The scenario approach as a general tool for risk control in data-driven optimization

speaker: **Simone Garatti**

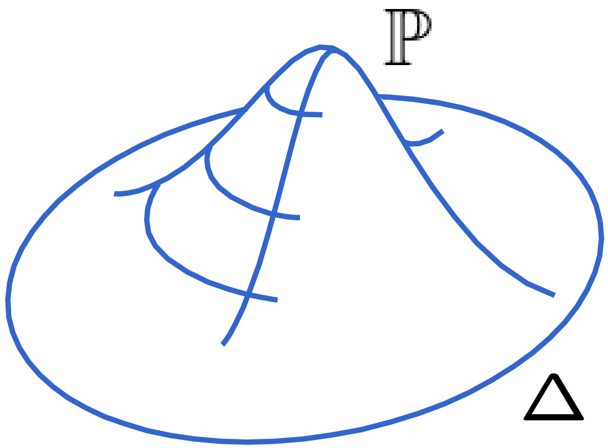
(Politecnico di Milano, Italy – email: simone.garatti@polimi.it)

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

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

Data-driven decision-making

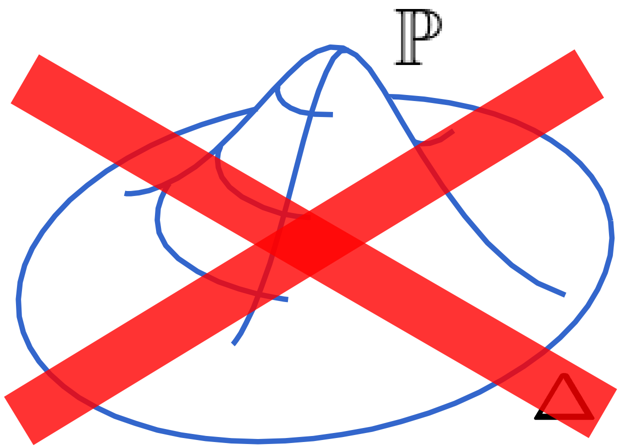
δ = uncertain element ⇒ exercise caution



decision x^*

Data-driven decision-making

δ = uncertain element ⇨ exercise caution

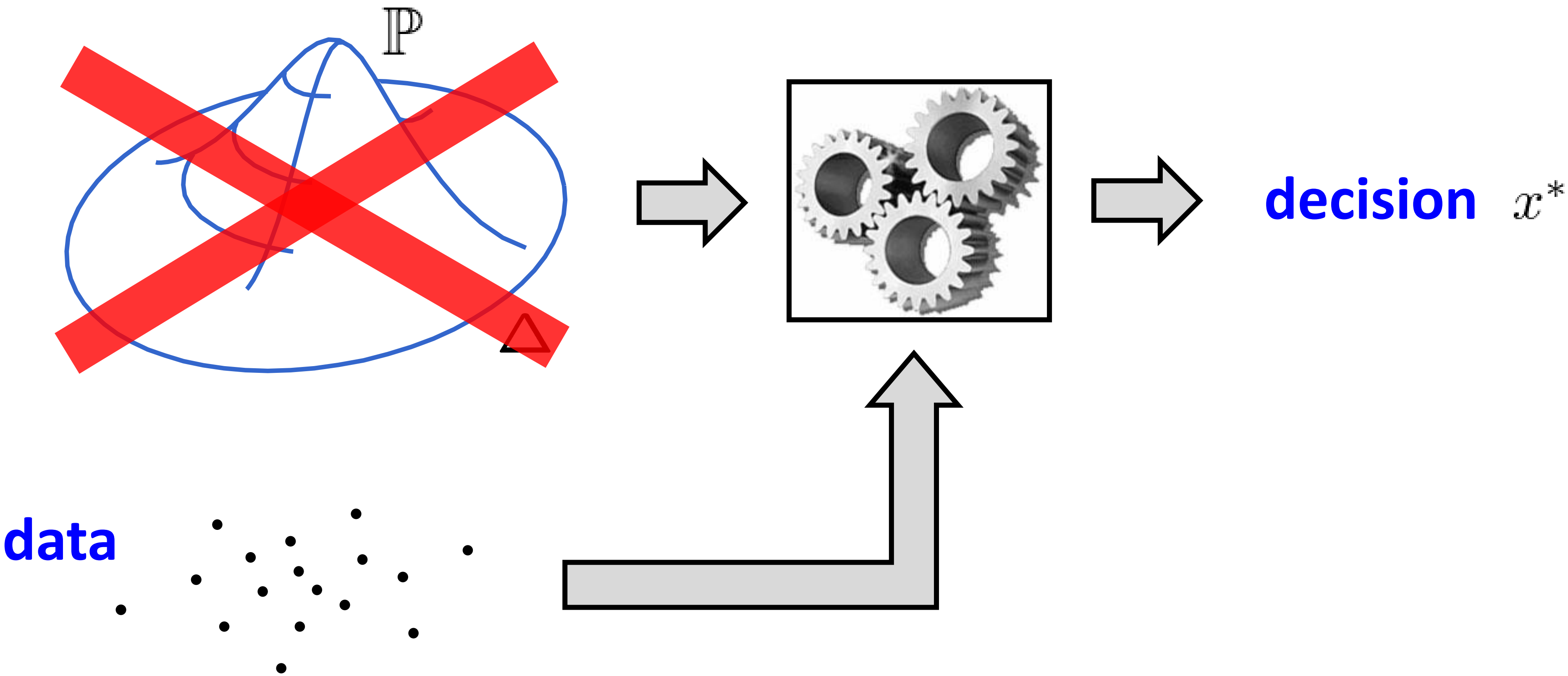


⇨

⇨ **decision** x^*

Data-driven decision-making

δ = uncertain element $\Rightarrow$ exercise caution



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

Desiderata

Effective ⇨ no over-conservatism

Dependable ⇨ certified decisions

Agnostic ⇨ no information other than data

Desiderata

Effective → no over-conservatism

Dependable → certified decisions

Agnostic → no information other than data

The way of the scenario approach: enforce design goals heuristically, possibly in various attempts (tunable schemes); provide the user with a solid theory to assess the quality of the solution(s) to decide when goals are met

Ingredients of the decision problem

Decision vector: $x \in \mathcal{X}$

Convex cost function: $c(x)$

Family of convex constraint sets: $\mathcal{X}_\delta$

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

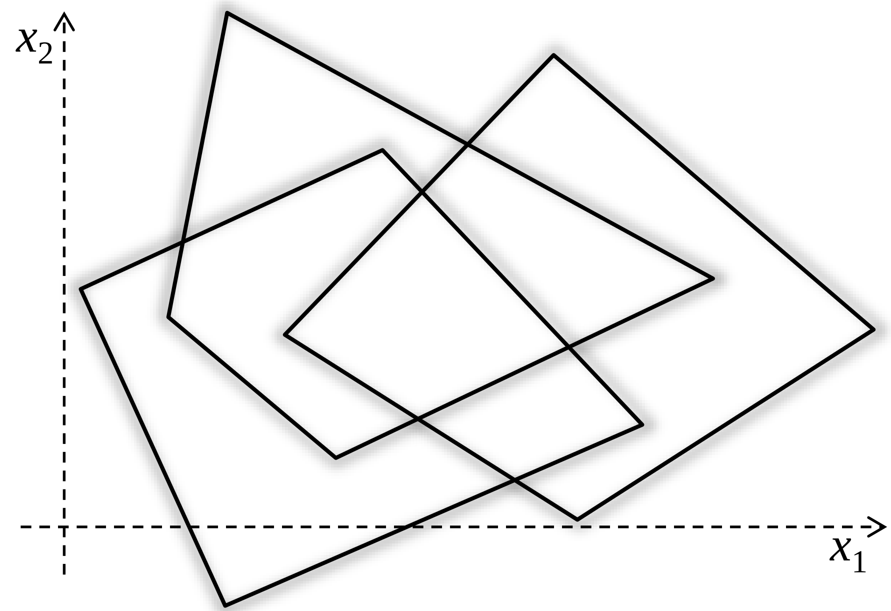
Worst-case scenario optimization

$$\delta^{(1)} \quad \rightarrow \quad \mathcal{X}_{\delta^{(1)}}$$

$$\delta^{(2)} \quad \rightarrow \quad \mathcal{X}_{\delta^{(2)}}$$

$$\vdots$$

$$\delta^{(N)} \quad \rightarrow \quad \mathcal{X}_{\delta^{(N)}}$$

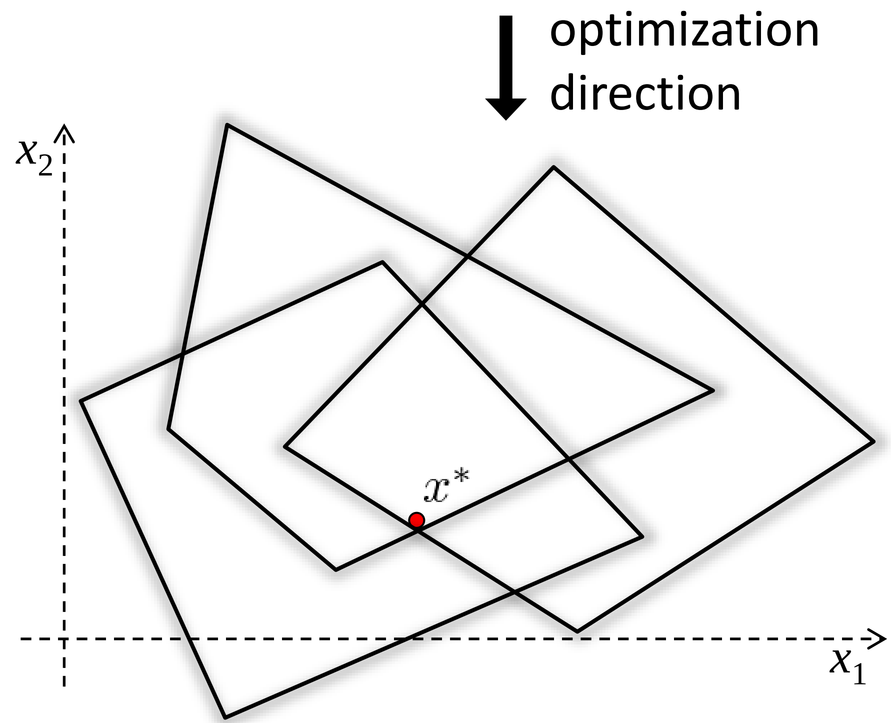


Worst-case scenario optimization

$$
\begin{array}{ccc}
\delta^{(1)} & \to & \mathcal{X}_{\delta^{(1)}} \\
\delta^{(2)} & \to & \mathcal{X}_{\delta^{(2)}} \\
 & \vdots & \\
\delta^{(N)} & \to & \mathcal{X}_{\delta^{(N)}}
\end{array}
$$

$$\Downarrow$$

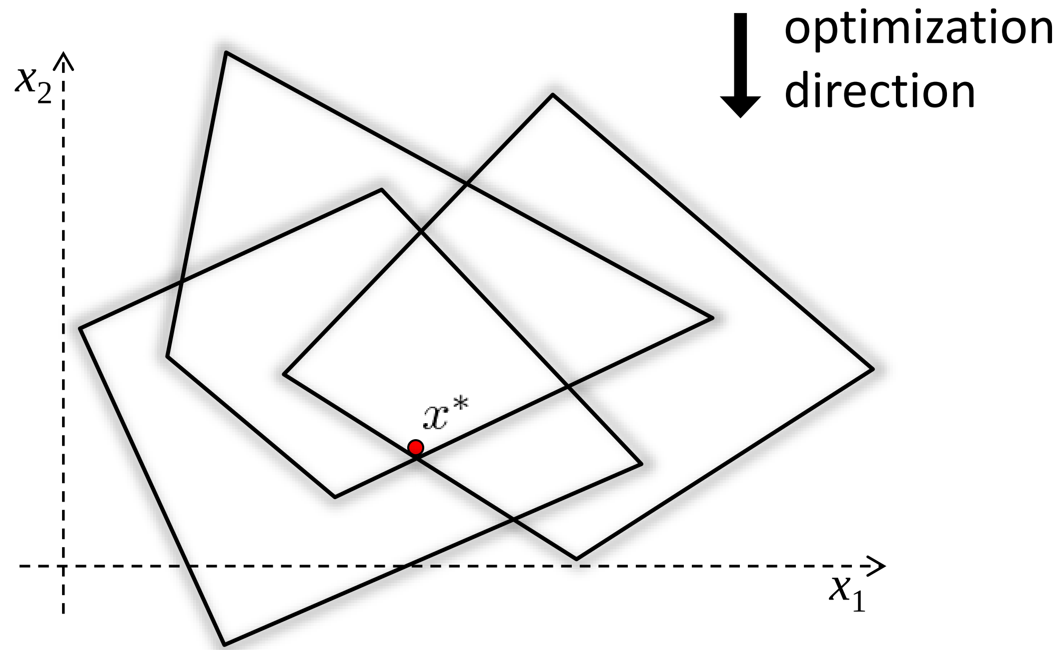
$$
\begin{aligned}
\min_{x \in \mathbb{R}^d} \quad & c(x) \\
\text{s.t.} \quad & x \in \mathcal{X}_{\delta^{(i)}}, \\
& i = 1, \dots, N
\end{aligned}
$$



solution = x^*

Scenario optimization with constraints relaxation

$$\begin{aligned} \min_{x \in \mathbb{R}^d} \quad & c(x) \\ \text{s.t.} \quad & x \in \mathcal{X}_{\delta^{(i)}}, \\ & i = 1, \ldots, N \end{aligned}$$

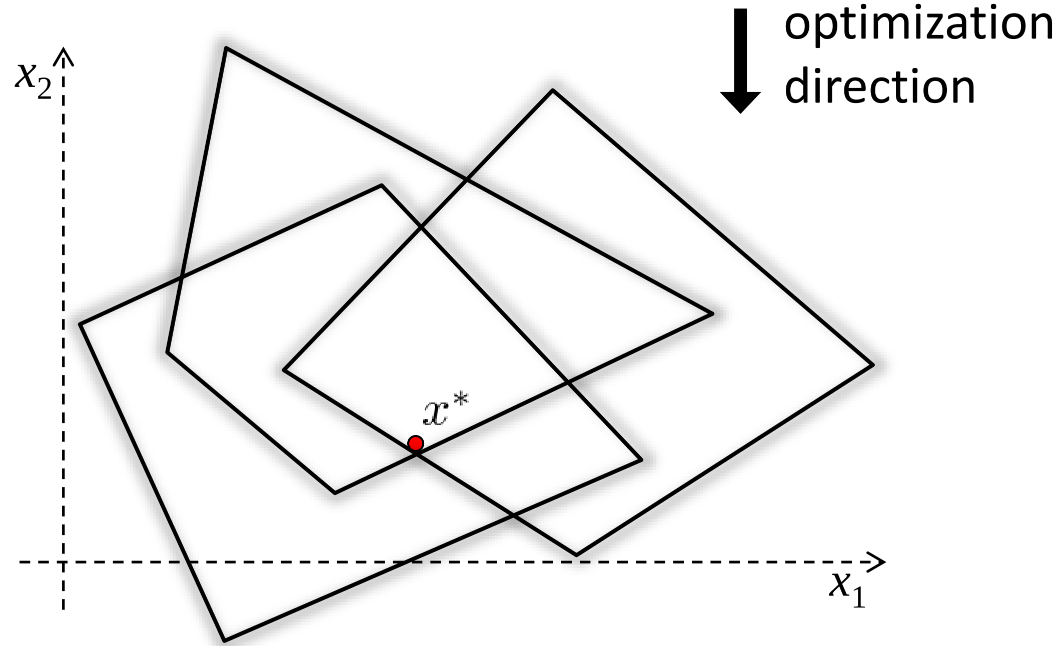


Scenario optimization with constraints relaxation

$$\min_{x \in \mathbb{R}^d} \quad c(x)$$

$$\text{s.t.} \quad f(x, \delta^{(i)}) \leq 0,$$

$$i = 1, \ldots, N$$

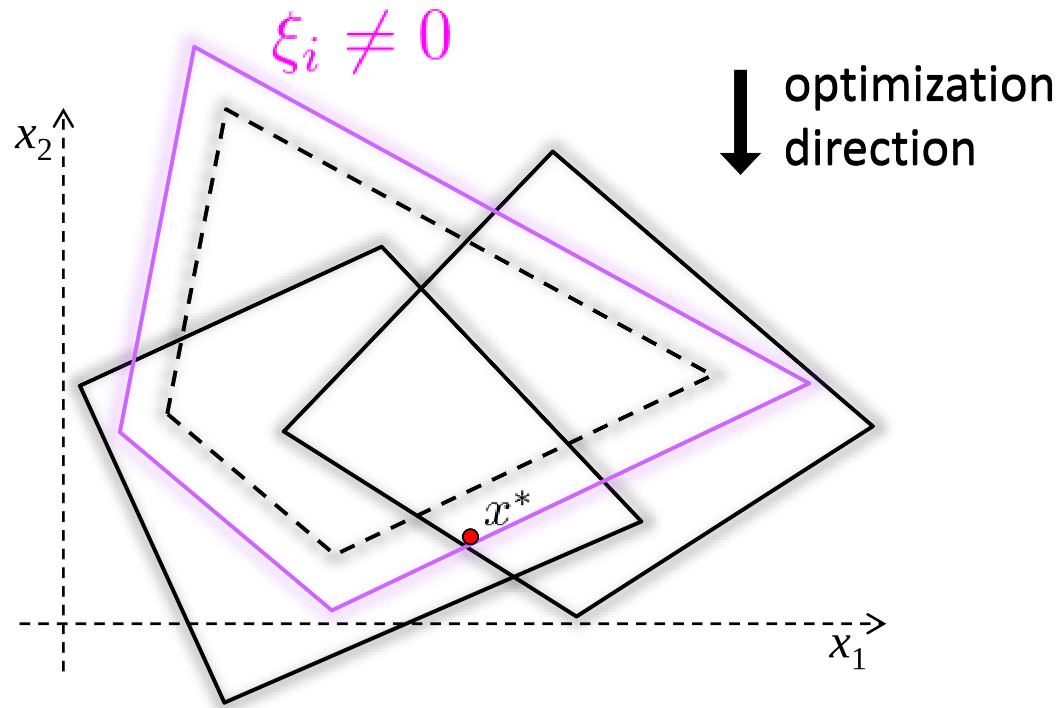


Scenario optimization with constraints relaxation

$$\min_{x\in\mathbb{R}^d,\xi_i\geq 0} \quad c(x)+\rho\sum_{i=1}^{N}\xi_i$$

$$\text{s.t.} \quad f(x,\delta^{(i)}) \leq \xi_i,$$

$$i = 1,\ldots,N$$



solution: x^*, $\{\xi_i^* : \xi_i^* \neq 0\}$

ρ = tunable tradeoff parameter

A general scenario decision-making framework

Decision map $M: \delta^{(1)}, \ldots, \delta^{(N)} \rightarrow (x^*, w^*)$ such that when new scenarios $\delta^{(N+1)}, \ldots, \delta^{(N+H)}$ are added:

- if $\underbrace{x^* \in \mathcal{X}_{\delta^{(N+i)}}}_{x^* \text{ already feasible}}$ for all *i*, then solution does <u>not</u> change

- if $\underbrace{x^* \notin \mathcal{X}_{\delta^{(N+i)}}}_{x^* \text{ unfeasible for some cases}}$ for some *i*, then solution must change

A general scenario decision-making framework

Decision map $M : \delta^{(1)}, \ldots, \delta^{(N)} \to (x^*, w^*)$ such that when new scenarios $\delta^{(N+1)}, \ldots, \delta^{(N+H)}$ are added:

- if $\underbrace{x^* \in \mathcal{X}_{\delta^{(N+i)}}}_{x^* \text{ already feasible}}$ for all *i*, then solution does <u>not</u> change

consistency

- if $\underbrace{x^* \notin \mathcal{X}_{\delta^{(N+i)}}}_{x^* \text{ unfeasible for some cases}}$ for some *i*, then solution must change

⇨ worst-case optimization, opt. with constraint relaxation, expected shortfall optimization, variational inequalities, …

Scenario approach: main features

- easy (algorithmically speaking) and widely applicable

- data used to directly target the objective

  ⇨ effective solutions!

Scenario approach: main features

- easy (algorithmically speaking) and widely applicable

- data used to directly target the objective

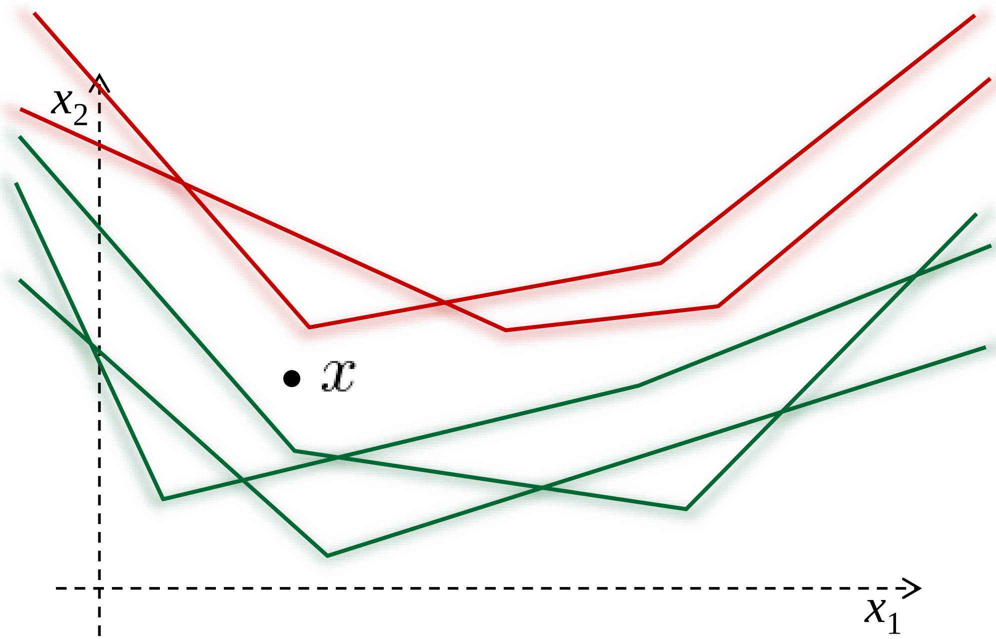
  ⇨ effective solutions!

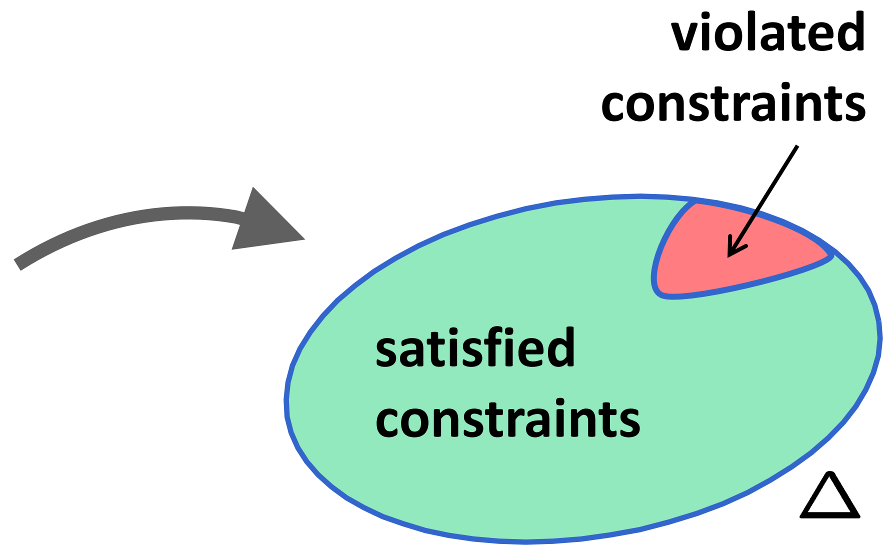
- feasibility addressed empirically

  ⇨ dependability of the scenario approach rests on our capability to keep control of the actual feasibility (risk)

Risk

$V(x) = \mathbb{P}\{\delta \in \Delta : \ x \notin \mathcal{X}_\delta\}$ out-of-sample constraint violation





$V(x)$ = “size” of red region

Solution certification

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

cost risk

$\mathbb{P}$ = mechanism by which δ is generated

scenario decision certification

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

Solution certification

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

cost — risk

$\mathbb{P}$ = mechanism by which δ is generated ??????????

scenario decision certification

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

Solution certification

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

cost risk

??????????

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

main issue: to evaluate $V(x^*)$

scenario solution certification

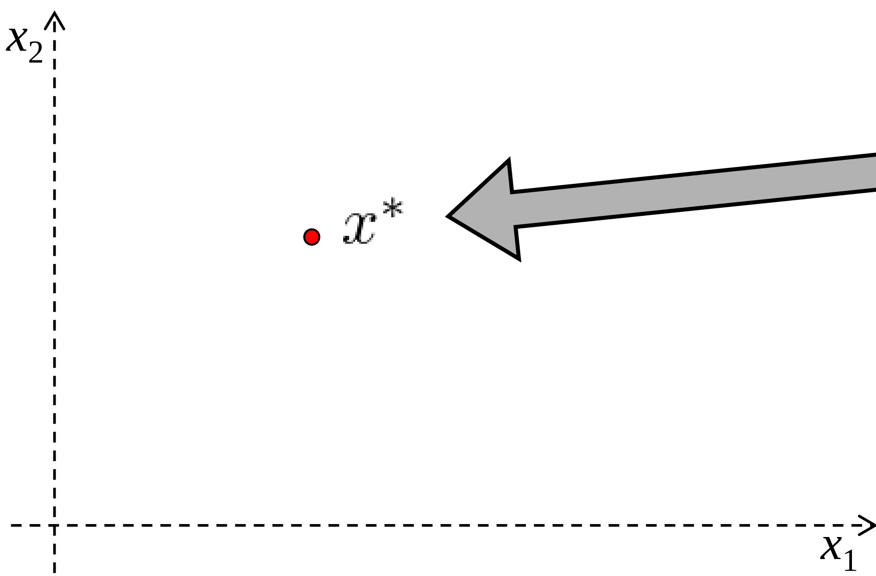
- $c(x^*)$ accessible (once x^* is computed)
- $V(x^*)$ **not accessible**

Why not validation

- 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!

Risk of the scenario decision

Problem: assess $V(x^*)$

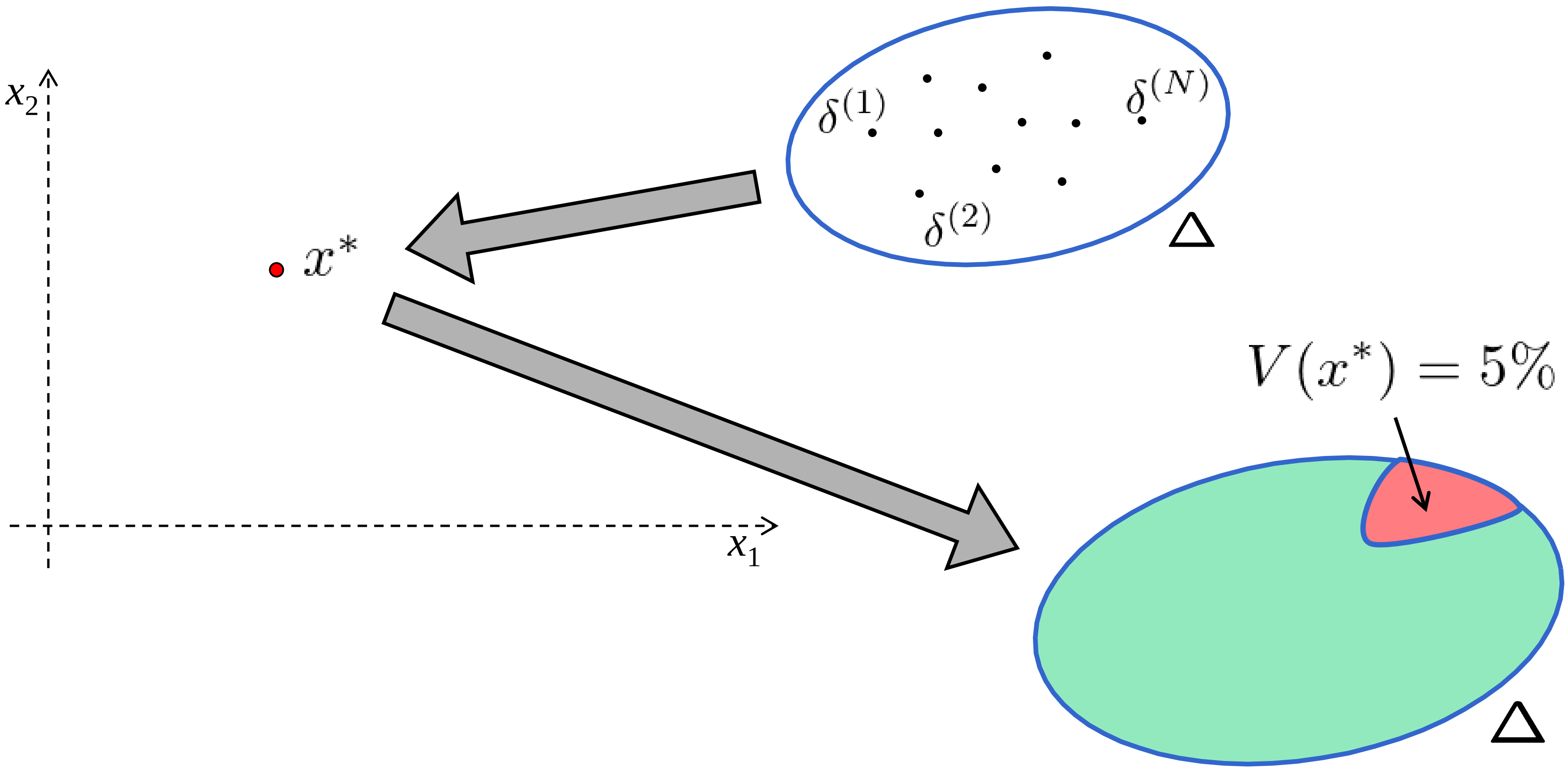


$$\min_{x \in \mathbb{R}^d, \xi_i \geq 0} \quad c(x) + \rho \sum_{i=1}^{N} \xi_i$$

$$\text{s.t.} \quad f(x, \delta^{(i)}) \leq \xi_i, \quad i = 1, \ldots, N$$

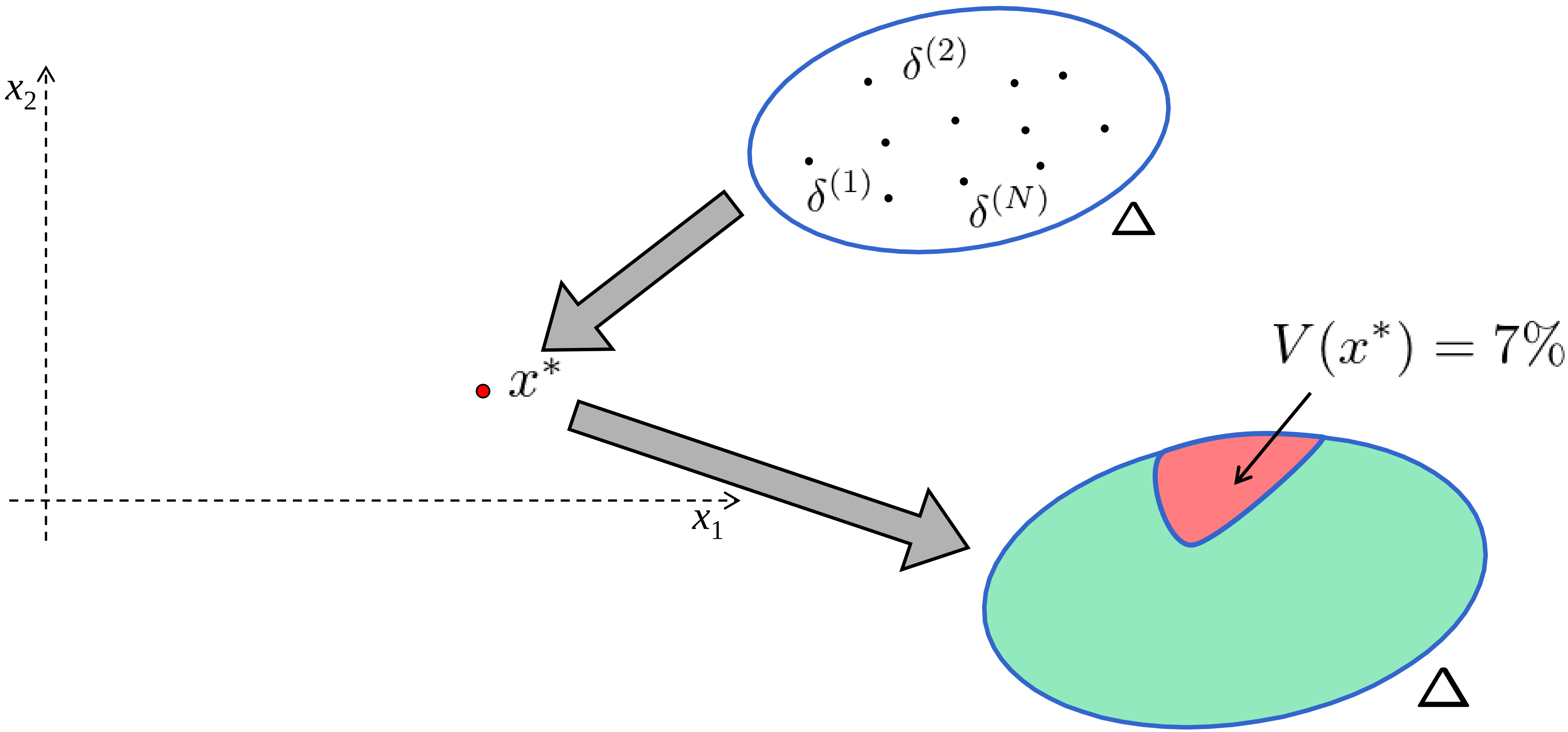
Risk of the scenario decision

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



Risk of the scenario decision

Problem: assess $V(x^*) = V(x^*(\delta^{(1)}, \delta^{(2)}, \dots, \delta^{(N)})$



Distribution of the risk

$V(x^*)$ is a random variable

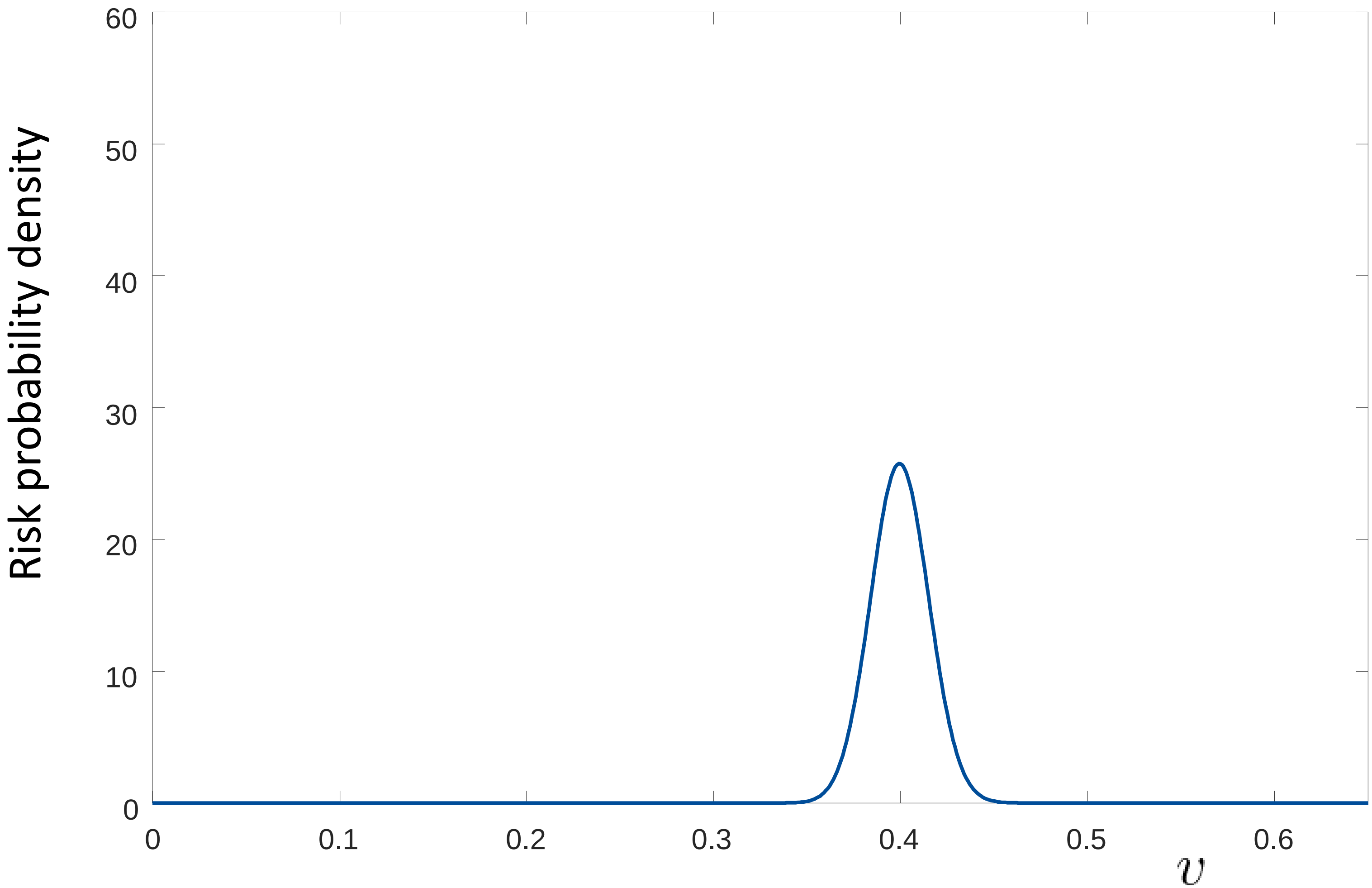
What about its probability distribution?

How does it change with $\mathbb{P}$, the mechanism generating δ ?

Is it concentrated?

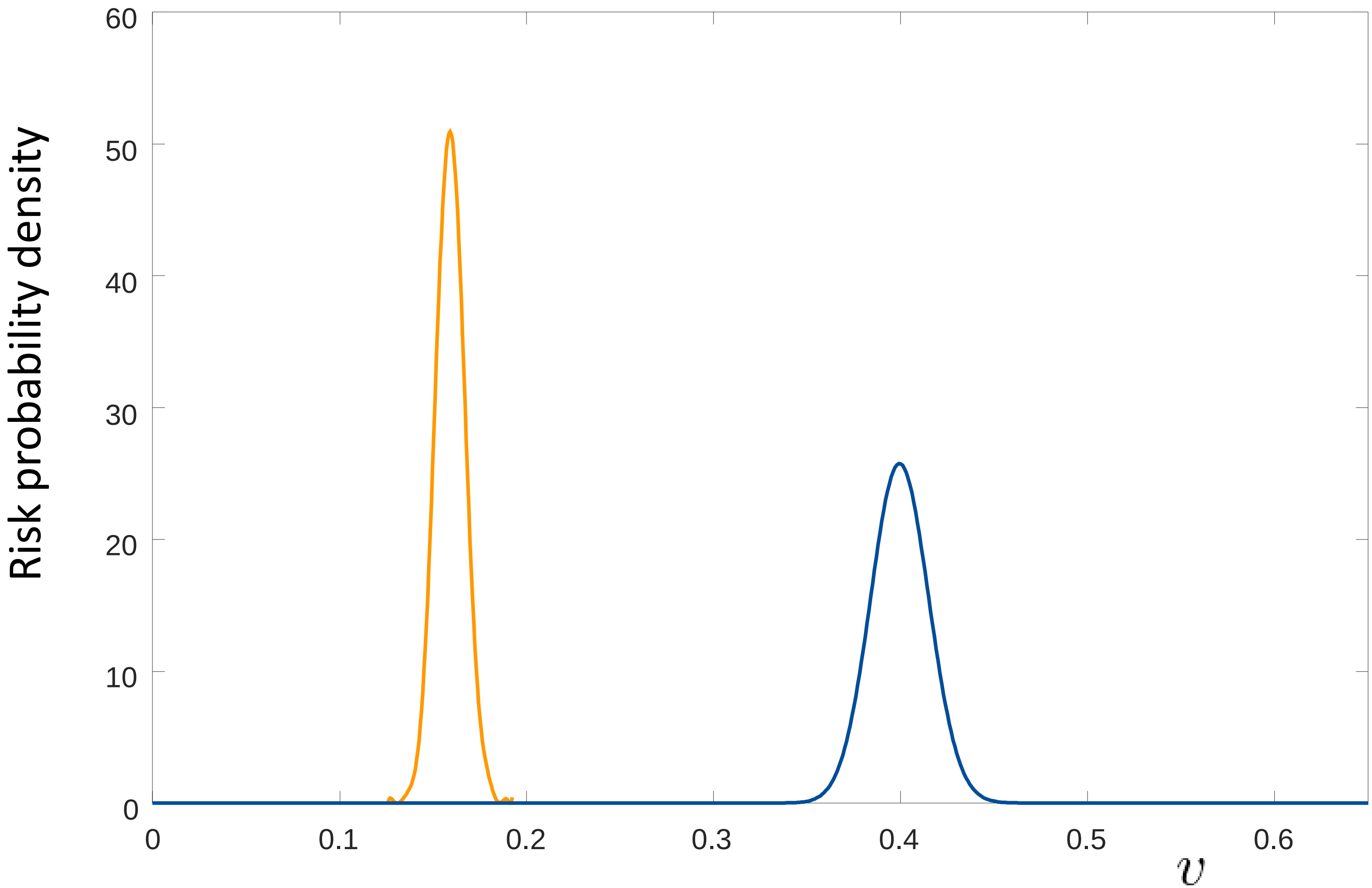
Distribution of the risk: examples

Same decision problem with $N = 1000$ for various $\mathbb{P}$



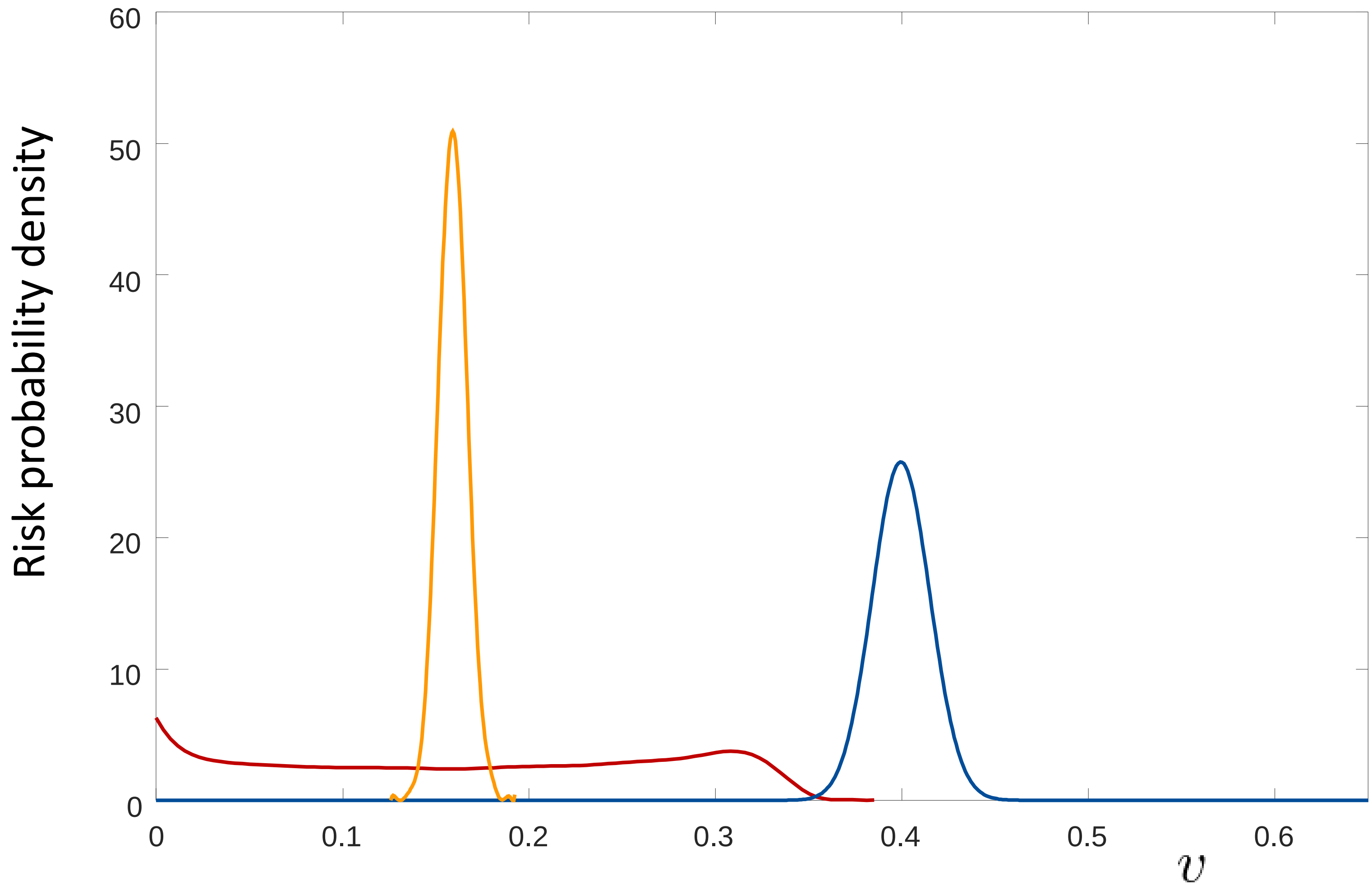
Distribution of the risk: examples

Same decision problem with N = 1000 for various $\mathbb{P}$



Distribution of the risk: examples

Same decision problem with $N = 1000$ for various $\mathbb{P}$



Support set and complexity

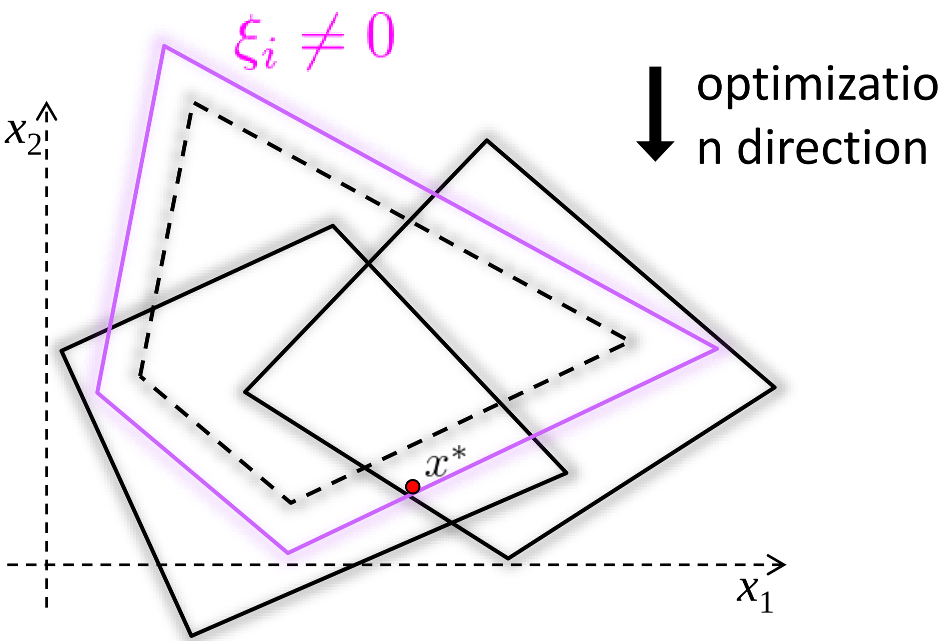
Support set: $\left\{\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right\}$ such that

1. $\mathrm{sol}\left(\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right) = \mathrm{sol}\left(\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}\right)$
2. no $\delta^{(i_j)}$ can be further removed without changing the solution

Support set and complexity

Support set: $\left\{\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right\}$ such that

1. $\mathrm{sol}\left(\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right) = \mathrm{sol}\left(\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}\right)$
2. no $\delta^{(i_j)}$ can be further removed without changing the solution

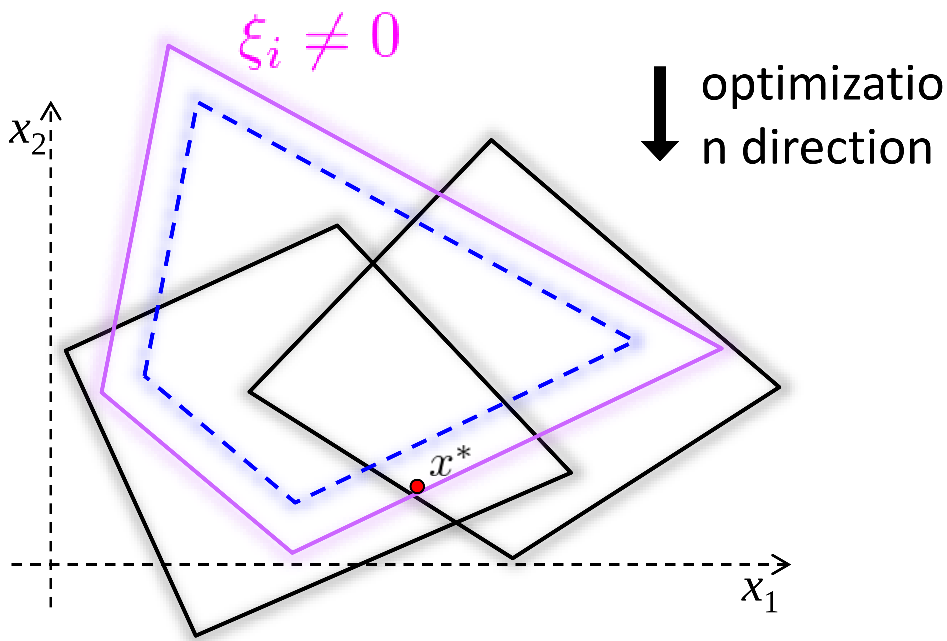


solution: $x^*, \{\xi_i^* : \xi_i^* \neq 0\}$

Support set and complexity

Support set: $\left\{\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right\}$ such that

1. $\mathrm{sol}\left(\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right) = \mathrm{sol}\left(\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}\right)$
2. no $\delta^{(i_j)}$ can be further removed without changing the solution



support set =

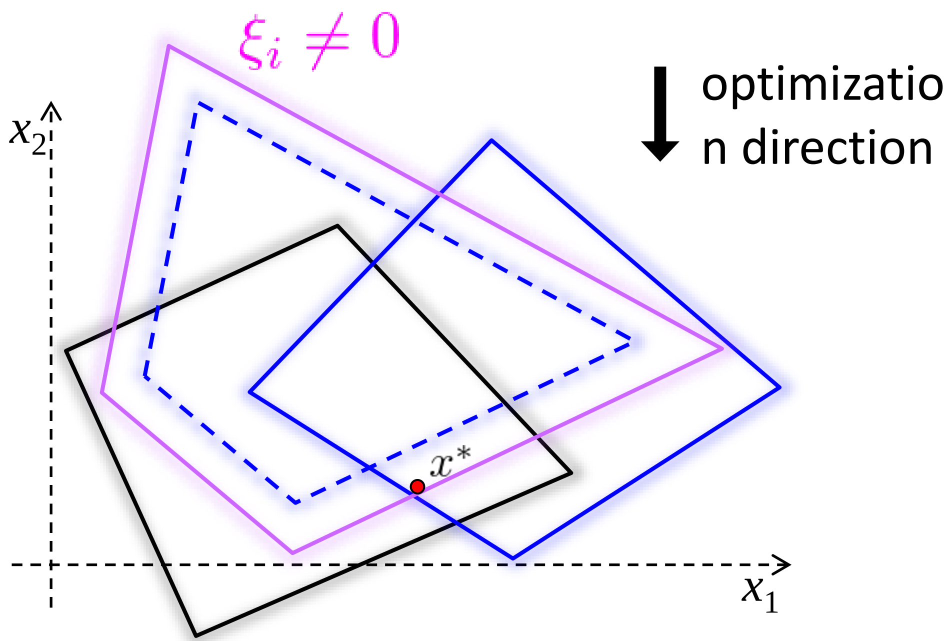
violated

solution: x^*, $\{\xi_i^* : \xi_i^* \neq 0\}$

Support set and complexity

Support set: $\left\{\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right\}$ such that

1. $\mathrm{sol}\left(\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right) = \mathrm{sol}\left(\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}\right)$
2. no $\delta^{(i_j)}$ can be further removed without changing the solution



support set =

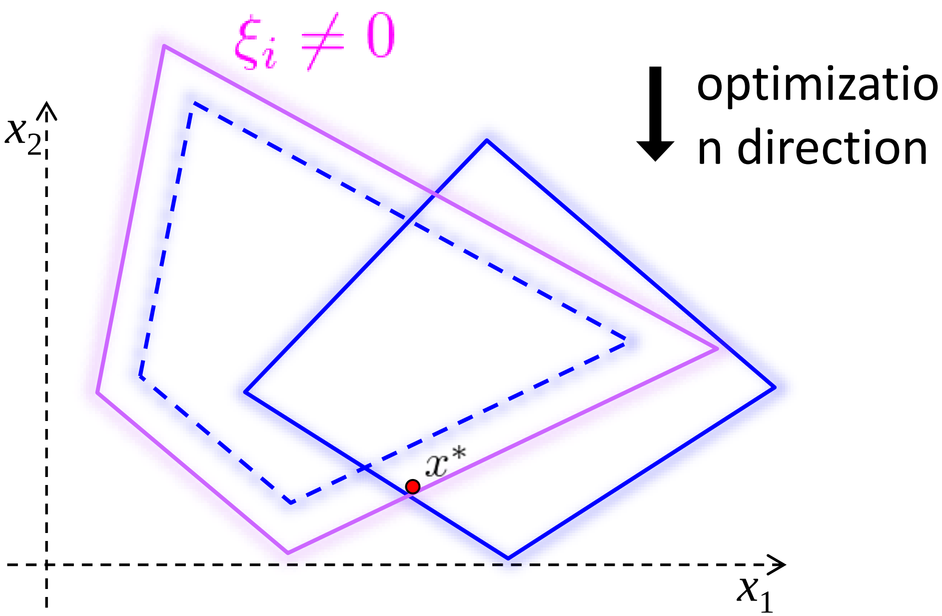
violated + active

solution: x^*, $\{\xi_i^* : \xi_i^* \neq 0\}$

Support set and complexity

Support set: $\left\{\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right\}$ such that

1. $\mathrm{sol}\left(\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right) = \mathrm{sol}\left(\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}\right)$
2. no $\delta^{(i_j)}$ can be further removed without changing the solution



support set =

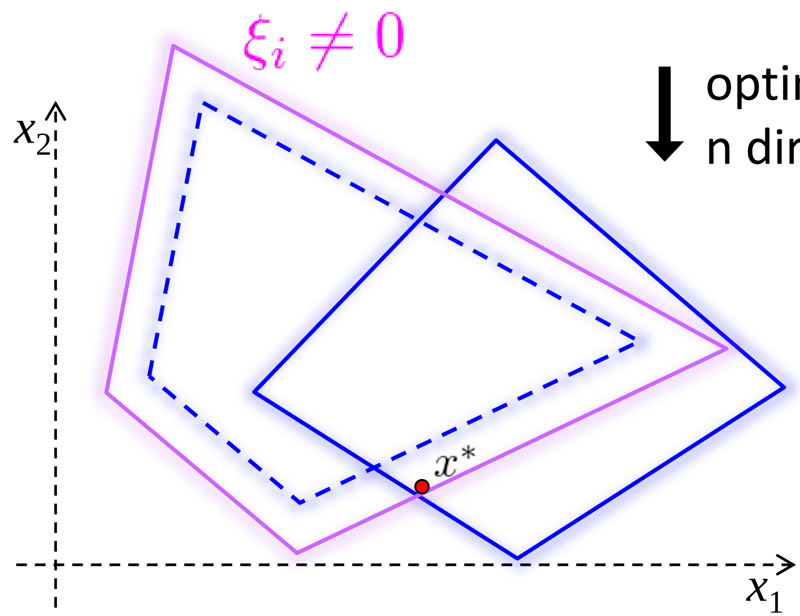
violated + active

solution: x^*, $\{\xi_i^* \,:\, \xi_i^* \neq 0\}$

Support set and complexity

Support set: $\left\{\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right\}$ such that

1. $\mathrm{sol}\left(\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right) = \mathrm{sol}\left(\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}\right)$
2. no $\delta^{(i_j)}$ can be further removed without changing the solution



Assumption: the support set is unique with probability 1

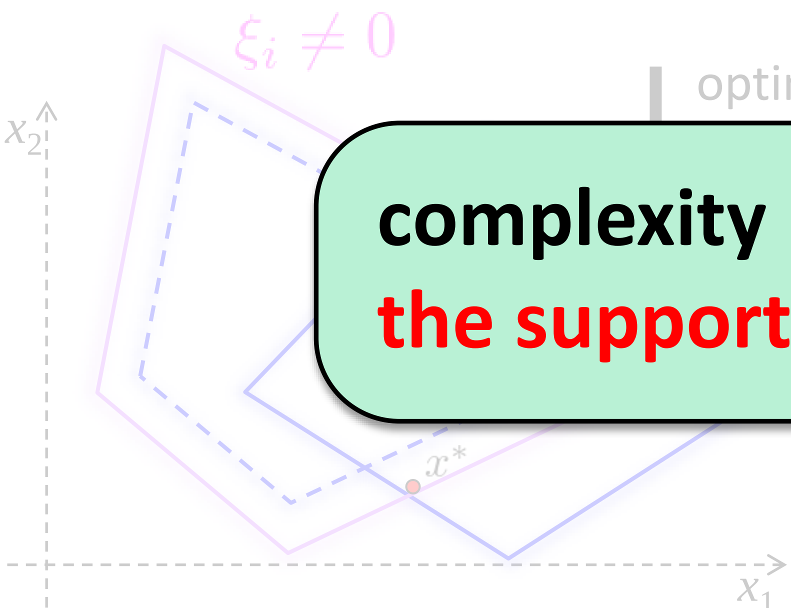
(non-accumulation of constraints in a convex setup)

solution: x^*, $\{\xi_i^* \,:\, \xi_i^* \neq 0\}$

Support set and complexity

Support set: $\left\{\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right\}$ such that

1. $\mathrm{sol}\left(\delta^{(i_1)}, \delta^{(i_2)}, \ldots, \delta^{(i_k)}\right) = \mathrm{sol}\left(\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)}\right)$
2. no $\delta^{(i_j)}$ can be further removed without changing the solution



optimizatio

complexity π^* = cardinality of the support set

pport set is
ility 1
(non-accumulation of constraints in a convex setup)

solution: $x^*, \{\xi_i^* : \xi_i^* \neq 0\}$

A new perspective

π^* is a random variable (integer, $\pi^* = k,\ k \in \{0, 1, \ldots, N\}$)

A new perspective

π^* is a random variable (integer, $\pi^* = k,\ k \in \{0, 1, \ldots, N\}$)

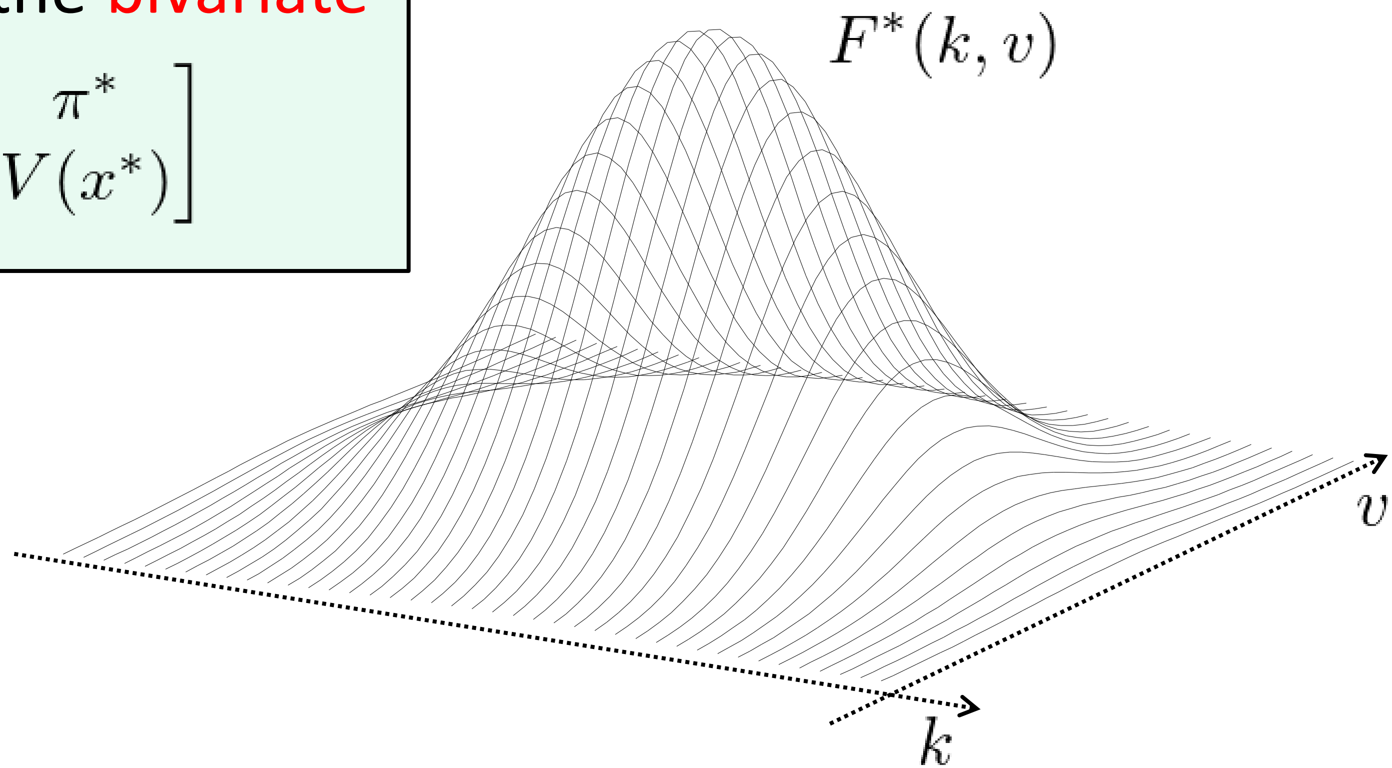
$V(x^*)$ is a random variable (real, $V(x^*) = v, \quad v \in [0, 1]$)

A new bivariate perspective

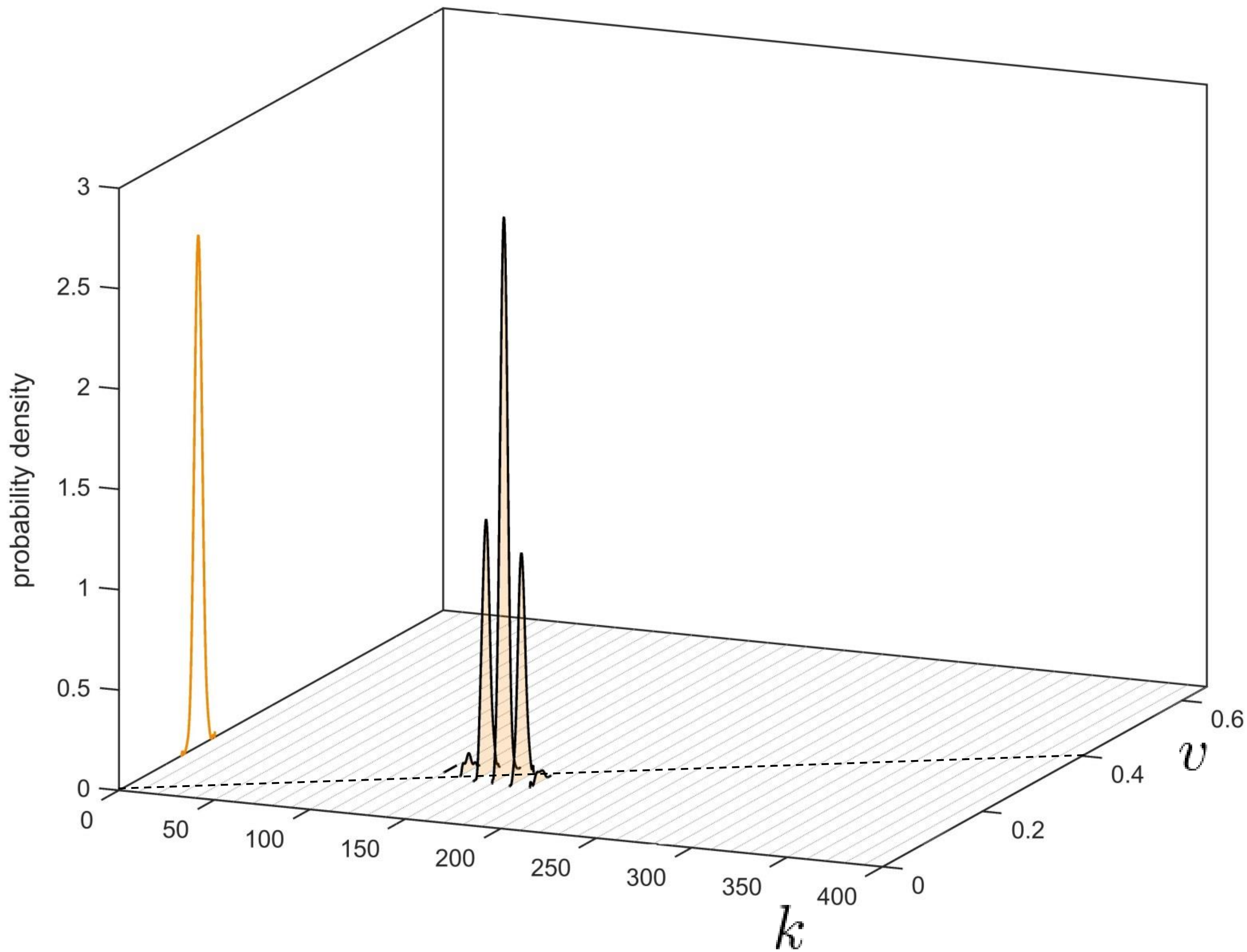
π^* is a random variable (integer, $\pi^* = k, \; k \in \{0, 1, \ldots, N\}$)

$V(x^*)$ is a random variable (real, $V(x^*) = v, \quad v \in [0, 1]$)

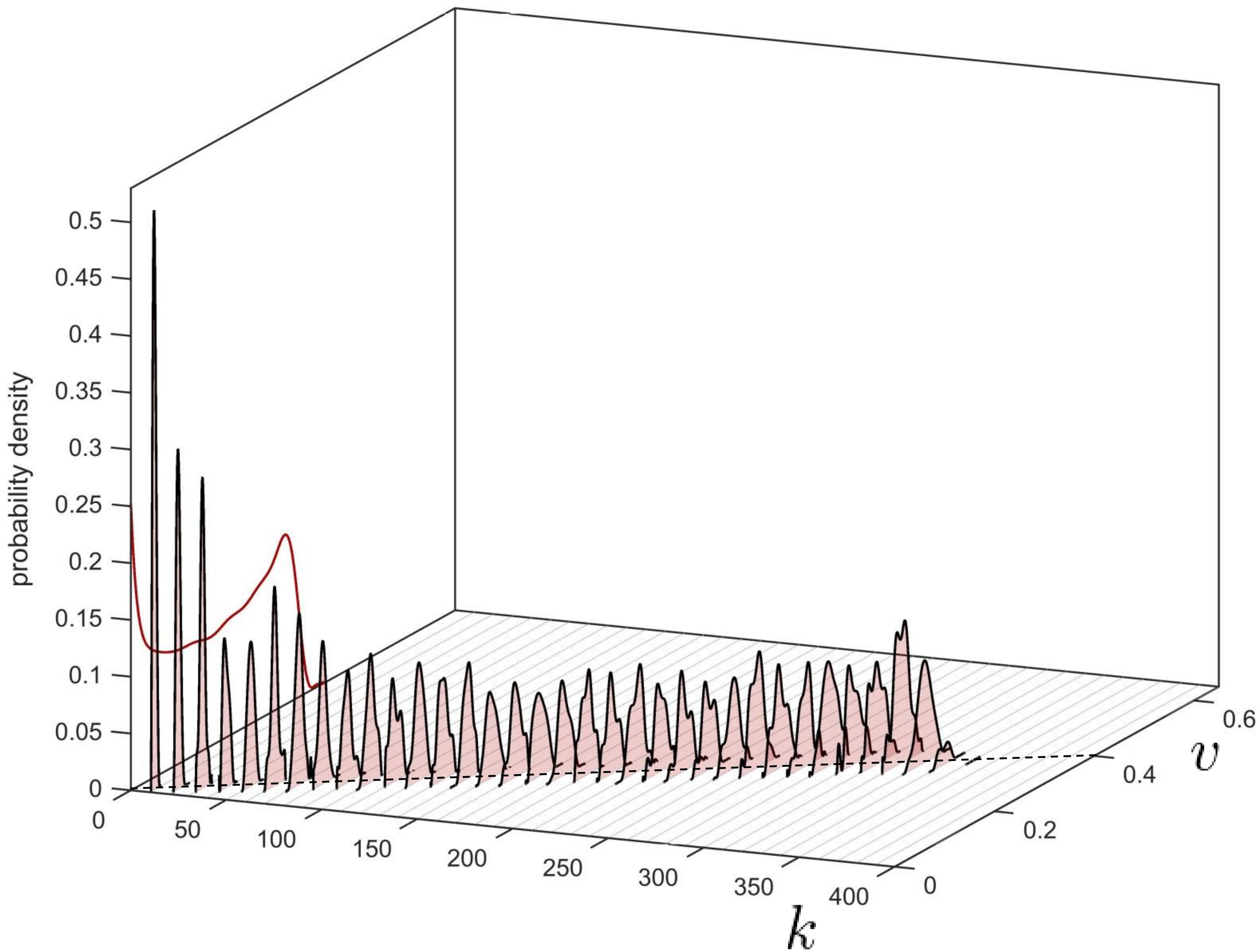
Study $F^*(k, v)$, the bivariate distribution of $\begin{bmatrix} \pi^* \\ V(x^*) \end{bmatrix}$



Bivariate risk-complexity distribution – Example 1



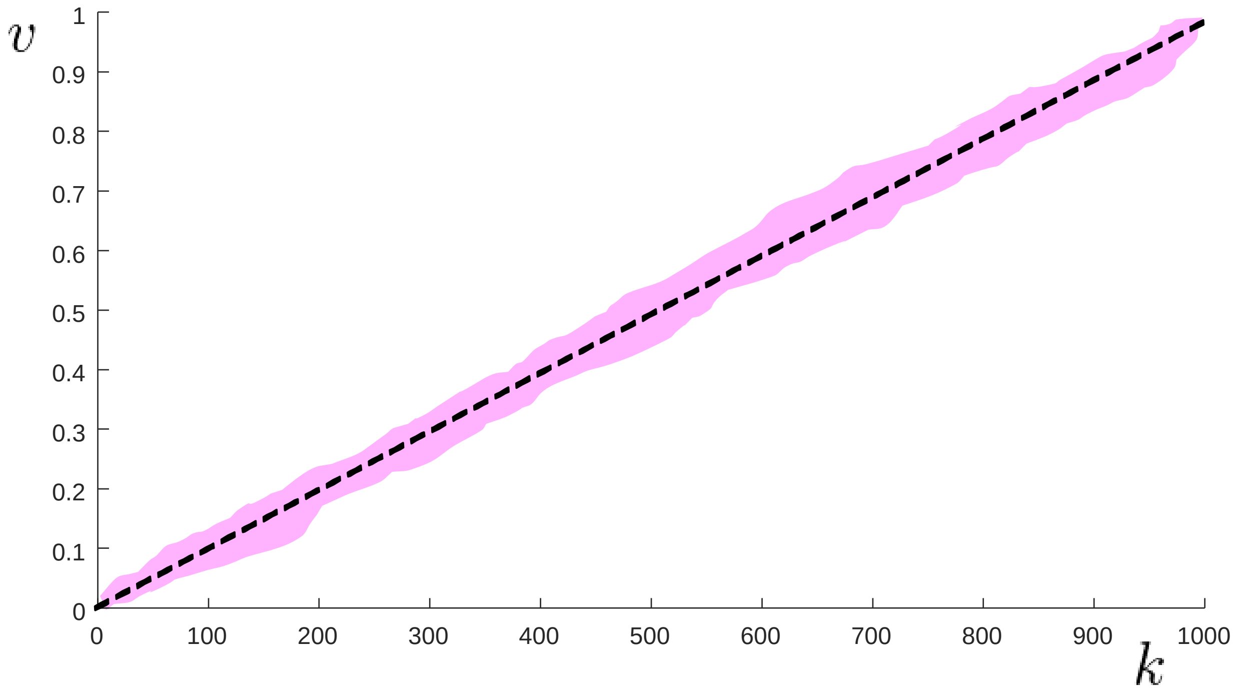
Bivariate risk-complexity distribution – Example 2



Main result (take-home message)

For all consistent decision schemes and distribution-free,

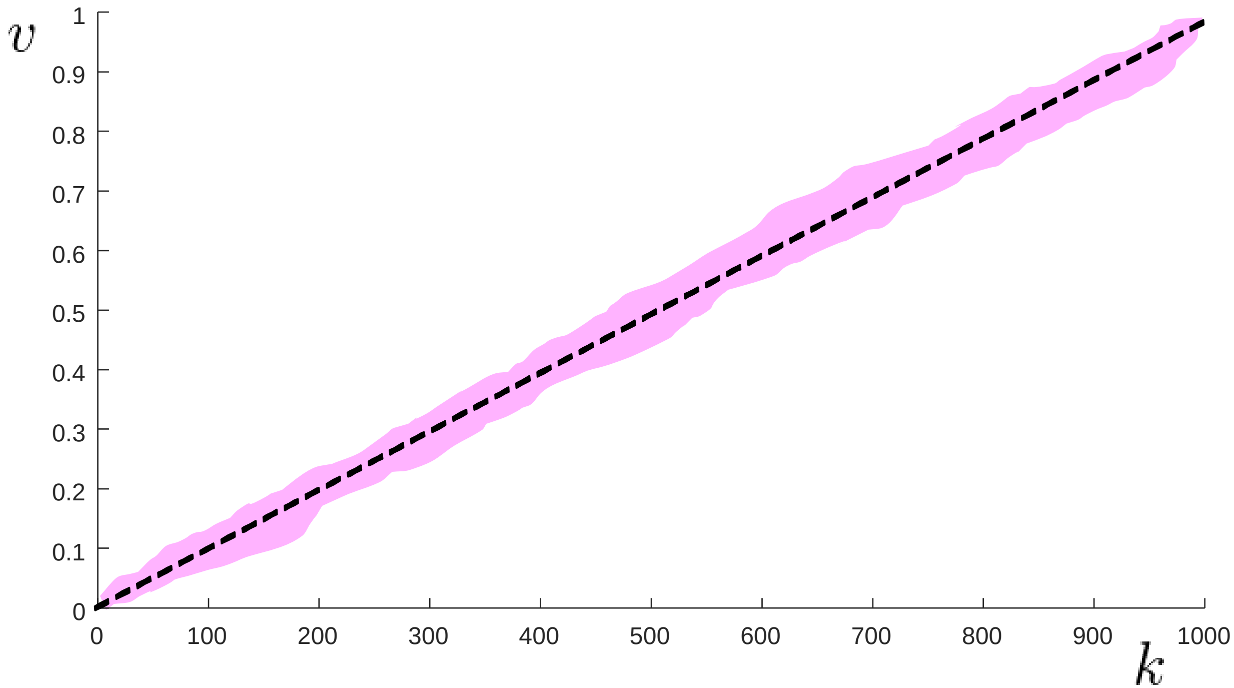
$F^*(k, v)$ concentrates around $v = \dfrac{k}{N+1}, \; k = 0, 1, \ldots, N$



Main result (take-home message)

For all consistent decision schemes and distribution-free,

$F^*(k,v)$ concentrates around $v = \dfrac{k}{N+1}, \; k = 0, 1, \ldots, N$

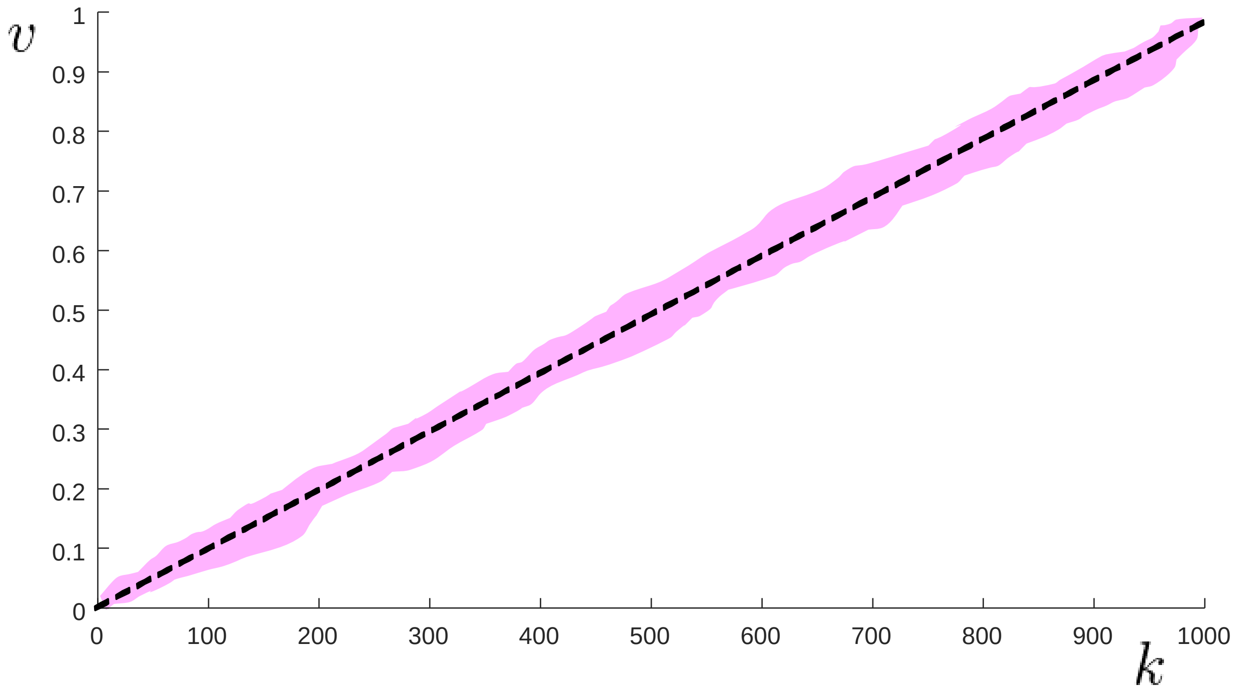


$V(x^*)$ **can be accurately estimated from** π^*

Main result (take-home message)

For all consistent decision schemes and distribution-free,

$F^*(k,v)$ concentrates around $v = \dfrac{k}{N+1}, \; k = 0, 1, \ldots, N$



observable!

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

Main result

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

Let $\epsilon_L(k), \epsilon^U(k)$ be the unique roots in (0,1) of polynomials

▷ $$\binom{N}{k}(1-\epsilon)^{N-k} - \frac{\beta}{2N}\sum_{m=k}^{N-1}\binom{m}{k}(1-\epsilon)^{m-k}$$

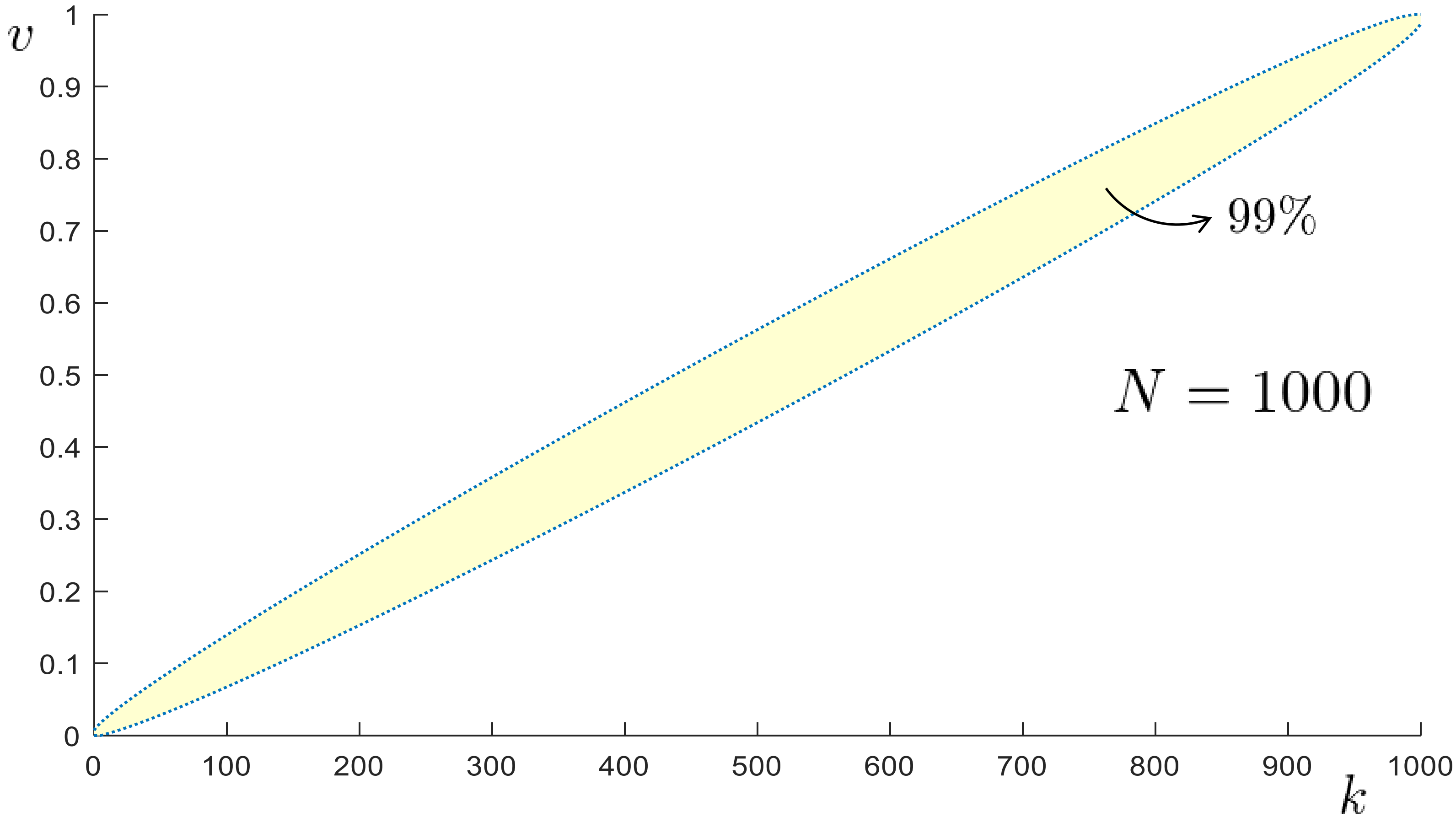
▷ $$\binom{N}{k}(1-\epsilon)^{N-k} - \frac{\beta}{2N}\sum_{m=N+1}^{2N}\binom{m}{k}(1-\epsilon)^{m-k}$$

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

$$\mathbb{P}^N\left\{\delta^{(1)},\ldots,\delta^{(N)}:\ \epsilon_L(\pi^*) \le V(x^*) \le \epsilon^U(\pi^*)\right\} \ge 1-\beta$$

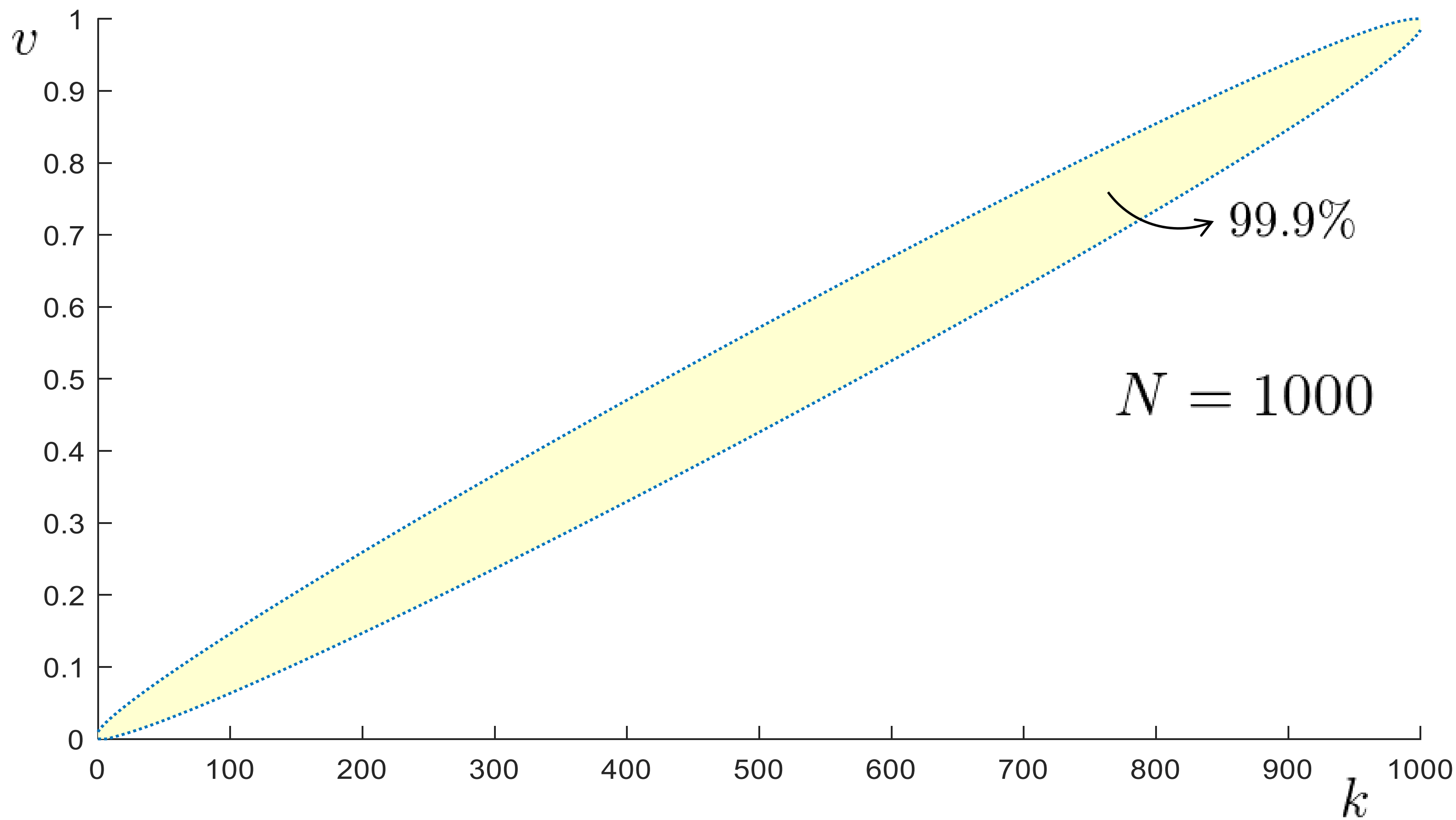
Main result

$$\mathbb{P}^N\left\{\delta^{(1)},\ldots,\delta^{(N)}:\ \epsilon_L(\pi^*)\leq V(x^*)\leq \epsilon^U(\pi^*)\right\}\geq 1-\beta$$



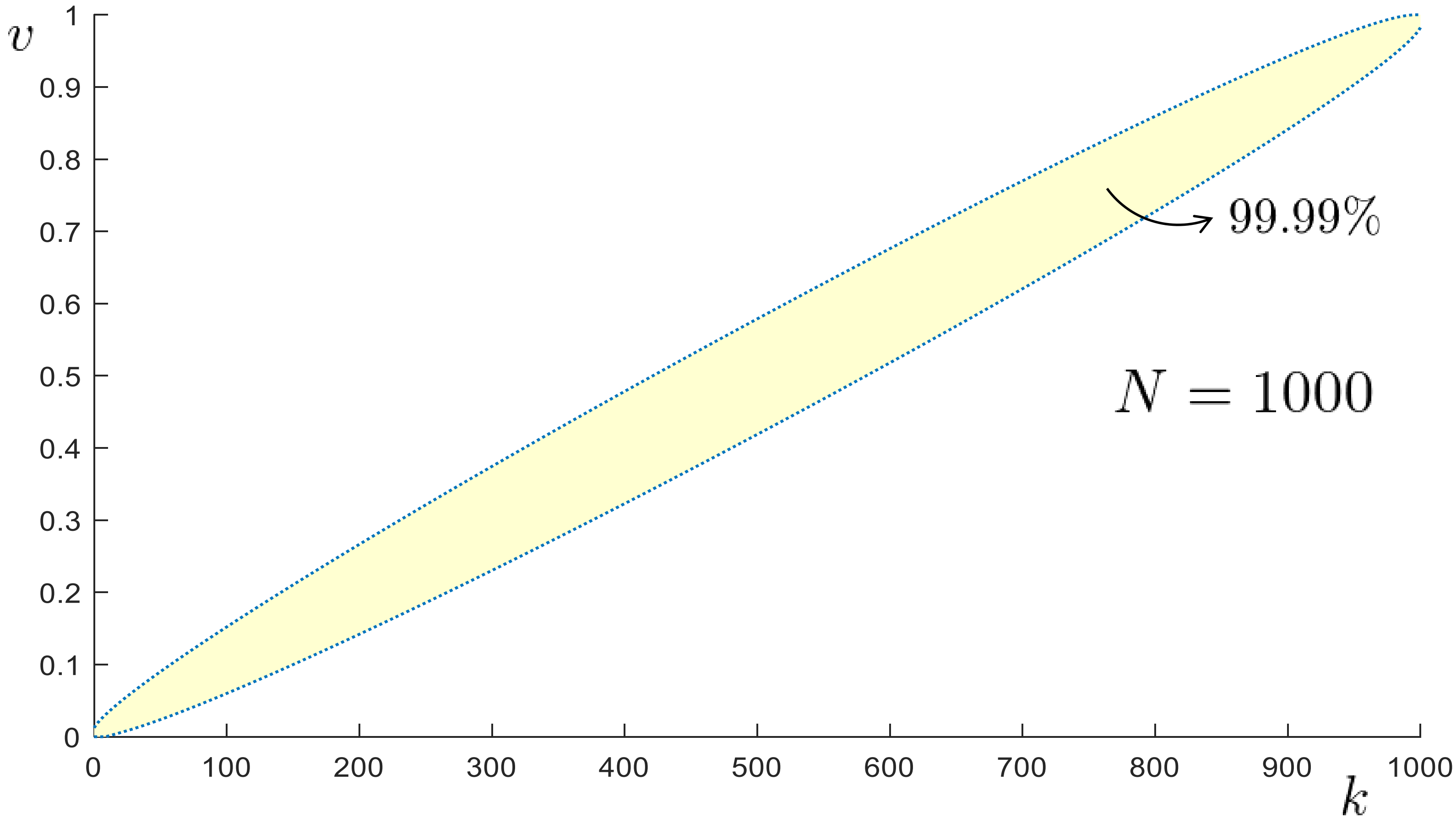
Main result

$$\mathbb{P}^N \left\{ \delta^{(1)}, \ldots, \delta^{(N)} : \ \epsilon_L(\pi^*) \leq V(x^*) \leq \epsilon^U(\pi^*) \right\} \geq 1 - \beta$$



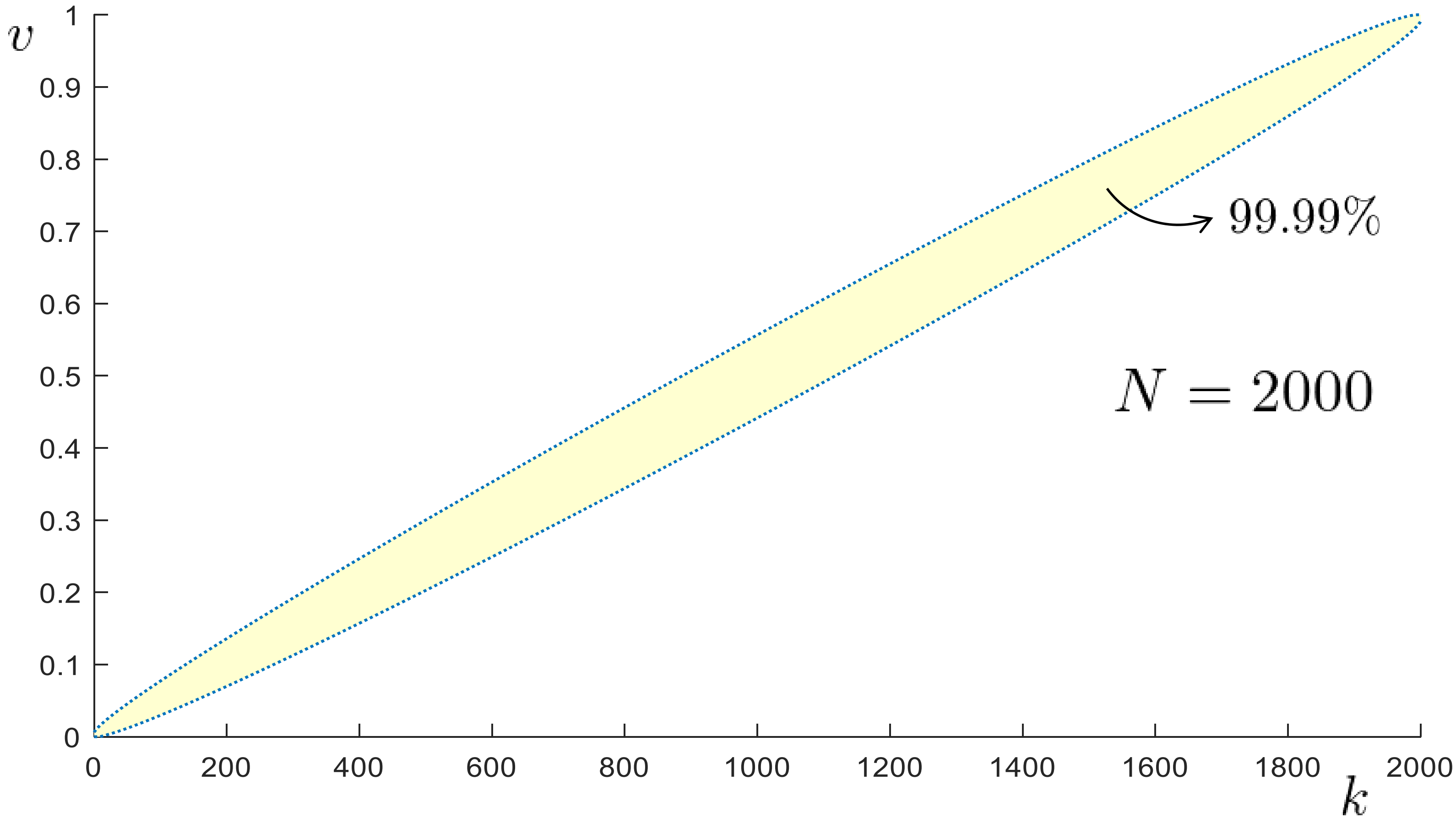
Main result

$$\mathbb{P}^N\left\{\delta^{(1)},\ldots,\delta^{(N)}:\ \epsilon_L(\pi^*)\leq V(x^*)\leq \epsilon^U(\pi^*)\right\}\geq 1-\beta$$



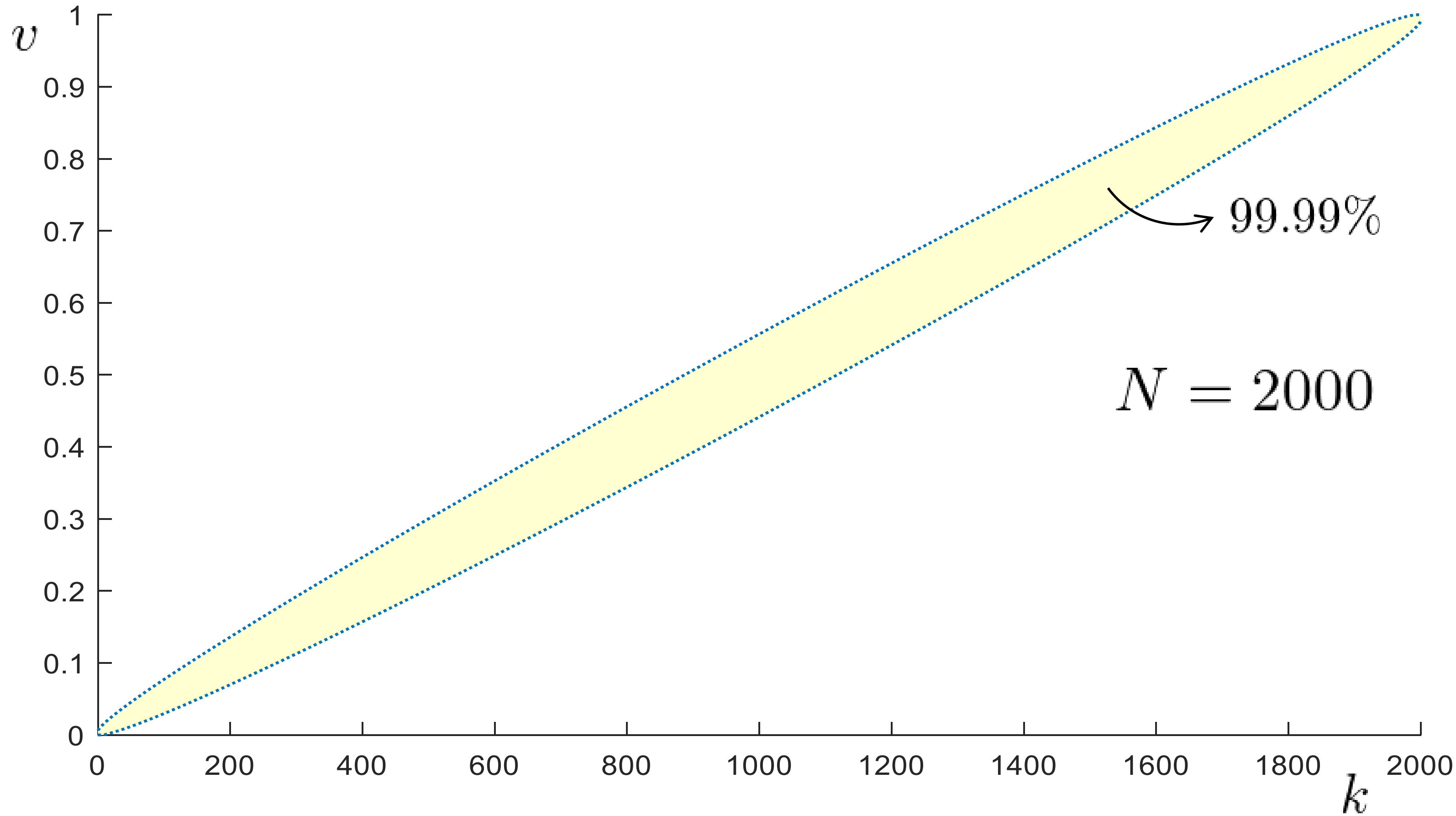
Main result

$$\mathbb{P}^N\left\{\delta^{(1)},\ldots,\delta^{(N)} :\ \epsilon_L(\pi^*) \leq V(x^*) \leq \epsilon^U(\pi^*)\right\} \geq 1-\beta$$



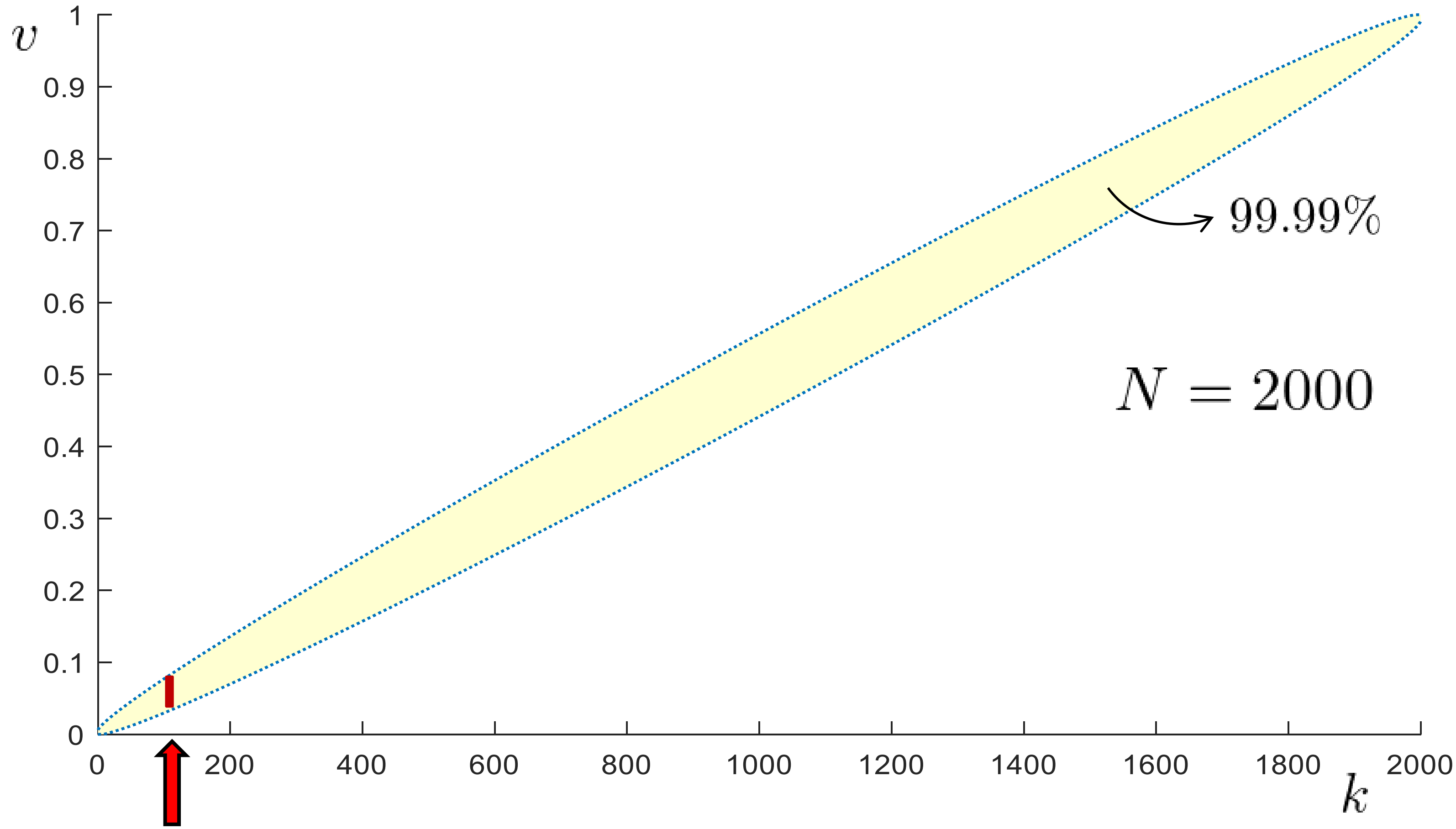
Main result

$\epsilon_L(\pi^*) \leq V(x^*) \leq \epsilon^U(\pi^*)$ **is true with confidence** $1-\beta$



Main result

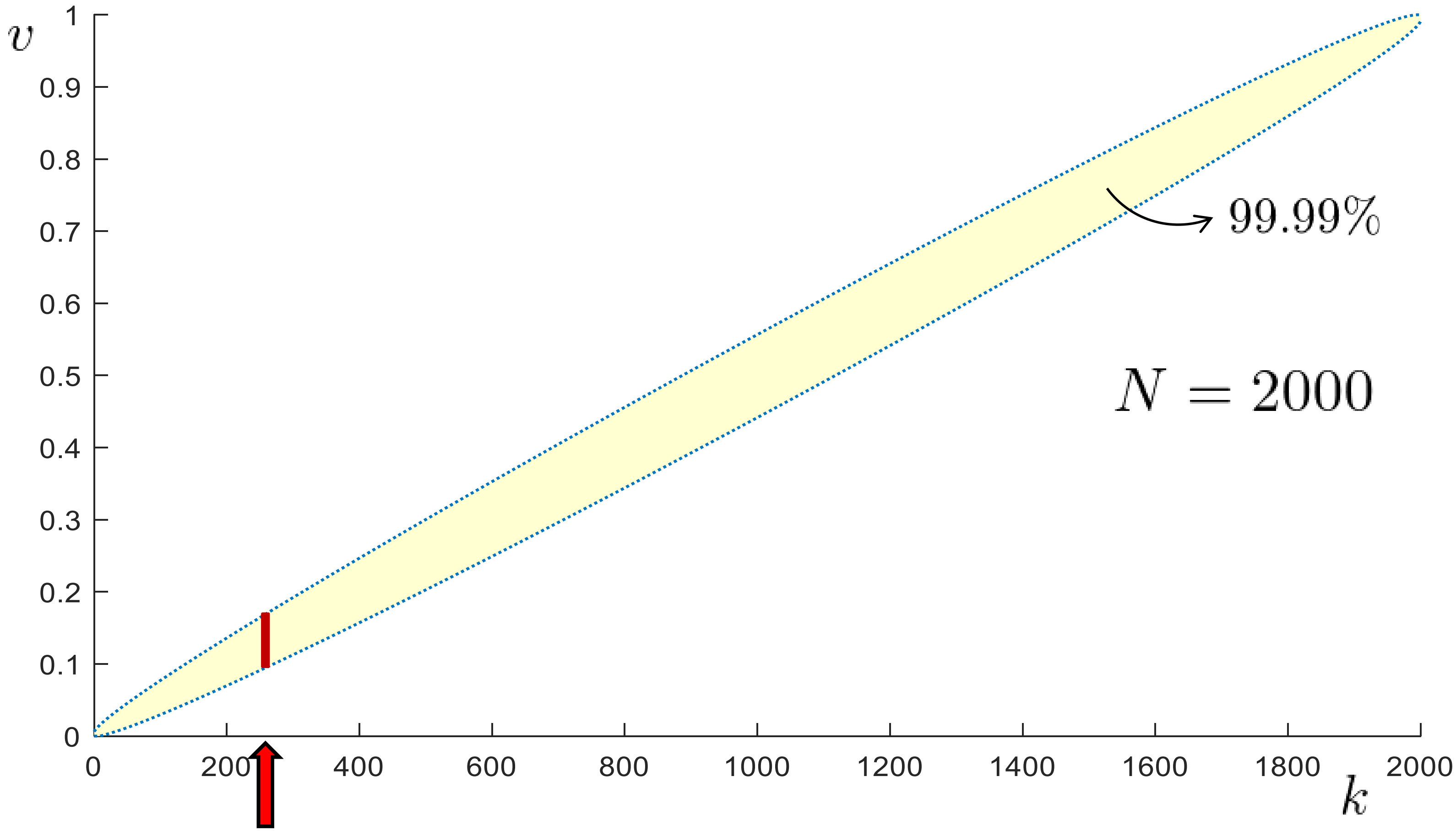
$\epsilon_L(\pi^*) \leq V(x^*) \leq \epsilon^U(\pi^*)$ **is true with confidence** $1 - \beta$



$\pi^* = 90 \to 0.026 \leq V(x^*) \leq 0.071$

Main result

$\epsilon_L(\pi^*) \leq V(x^*) \leq \epsilon^U(\pi^*)$ **is true with confidence** $1-\beta$



$$\pi^* = 260 \rightarrow 0.095 \leq V(x^*) \leq 0.17$$

Solution assessment

$c(x^*)$	accessible
$V(x^*)$	not accessible

Solution assessment

$c(x^*)$	accessible
$[\epsilon_L(\pi^*), \epsilon^U(\pi^*)]$	accessible

Solution assessment

$c(x^*)$ accessible

$[\epsilon_L(\pi^*), \epsilon^U(\pi^*)]$ accessible

accept/reject
the solution

make further decisions

compare various
“solutions”

Cost vs. risk tradeoff

$$\min_{x \in \mathbb{R}^d, \xi_i \geq 0} \quad c(x) + \rho \sum_{i=1}^{N} \xi_i$$

$$\text{s.t.} \quad f(x, \delta^{(i)}) \leq \xi_i,$$

$$i = 1, \ldots, N$$

$$\left.\begin{array}{cccc} \rho_1 & \rho_2 & \rho_3 & \cdots \\ \downarrow & \downarrow & \downarrow & \cdots \\ x_1^*, \pi_1^* & x_2^*, \pi_2^* & x_3^*, \pi_3^* & \cdots \end{array}\right\} \quad \text{cost vs. risk tradeoffs}$$

Cost vs. risk tradeoff (cont'd)

$$\min_{x \in \mathbb{R}^d, \xi_i \geq 0} \quad c(x) + \rho \sum_{i=1}^{N} \xi_i$$

$$\text{s.t.} \quad f(x, \delta^{(i)}) \leq \xi_i,$$

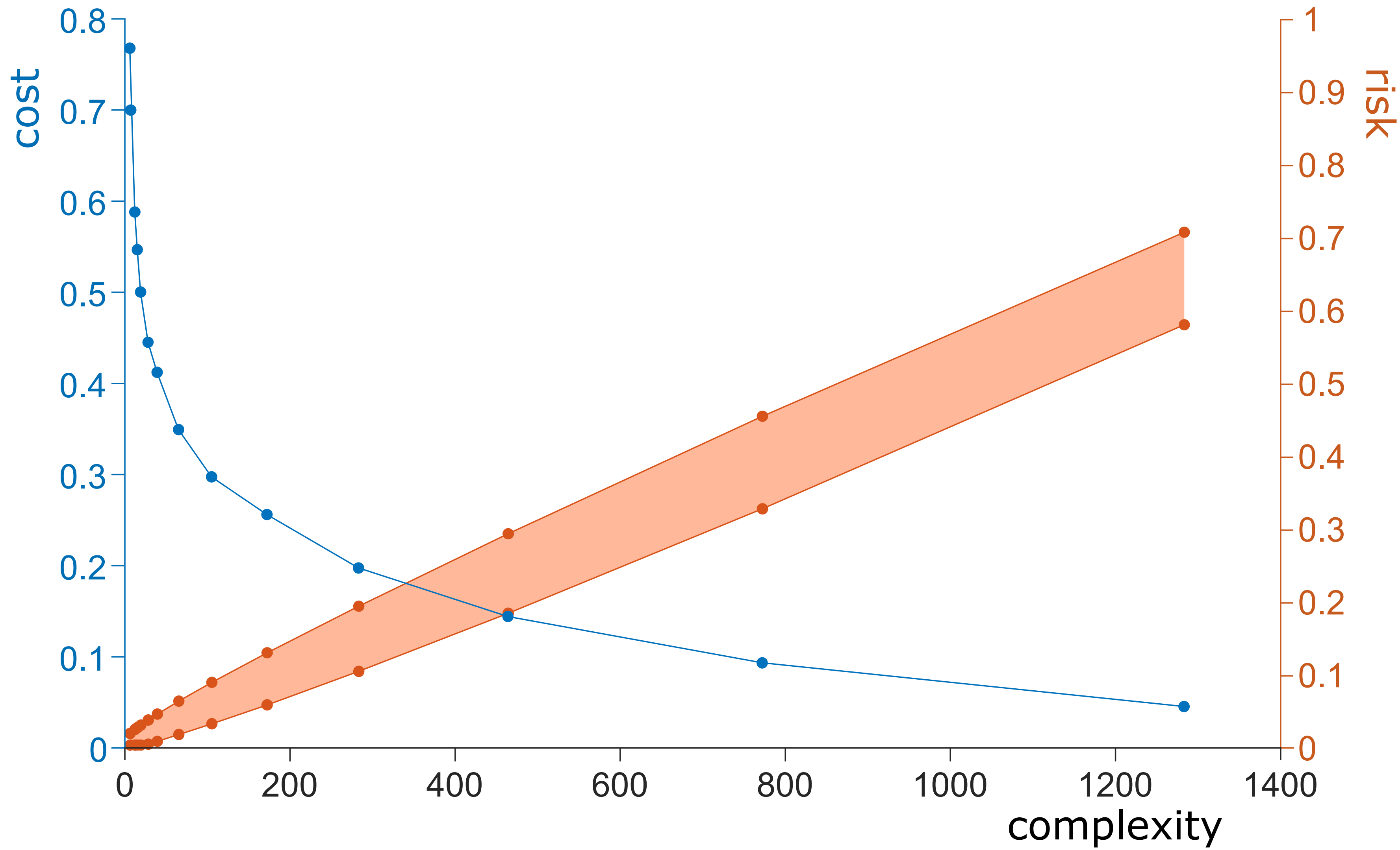
$$i = 1, \ldots, N$$

$$\begin{array}{cccc} \rho_1 & \rho_2 & \rho_3 & \cdots \\ \downarrow & \downarrow & \downarrow & \cdots \\ x_1^*, \pi_1^* & x_2^*, \pi_2^* & x_3^*, \pi_3^* & \cdots \end{array} \left. \vphantom{\begin{array}{c} a \\ a \\ a \end{array}} \right\}$$

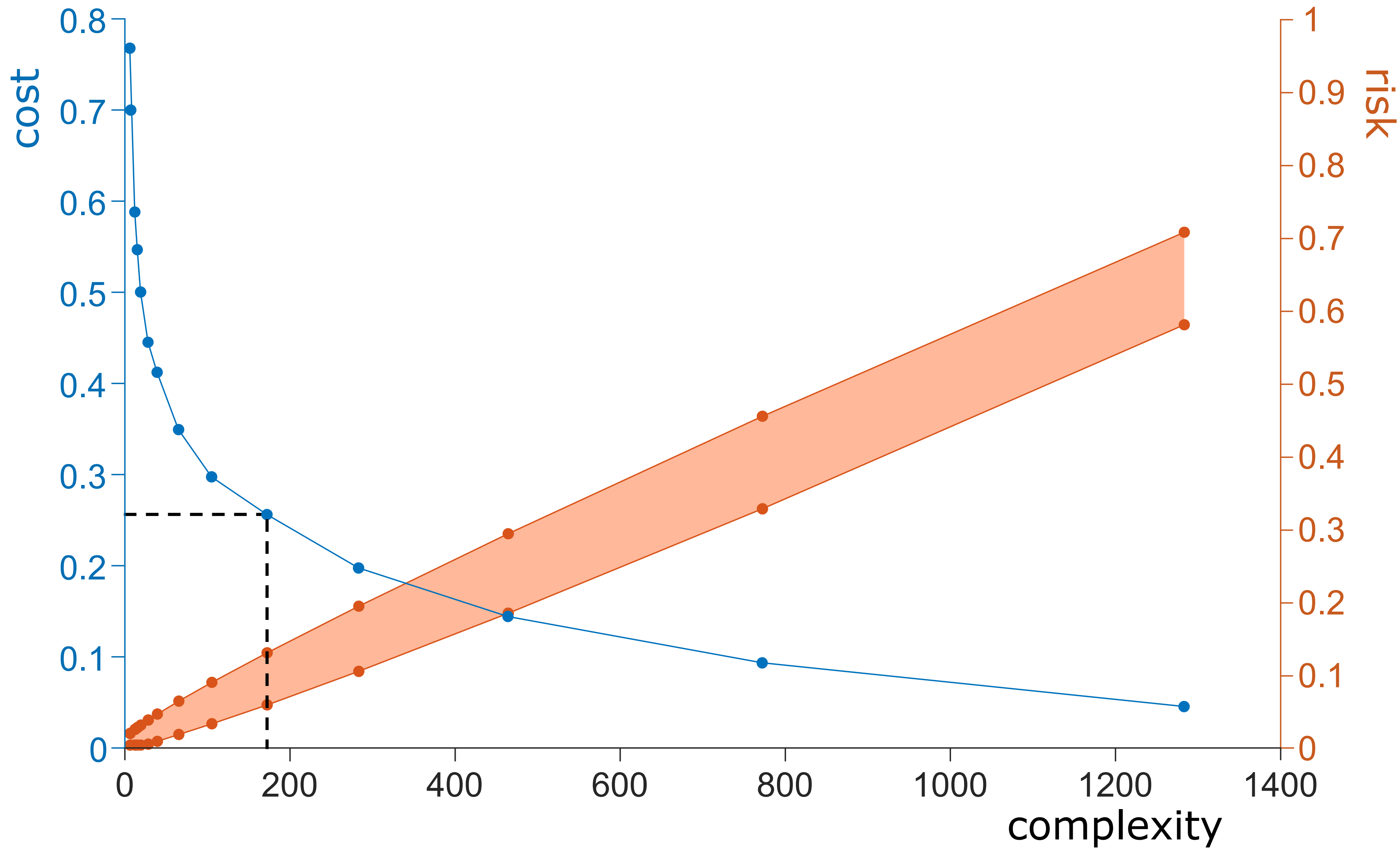
cost vs. risk tradeoffs

quantitative comparison via $c(x_i^*)$ and $[\epsilon_L(\pi_i^*), \epsilon^U(\pi_i^*)]$

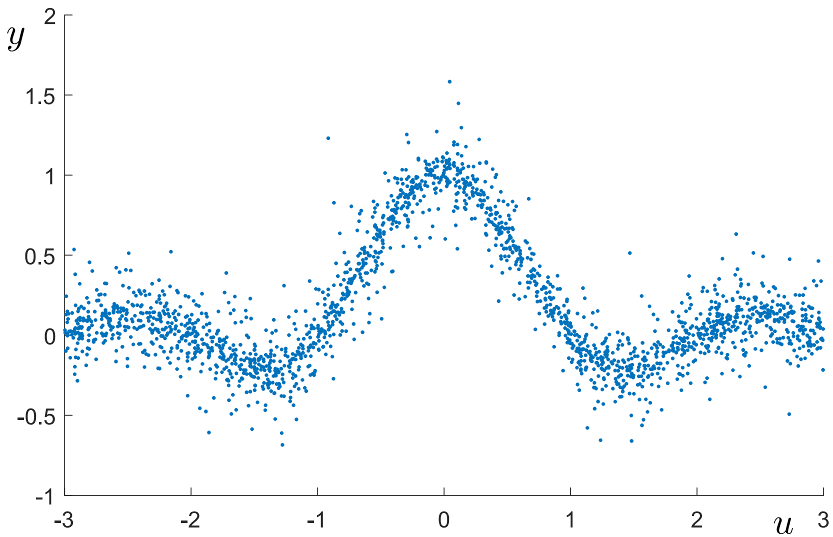
Cost vs. risk plot



Cost vs. risk plot



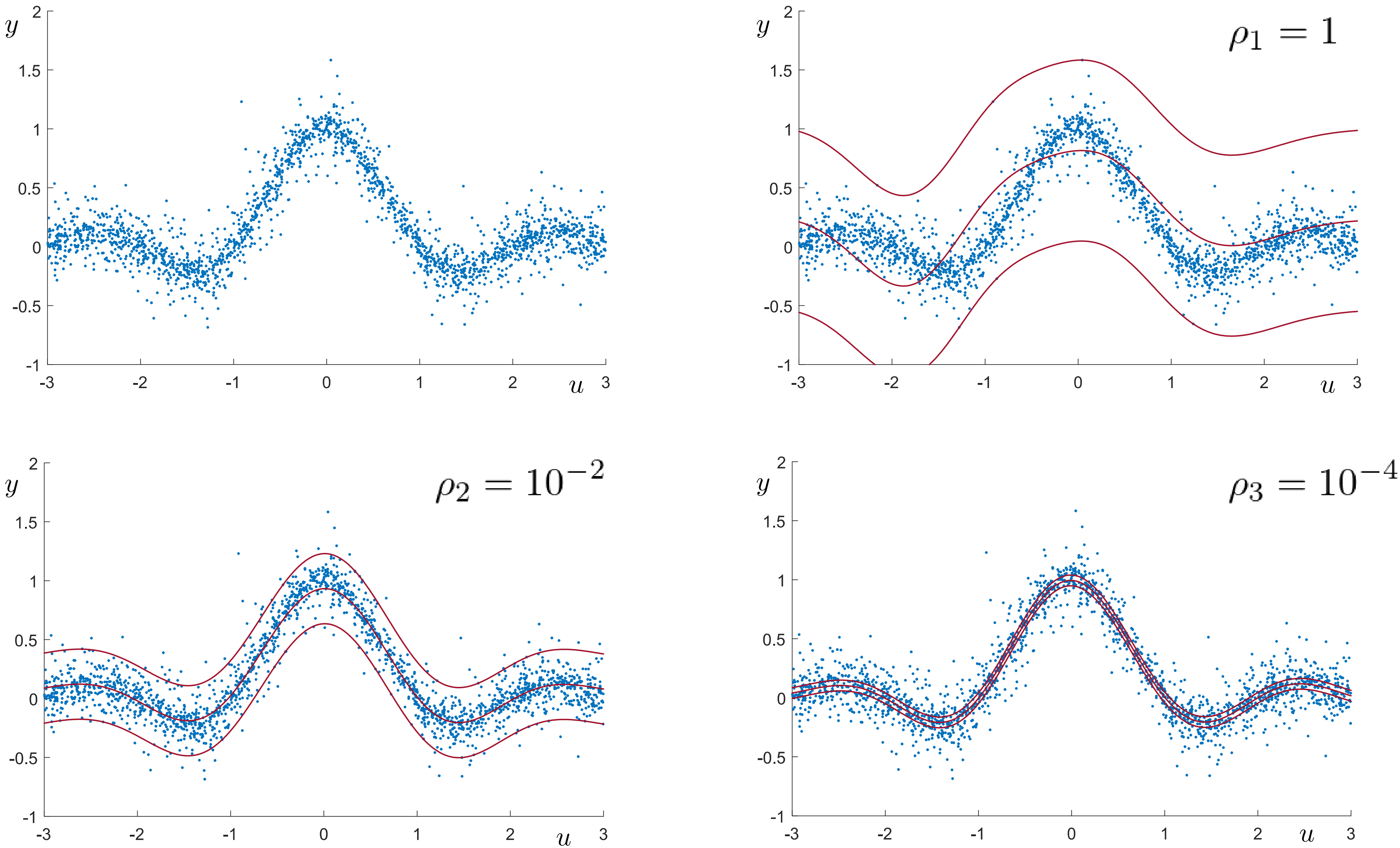
Example: Support Vector Regression



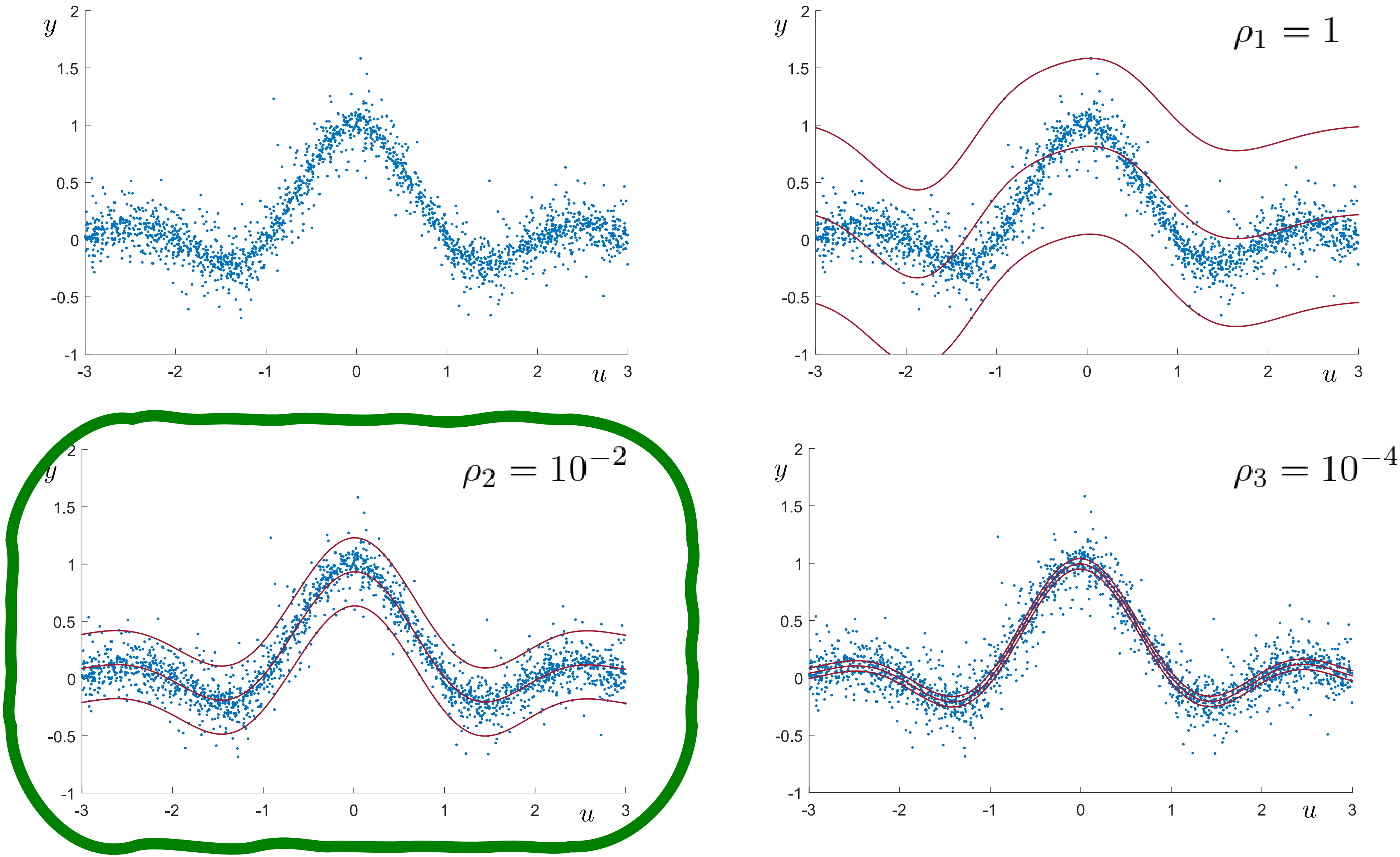
$$\min_{\substack{w \in \mathcal{U}, \gamma \geq 0, b \in \mathbb{R} \\ \xi_i \geq 0, i=1,\ldots,N}} \quad (\gamma + \tau \|w\|^2) + \rho \sum_{i=1}^{N} \xi_i$$

$$\text{subject to:} \quad |y_i - \langle w, \mathbf{u}_i \rangle - b| - \gamma \leq \xi_i, \quad i = 1, \ldots, N$$

Example: Support Vector Regression



Example: Support Vector Regression



Conclusions

- Data are a "gold mine" for decision-making, but good theories are needed for a reliable exploitation

- Scenario approach: a flexible and effective setup for data-driven decision making with a good theory to assess the reliability of the solution

- At a very general level, the complexity π^* (visible) carries fundamental information on the risk $V(x^*)$ (hidden)

- The risk can be estimated from the complexity, without resorting to (possibly unreliable) prior information

Thank you !

M.C. Campi, S. Garatti. Wait-and-judge scenario optimization. Mathematical Programming, 167(1):155-189, 2018. https://doi.org/10.1007/s10107-016-1056-9

S. Garatti , M.C. Campi. Risk and complexity in scenario optimization. Mathematical Programming, 191(1): 243-279, 2022. https://doi.org/10.1007/s10107-019-01446-4

M.C. Campi, S. Garatti. A Theory of the Risk for Optimization with Relaxation and its Application to Support Vector Machines. Journal of Machine Learning Research, 288:1-38, 2021.