THE SCENARIO APPROACH
to
STOCHASTIC OPTIMIZATION

Marco C. Campi

University of Brescia

Italy

“What I like about experience is that it is such an honest thing. … You may have deceived yourself, but experience is not trying to deceive you.”

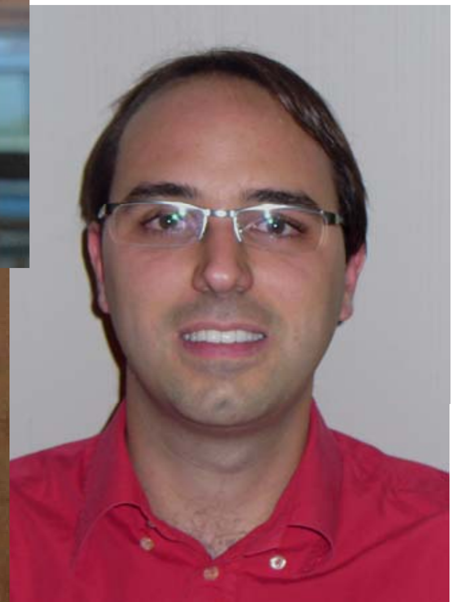
C.S. Lewis

thanks to:

Algo
Care’

Maria Prandini

Simone
Garatti



Bernardo
Pagnoncelli

Giuseppe
Calafiore

Federico
Ramponi

optimization

- controller synthesis
- classification
- portfolio selection

⋮

$\longrightarrow$ optimization program

optimization

- controller synthesis
- classification
- portfolio selection

  ⋮

optimization program

uncertain environment

- exercise caution

uncertain optimization:

$$\underset{\theta}{\text{optimize}} \; J(\theta, \delta)$$

uncertain optimization:

$$\underset{\theta}{\text{optimize}} \; J(\theta, \delta)$$

not a valid mathematical formulation

uncertain optimization:

$$\underset{\theta}{\text{optimize}} \; J(\theta, \delta)$$

not a valid mathematical formulation

often, a description of uncertainty is not available, or it is only partially available scenario-based knowledge:

scenario-based knowledge:

knowledge about uncertainty can be acquired through experience,

that is, we look at previous cases, or scenarios, of the same problem

example: portfolio optimization

[with B. Pagnoncelli & D. Reich]

example: portfolio optimization

1$

[with B. Pagnoncelli & D. Reich]

example: portfolio optimization

1\$

d assets

[with B. Pagnoncelli & D. Reich]

example: portfolio optimization

1$

d assets

p_k = percentage of capital invested on asset k

[with B. Pagnoncelli & D. Reich]

example: portfolio optimization

1$

d assets

p_k = percentage of capital invested on asset k

$\theta = [p_1 \cdots p_d]^T$

[with B. Pagnoncelli & D. Reich]

example: portfolio optimization

1$

d assets

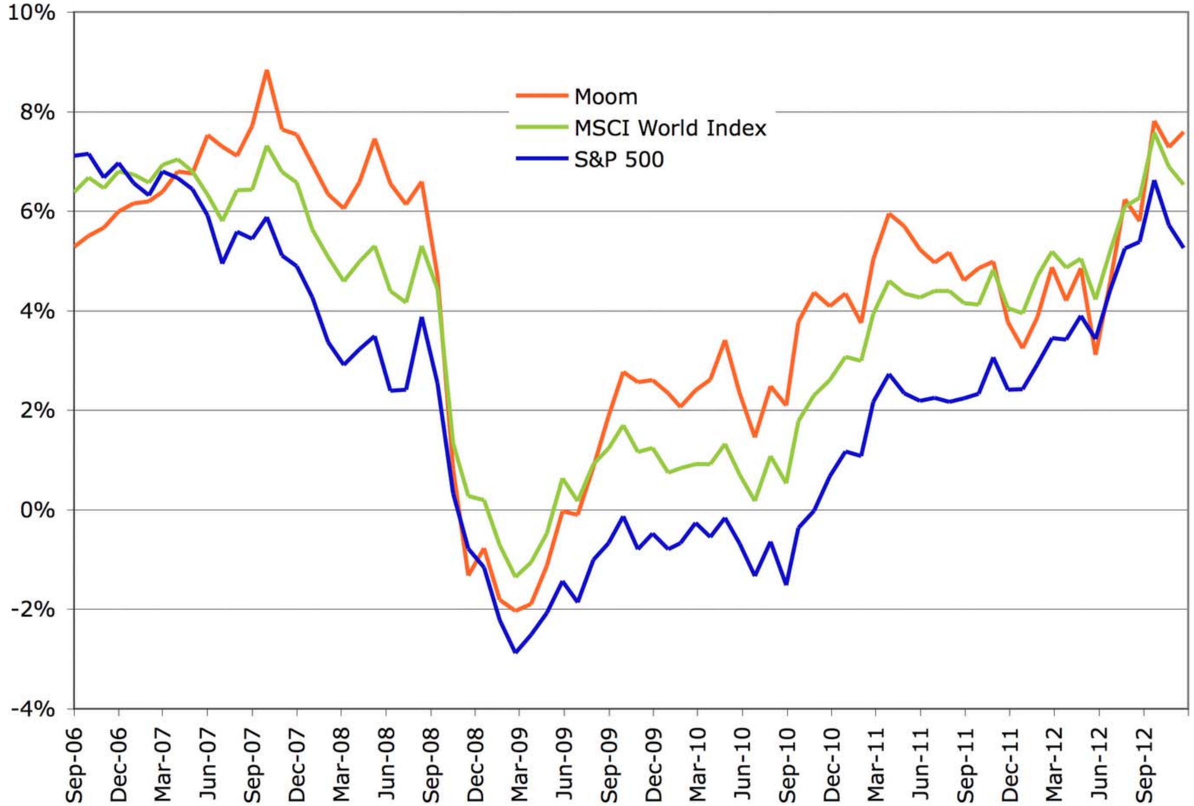
p_k = percentage of capital invested on asset k

$\theta = [p_1 \cdots p_d]^T$

$$J(\theta, \delta) = \sum_{k=1}^{d} p_k R_k \qquad \delta = [R_1 \cdots R_d]^T$$

[with B. Pagnoncelli & D. Reich]

example: portfolio optimization



[with B. Pagnoncelli & D. Reich]

example: portfolio optimization

record of past rate of returns:

$R_k(i)$ = return of asset k over period i (scenarios)

[with B. Pagnoncelli & D. Reich]

example: portfolio optimization

record of past rate of returns:

$R_k(i)$ = return of asset k over period i (scenarios)

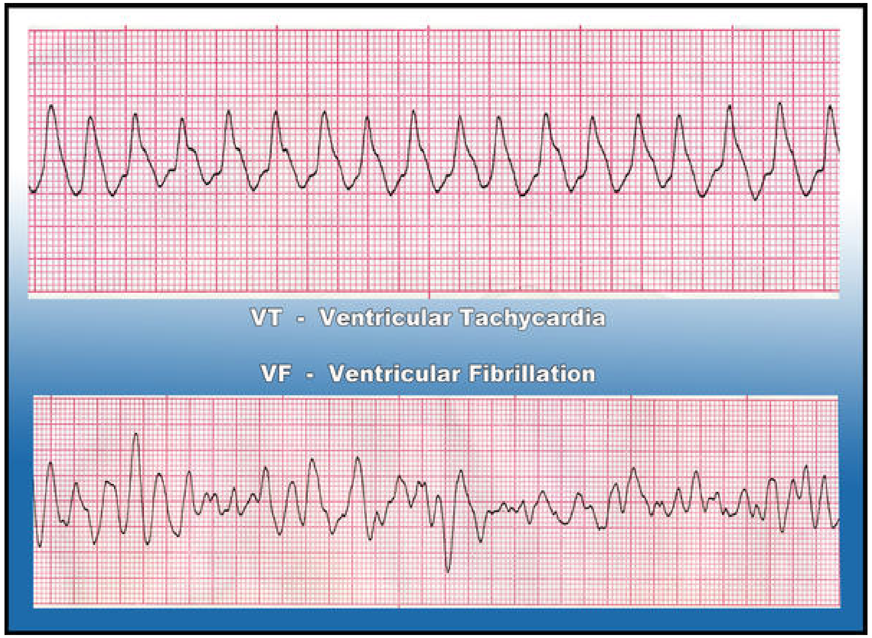
$$J(\theta, \delta^{(i)}) = \sum_{k=1}^{d} p_k R_k(i), \quad i = 1 \ldots, N$$

[with B. Pagnoncelli & D. Reich]

example: classification - defibrillation

[with A. Caré]

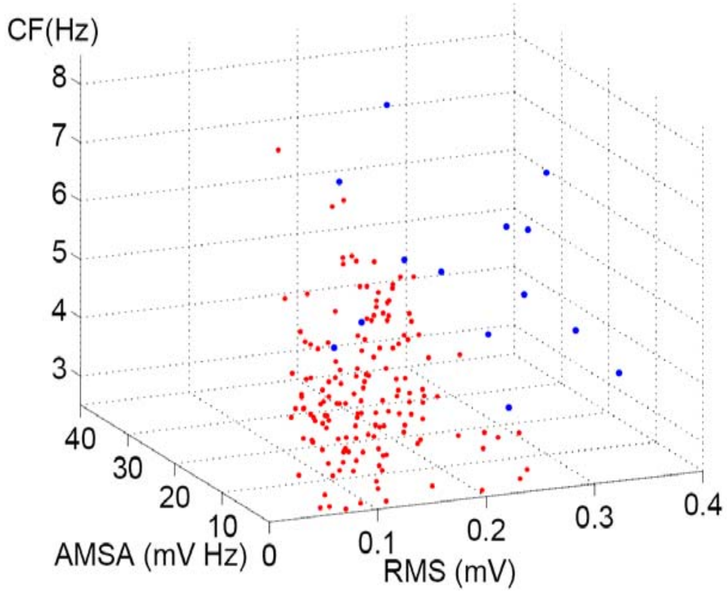
example: classification - defibrillation



decide whether a defibrillator has to be applied

[with A. Caré]

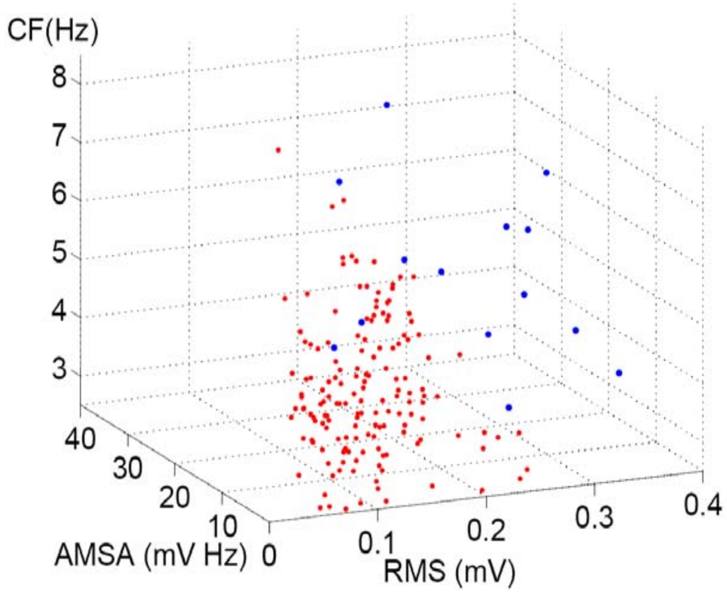
example: classification - defibrillation



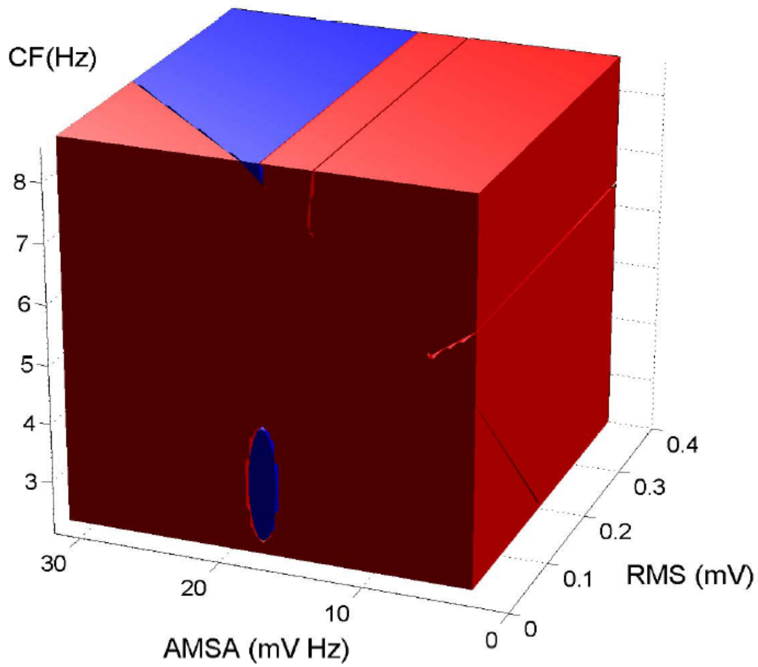
(scenarios)

[with A. Caré]

example: classification - defibrillation



(scenarios)



[with A. Caré]

“scenario” optimization

(convex case)

min-max “scenario” optimization

[with G. Calafiore]

min-max “scenario” optimization

$J(\theta, \delta)$ convex in θ

[with G. Calafiore]

min-max "scenario" optimization

$J(\theta, \delta)$ convex in θ

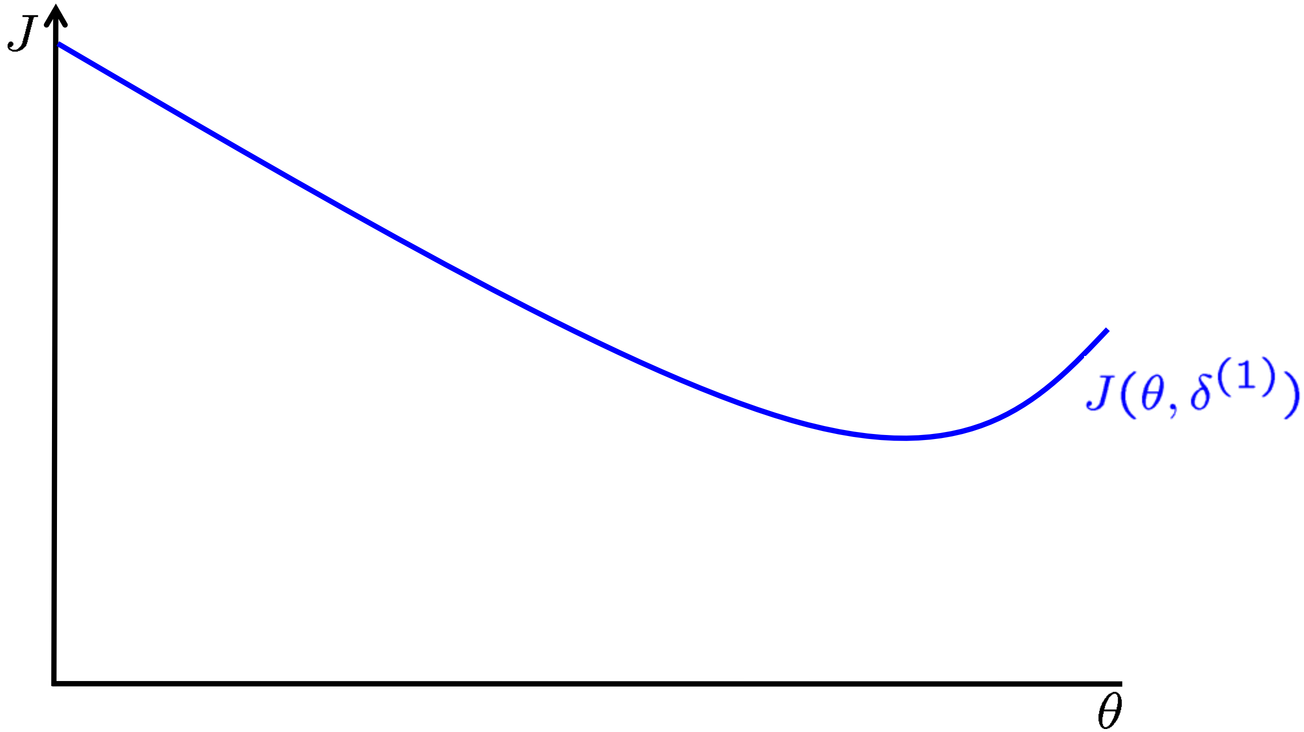
$$\delta^{(1)}, \delta^{(2)}, \ldots, \delta^{(N)} \longrightarrow J(\theta, \delta^{(1)}), J(\theta, \delta^{(2)}), \ldots, J(\theta, \delta^{(N)})$$

[with G. Calafiore]

min-max "scenario" optimization

$J(\theta, \delta)$ convex in θ

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

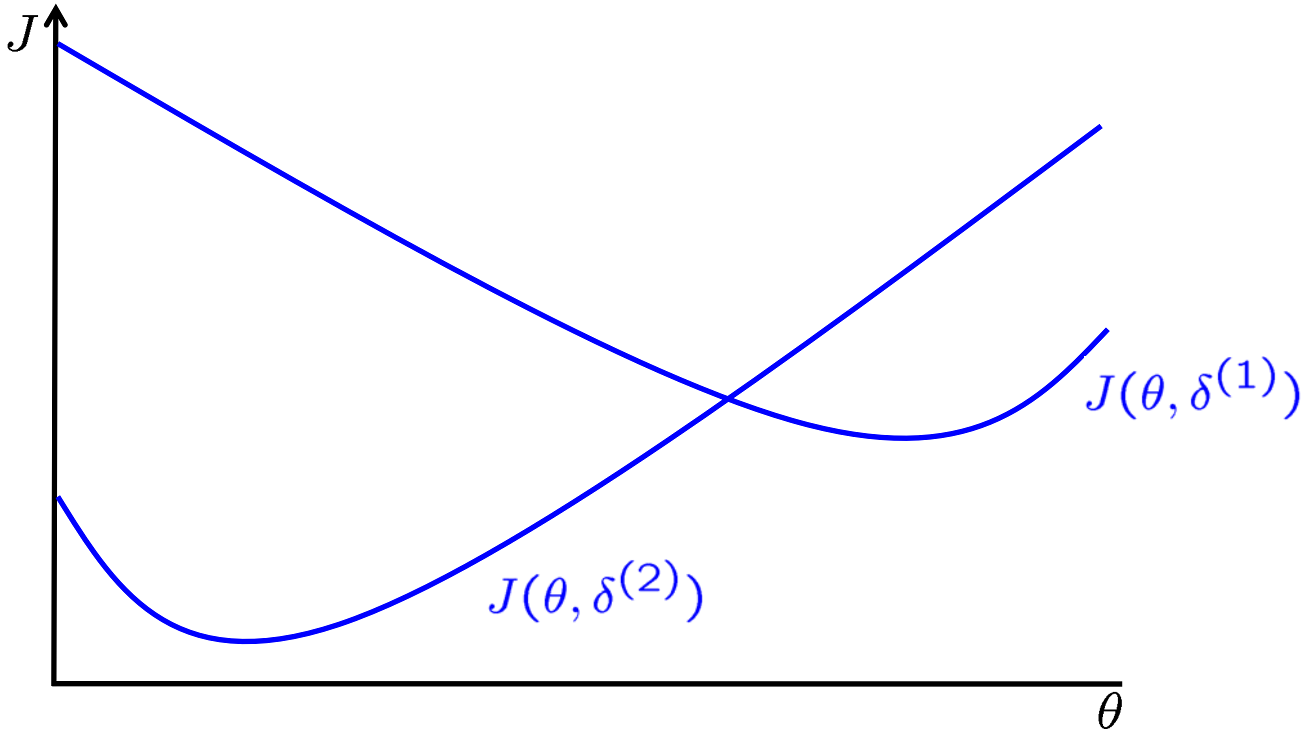


[with G. Calafiore]

min-max "scenario" optimization

$J(\theta, \delta)$ convex in θ

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

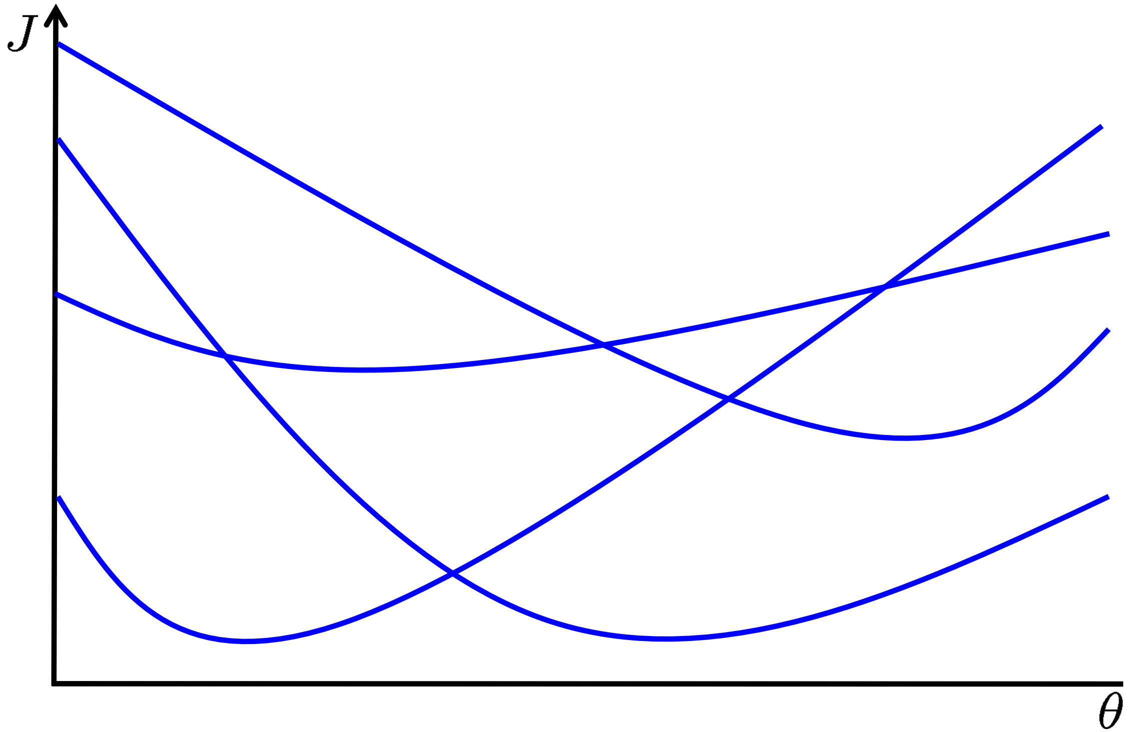


[with G. Calafiore]

min-max “scenario” optimization

$J(\theta, \delta)$ convex in θ

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

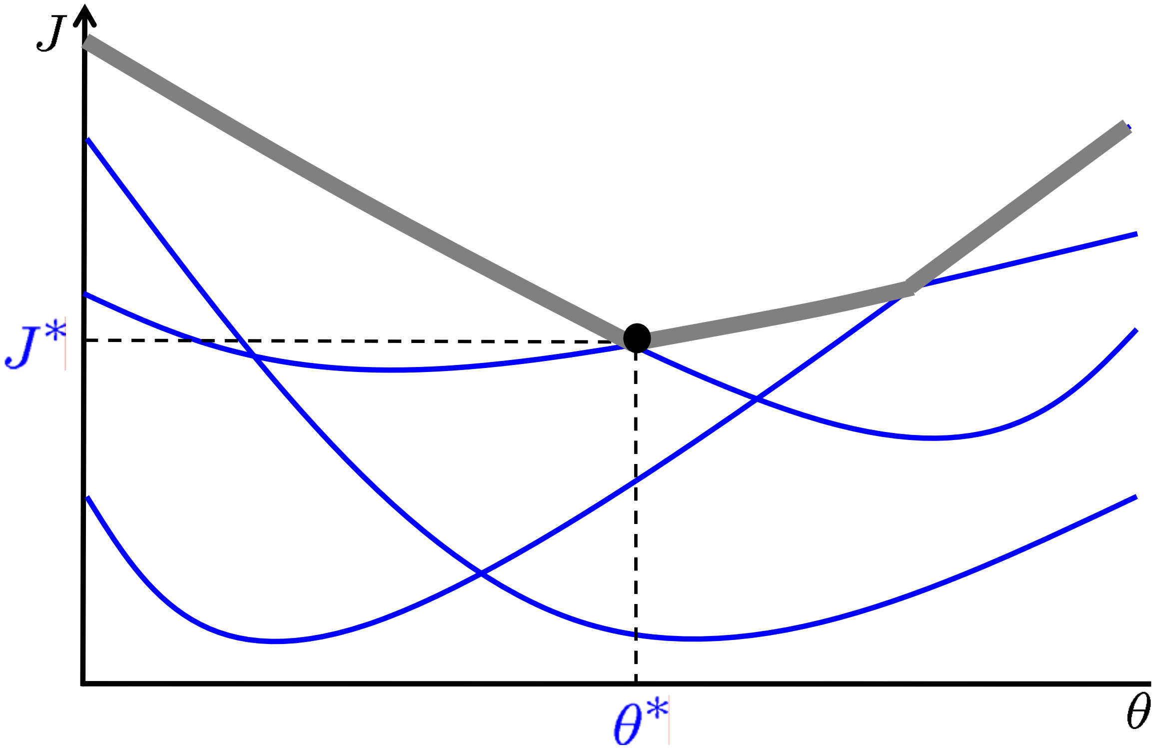


[with G. Calafiore]

min-max "scenario" optimization

$J(\theta, \delta)$ convex in θ

$$\delta^{(1)}, \delta^{(2)}, \dots, \delta^{(N)} \longrightarrow J(\theta, \delta^{(1)}), J(\theta, \delta^{(2)}), \dots, J(\theta, \delta^{(N)})$$



$\mathrm{SP}_N =$ Senario Program

[with G. Calafiore]

what are the guarantees
for unseen situations?

what are the guarantees for unseen situations?

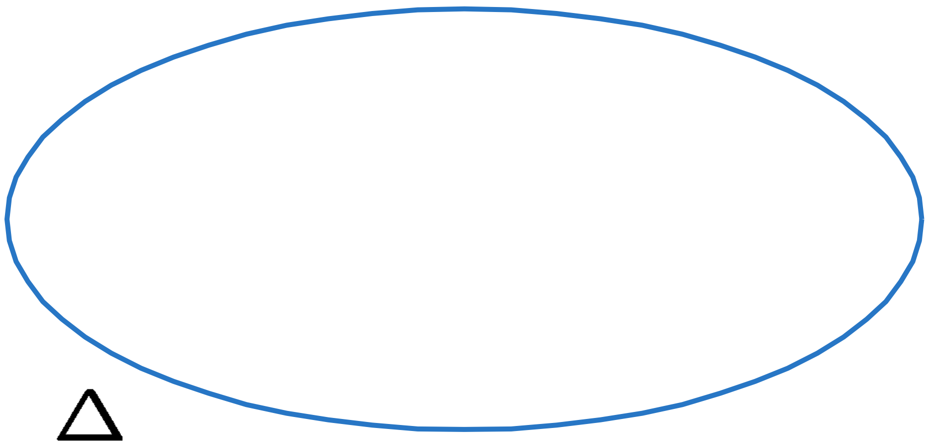
how guaranteed is J^* for another δ?

what are the guarantees
for unseen situations?

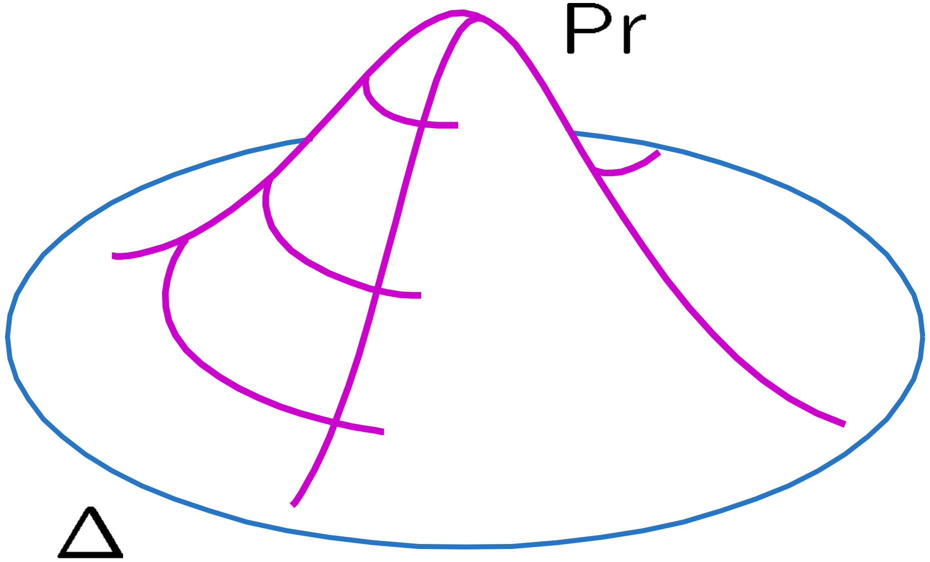
how guaranteed is J^* for another δ?

from the “<u>visible</u>” to the “<u>invisible</u>”

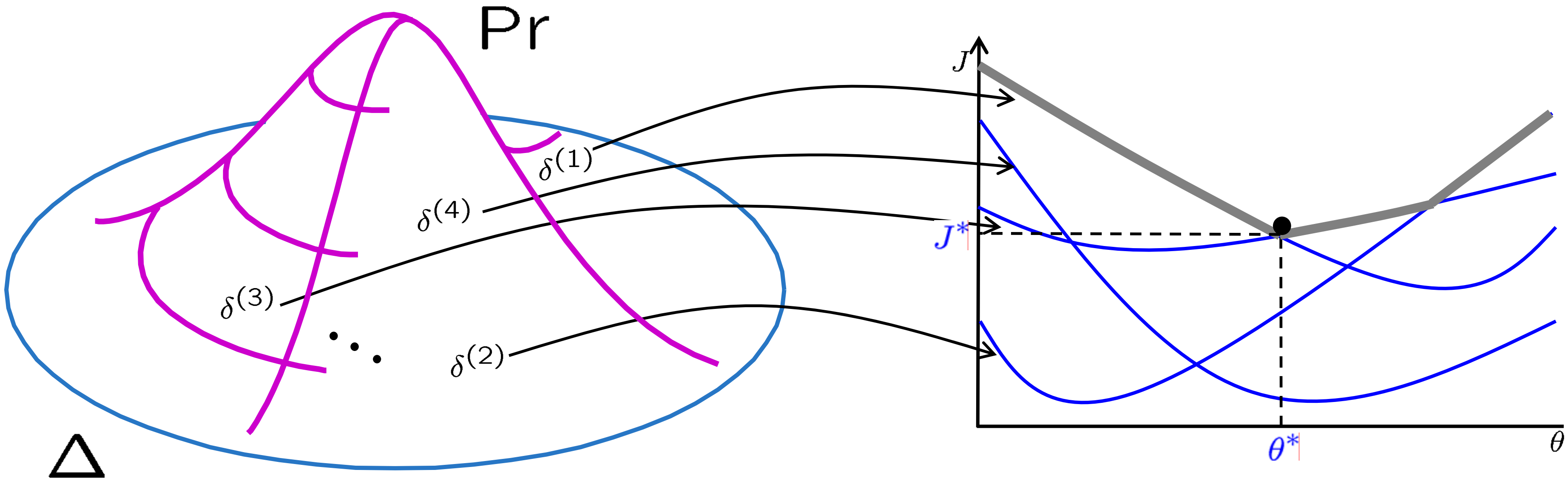
about uncertainty



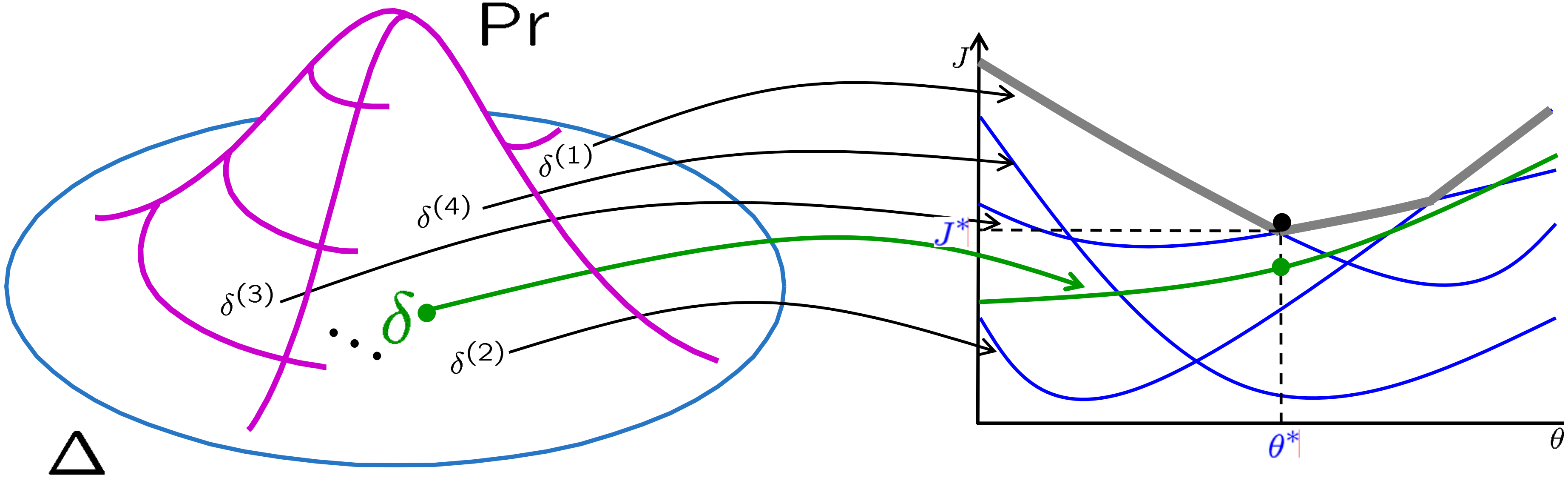
about uncertainty



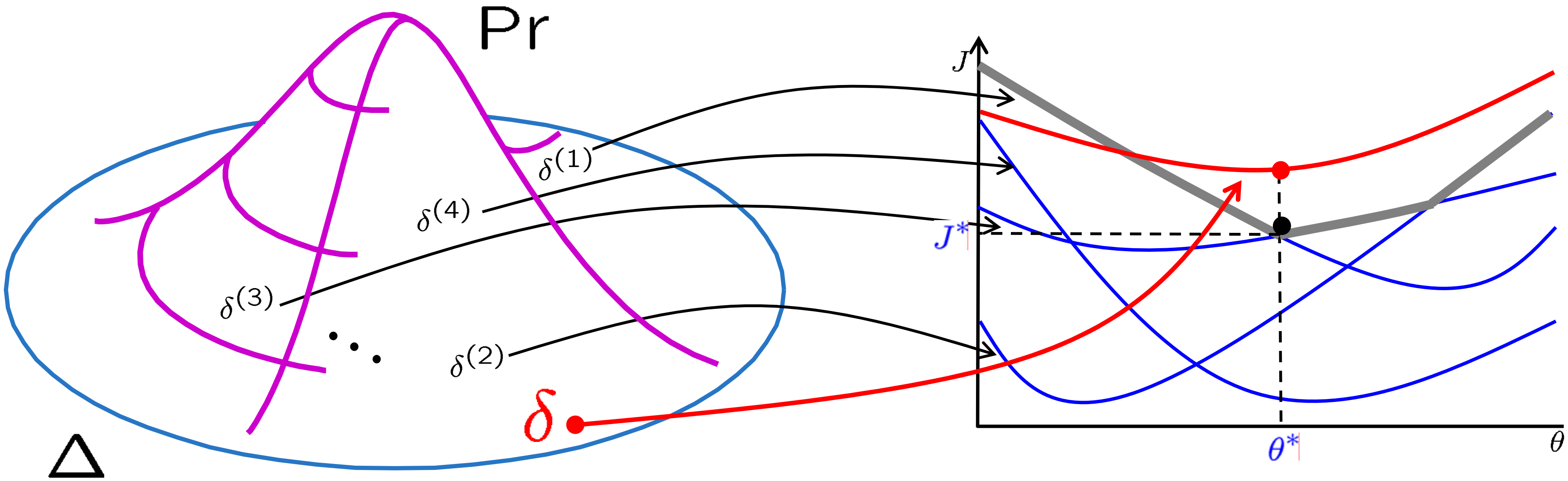
about uncertainty



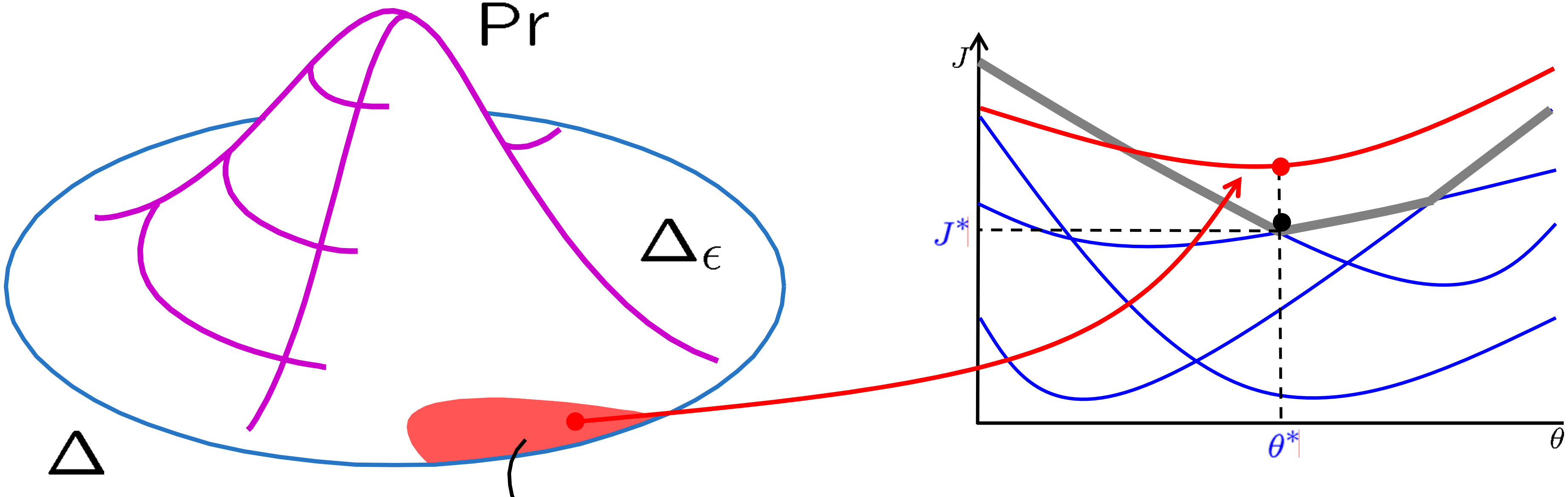
about uncertainty



about uncertainty



about uncertainty



$prob =$ risk

Theorem (with S. Garatti)

Fix $\epsilon \in (0,1)$ (risk parameter)
$\beta \in (0,1)$ (confidence parameter)

If $N = \frac{2}{\epsilon}\left(\ln \frac{1}{\beta} + d\right)$,
then,
with probability $\geq 1-\beta$,
risk $\leq \epsilon$.

Theorem (with S. Garatti)

Fix $\epsilon \in (0,1)$ (risk parameter)

If $N = \frac{2}{\epsilon}\left(\ln \frac{1}{\beta} + d\right)$,
then,

risk $\leq \epsilon$.

Theorem (with S. Garatti)

Fix $\epsilon \in (0,1)$ (risk parameter)
$\beta \in (0,1)$ (confidence parameter)

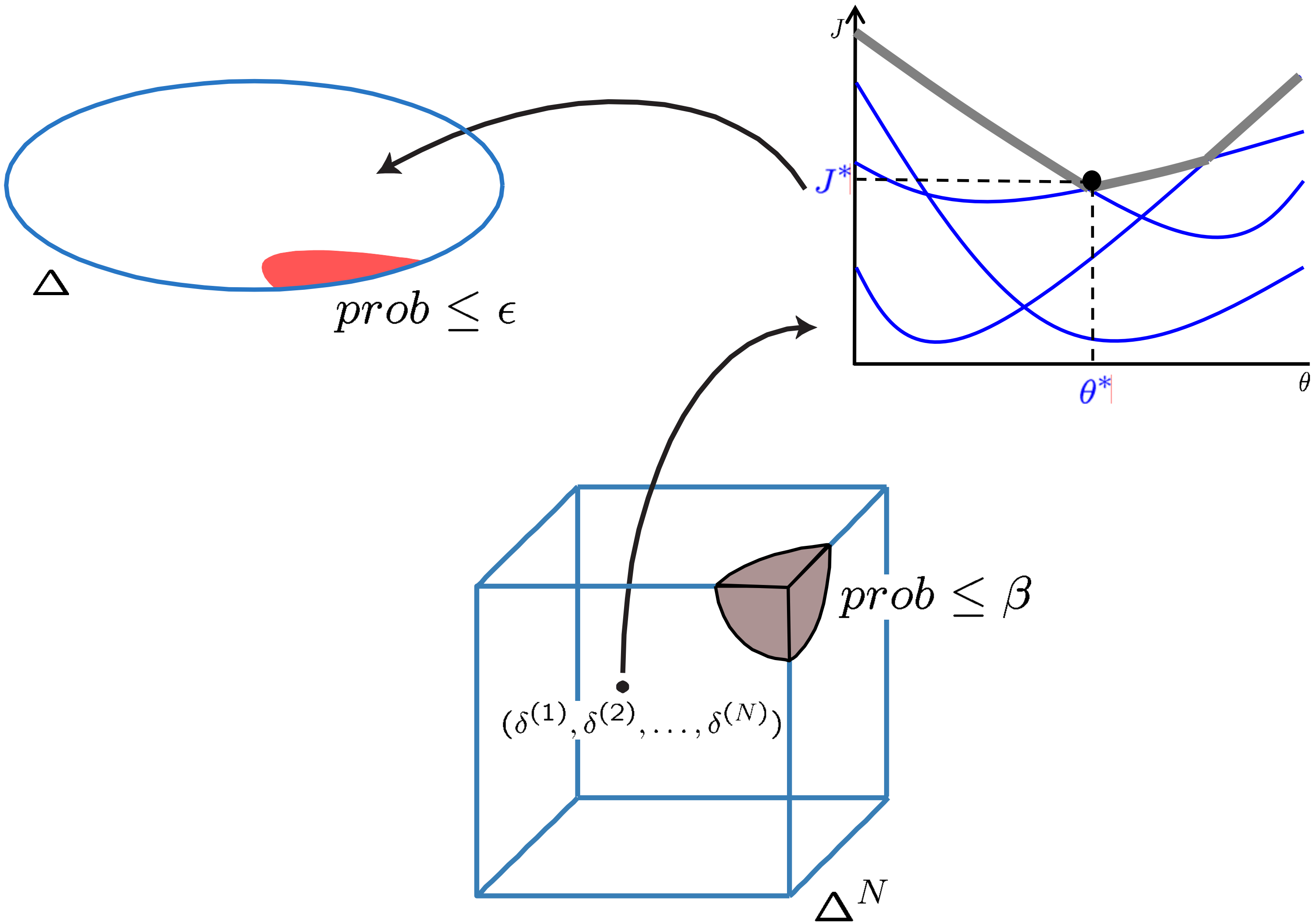
If $N = \frac{2}{\epsilon}\left(\ln \frac{1}{\beta} + d\right)$,
then,
with probability $\geq 1 - \beta$,
risk $\leq \epsilon$.

Theorem (with S. Garatti)

Fix $\epsilon \in (0, 1)$ (risk parameter)

If $N = \frac{2}{\epsilon}(7 \ln 10 + d)$,
then,

risk $\leq \epsilon$.



J
J^*
θ^*
θ
Δ
$prob \leq \epsilon$
$prob \leq \beta$
$(\delta^{(1)}, \delta^{(2)}, \dots, \delta^{(N)})$
Δ^N

comments

generalization $\longrightarrow$ need for structure

good news: the structure we need
is only convexity

... more comments

$$N = \frac{2}{\epsilon}\left(\ln \frac{1}{\beta} + d\right)$$

- *N* depends on how complex the decision is via d
- *N* does not depend on how complex the “real world” is

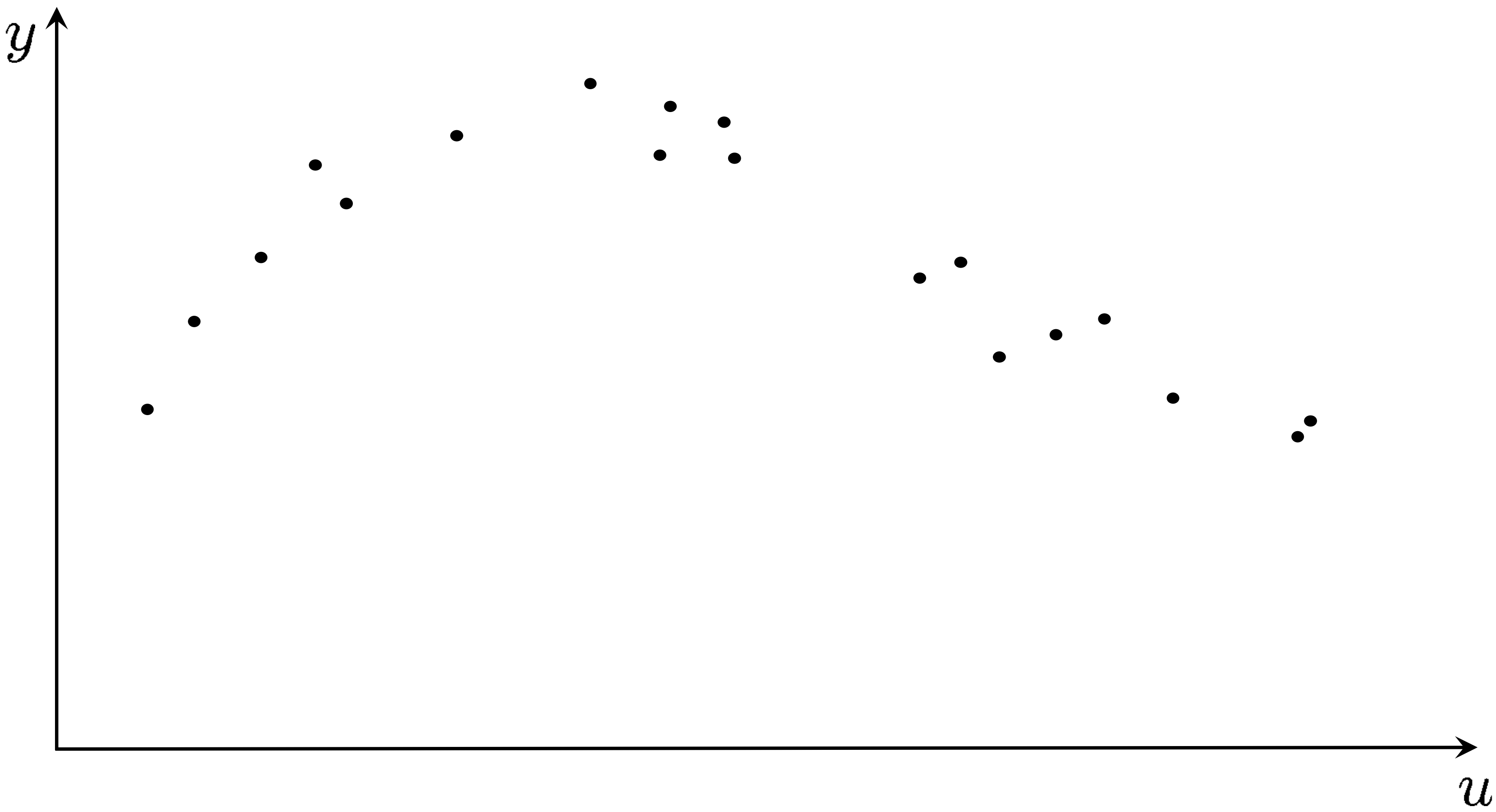
… more comments

$$N = \frac{2}{\epsilon}\left(\ln\frac{1}{\beta} + d\right)$$

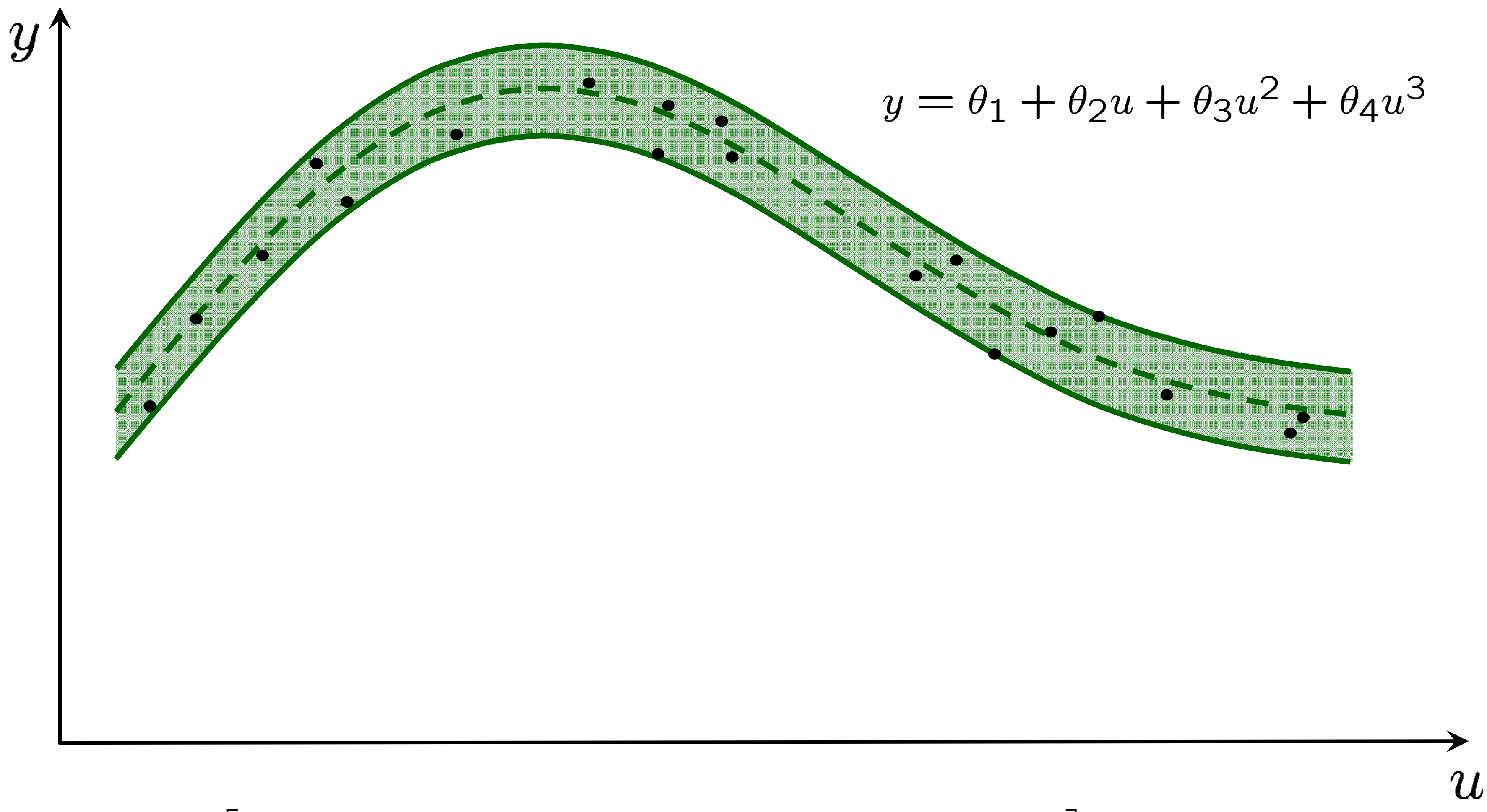
- *N* depends on how complex the decision is via d
- *N* does not depend on how complex the "real world" is

don't try to reconstruct the real world to answer easy questions!

... more comments (IPM = Interval Prediction Model)

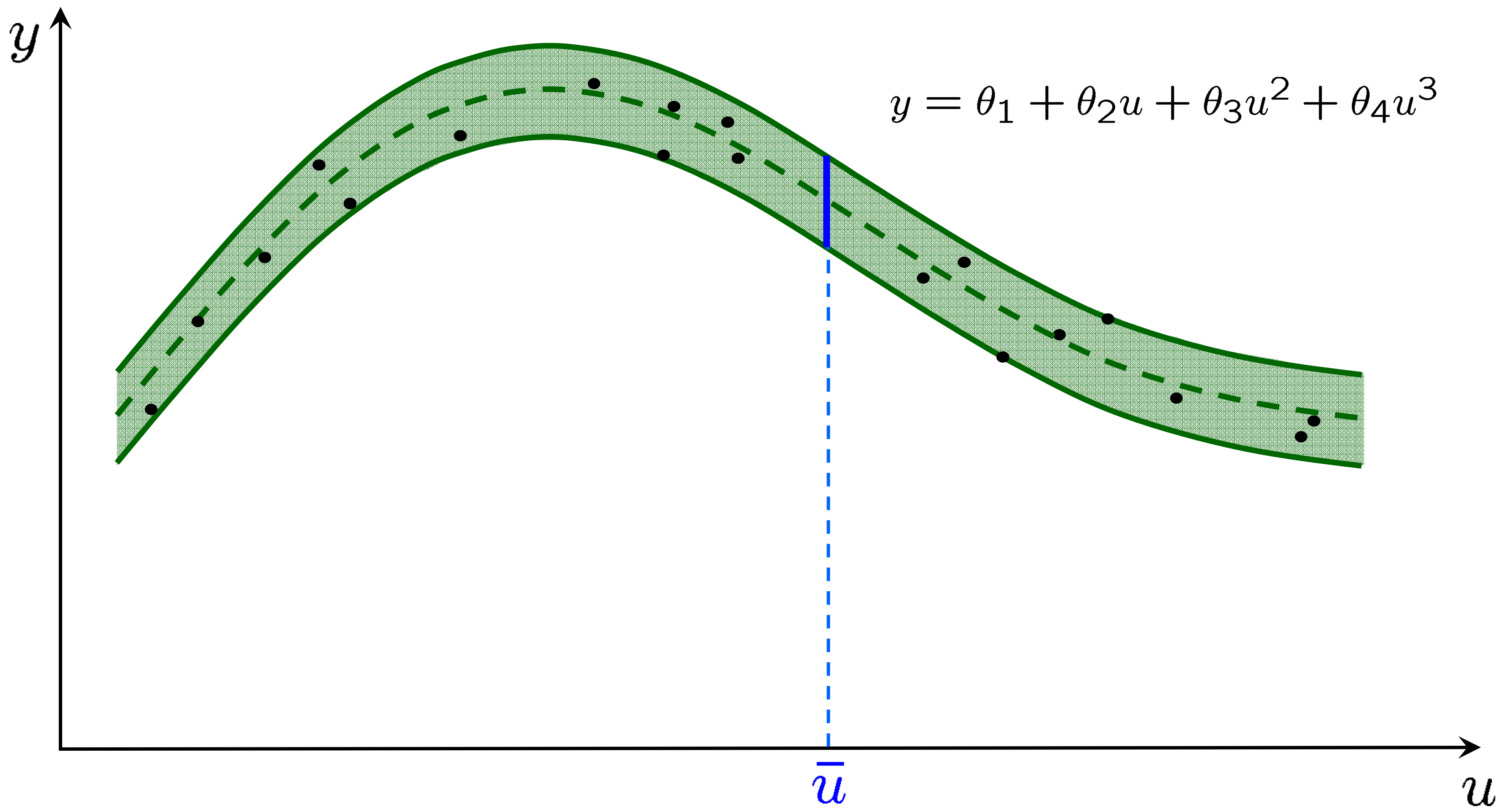


… more comments (IPM = Interval Prediction Model)



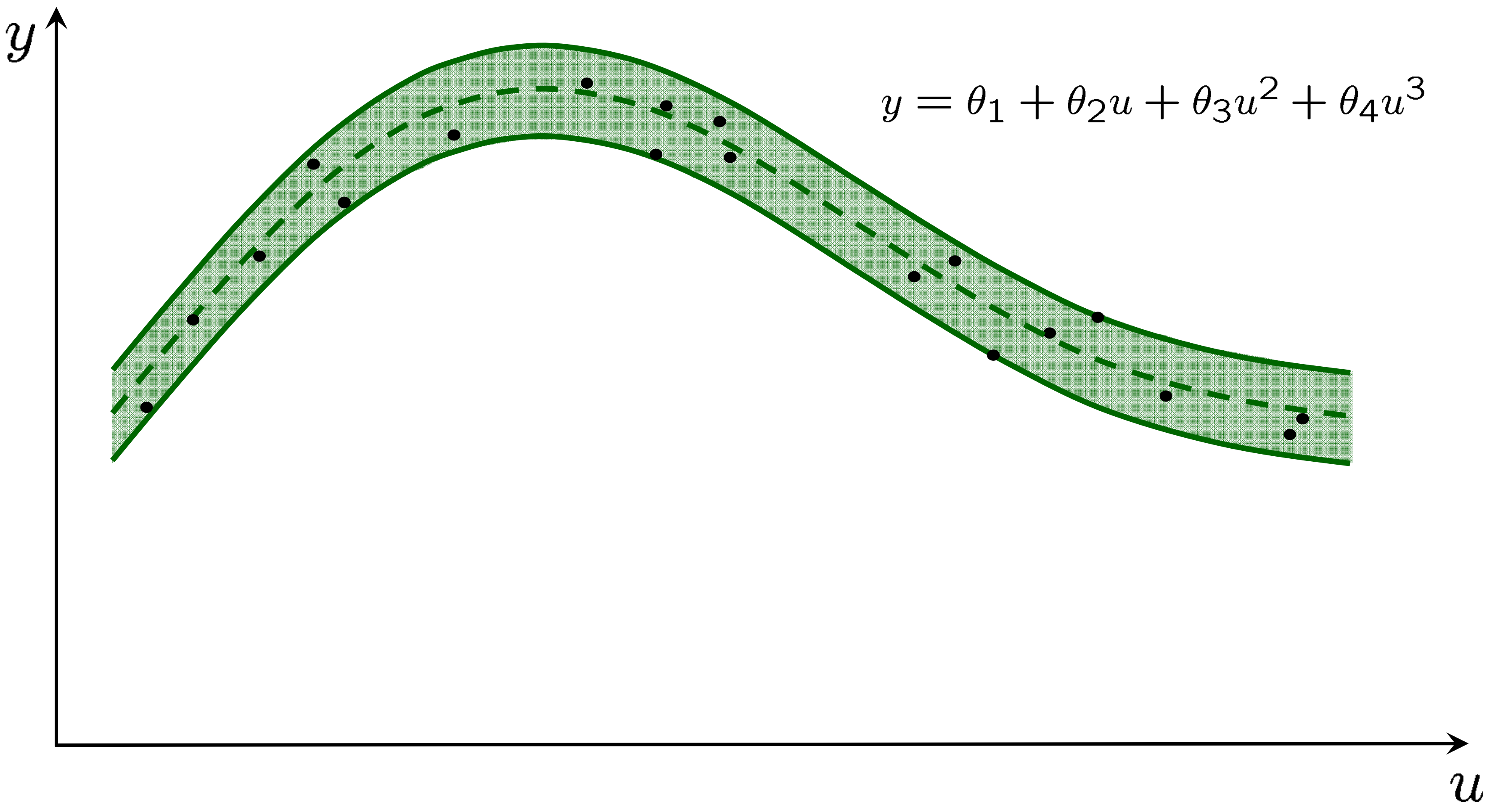
$$\min_{\theta_1,\theta_2,\theta_3,\theta_4} \left[\max_i |y_i - [\theta_1 + \theta_2 u_i + \theta_3 u_i^2 + \theta_4 u_i^3]| \right]$$

… more comments (IPM = Interval Prediction Model)



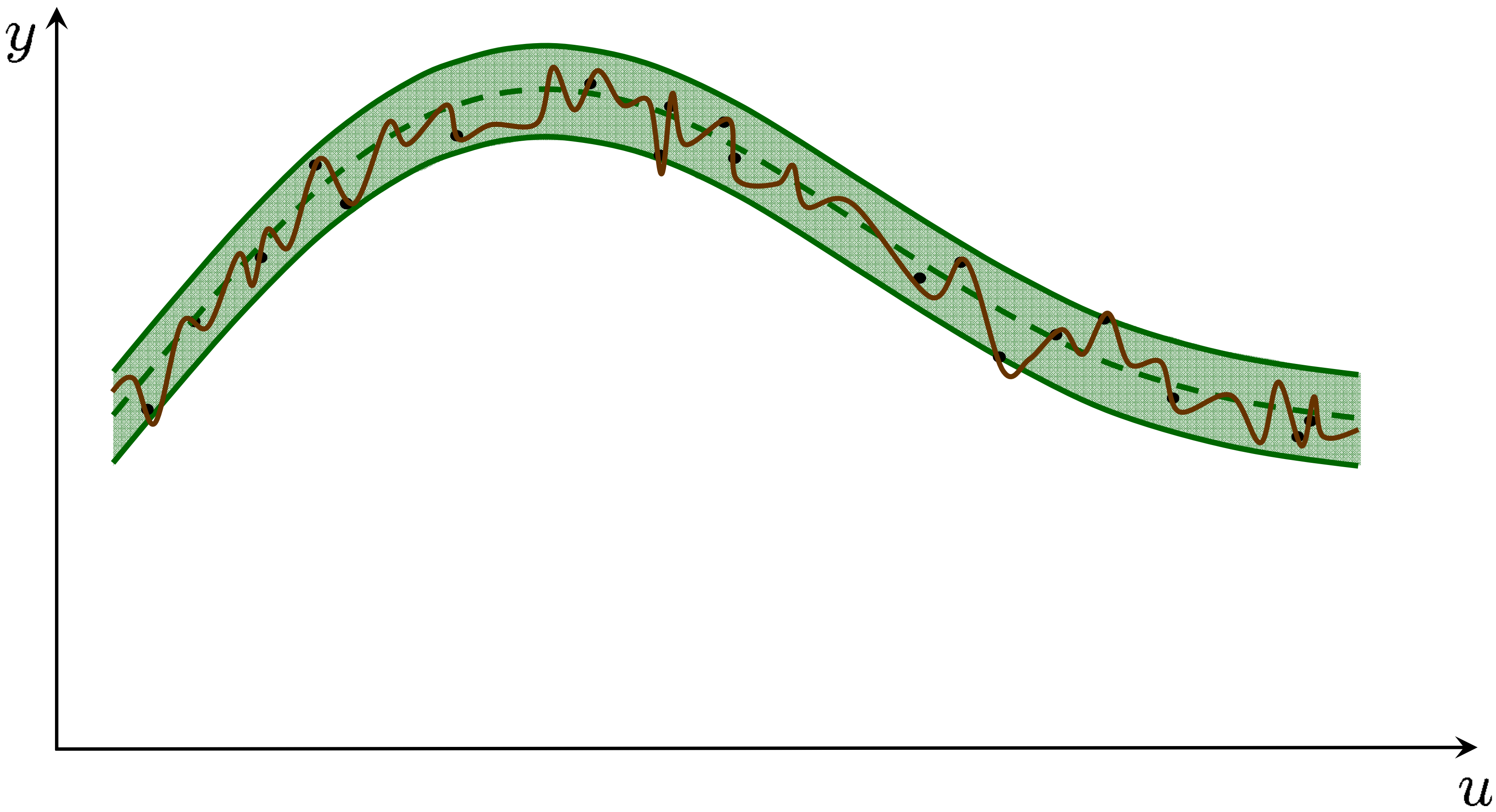
$$\min_{\theta_1,\theta_2,\theta_3,\theta_4} \left[\max_i |y_i - [\theta_1 + \theta_2 u_i + \theta_3 u_i^2 + \theta_4 u_i^3]| \right]$$

… more comments (IPM = Interval Prediction Model)



risk = prob. that next point is outside IPM

… more comments (IPM = Interval Prediction Model)



… more comments

$$N = \frac{2}{\epsilon}\left(\ln\frac{1}{\beta} + d\right)$$

- *N* is independent of *Pr* *(distribution-free result)*

… more comments

$$N = \frac{2}{\epsilon}\left(\ln\frac{1}{\beta} + d\right)$$

- *N* is independent of *Pr*   *(distribution-free result)*

"What I like about experience is that it is such an honest thing. … You may have deceived yourself, but experience is not trying to deceive you."

C.S. Lewis

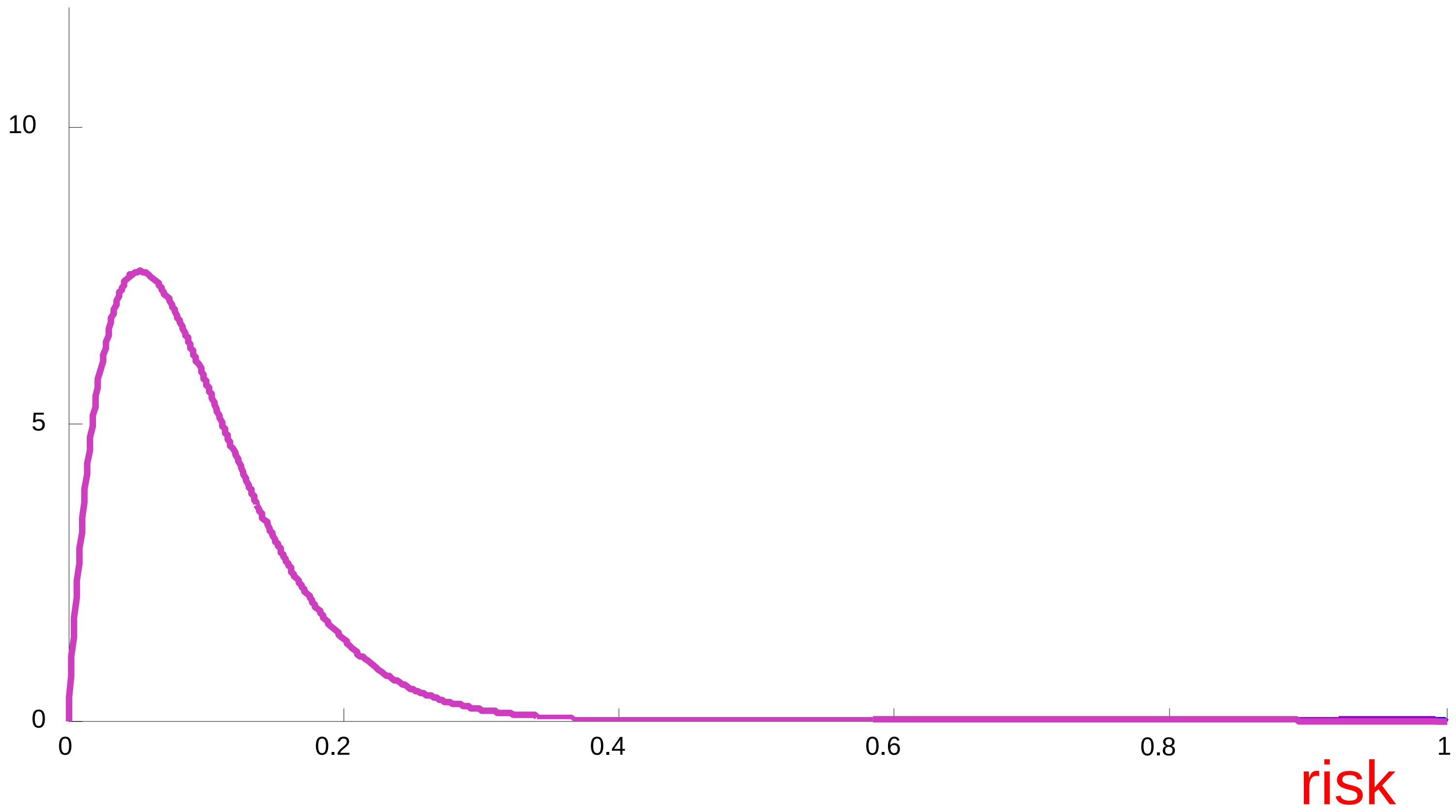
a more general
theoretical result

Theorem (with S. Garatti)

The "risk" has a beta-distribution (universal).

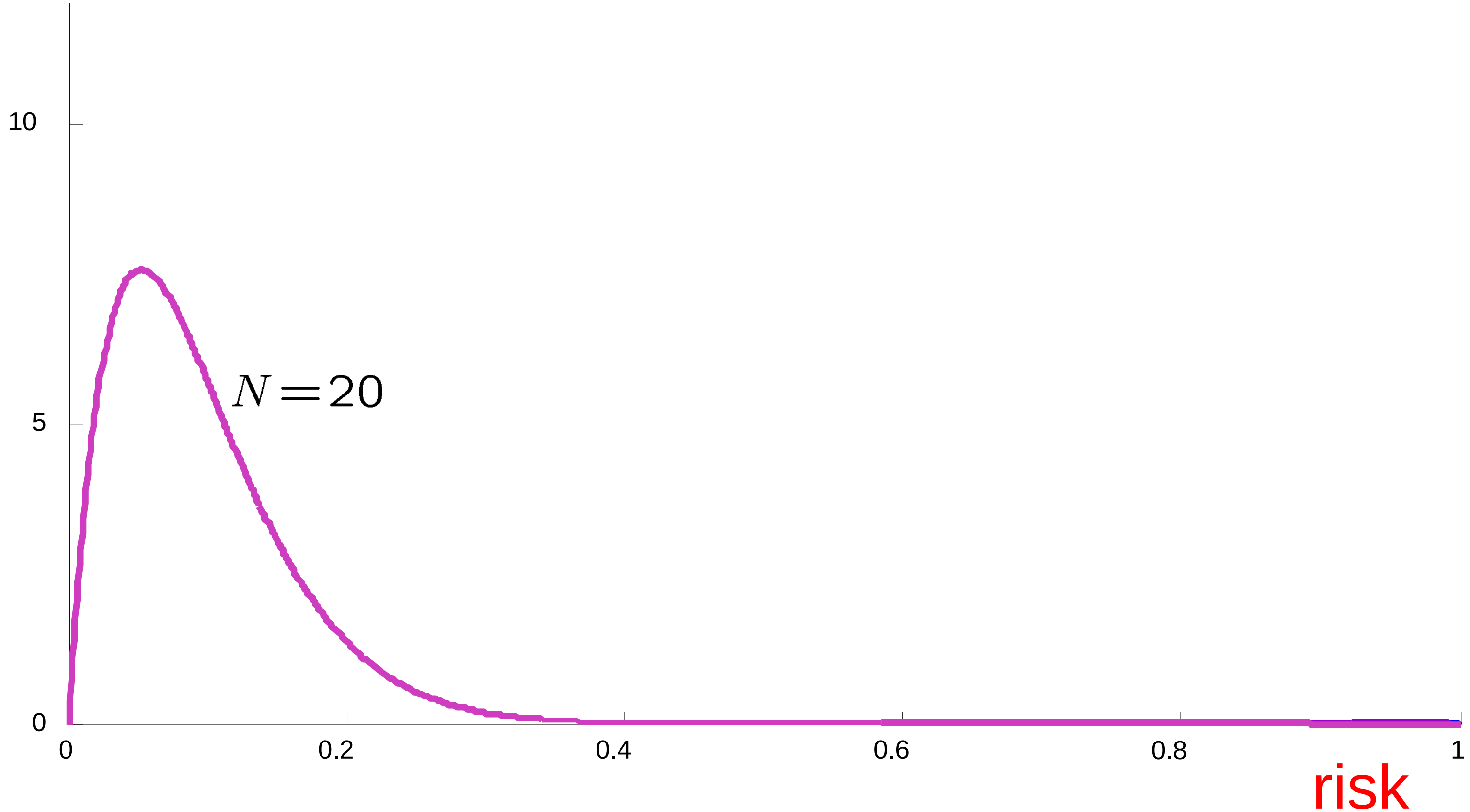
Theorem (with S. Garatti)

The "risk" has a beta-distribution (universal).



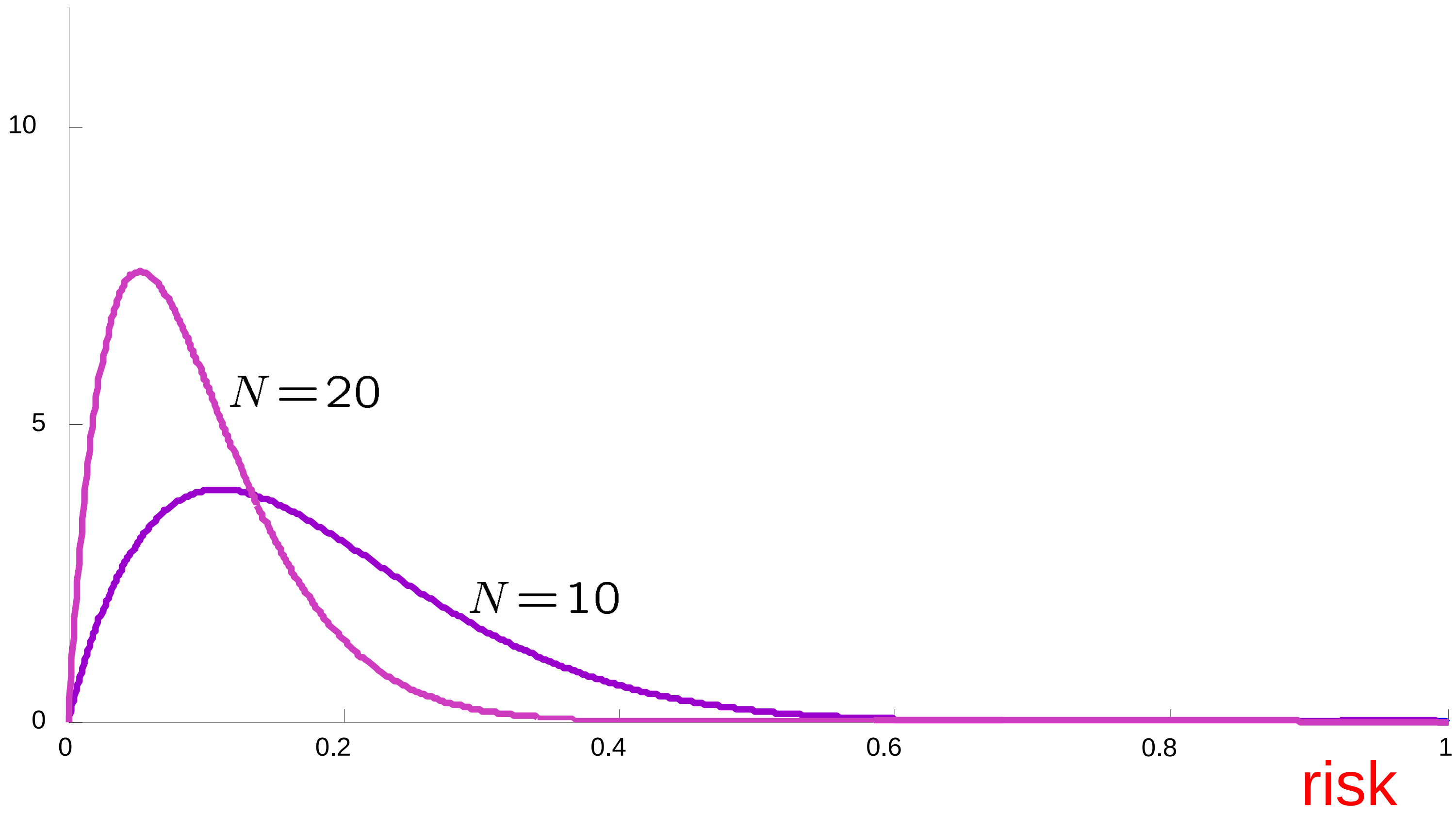
Theorem (with S. Garatti)

The "risk" has a beta-distribution (universal).



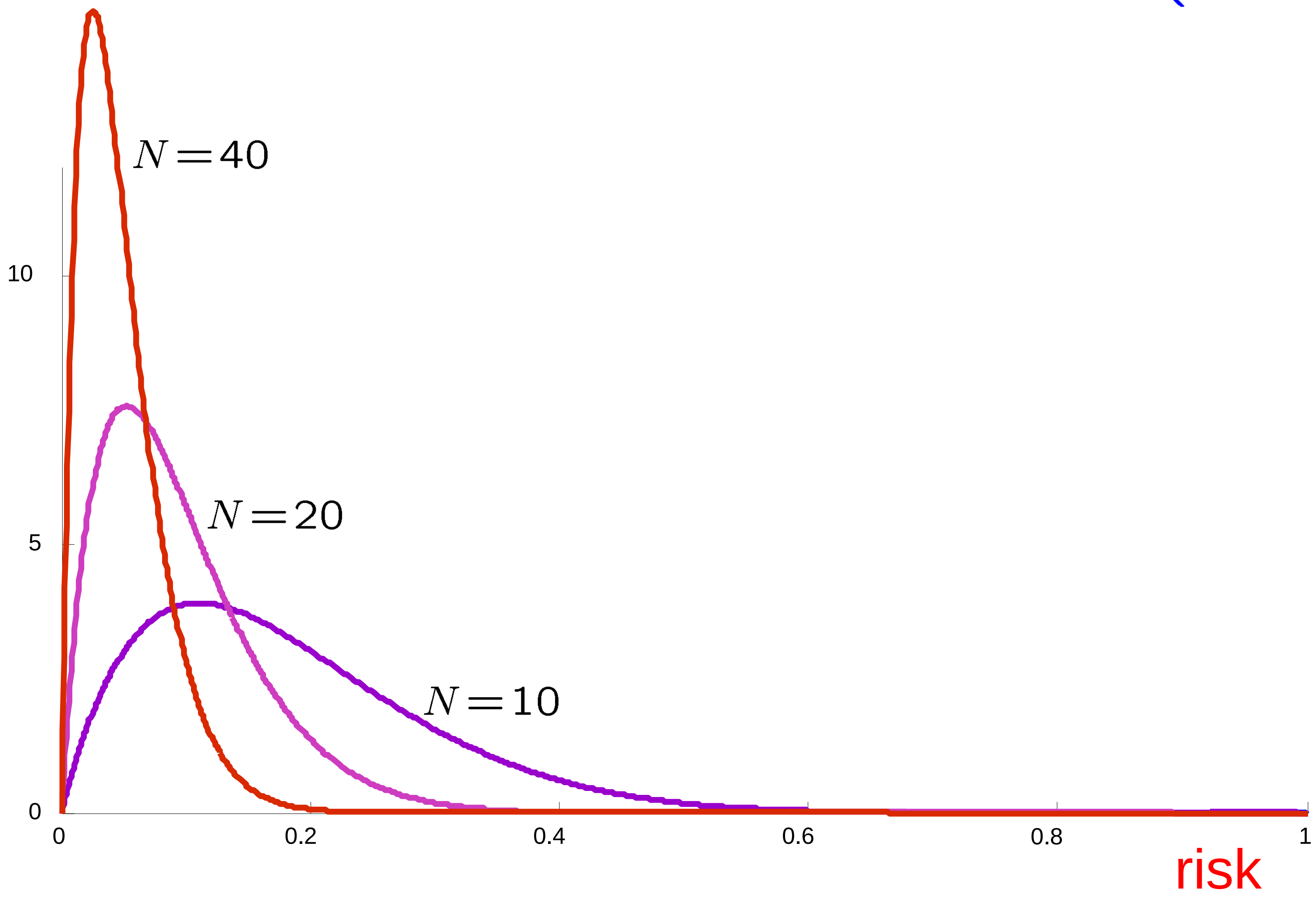
Theorem (with S. Garatti)

The “risk” has a beta-distribution (universal).



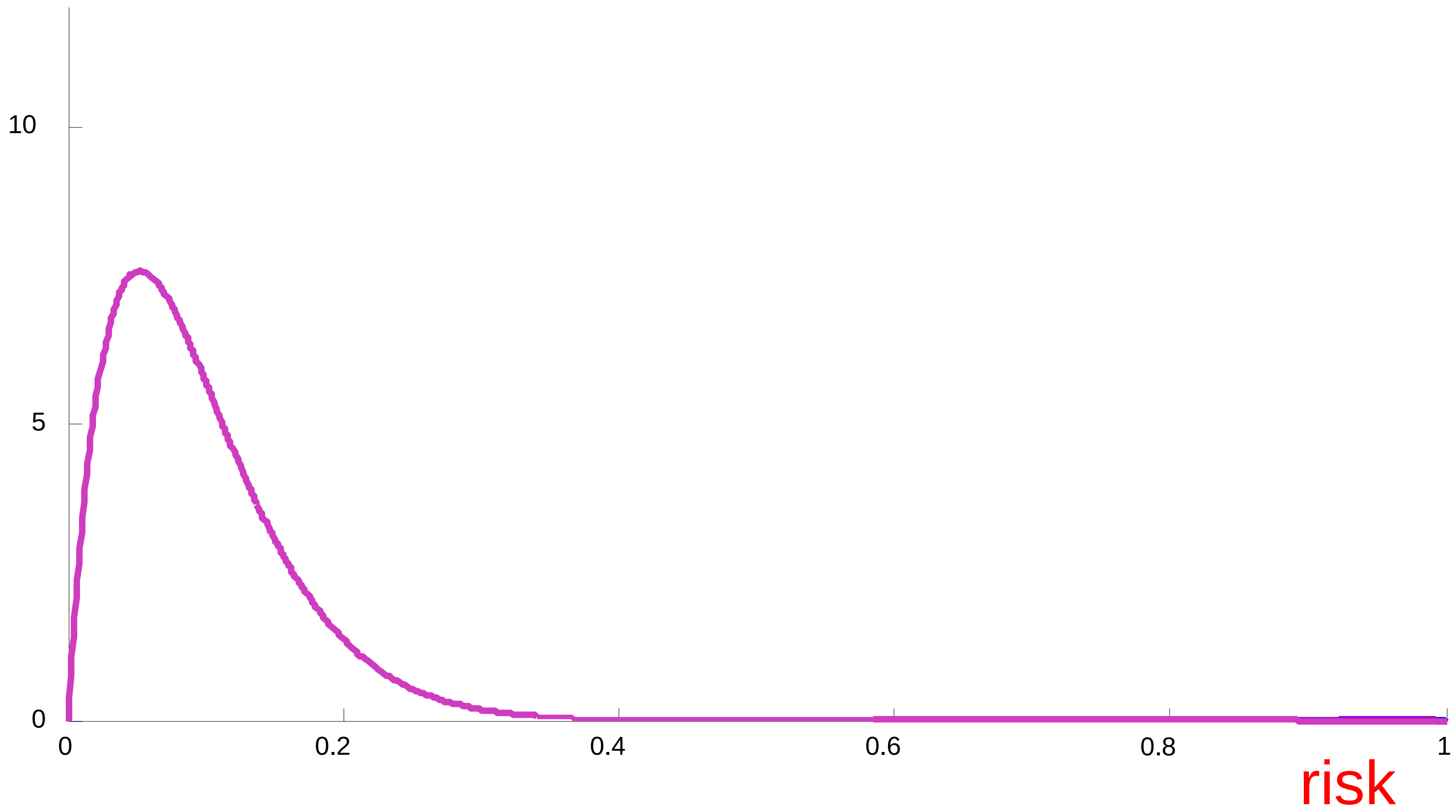
Theorem (with S. Garatti)

The "risk" has a beta-distribution (universal).



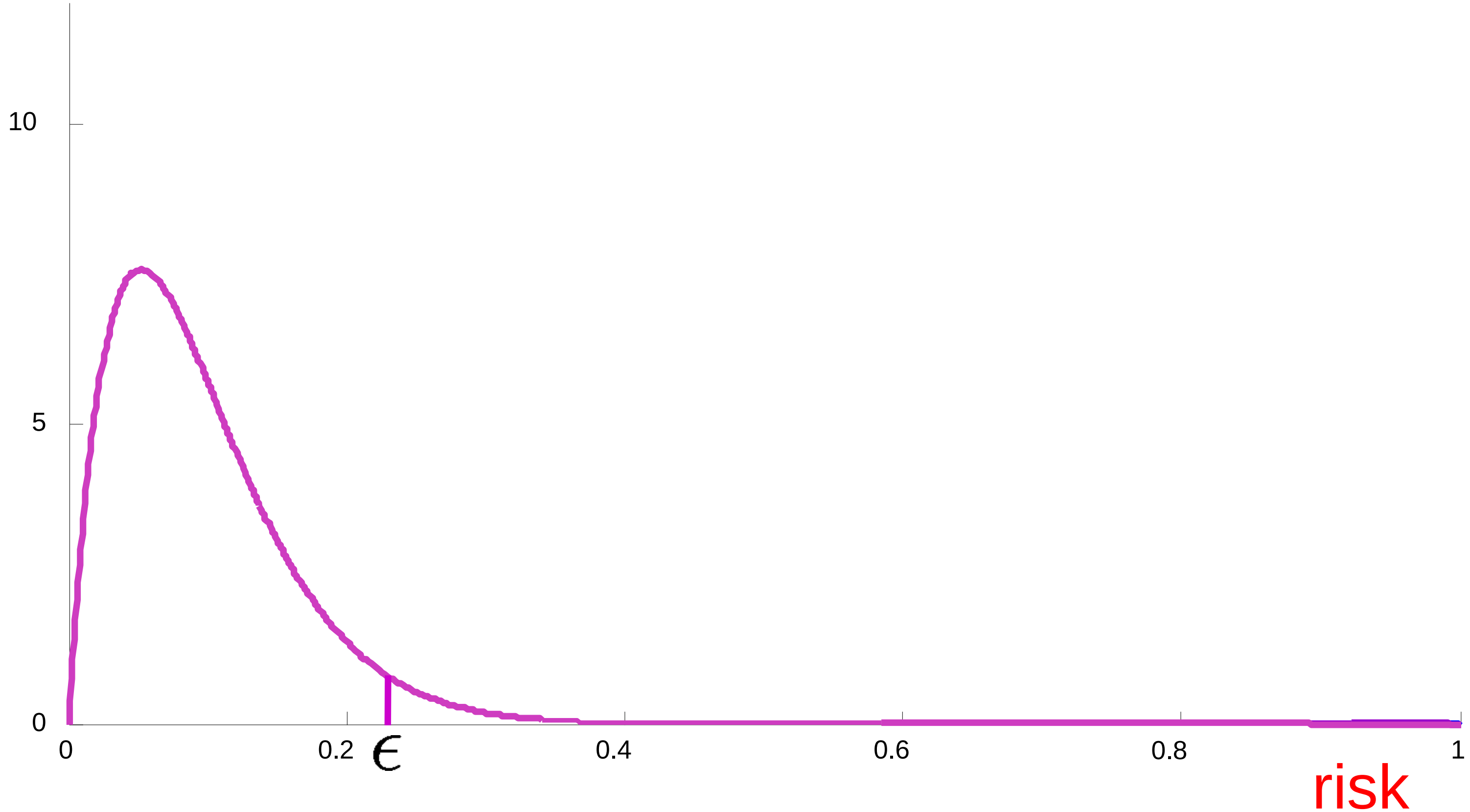
Theorem (with S. Garatti)

The "risk" has a beta-distribution (universal).



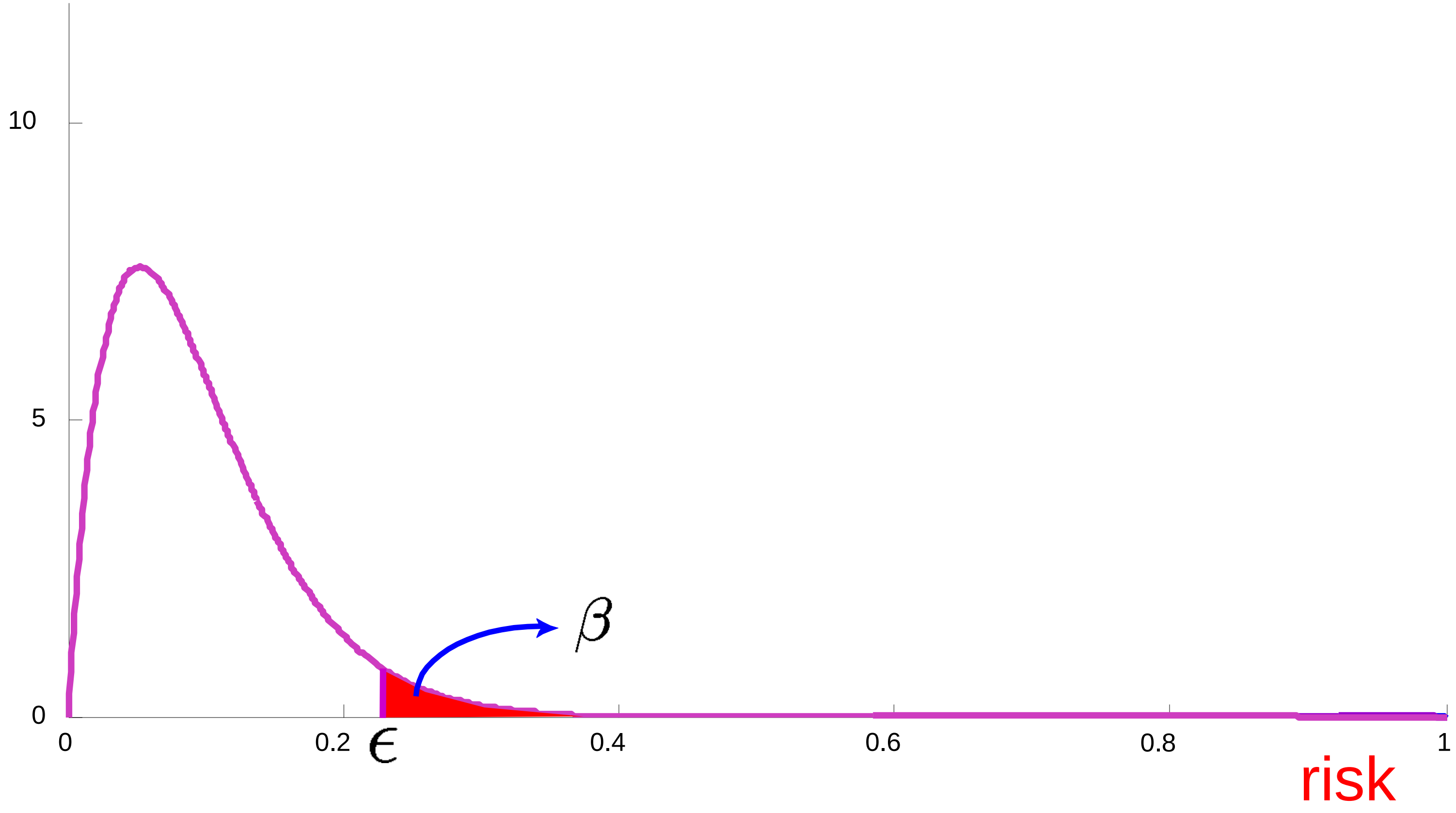
Theorem (with S. Garatti)

The "risk" has a beta-distribution (universal).



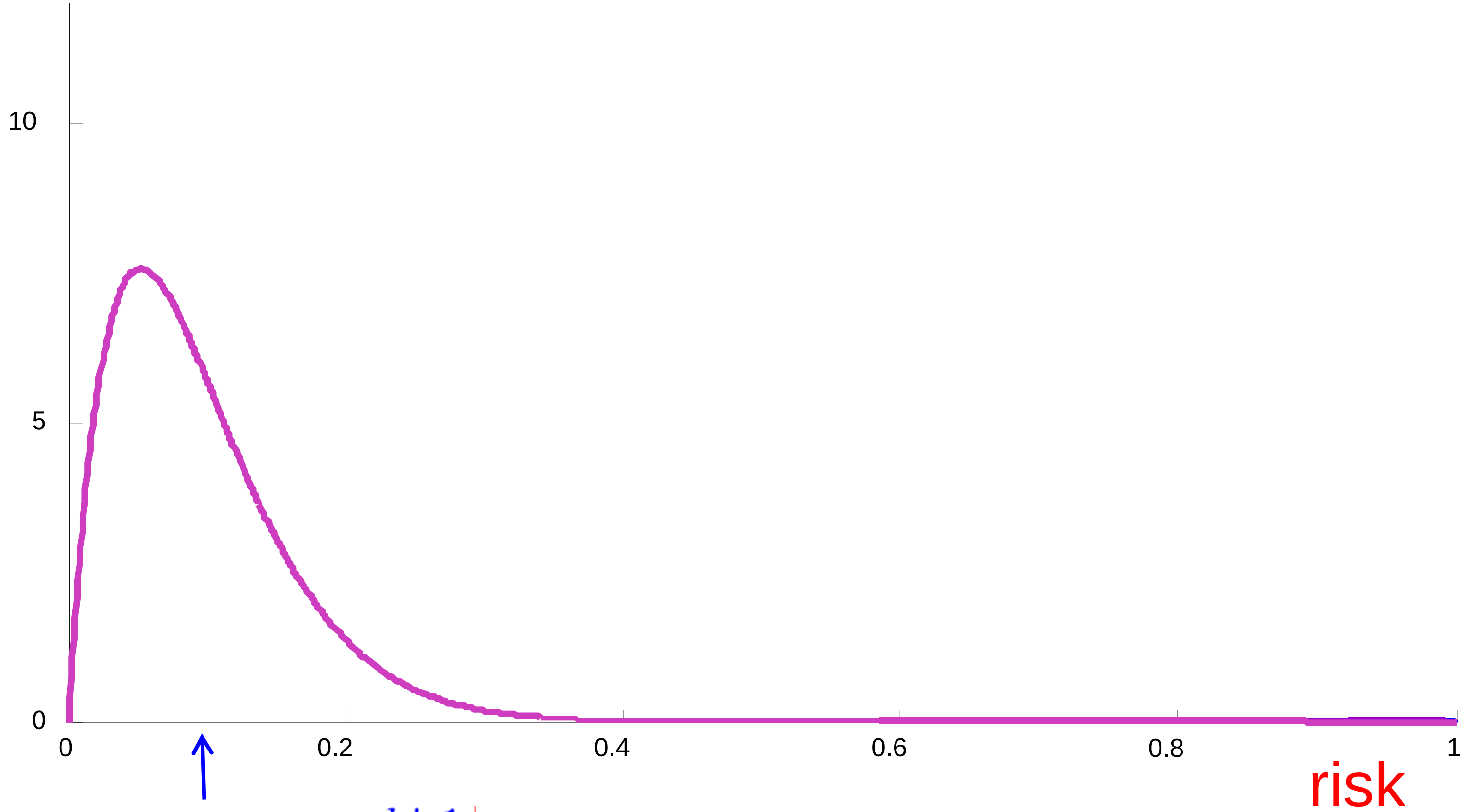
Theorem (with S. Garatti)

The "risk" has a beta-distribution (universal).



Theorem (with S. Garatti)

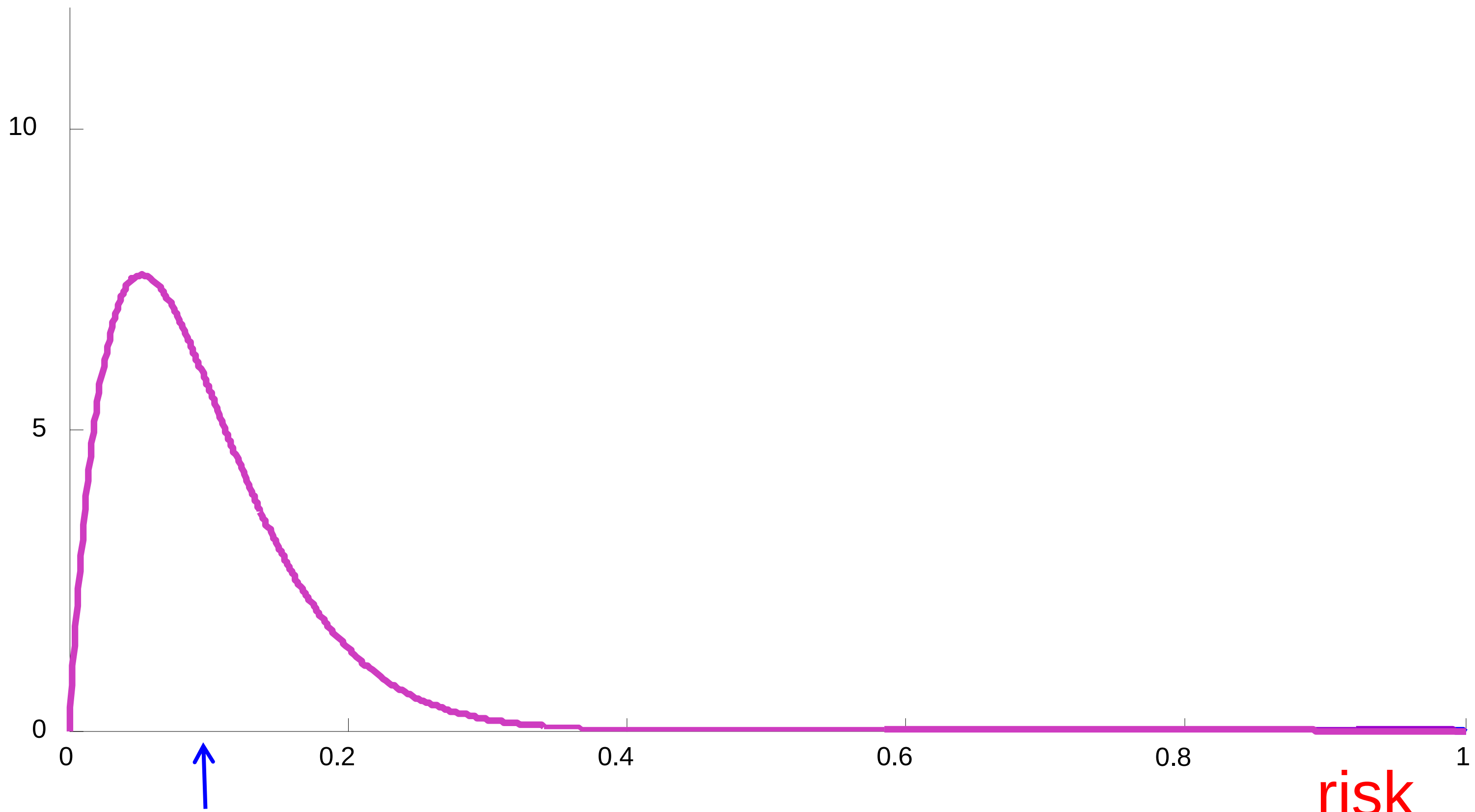
The "risk" has a beta-distribution (universal).



$E[\text{risk}] = \frac{d+1}{N+1}$

Theorem (with S. Garatti)

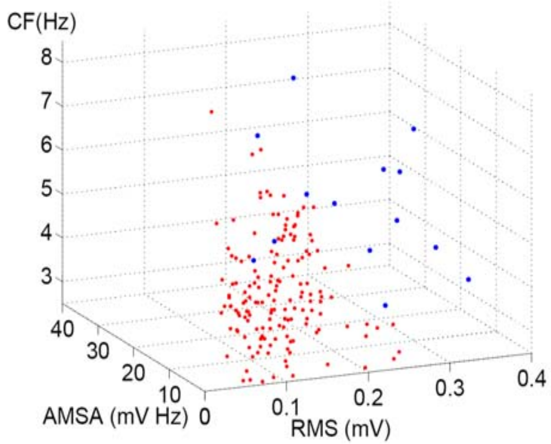
The “risk” has a beta-distribution (universal).



$E[\text{risk}] = \frac{d+1}{N+1}$ = total probability of error

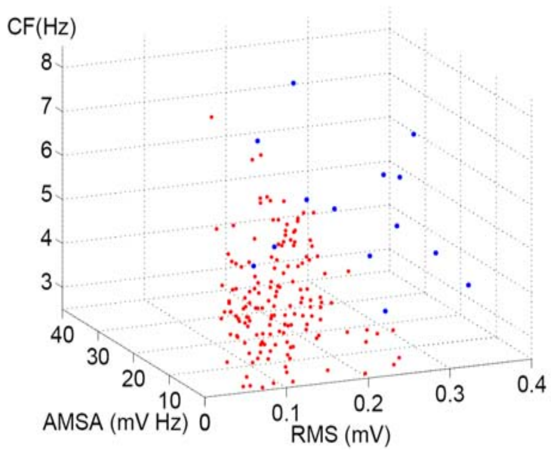
application: classification - defibrillation

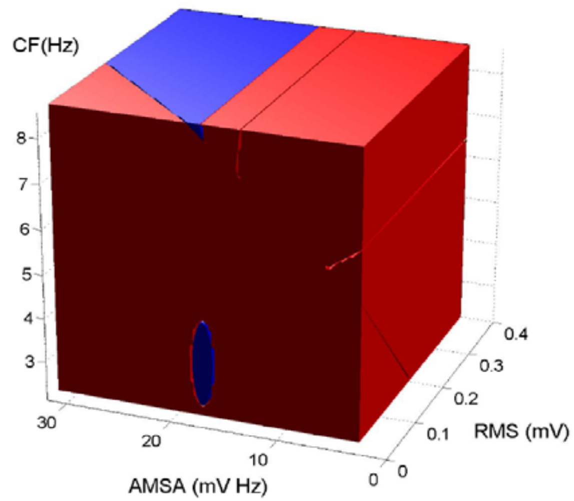
\# Instances = 170
\# NoRosc = 155
\# Rosc = 15



application: classification - defibrillation

$\#$ Instances $= 170$
$\#$ NoRosc $= 155$
$\#$ Rosc $= 15$

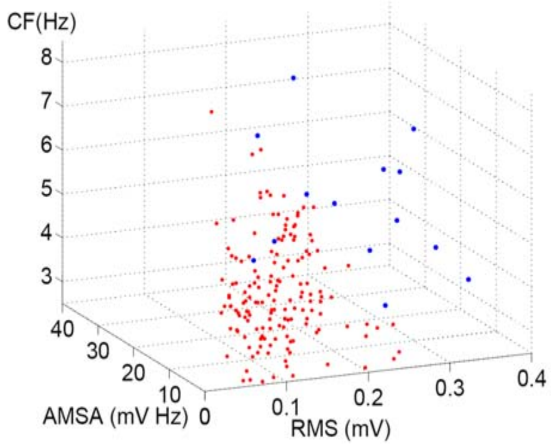


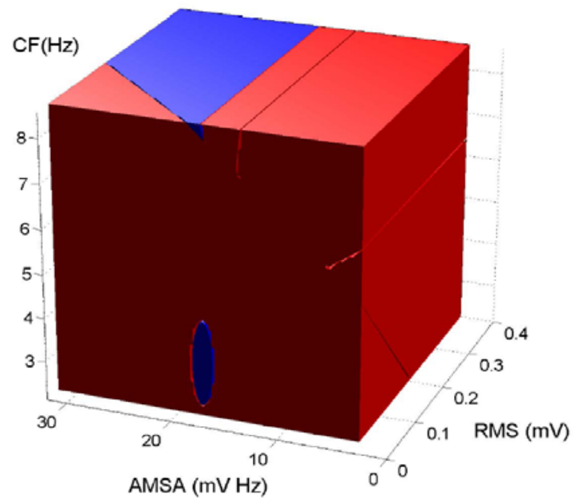


$d =$ complexity of classifier

application: classification - defibrillation

$\#$ Instances $= 170$
$\#$ NoRosc $= 155$
$\#$ Rosc $= 15$



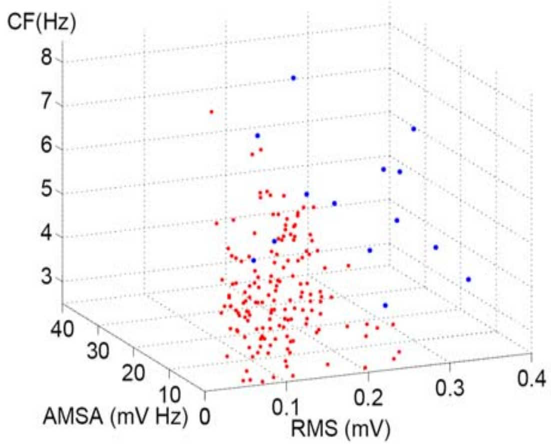


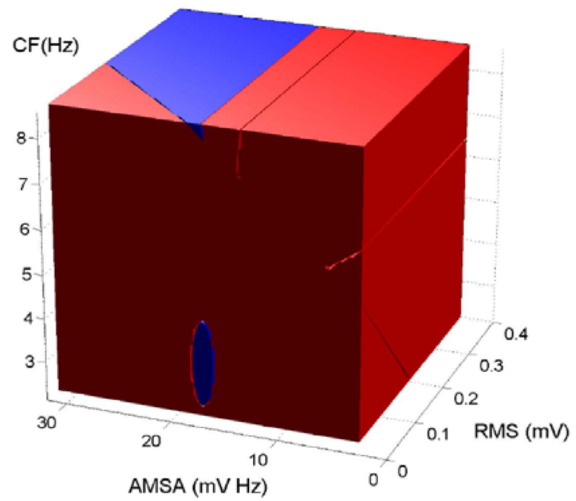
$d =$ complexity of classifier

10-fold cross-validation

application: classification - defibrillation

\# Instances = 170
\# NoRosc = 155
\# Rosc = 15





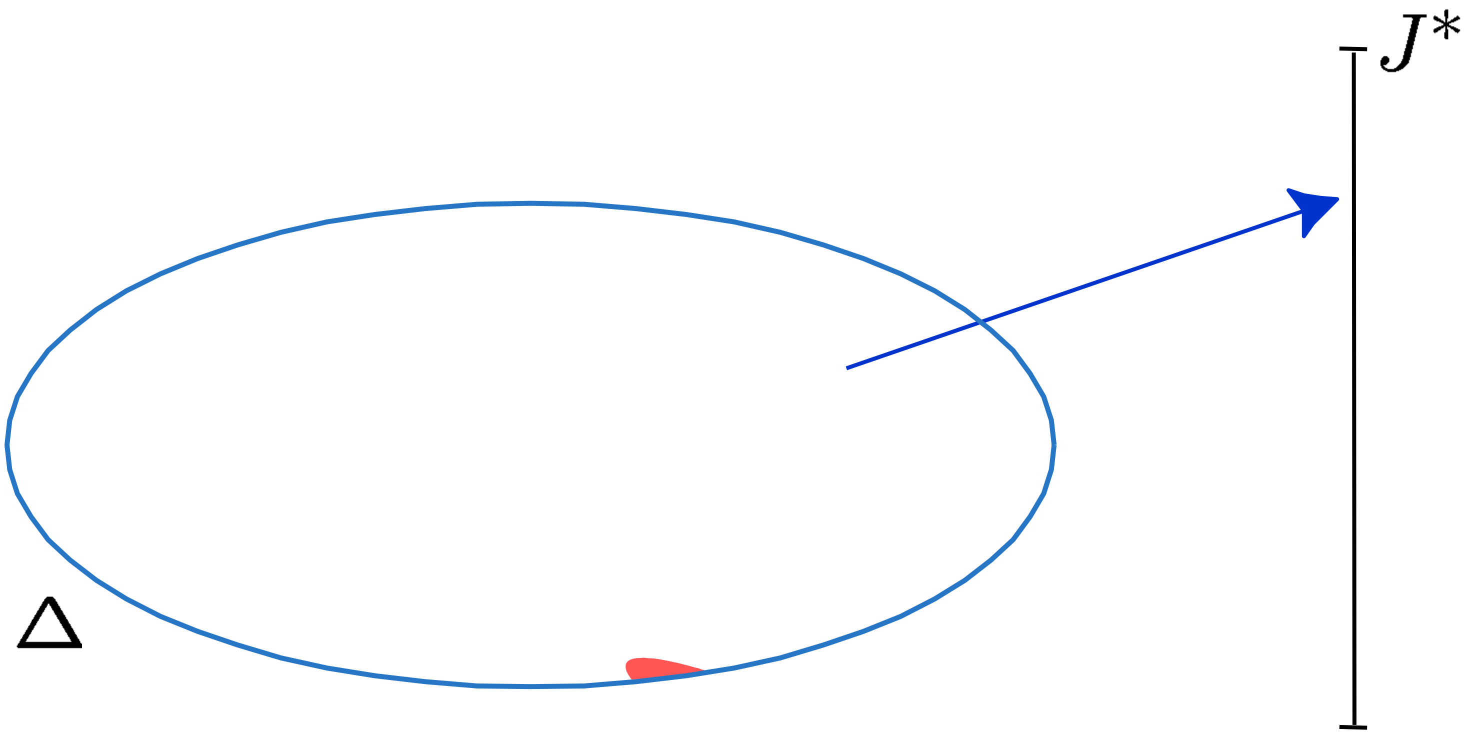
d = complexity of classifier

10-fold cross-validation

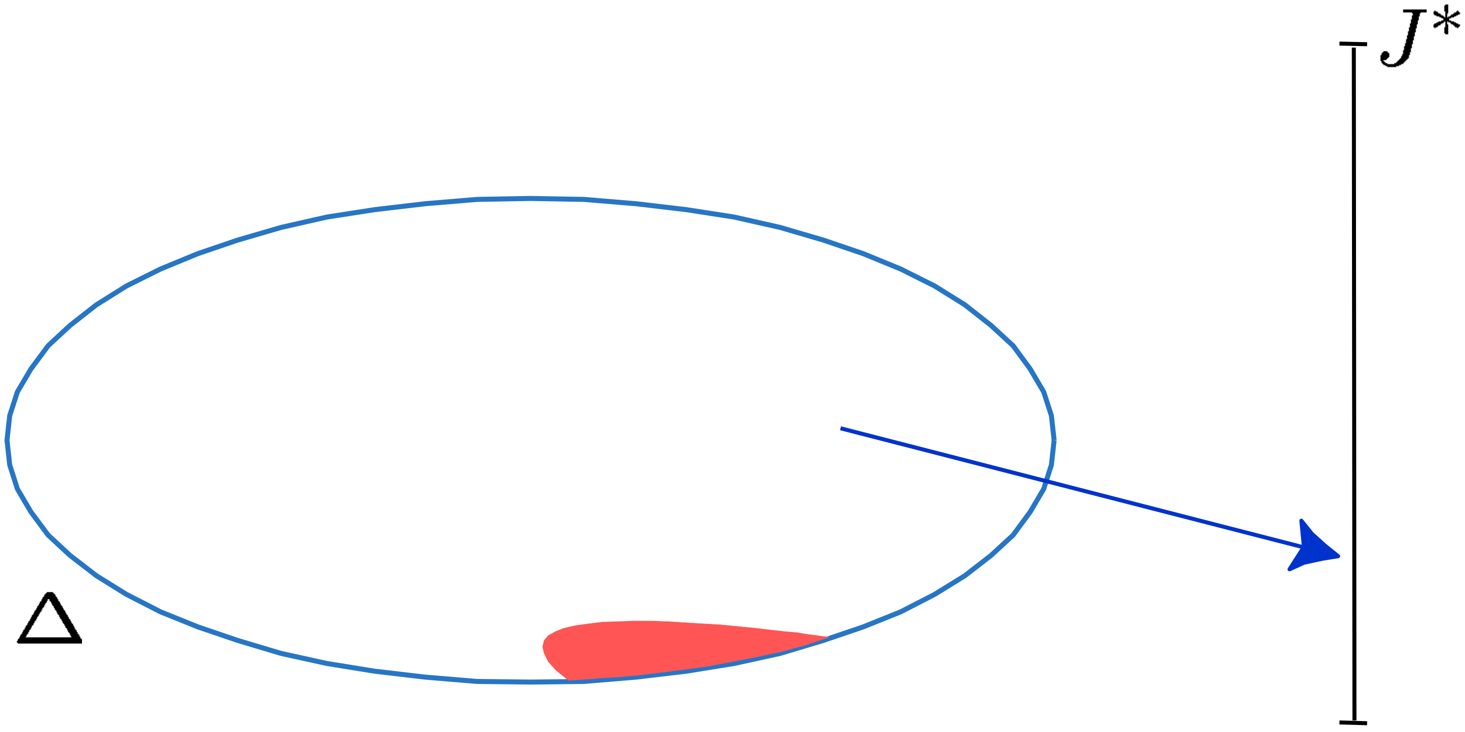
d	3	9	15	21	27
# of errors	4 (2.35%)	10 (5.88%)	18 (10.59%)	23 (13.53%)	27 (15.88%)
# of unknowns	143	132	73	13	0
E[risk]	1.96%	5.88%	9.80%	13.73%	17.65%

generalizations and beyond

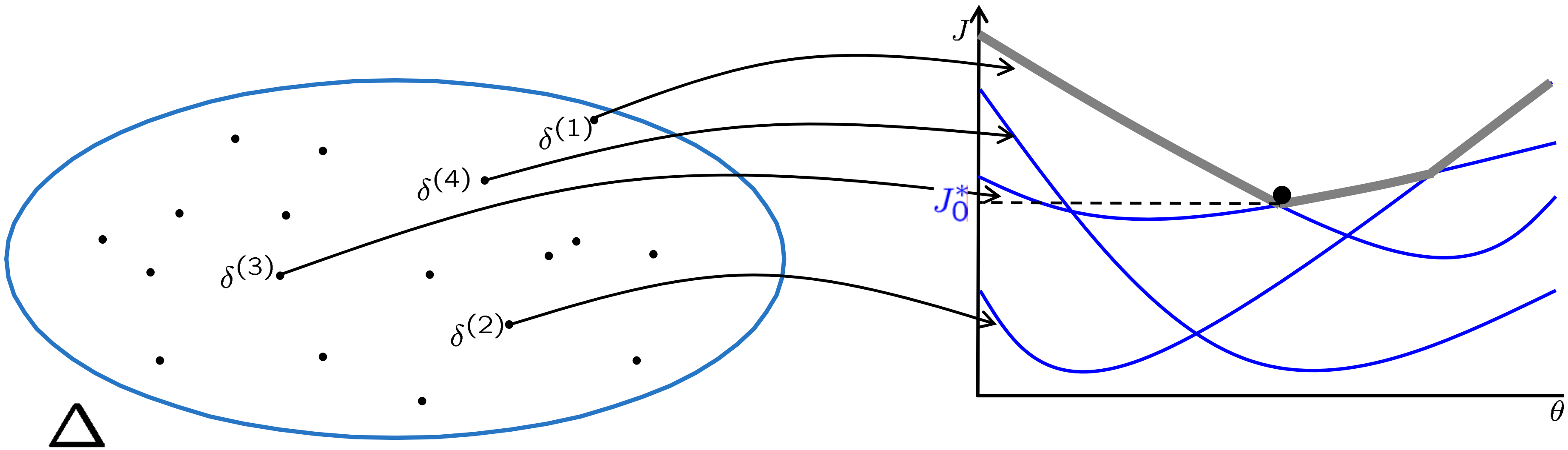
generalizations: risk-return tradeoff



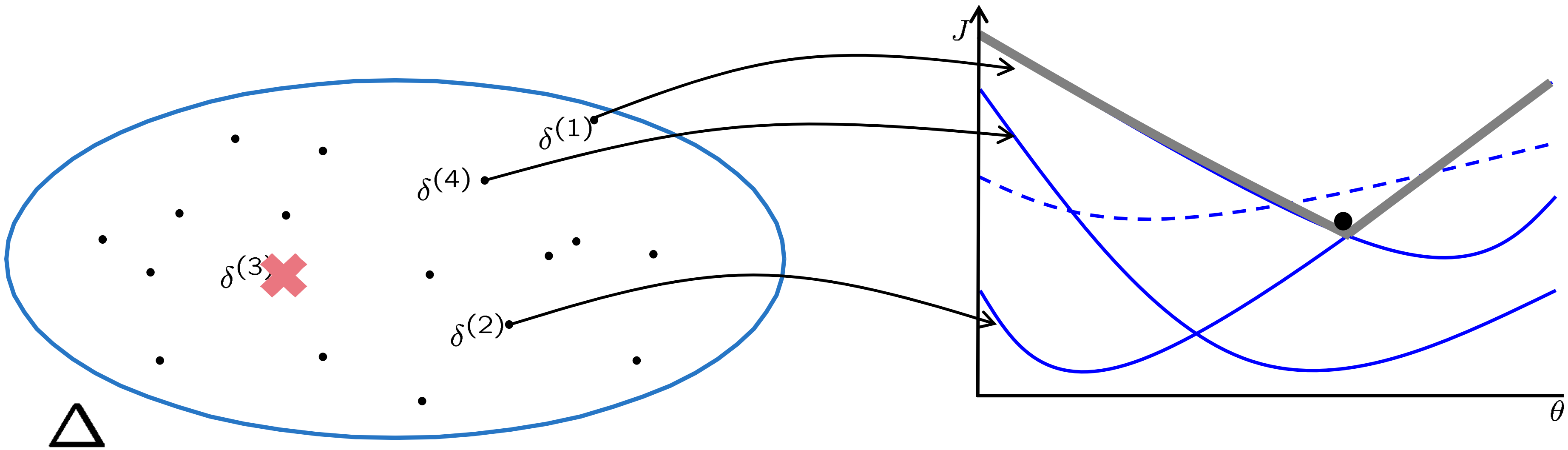
generalizations: risk-return tradeoff



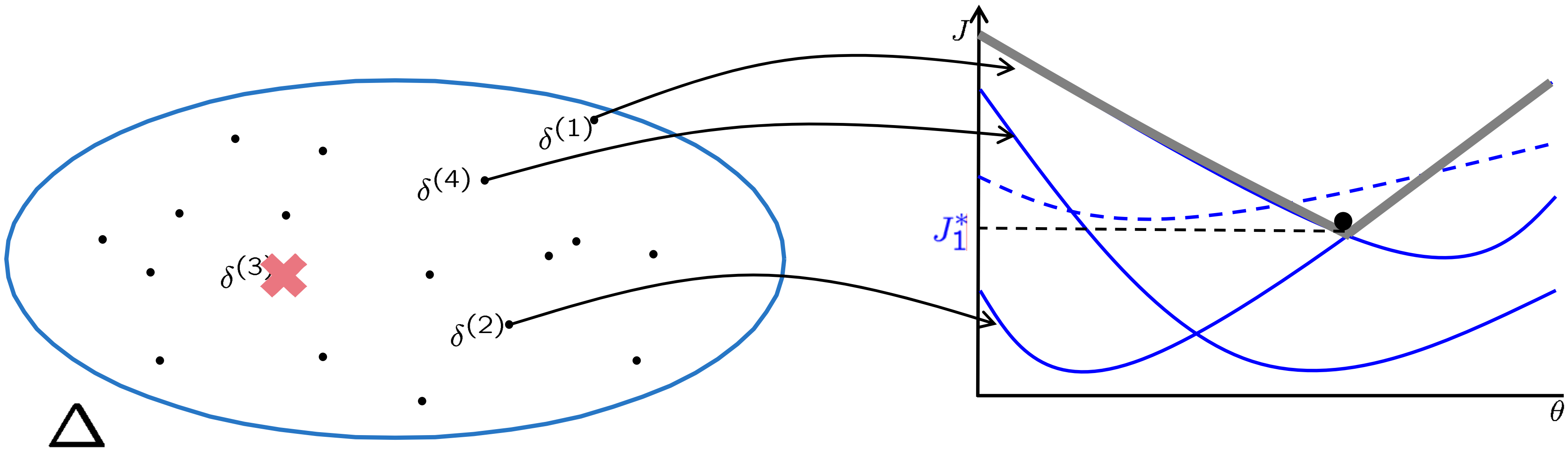
generalizations: risk-return tradeoff



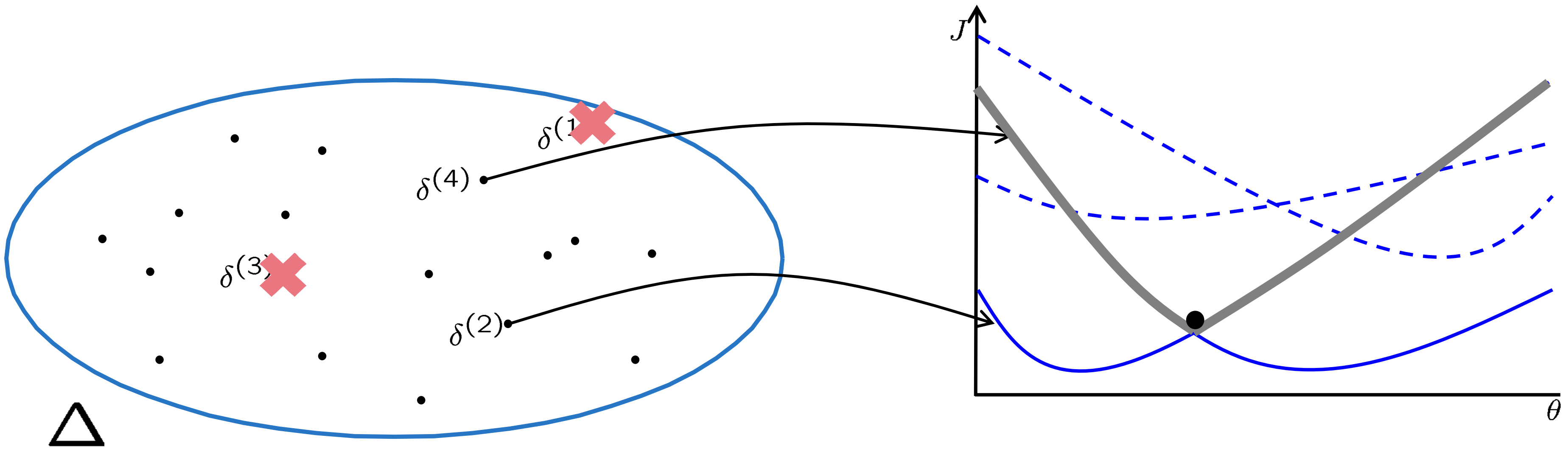
generalizations: risk-return tradeoff



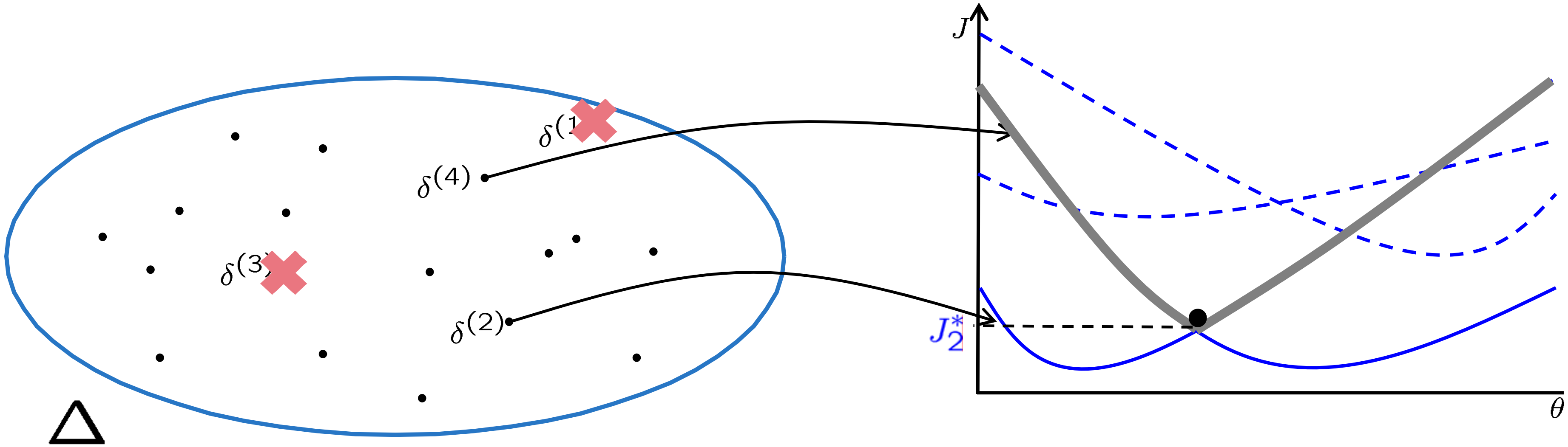
generalizations: risk-return tradeoff

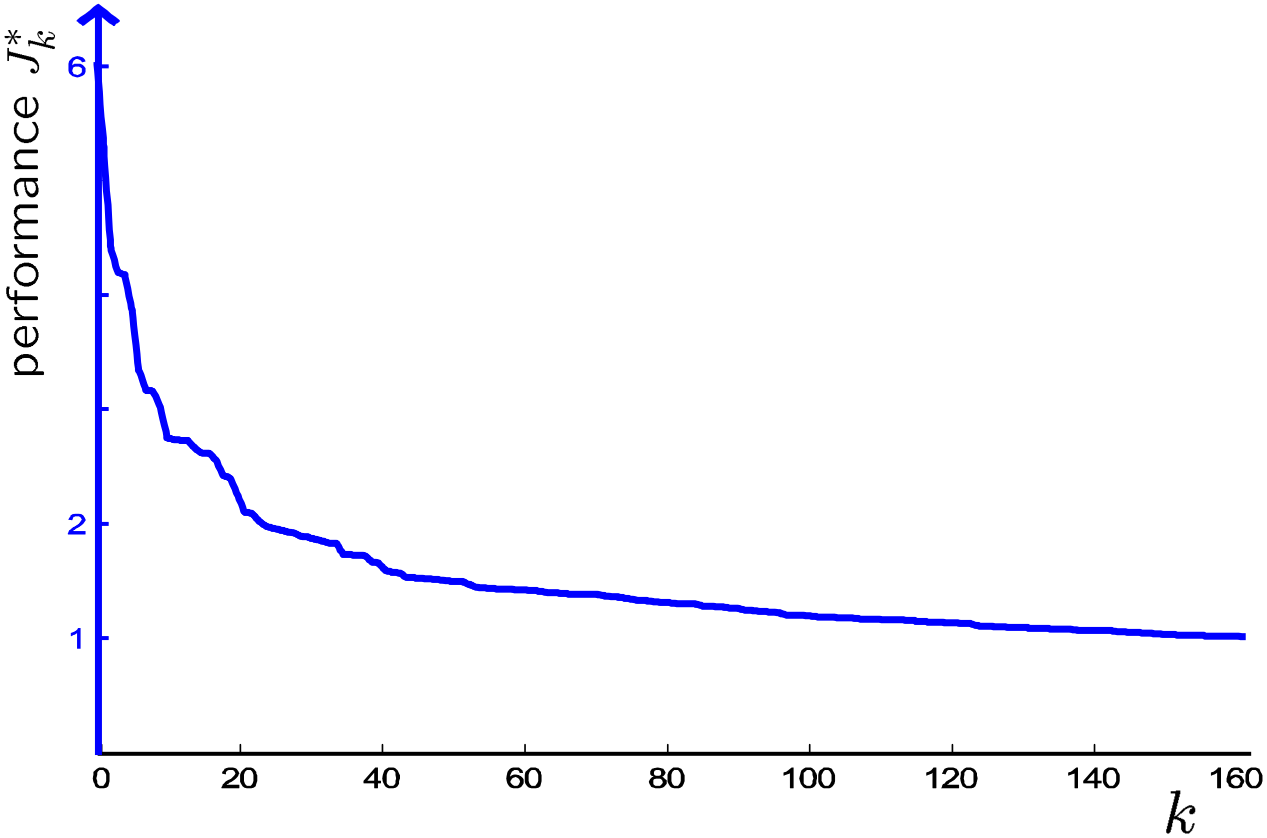


generalizations: risk-return tradeoff



generalizations: risk-return tradeoff





performance J_k^*
6
2
1
0
20
40
60
80
100
120
140
160
k

Theorem (risk-return trade-off)

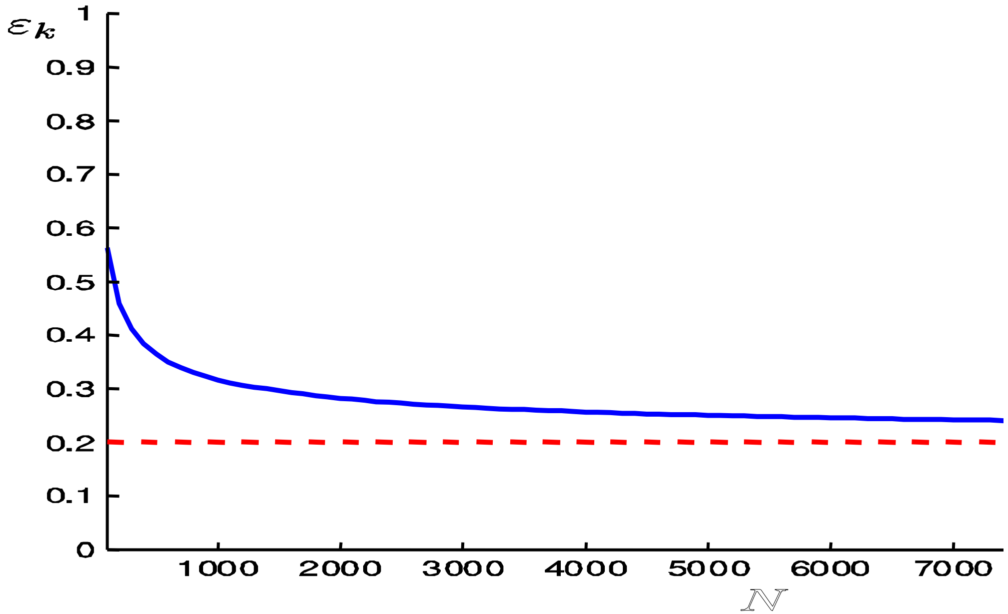
With probability $\geq 1-\beta$,
risk of $J_k^* \leq \epsilon_k$ where:

$$\epsilon_k = \frac{k}{N} + O\left(\frac{1}{\sqrt{N}}\right)$$

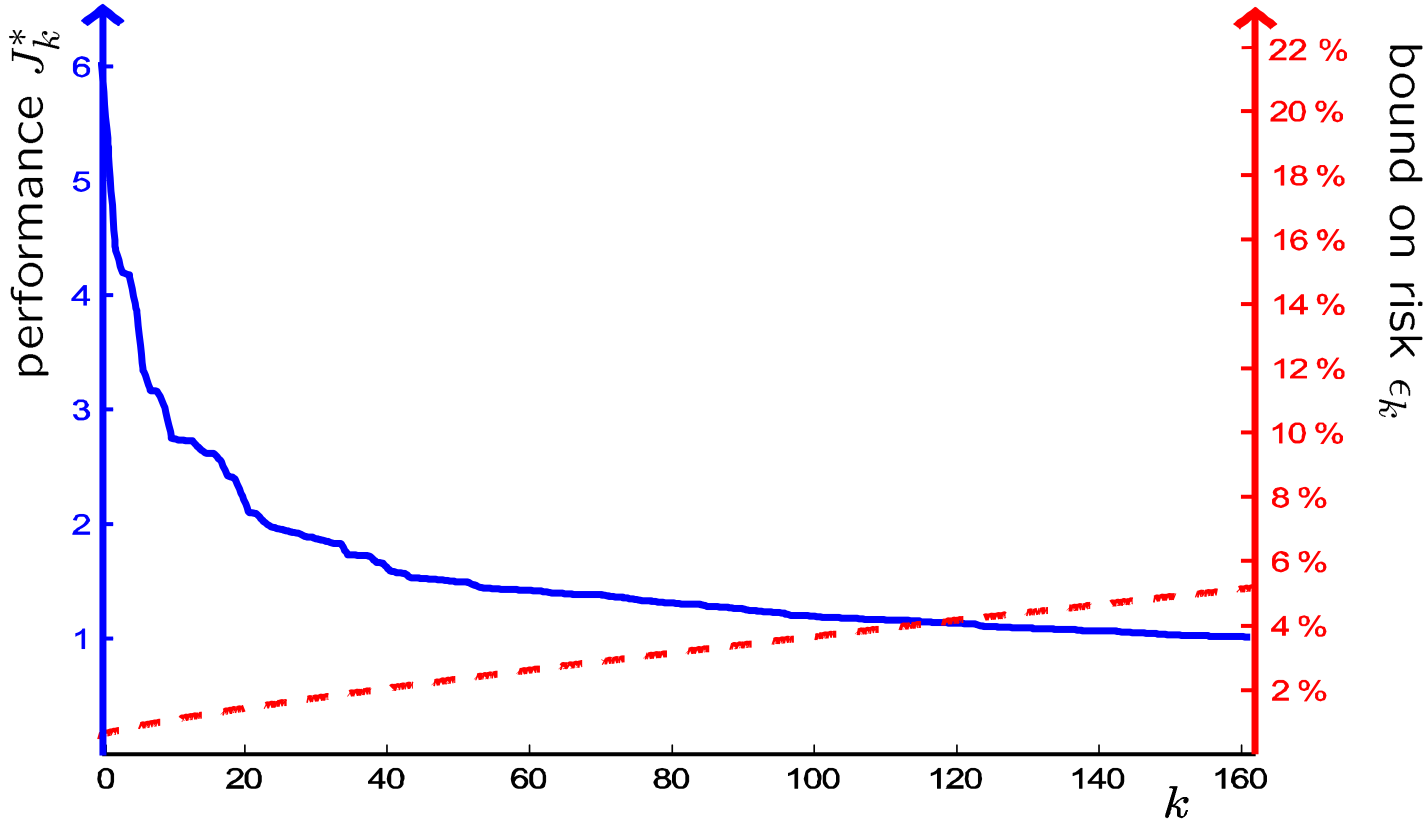
Theorem (risk-return trade-off)

With probability $\geq 1 - \beta$,
risk of $J_k^* \leq \epsilon_k$ where:

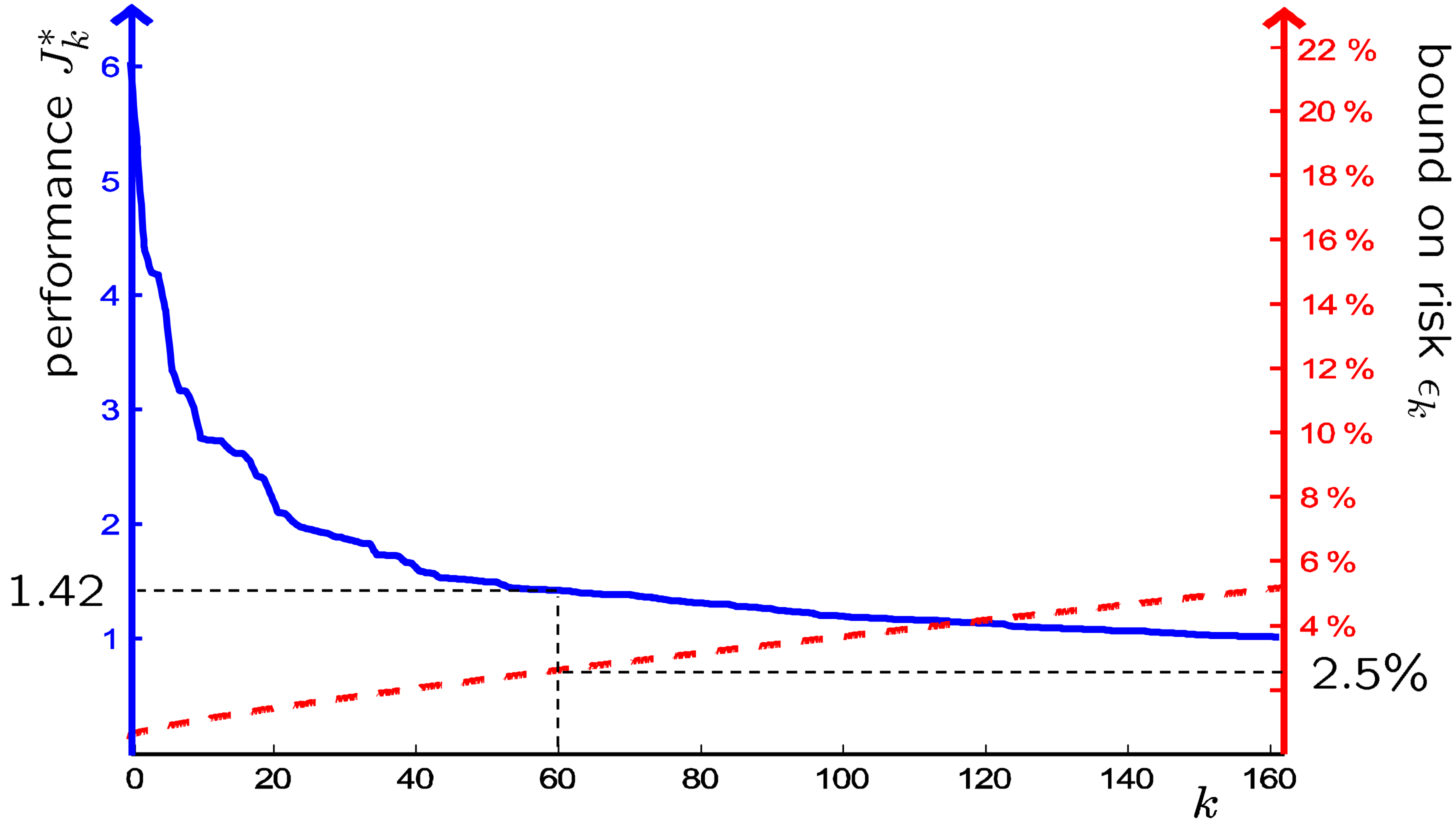
$$\epsilon_k = \frac{k}{N} + O\left(\frac{1}{\sqrt{N}}\right)$$



performance - risk plot



performance - risk plot

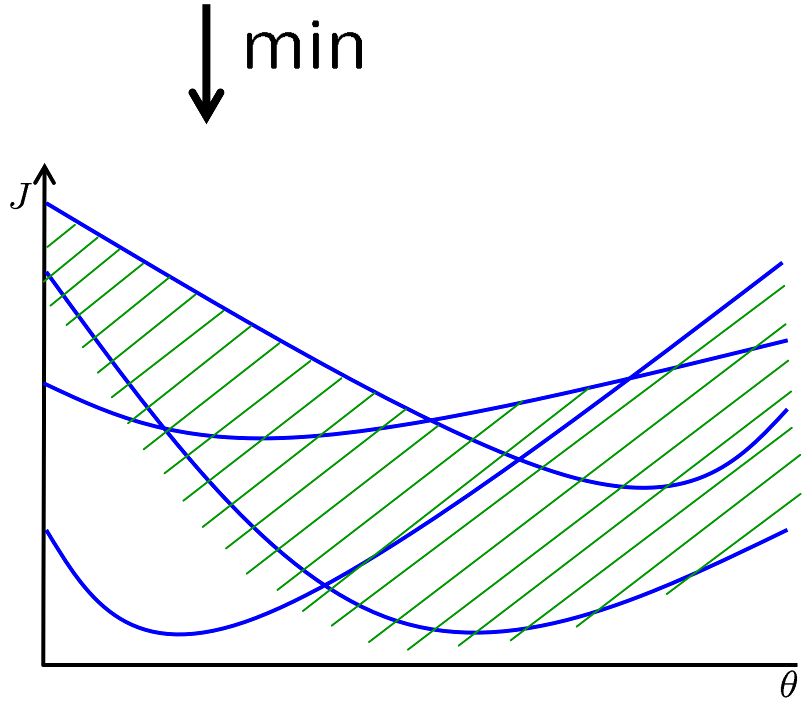


generalizations

$$\min_{\theta} \max_{i} J(\theta, \delta^{(i)})$$

generalizations

$$\min_{\theta} \max_{i} J(\theta, \delta^{(i)})$$



$$\min_{\theta, J} \quad J$$

subject to: $J \geq J(\theta, \delta^{(i)}), \quad i = 1, \ldots, N$

generalizations

$$\min_{\theta} \; c^T\theta$$

$$\text{subject to: } \theta \in \Theta_i, \quad i = 1, \ldots, N$$

generalizations

$$\min_{\theta} \ c^T\theta$$
$$\text{subject to: } \theta \in \Theta_i, \quad i = 1, \ldots, N$$

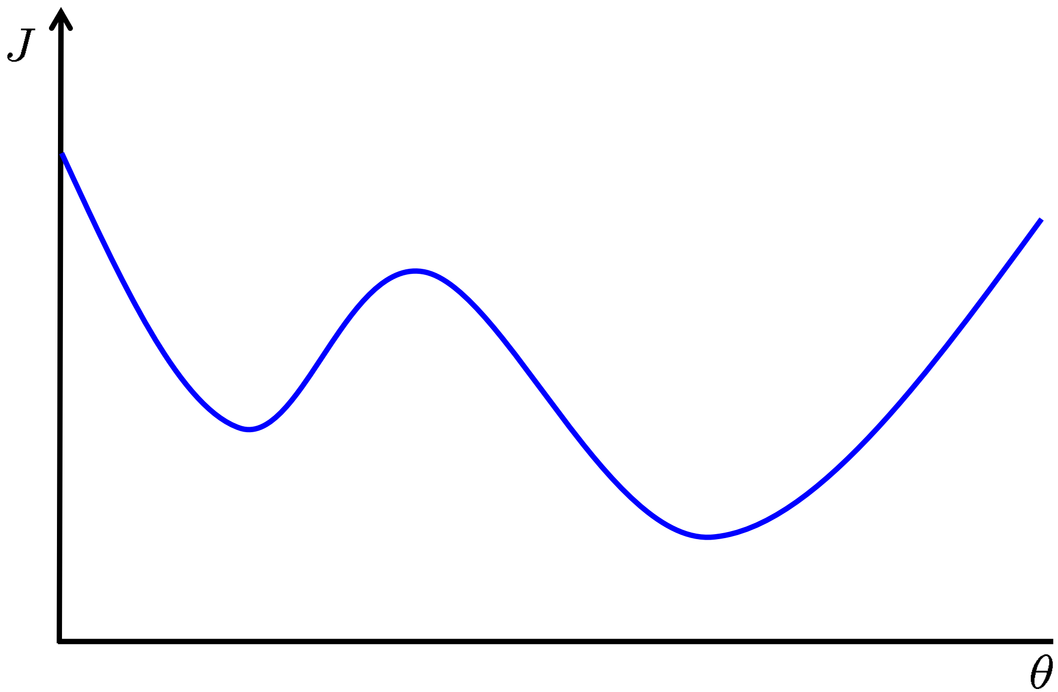
relevanto to:
- quantitative finance (minimum return)
- control with constraints (MPC)

⋮

Many have given a contribution:

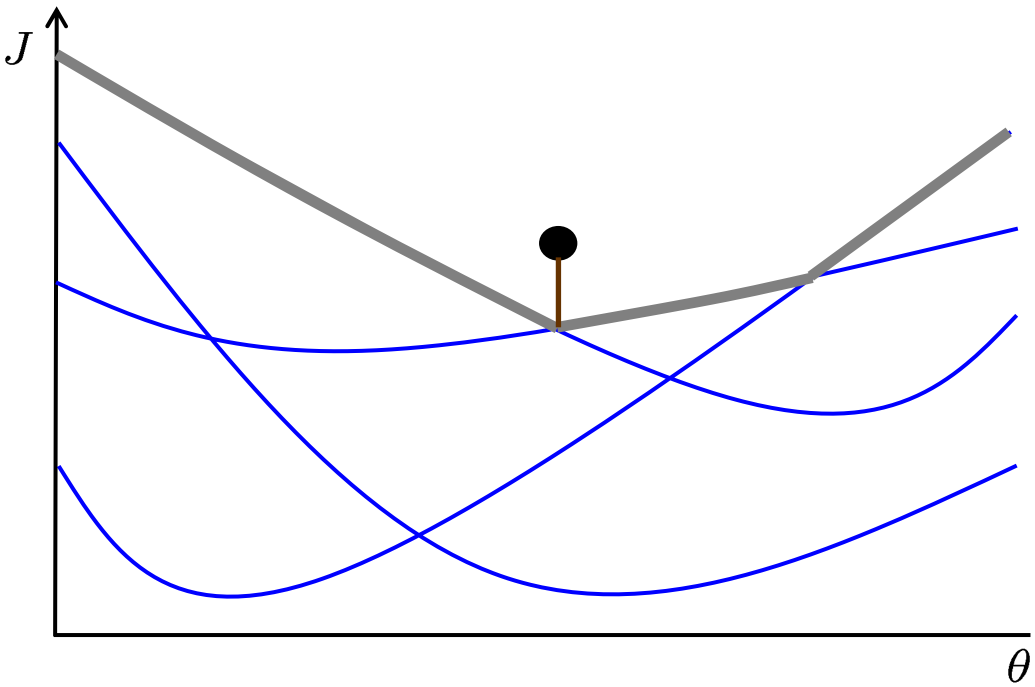
T. Álamo, G. Andersson, F. Borrelli, G. Calafiore, A. Carè, F. Dabbene, L. Fagiano, S. Garatti, P. Goulart, O. Granichin, J. Hespanha, H. Hjalmarsson, G. Iyengar, T. Kanamori, D. Kuhn, C. Lagoa, J. Luedtke, A. Luque, J. Lygeros, K. Margellos, M. Morari, J. Matuško, B. Pagnoncelli, M. Prandini, F. Ramponi, D, Reich, C. Rojas, B. Rustem, G. Schildbach, M. Sznaier, A. Takeda, R. Tempo, B. Van Parys, P. Vayanos, M. Vrakopoulou, J. Welsh, W. Wiesemann the theory is still in its infancy

the theory is still in its infancy



non-convex

the theory is still in its infancy



margin

… concluding

the problem of extracting knowledge from observations is perhaps the most central issue of all science

… concluding

the problem of extracting knowledge from observations is perhaps the most central issue of all science

the scenario approach is one way, and a lot of work remains to be done

… concluding

the problem of extracting knowledge from observations is perhaps the most central issue of all science

the scenario approach is one way, and a lot of work remains to be done

certainly: it is a wonderful world to explore!

… concluding

the problem of extracting knowledge from observations is perhaps the most central issue of all science

the scenario approach is one way, and a lot of work remains to be done

certainly: it is a wonderful world to explore!

THANK YOU!

REFERENCES

M.C. Campi and S. Garatti.
The Exact Feasibility of Randomized Solutions of Uncertain Convex Programs.
SIAM J. on Optimization, 19, no.3: 1211-1230, 2008.

M.C. Campi and S. Garatti.
A Sampling-and-Discarding Approach to Chance-Constrained Optimization: Feasibility and Optimality.
J. of Optimization Theory and Applications, 148: 257-280, 2011.

B.K. Pagnoncelli, D. Reich and M.C. Campi
Risk-Return Trade-off with the Scenario Approach: A Case study in Portfolio Selection.
J. Of Optimization Theory and Applications, 155: 707-722, 2012.

G. Calafiore and M.C. Campi.
Uncertain Convex Programs: randomized Solutions and Confidence Levels.
Mathematical Programming, 102: 25-46, 2005.

G. Calafiore and M.C. Campi.
The Scenario Approach to Robust Control Design.
IEEE Trans. on Automatic Control, AC-51: 742-753, 2006.

M.C. Campi, G. Calafiore and S. Garatti.
Interval Predictor Models: Identification and Reliability.
Automatica, 45: 382-392, 2009.

M.C. Campi.
Classification with guaranteed probability of error.
Machine Learning, 80: 63-84, 2010.