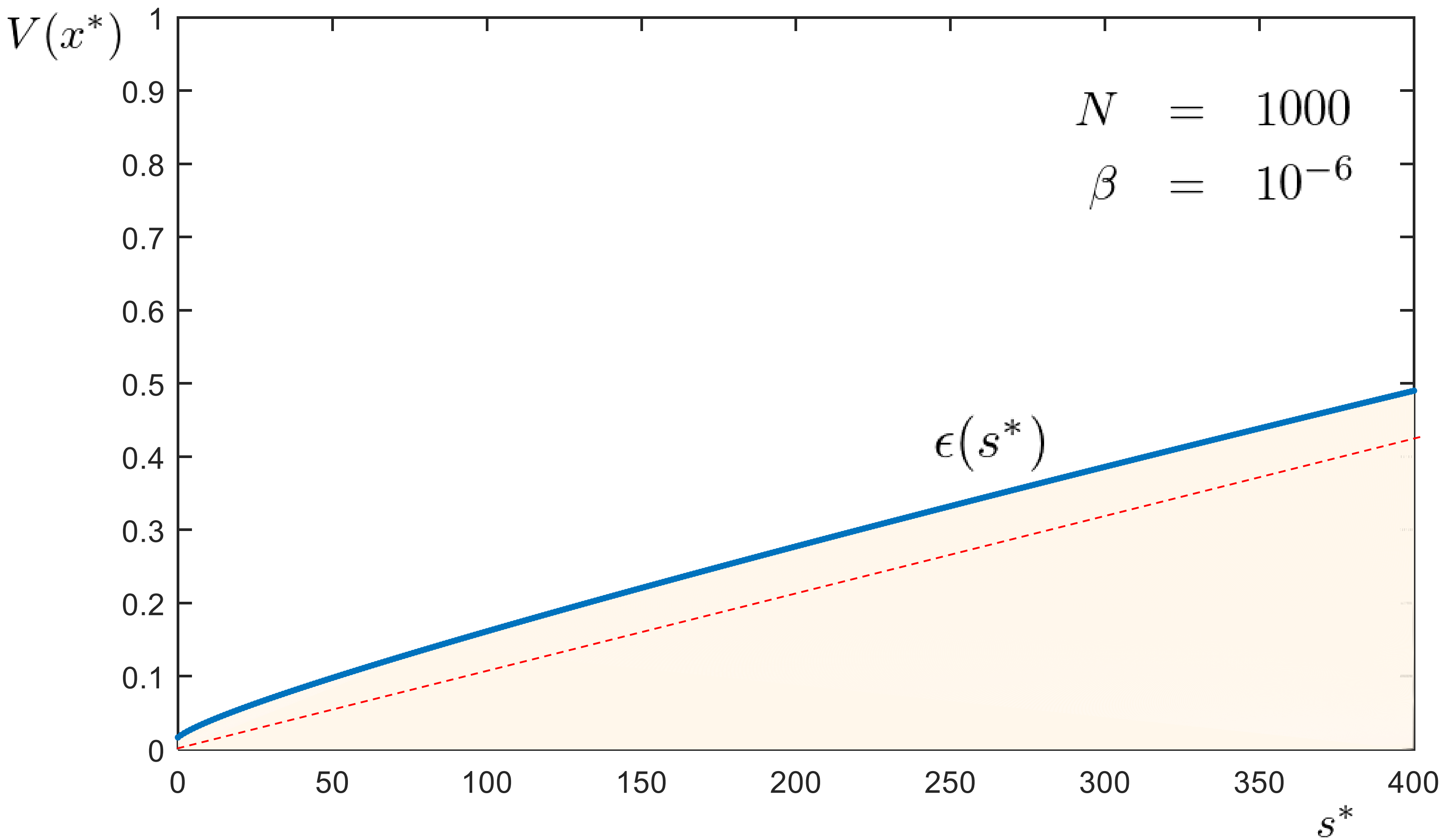
$$\triangleright \quad \binom{N}{k}(1-\epsilon)^{N-k} - \frac{\beta}{N}\sum_{m=k}^{N-1}\binom{m}{k}(1-\epsilon)^{m-k} \qquad k = 0, 1, \ldots$$

Then, irrespective of $\mathbb{P}$ (**distribution-free**),

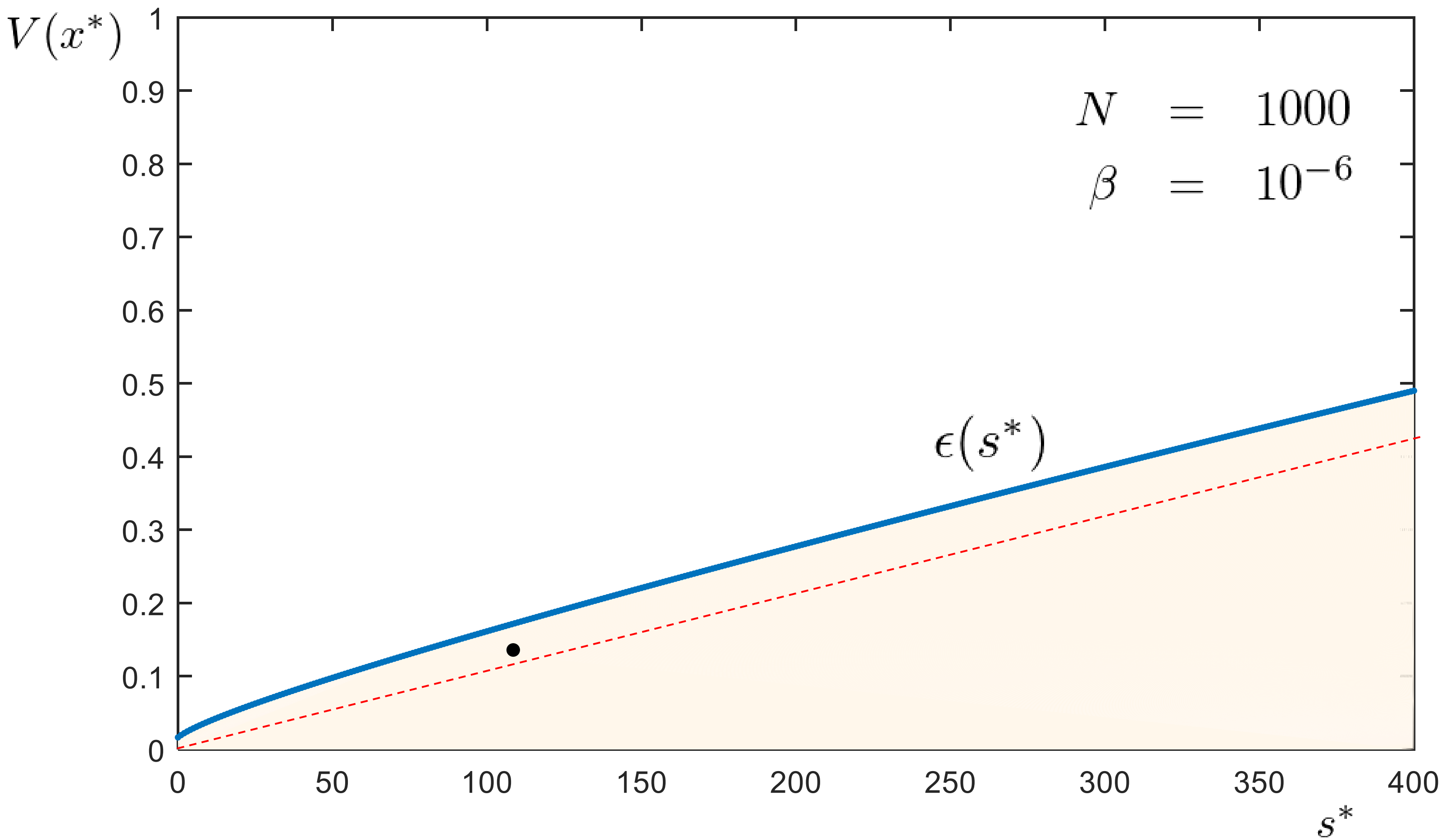
$$\mathbb{P}^N\{\ V(x^*) > \epsilon(s^*)\ \} \leq \beta$$

... use $\epsilon(s^*)$ as a certification of $V(x^*)$

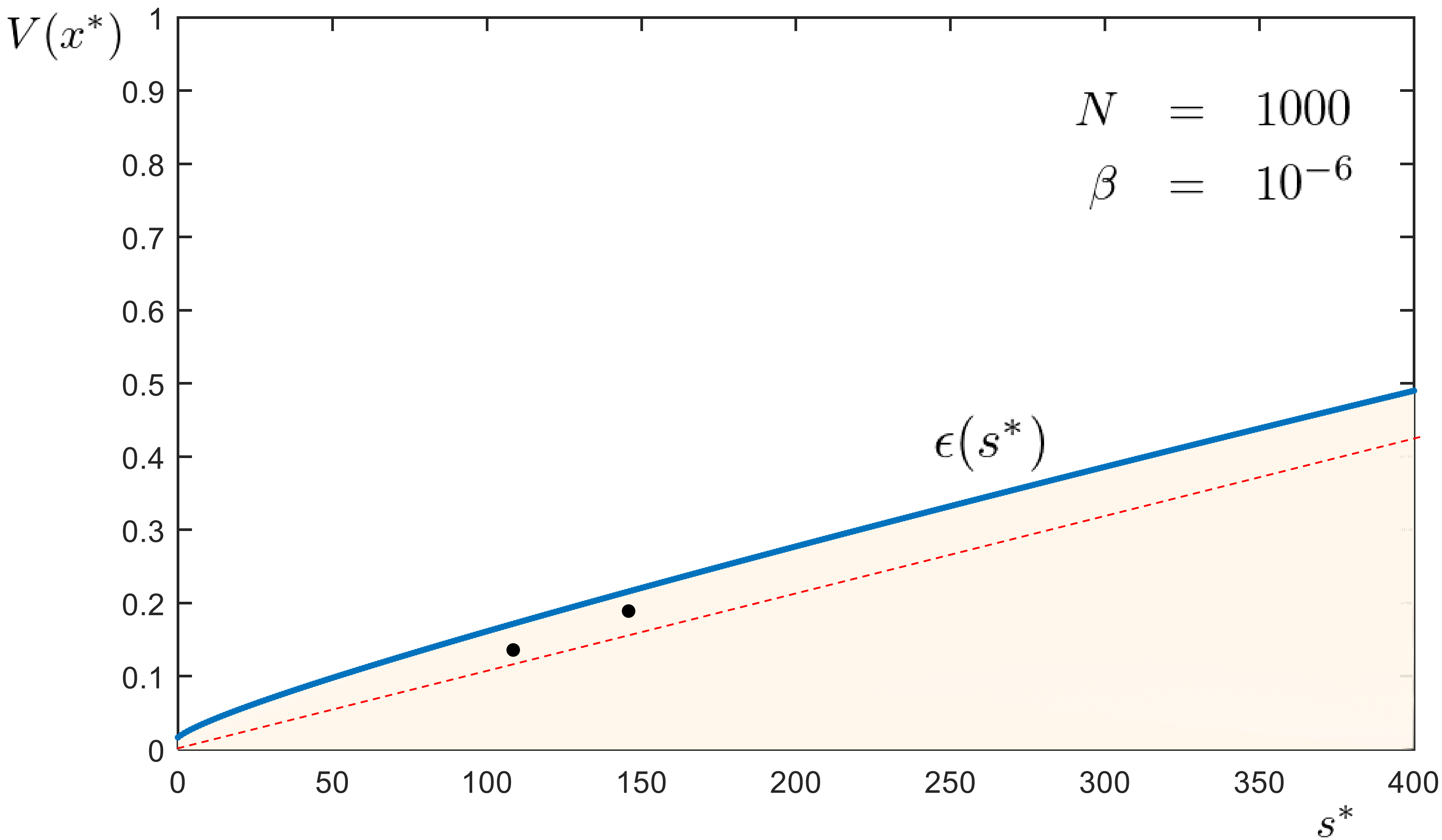
Main result – cont'd



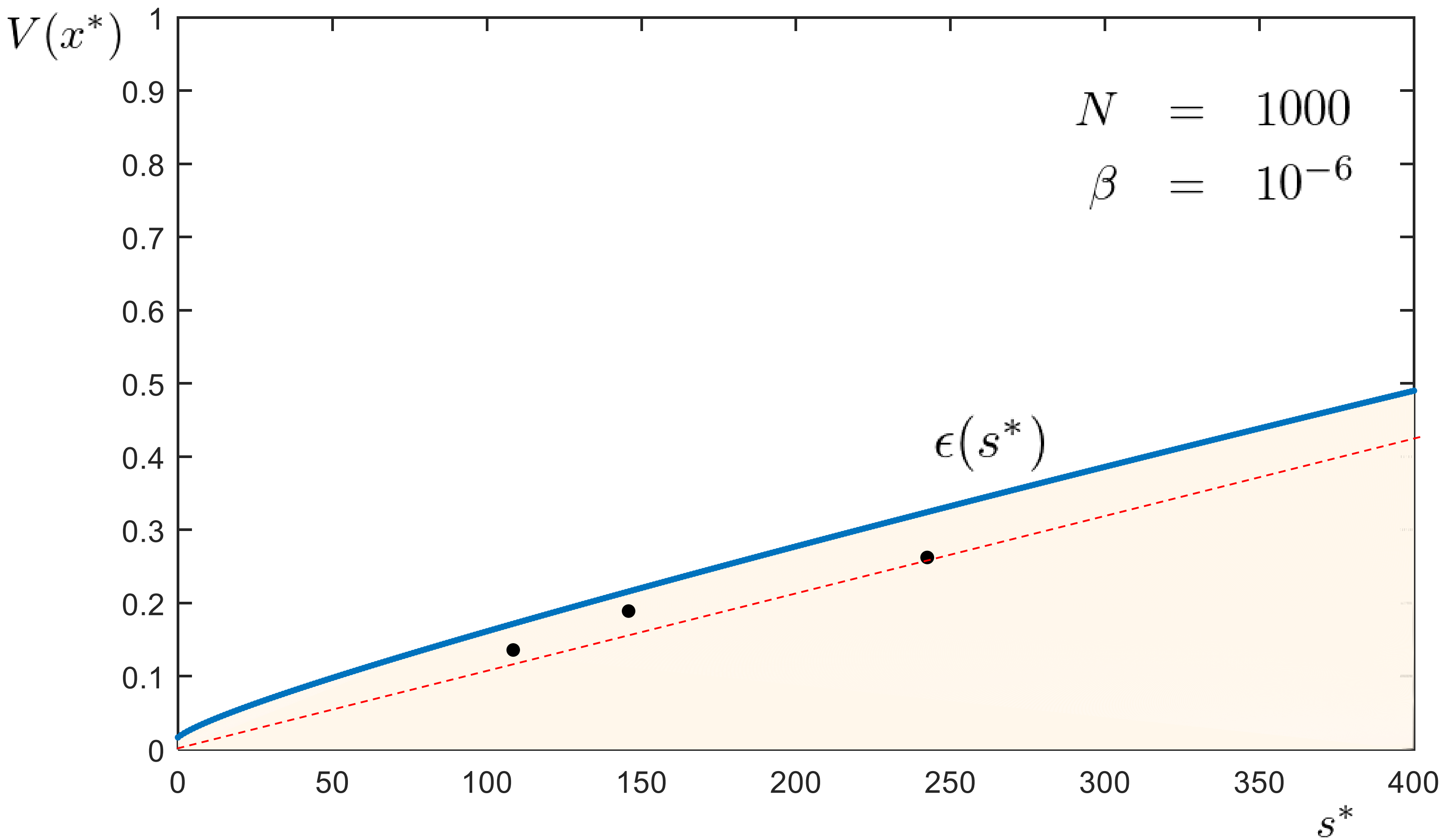
Main result – cont'd



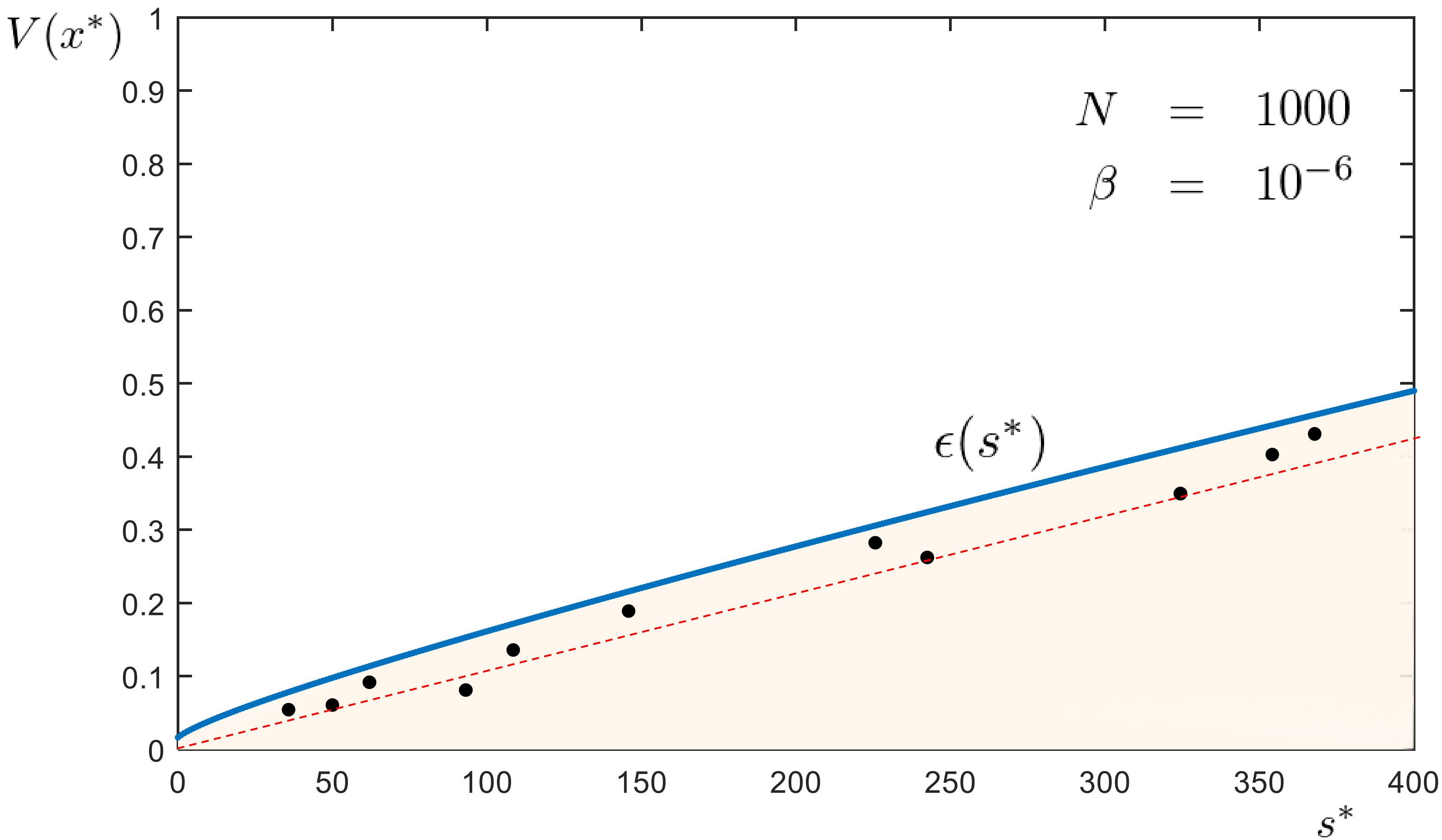
Main result – cont'd



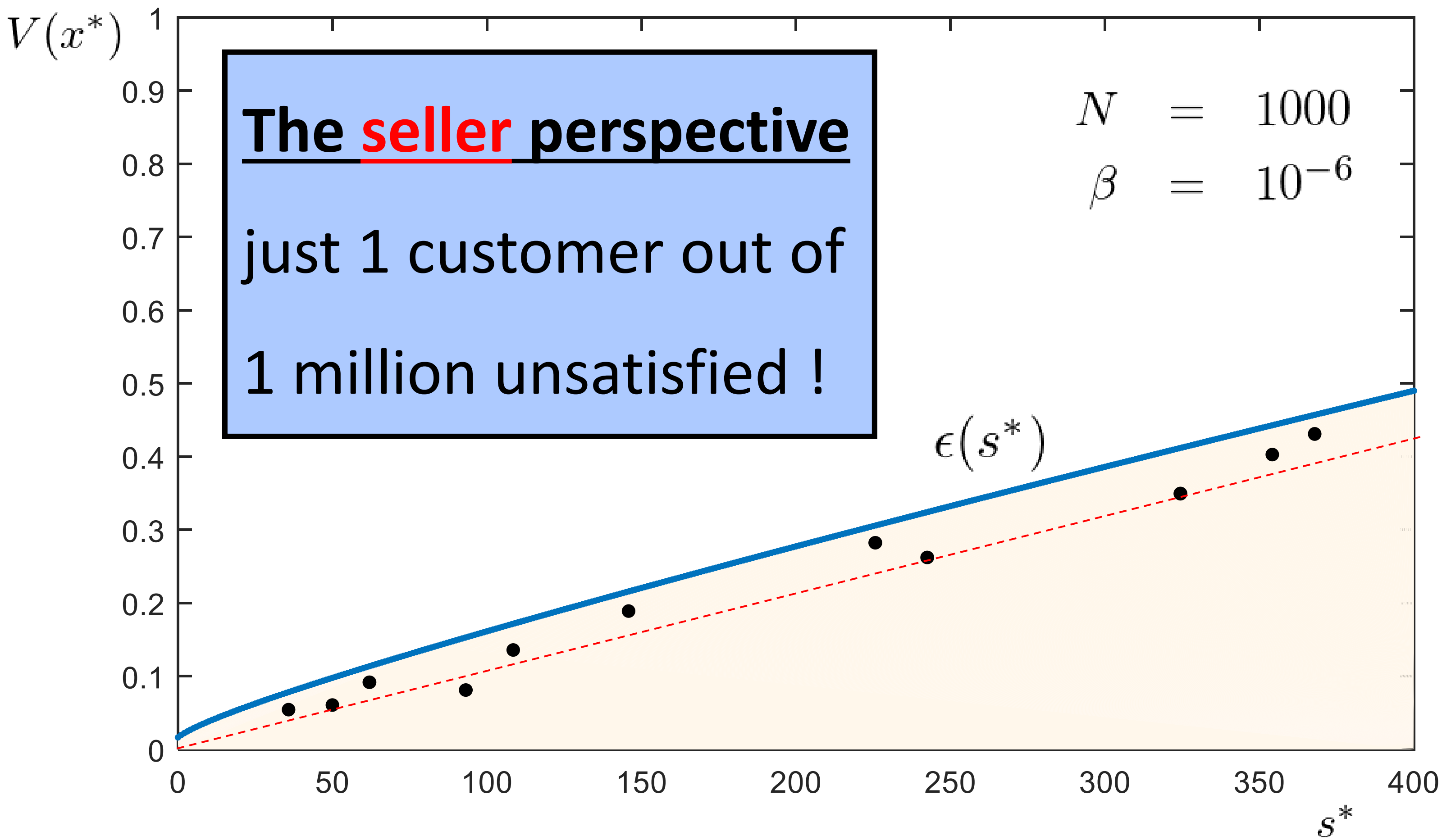
Main result – cont'd



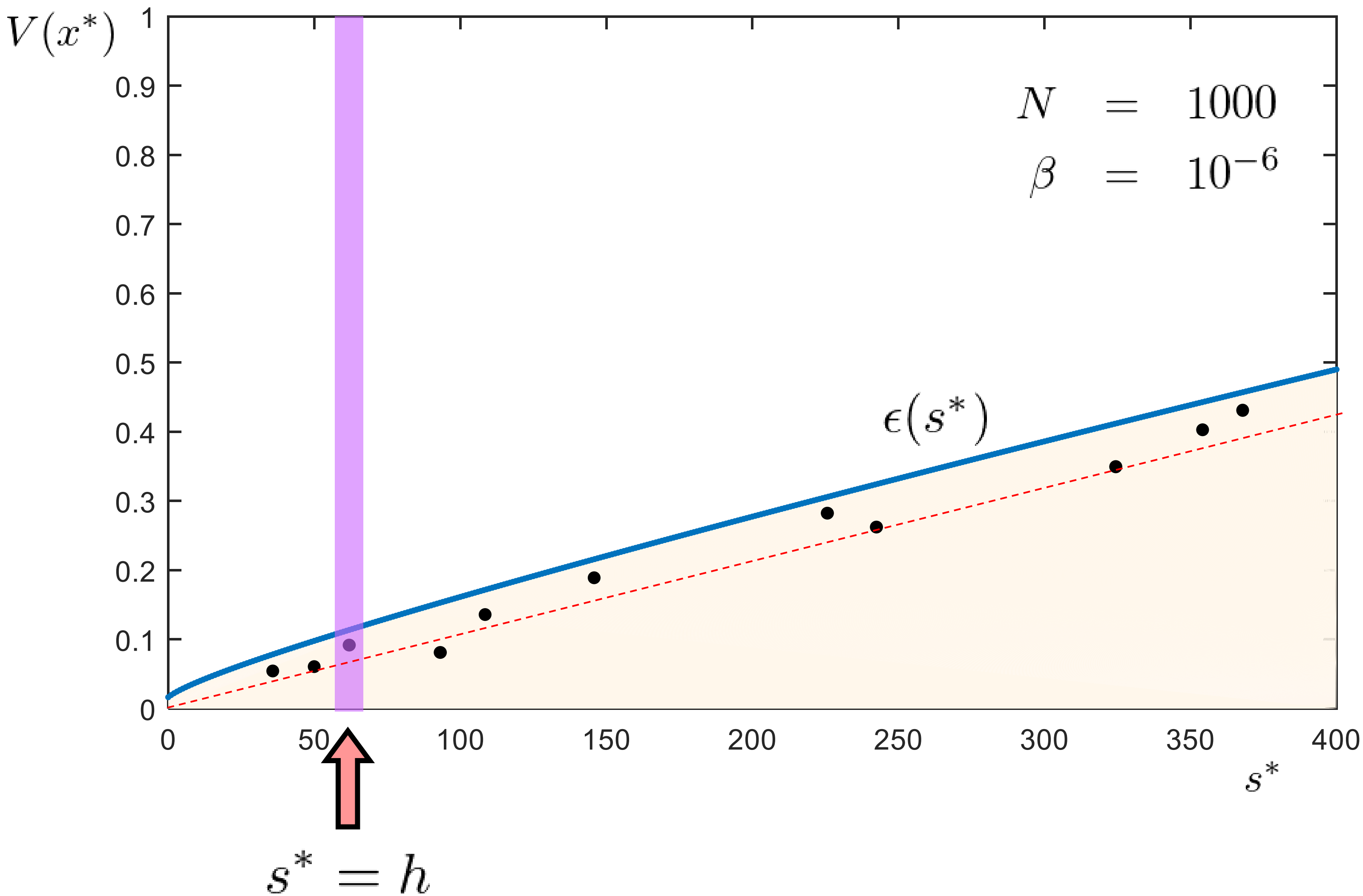
Main result – cont'd



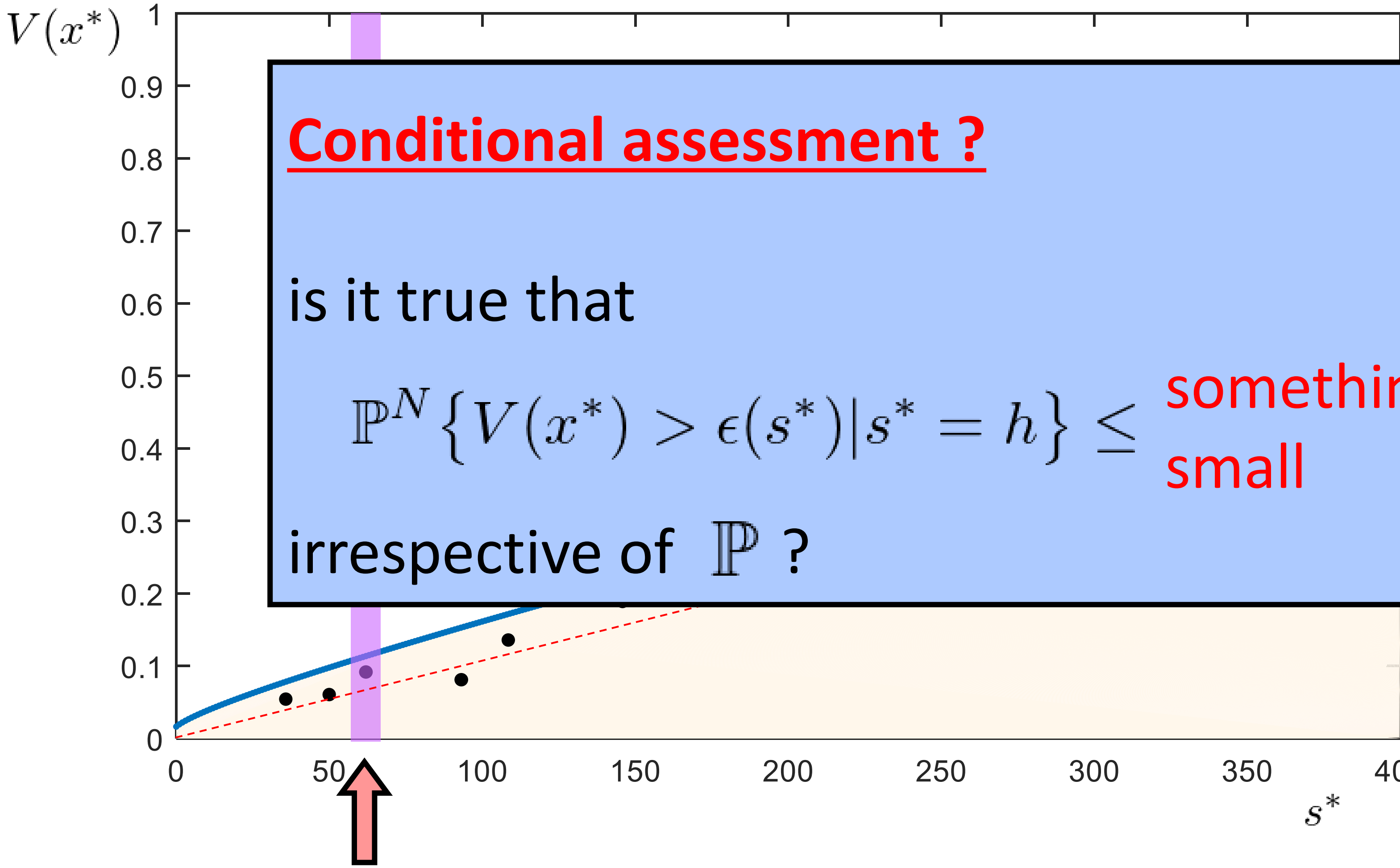
Main result – cont'd



The buyer perspective



The buyer perspective



Conditional assessment ?

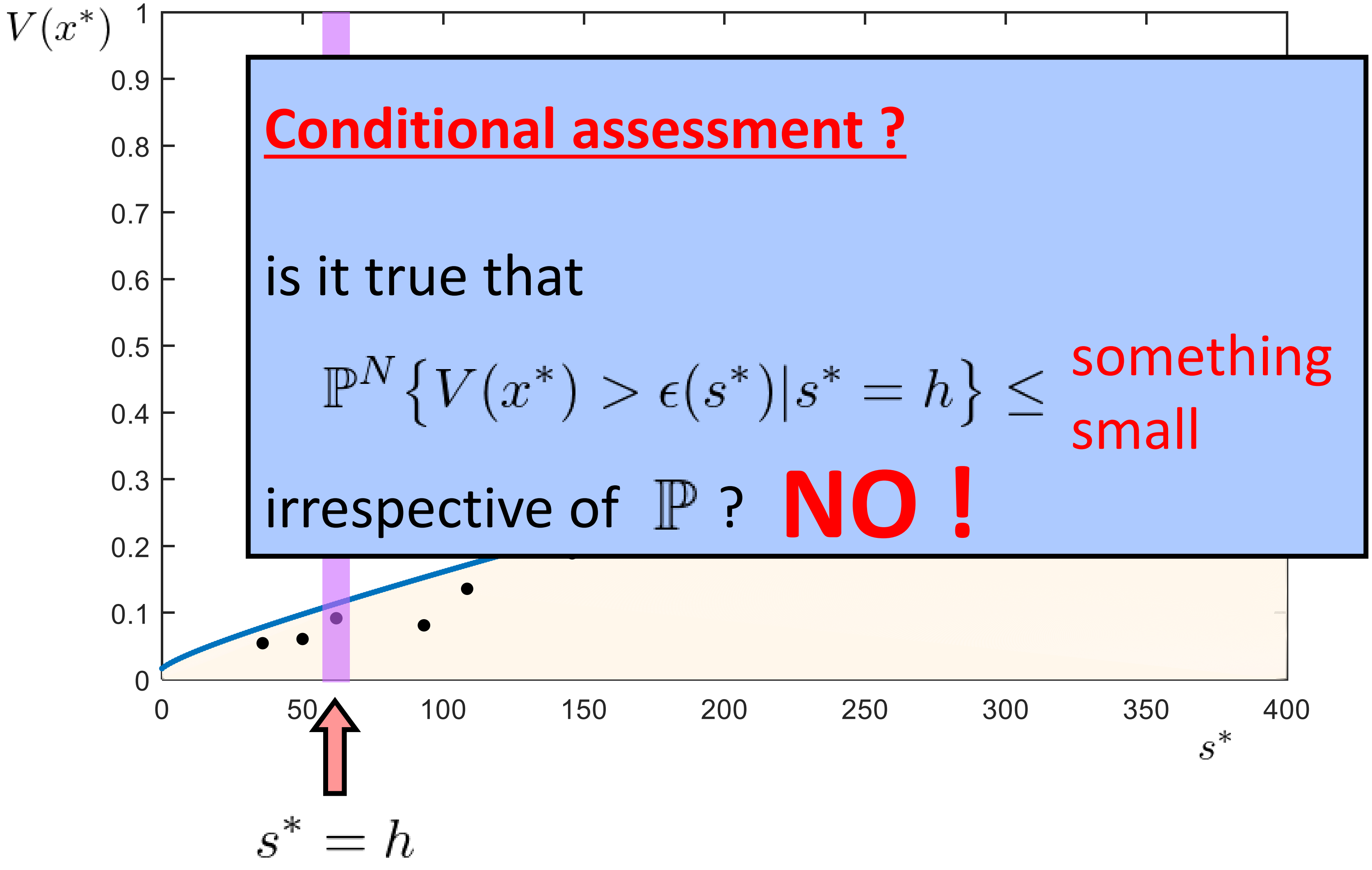
is it true that

$$\mathbb{P}^N\{V(x^*) > \epsilon(s^*) | s^* = h\} \leq$$ something small

irrespective of $\mathbb{P}$?

$s^* = h$

The buyer perspective



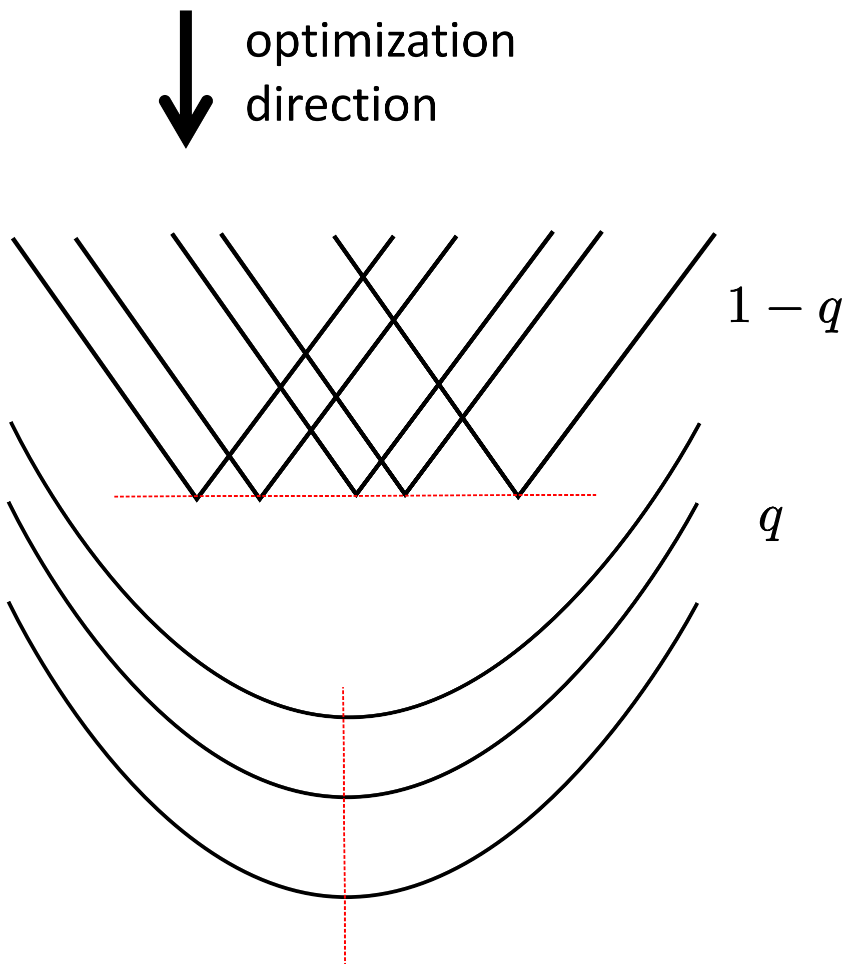
Conditional assessment ?

is it true that

$$\mathbb{P}^N\{V(x^*) > \epsilon(s^*) | s^* = h\} \leq$$ something small

irrespective of $\mathbb{P}$? **NO !**

An impossibility result



$$\mathbb{P}^N\{V(x^*) > \epsilon(s^*) | s^* = 1\}$$

$$= \frac{\mathbb{P}^N\{V(x^*) > \epsilon(1) \wedge s^* = 1\}}{\mathbb{P}^N\{s^* = 1\}}$$

$$= [\text{if } \epsilon(1) < 1-q]$$

$$= \frac{\mathbb{P}^N\{s^* = 1\}}{\mathbb{P}^N\{s^* = 1\}} = 1$$

A Bayesian framework

$\mathbb{P}_\theta$ = mechanism with which the δ's are generated
π = prior distribution on θ
$\mathbf{P}$ = total probability

Objective: to evaluate

$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\}$$

$$= \frac{\int_\Theta \mathbb{P}_\theta^N\{V_\theta(x^*) > \epsilon(h) \wedge s^* = h\}\pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$

A Bayesian framework

$\mathbb{P}_\theta$ = mechanism with which the δ's are generated

π = prior distribution on θ

$\mathbf{P}$ = total probability

Objective: to evaluate

$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\}$$

$$= \frac{\int_\Theta \mathbb{P}_\theta^N \{V_\theta(x^*) > \epsilon(h) \wedge s^* = h\} \pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N \{s^* = h\} \pi(\mathrm{d}\theta)}$$

... scary!

The idea (leveraging previous result)

$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\} = \frac{\int_\Theta \mathbb{P}_\theta^N\{V_\theta(x^*) > \epsilon(h) \wedge s^* = h\}\pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$

The idea (leveraging previous result)

$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\} = \frac{\int_\Theta \boxed{\mathbb{P}_\theta^N\{V_\theta(x^*) > \epsilon(h) \wedge s^* = h\}} \pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$

$$\boxed{\leq \begin{cases} \beta & \Leftarrow \text{previous result} \\ \mathbb{P}_\theta^N\{s^* = h\} & \Leftarrow \text{trivial} \end{cases}}$$

The idea (leveraging previous result)

$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\} = \frac{\int_\Theta \boxed{\mathbb{P}_\theta^N\{V_\theta(x^*) > \epsilon(h) \wedge s^* = h\}} \pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$

$$\boxed{\leq \begin{cases} \beta & \Longleftarrow \text{previous result} \\ \mathbb{P}_\theta^N\{s^* = h\} & \Leftarrow \text{trivial} \end{cases}}$$

$$\leq \quad \frac{\int_{\mathbb{P}_\theta^N\{s^*=h\}<\beta} \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta) + \beta \cdot \int_{\mathbb{P}_\theta^N\{s^*=h\}\geq\beta} \pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$

Main result (buyer perspective)

$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\}$$

$$\leq \frac{\int_{\mathbb{P}_\theta^N\{s^*=h\}<\beta} \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta) + \beta \cdot \int_{\mathbb{P}_\theta^N\{s^*=h\}\geq\beta} \pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$

Main result (buyer perspective)

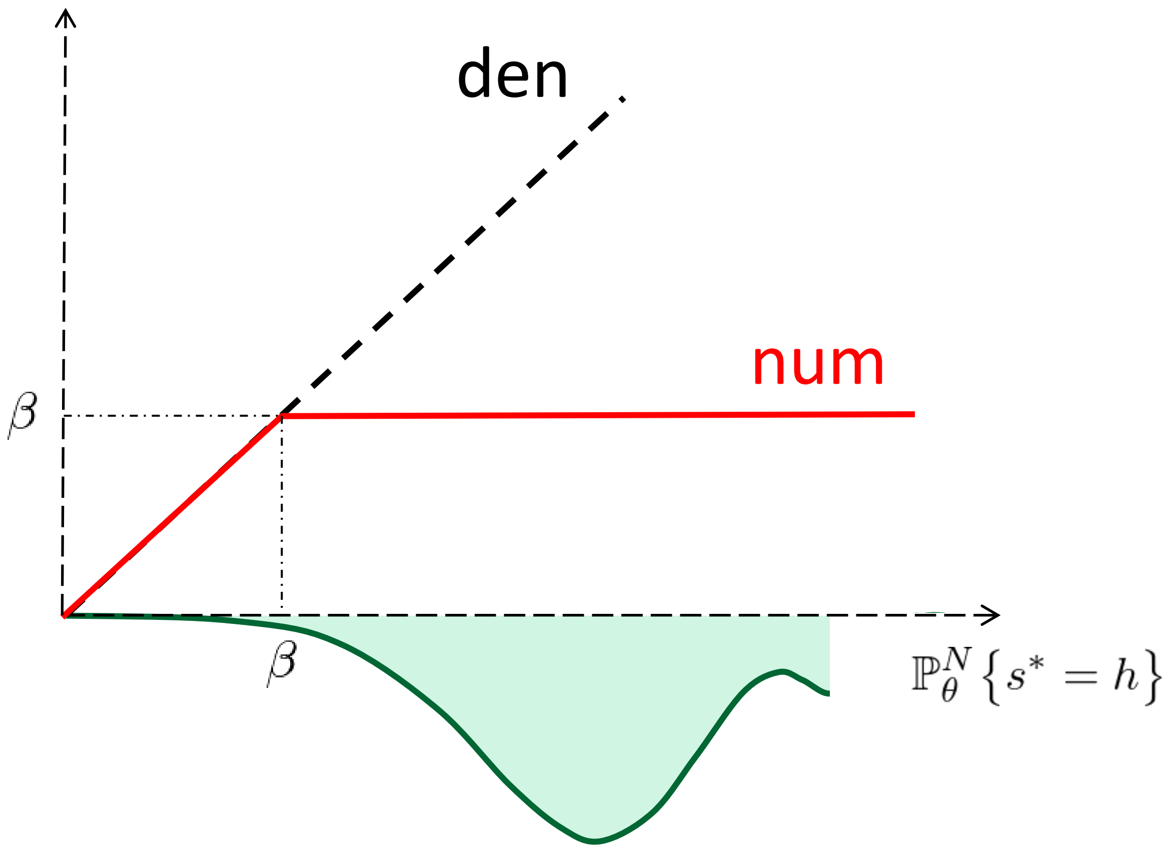
$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\}$$

$$\leq \frac{\int_{\mathbb{P}_\theta^N\{s^*=h\}<\beta} \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta) + \beta \cdot \int_{\mathbb{P}_\theta^N\{s^*=h\}\geq\beta} \pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$

Main result (buyer perspective)

$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\}$$

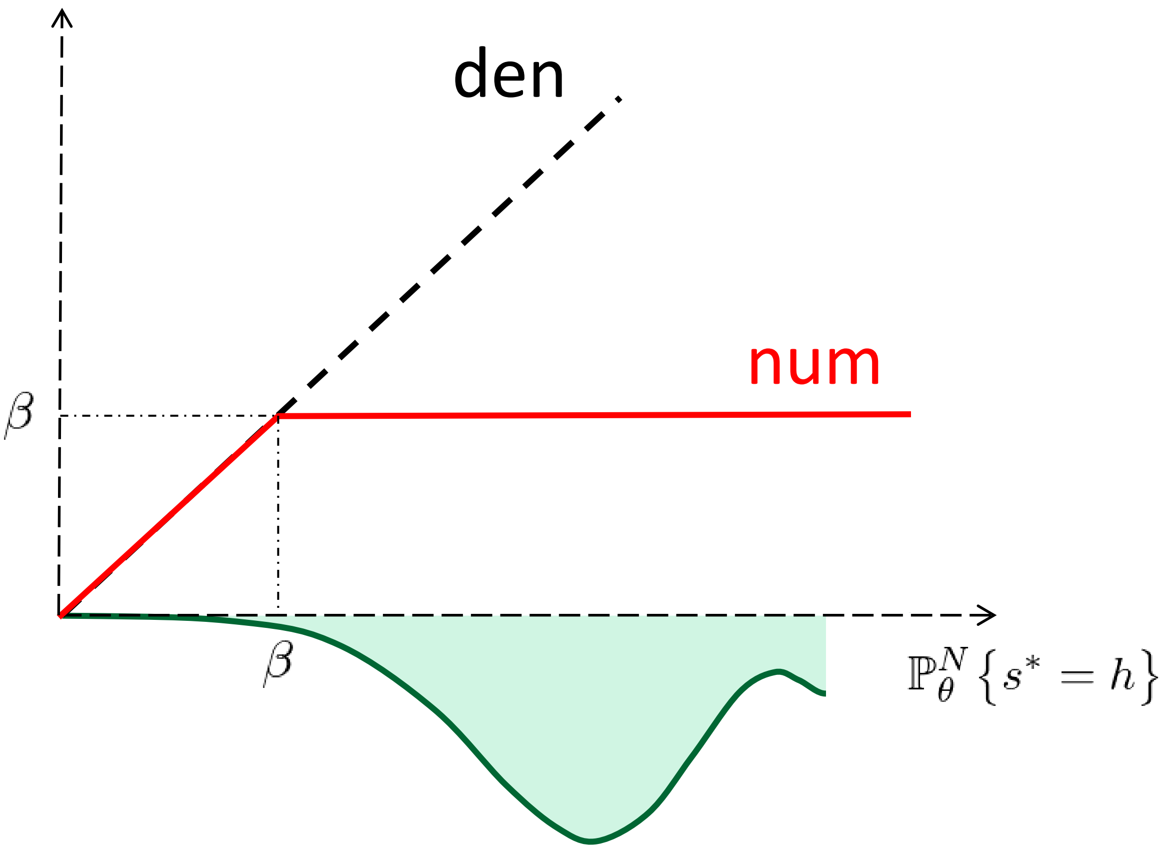
$$\leq \frac{\int_{\mathbb{P}_\theta^N\{s^*=h\}<\beta} \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta) + \beta \cdot \int_{\mathbb{P}_\theta^N\{s^*=h\}\geq\beta} \pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$



Main result (buyer perspective)

$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\}$$

$$\leq \frac{\int_{\mathbb{P}_\theta^N\{s^*=h\}<\beta} \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta) + \beta \cdot \int_{\mathbb{P}_\theta^N\{s^*=h\}\geq\beta} \pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$

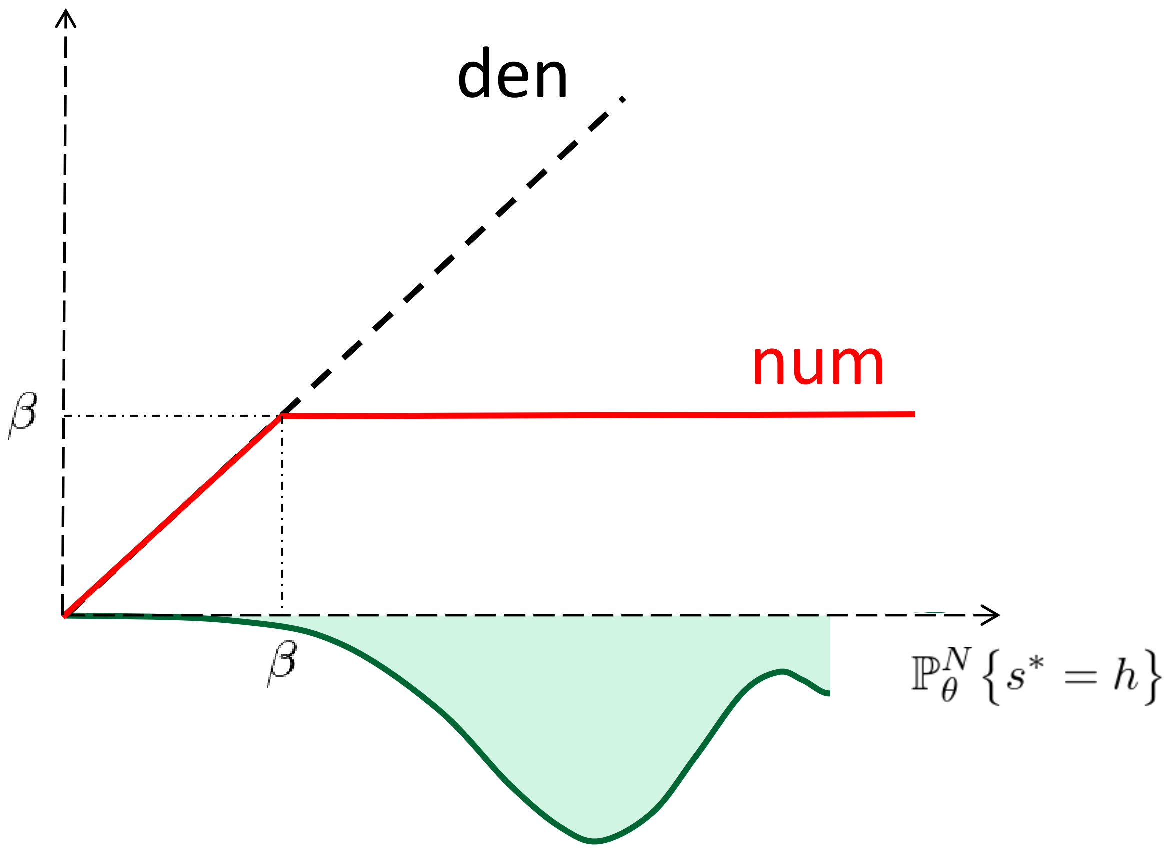


$\approx \beta$ if the distribution of $\mathbb{P}_\theta^N\{s^* = h\}$ is not too concentrated below and nearby β (insensitivity to prior)

Main result (buyer perspective)

$$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\}$$

$$\leq \frac{\int_{\mathbb{P}_\theta^N\{s^*=h\}<\beta} \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta) + \beta \cdot \int_{\mathbb{P}_\theta^N\{s^*=h\}\geq\beta} \pi(\mathrm{d}\theta)}{\int_\Theta \mathbb{P}_\theta^N\{s^* = h\}\pi(\mathrm{d}\theta)}$$



π not needed !

it is enough to specify

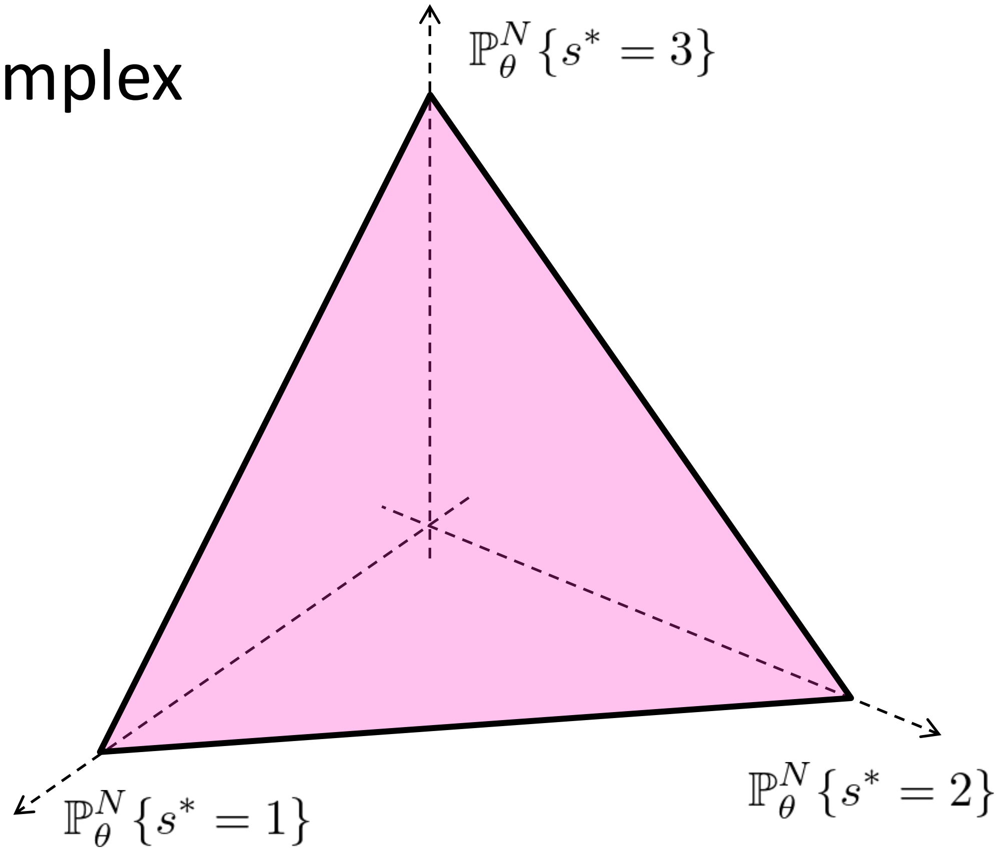
the distribution of

$\mathbb{P}_\theta^N\{s^* = h\}$

(mild prior)

Main result (buyer perspective) – cont'd

$\mathbb{P}_\theta^N\{s^* = h\}$ defined over the simplex



- Uniform prior (principle of indifference)
- Dirichlet priors

$\mathbf{P}\{V_\theta(x^*) > \epsilon(s^*) | s^* = h\}$ can be easily and explicitly evaluated

conditional distribution ($1 - \mathbf{P}\{V_\theta(x^*) > \epsilon | s^* = h\}$)

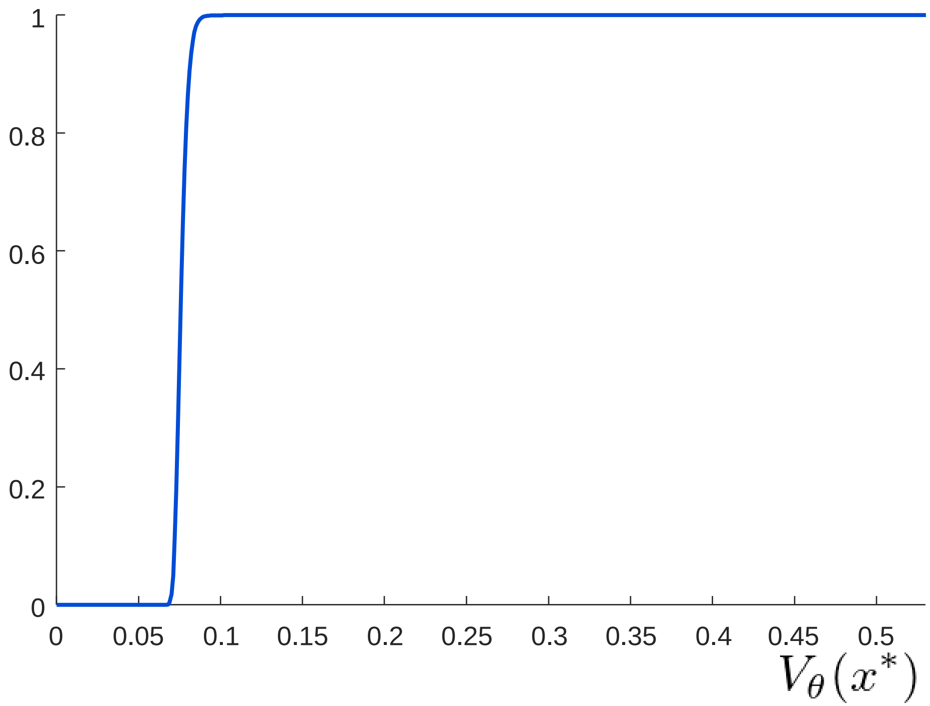
Real data example

Corel Image Features – UCI ML repository

⇨ 68040 images, 89 features each

orthant construction in $\mathbb{R}^{89}$ for N=1000 images ⇨ $s^* = 53$

conditional cdf given $s^* = 53$



uniform prior

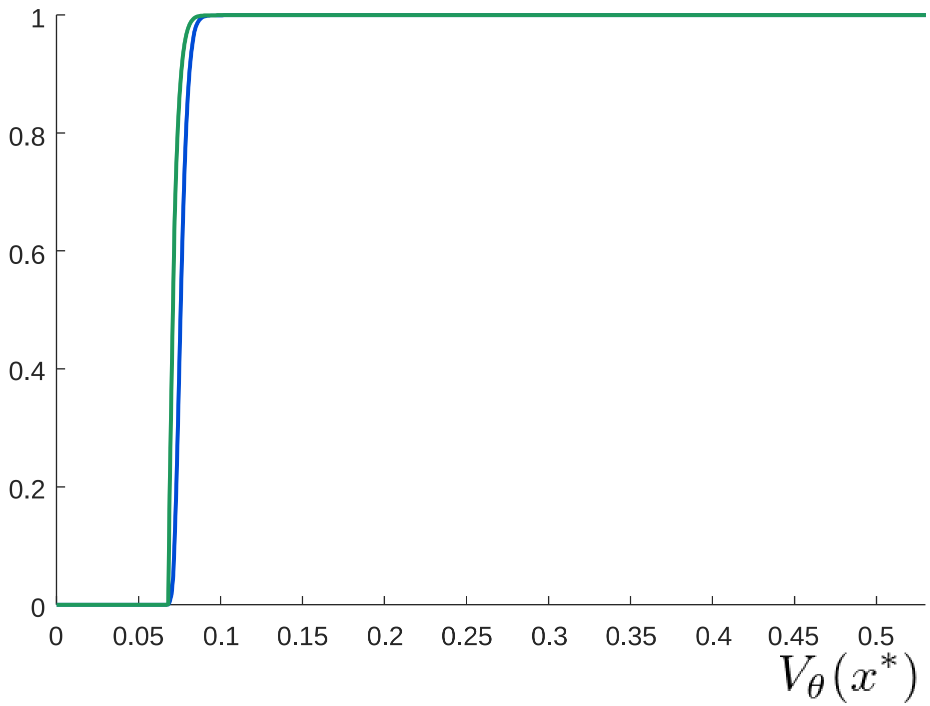
Real data example

Corel Image Features – UCI ML repository

→ 68040 images, 89 features each

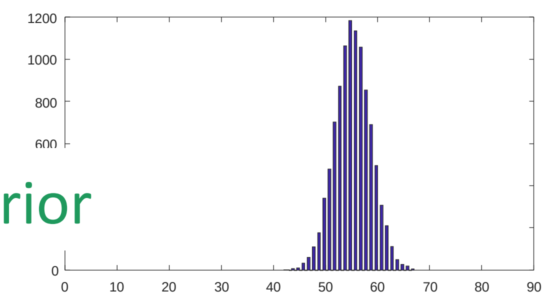
orthant construction in $\mathbb{R}^{89}$ for N=1000 images → $s^* = 53$

conditional cdf given $s^* = 53$



uniform prior

concentrated prior



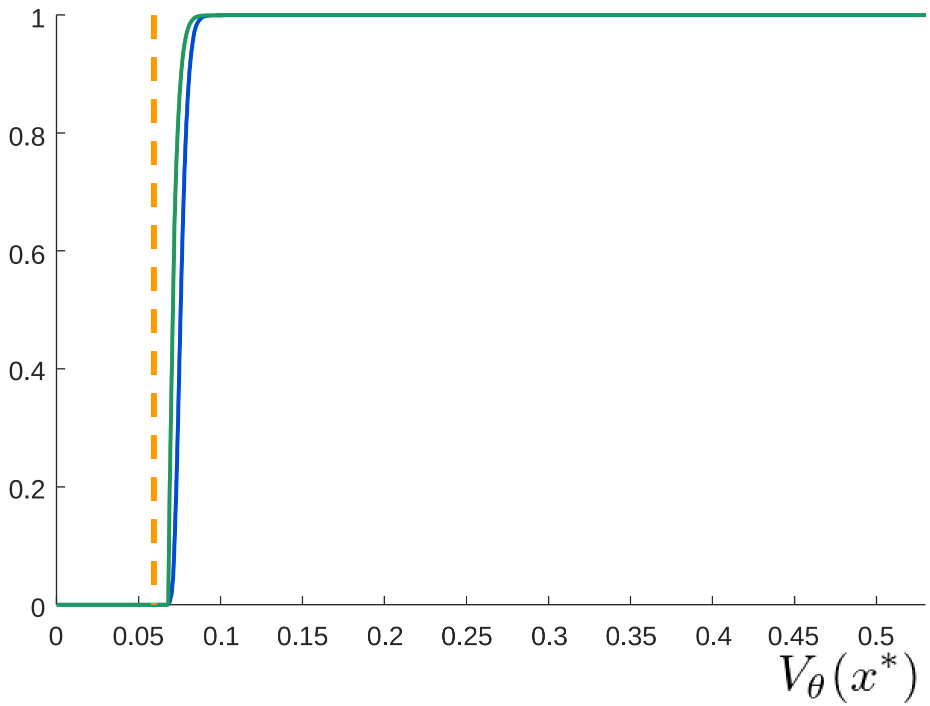
Real data example

Corel Image Features – UCI ML repository

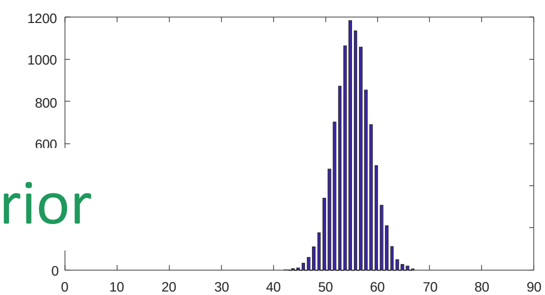
⇨ 68040 images, 89 features each

orthant construction in $\mathbb{R}^{89}$ for N=1000 images ⇨ $s^* = 53$

conditional cdf given $s^* = 53$



uniform prior



concentrated prior

“actual” violation

Conclusions

- Scenario Approach: a paradigm for effective, certified data-driven optimization

- Assessment of $V(x^*)$ (hidden) through s^* (observable)

- Non-conditional assessment $\Rightarrow$ distribution-free

- Conditional assessment $\Rightarrow$ requires some prior

Yet, strong assessment with mild prior!

Thank you !

S. Garatti and M.C. Campi, "On Conditional Risk Assessments in Scenario Optimization". SIAM Journal on Optimization, 33(2):455-480, 2023. https://doi.org/10.1137/21M1451385

The research of Simone Garatti is supported by FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence)

Why not out-of-sample testing?

- using some data for testing rather than designing… waste of information, questionable!
- scenarios (data) are often limited resources (collecting data can be time-consuming or burdensome, involving a monetary cost)
- in the present context validation is not necessary!

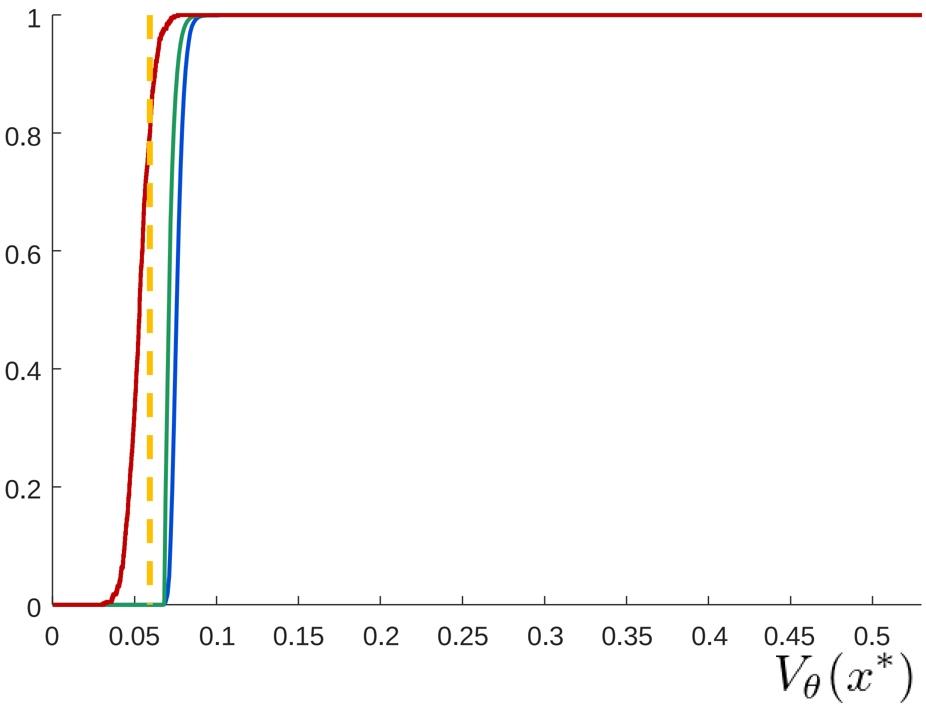
Real data example

Corel Image Features – UCI ML repository

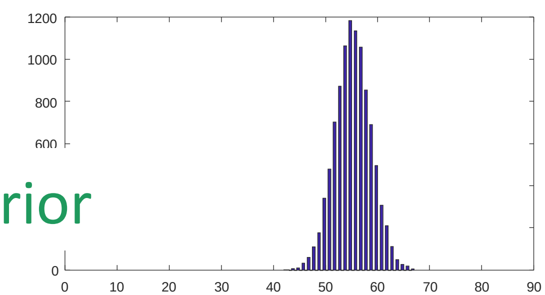
➡ 68040 images, 89 features each

orthant construction in $\mathbb{R}^{89}$ for N=1000 images ➡ $s^* = 53$

conditional cdf given $s^* = 53$



uniform prior



concentrated prior

“actual” violation

empirical frequency