Introduction to Theory of Safe Decision Making

Dr. Akhil Anand

1st AID Scientific Workshop, Trondheim

Forewords & Disclaimer

Objectives:

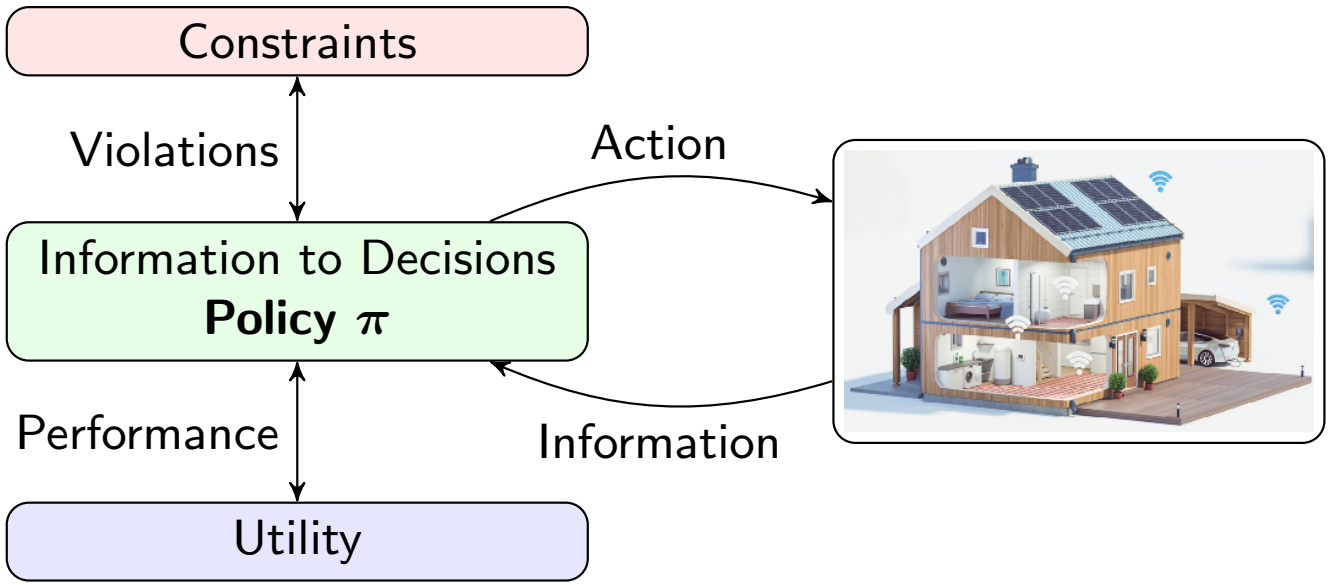
- Put in place some common concepts & language
- Identify some key points in safe decision making
- Connect to AI

Disclaimer: we are a broad group who needs to get to know each others scientifically.
Apologies if I don't "hit" the right level for all.
I have favored simplicity over "absolute" rigor.

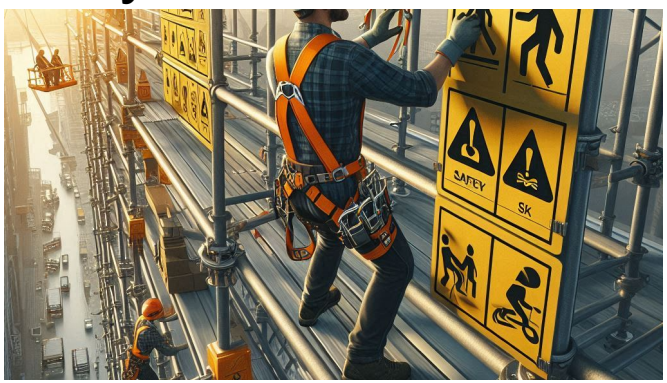
Outline

1. Some Basics of Safe Decision Making
2. Methods
3. Safe Decisions from Data & AI
4. Epistemic Uncertainty and Safe Decisions

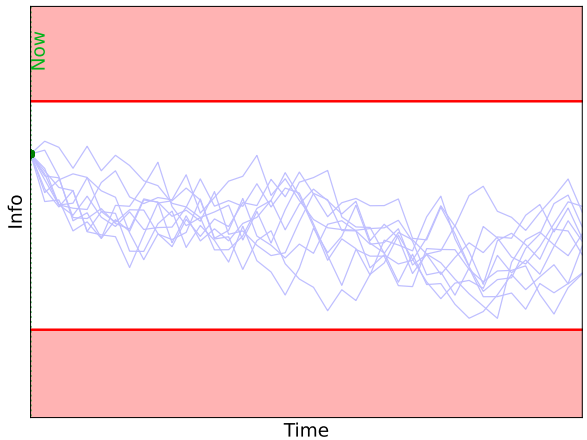
Formally Defining Safety?



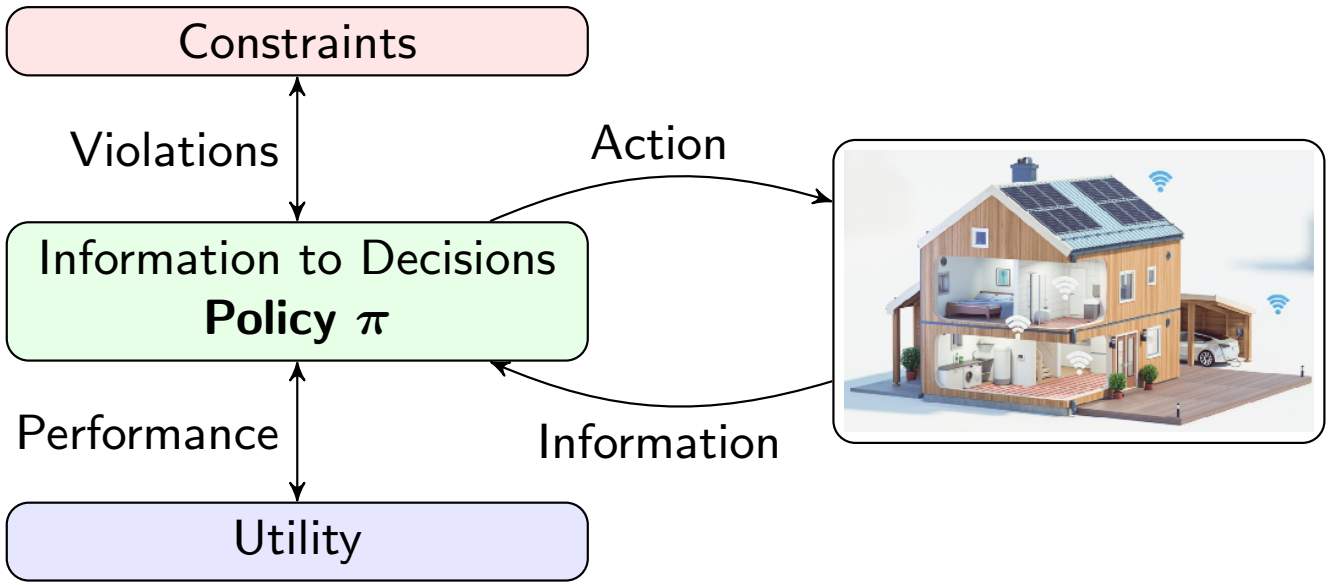
Safety in the real world



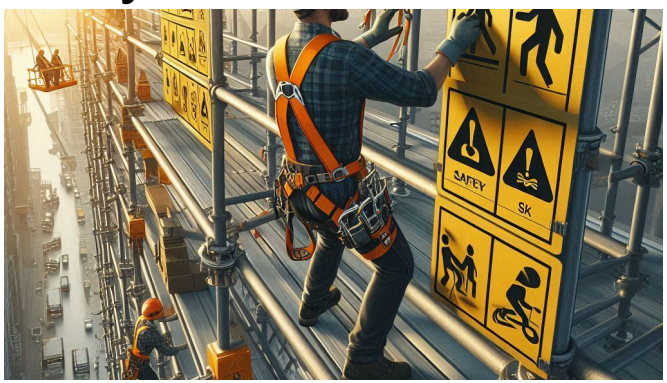
Safety in the math world



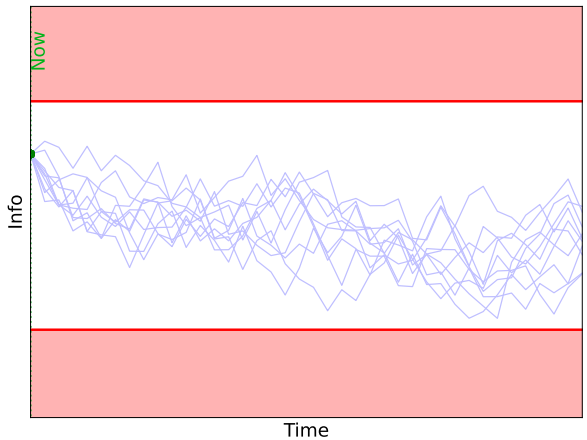
Formally Defining Safety?



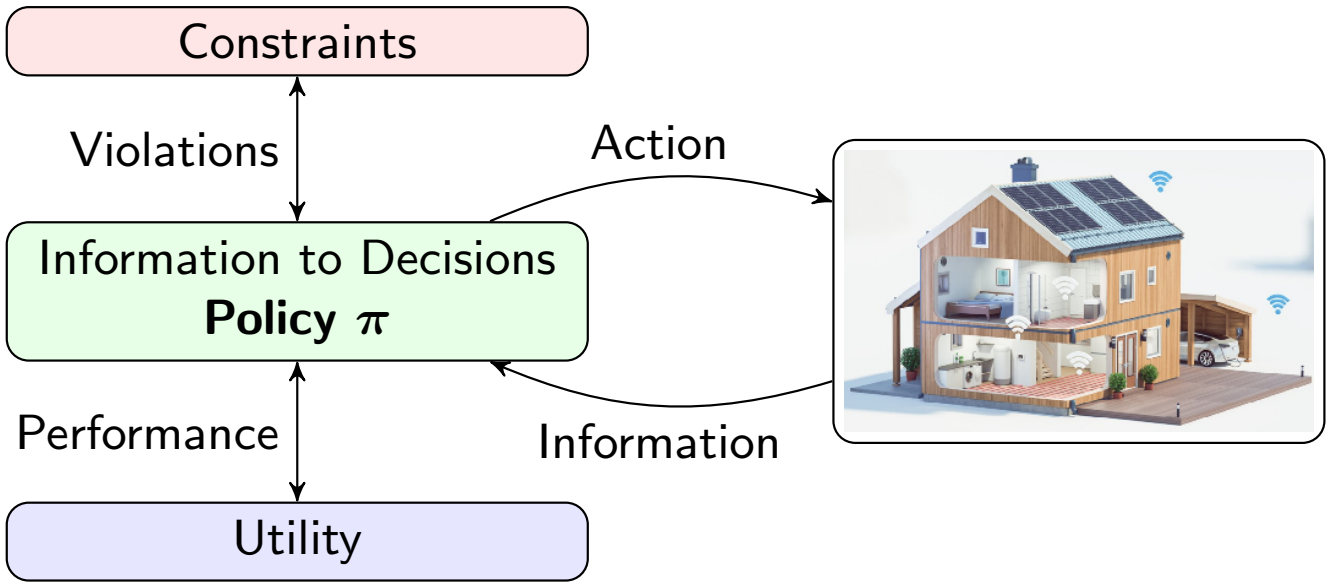
Safety in the real world



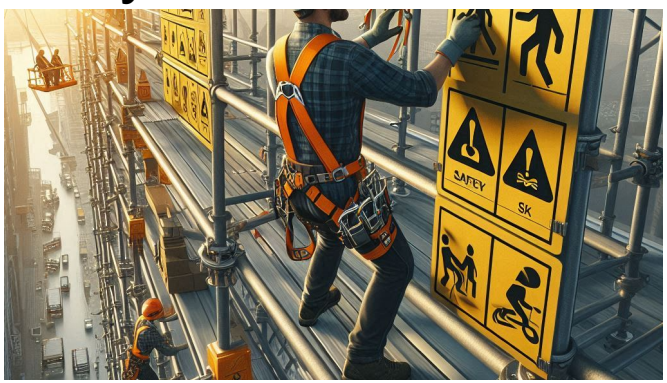
Safety in the math world



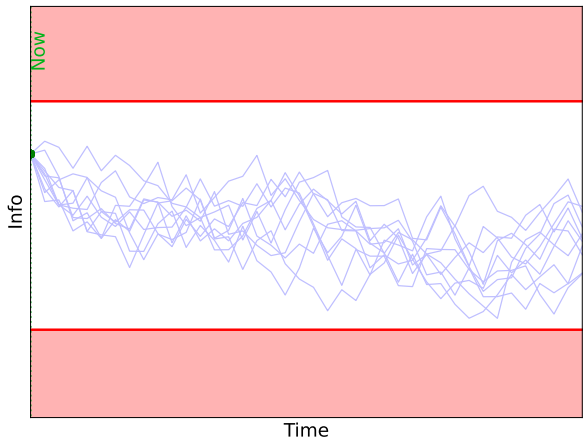
Formally Defining Safety?



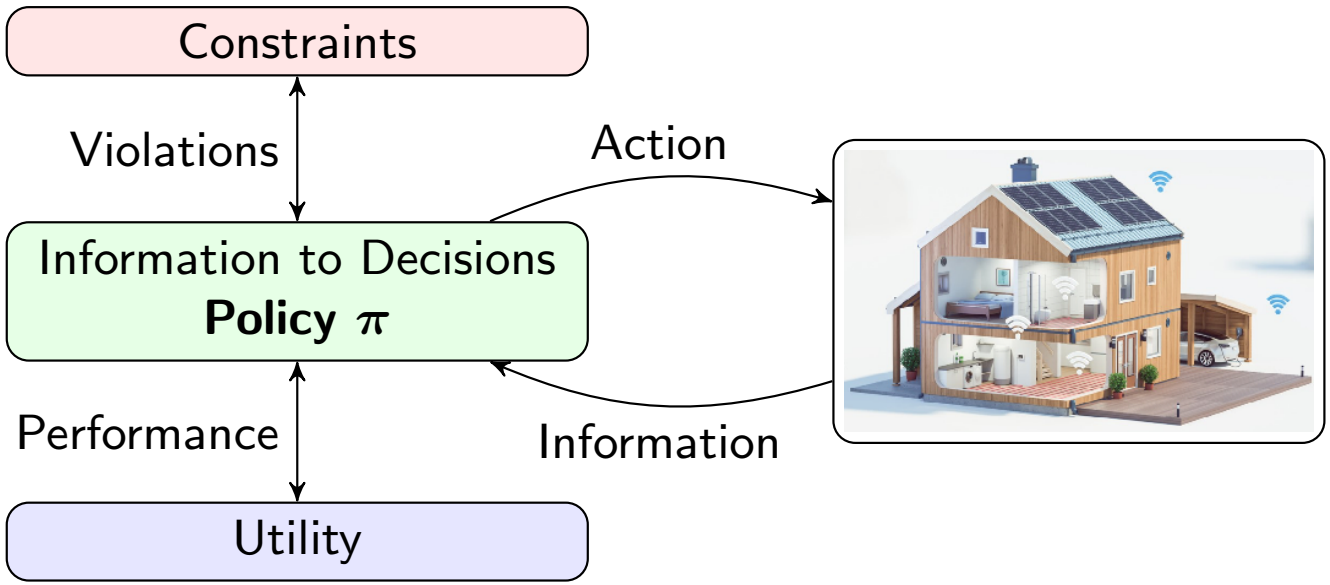
Safety in the real world



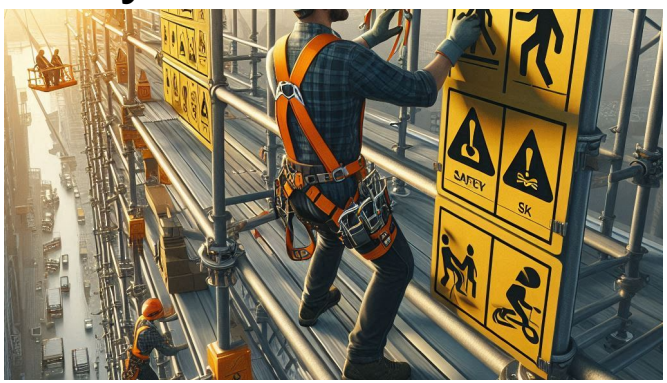
Safety in the math world



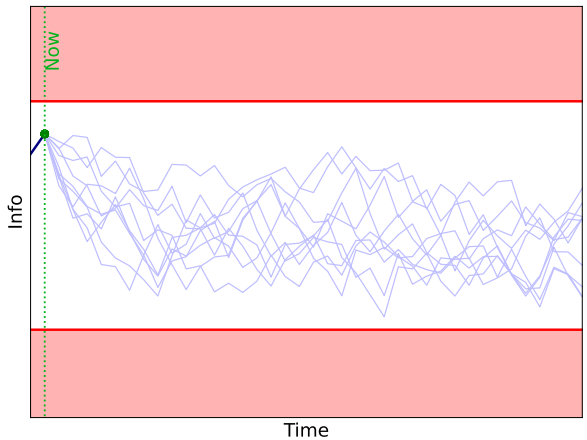
Formally Defining Safety?



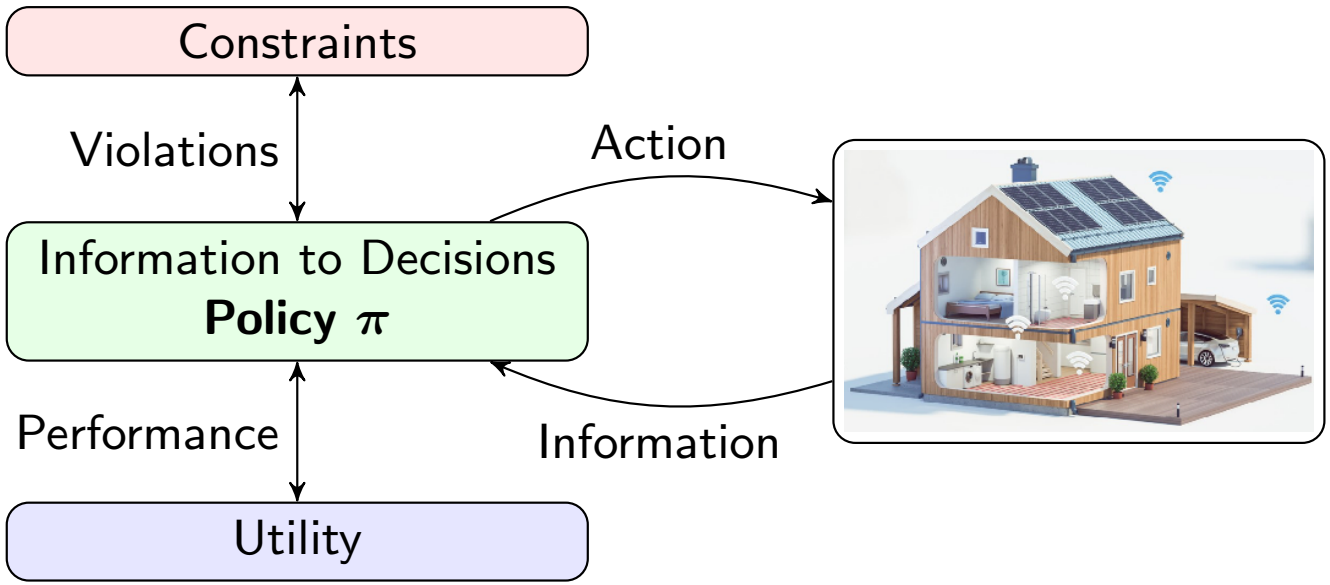
Safety in the real world



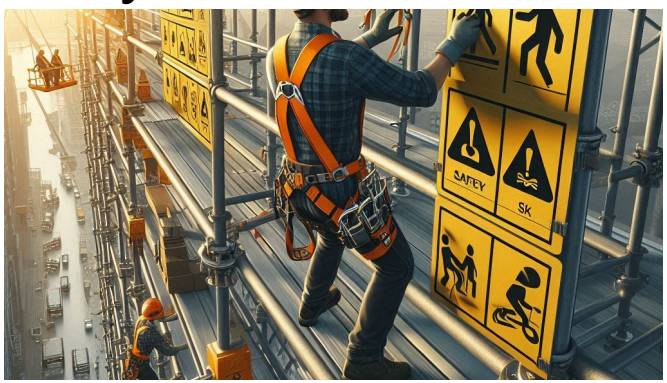
Safety in the math world



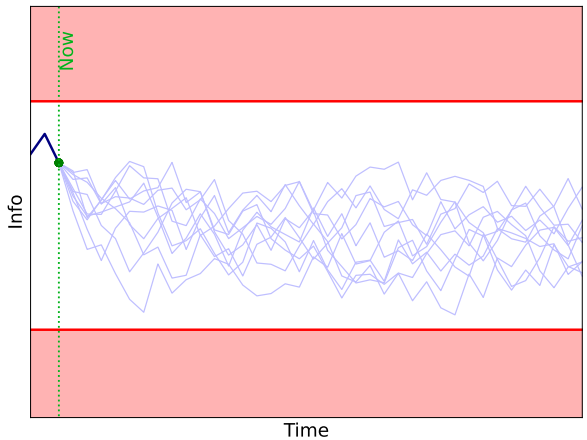
Formally Defining Safety?



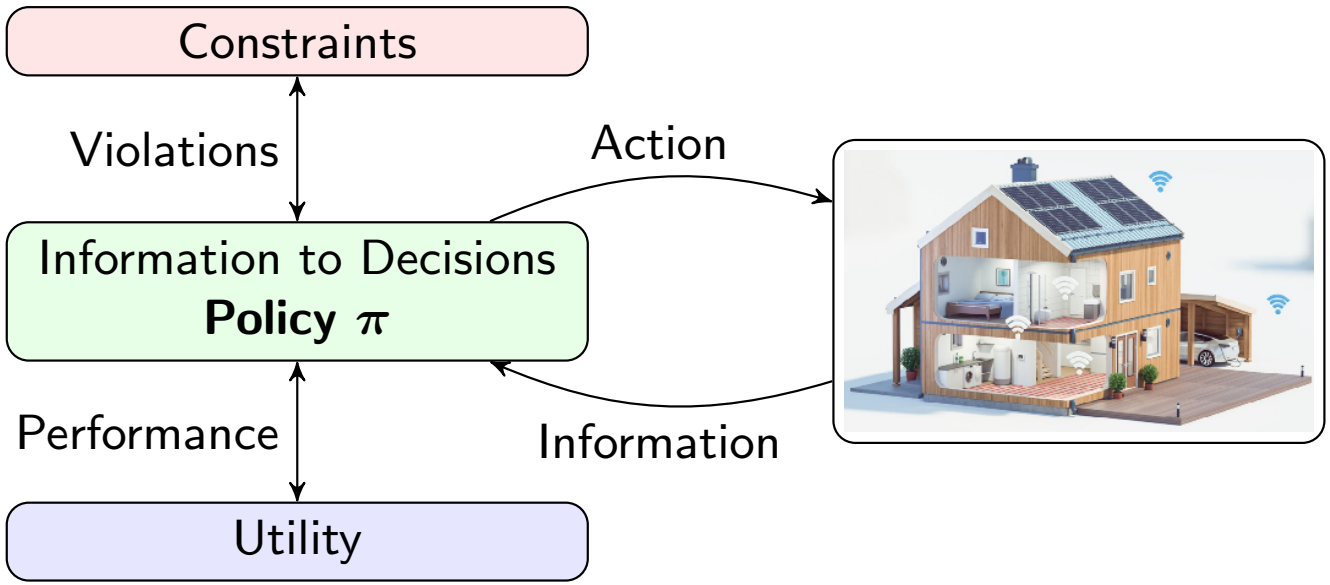
Safety in the real world



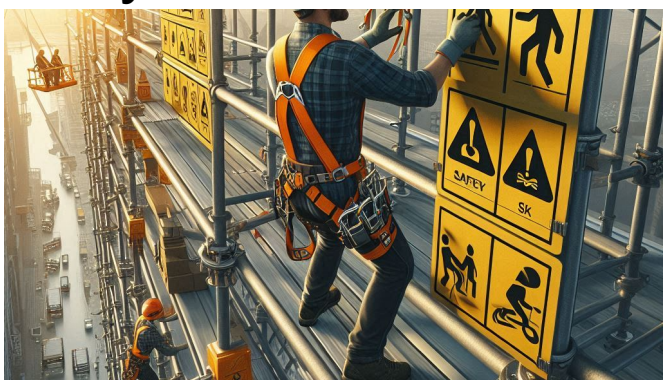
Safety in the math world



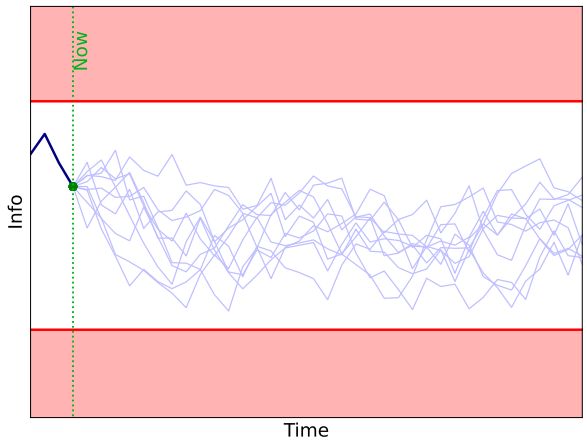
Formally Defining Safety?



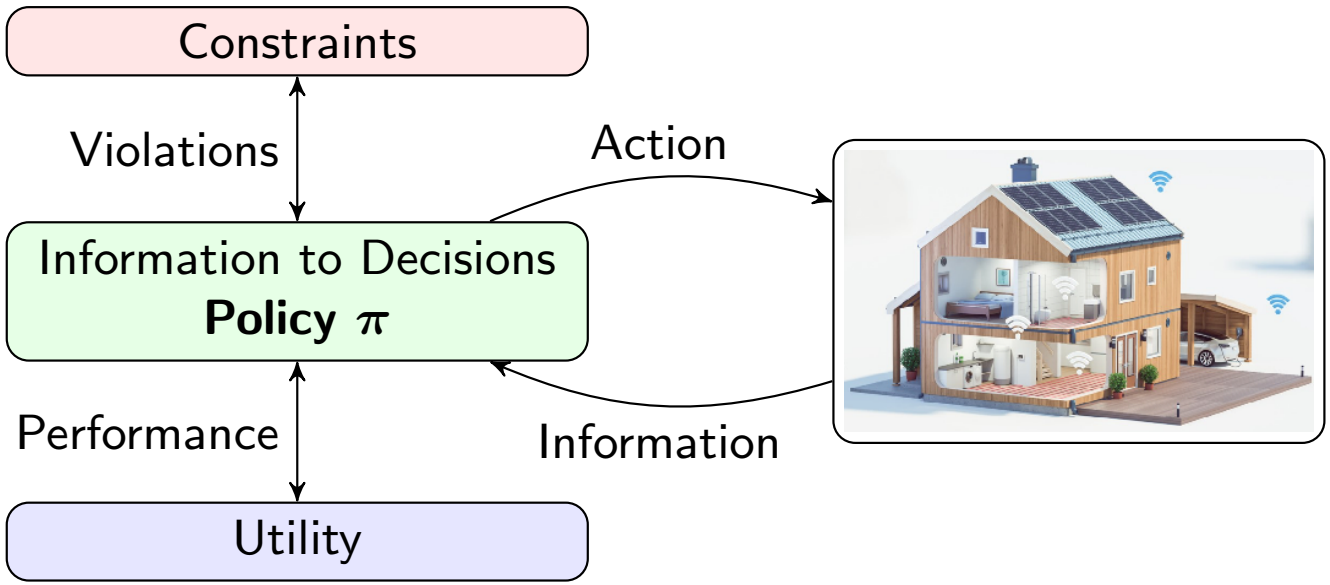
Safety in the real world



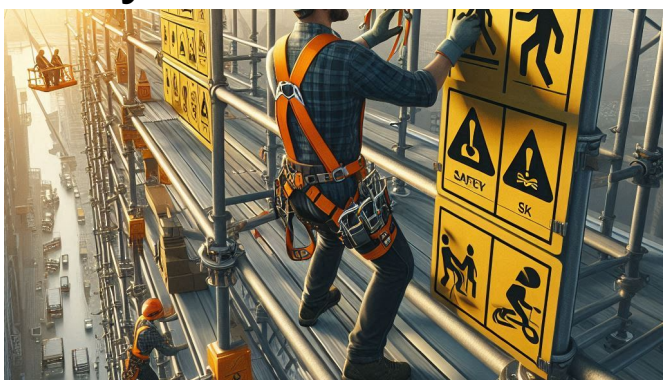
Safety in the math world



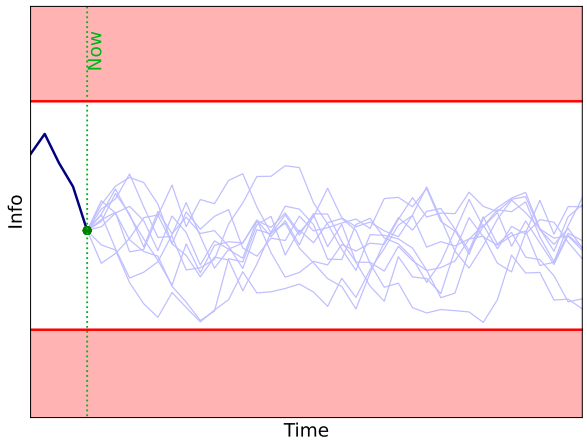
Formally Defining Safety?



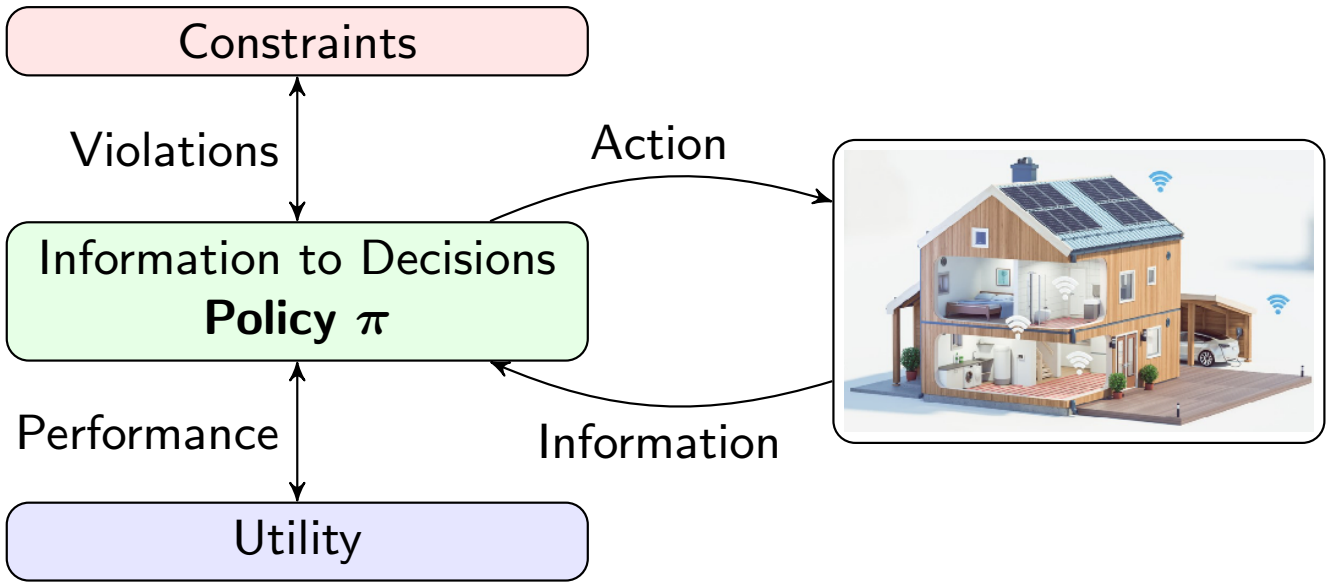
Safety in the real world



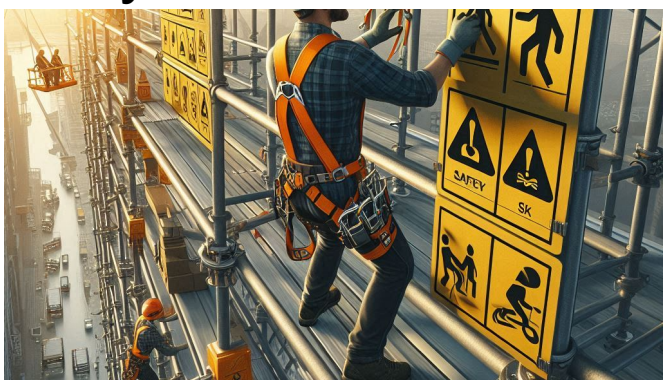
Safety in the math world



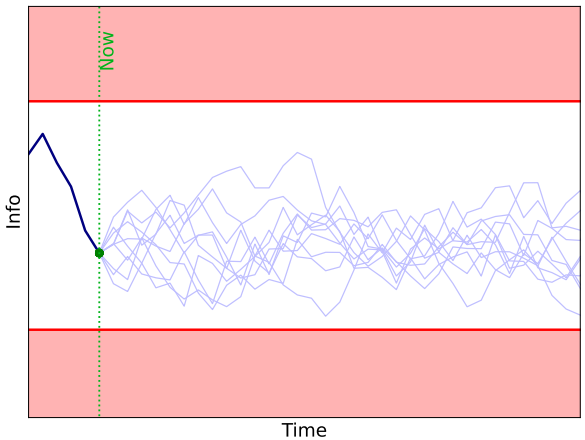
Formally Defining Safety?



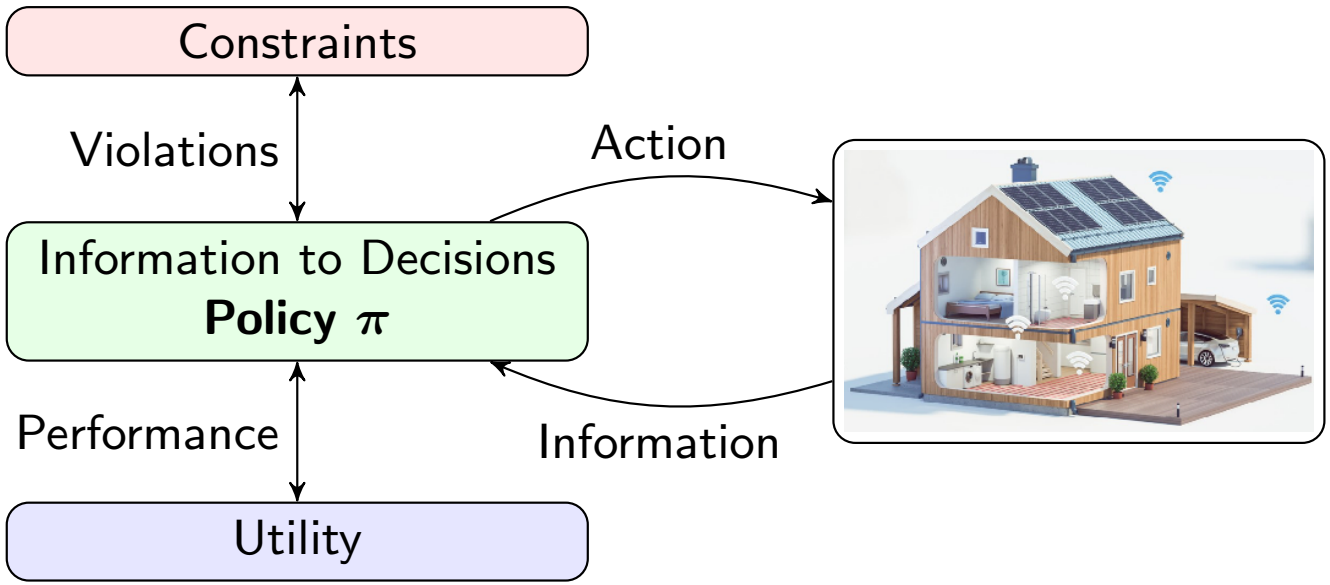
Safety in the real world



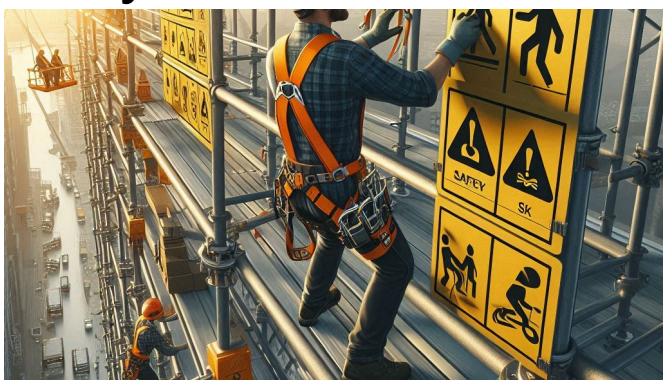
Safety in the math world



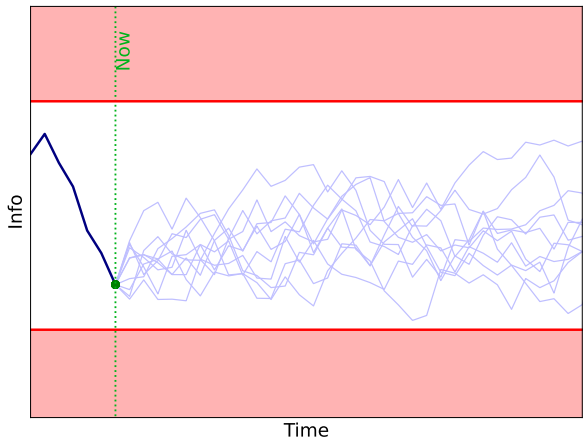
Formally Defining Safety?



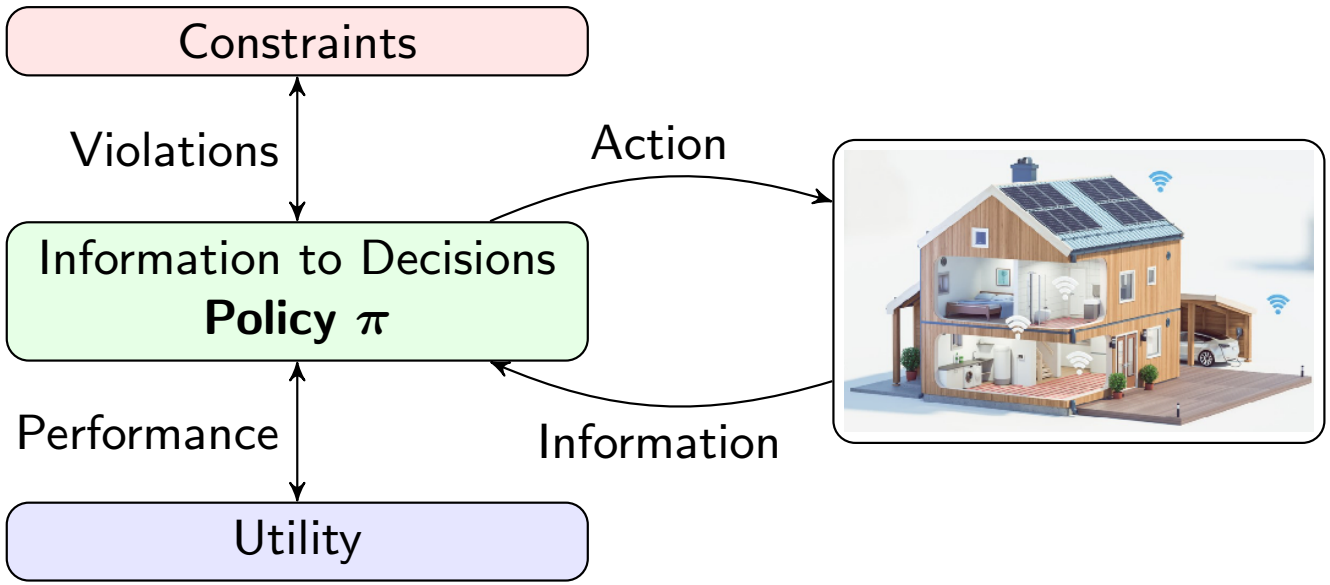
Safety in the real world



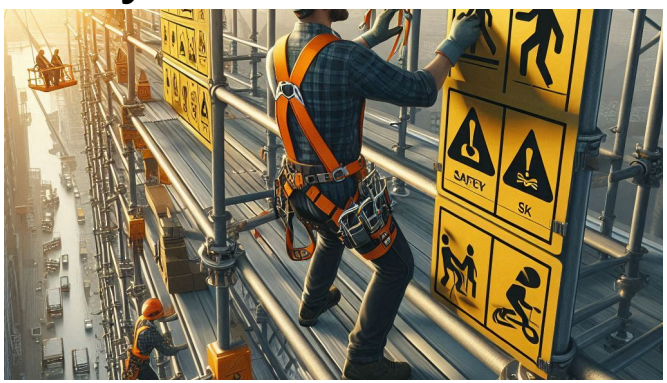
Safety in the math world



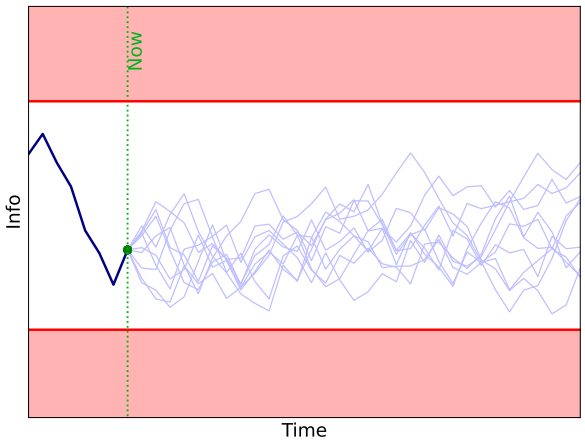
Formally Defining Safety?



Safety in the real world



Safety in the math world

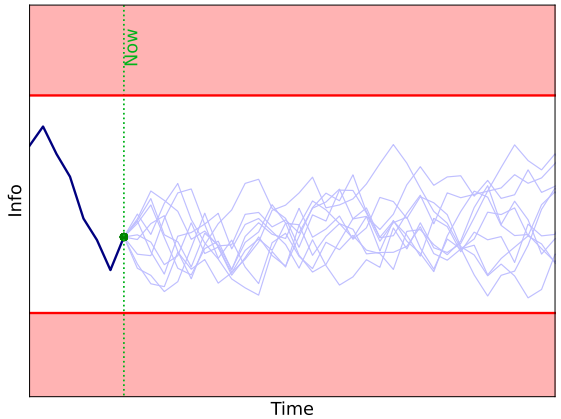


Constrained MDPs

In words

$$\boldsymbol{\pi}^\star = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

$$\text{s.t.} \quad \text{Constraints ok at all time}$$

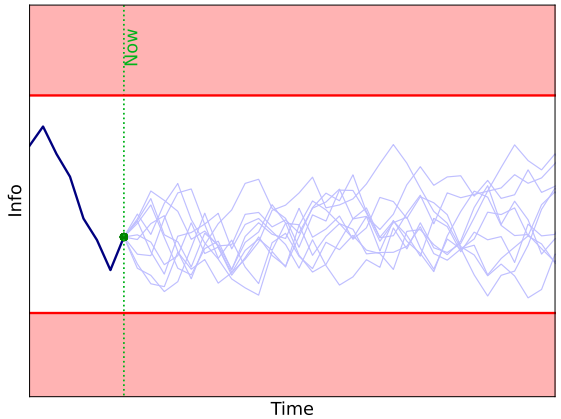


Constrained MDPs

In words

$$\boldsymbol{\pi}^\star = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

$$\text{s.t.} \quad \text{Constraints ok at all time}$$



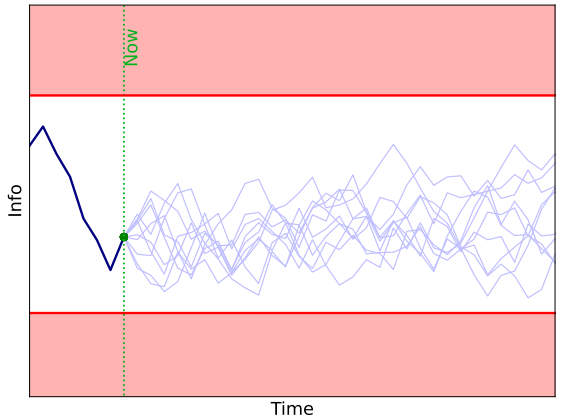
Formally

$$\boldsymbol{\pi}^\star = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \left(\text{Utility} - \begin{cases} 0 & \text{if} \quad \text{Constraint ok} \\ \infty & \text{if} \quad \text{Constraint not ok} \end{cases}\right)\right]$$

Constrained MDPs

In words

$$\boldsymbol{\pi}^\star = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$
$$\text{s.t.} \quad \text{Constraints ok at all time}$$



Formally

$$\boldsymbol{\pi}^\star = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \left(\text{Utility} - \begin{cases} 0 & \text{if} \quad \text{Constraint ok} \\ \infty & \text{if} \quad \text{Constraint not ok} \end{cases}\right)\right]$$

Build policy using

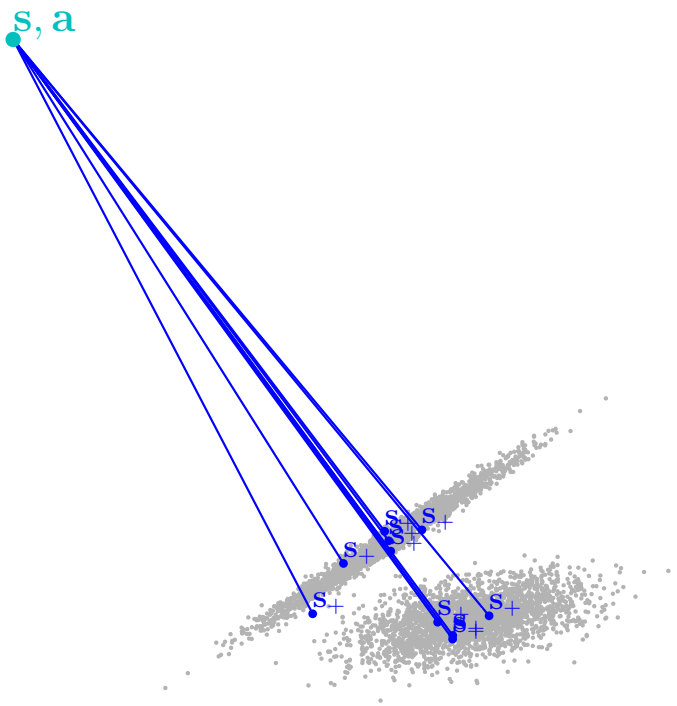
- Perfect model of the real world $\hat{\mathrm{P}}[\mathbf{s}_+|\mathbf{s},\mathbf{a}] = \mathrm{P}[\mathbf{s}_+|\mathbf{s},\mathbf{a}]$
- Model "pessimistic" about the uncertainties

... to evaluate "$\mathrm{E}[\cdot]$"

Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

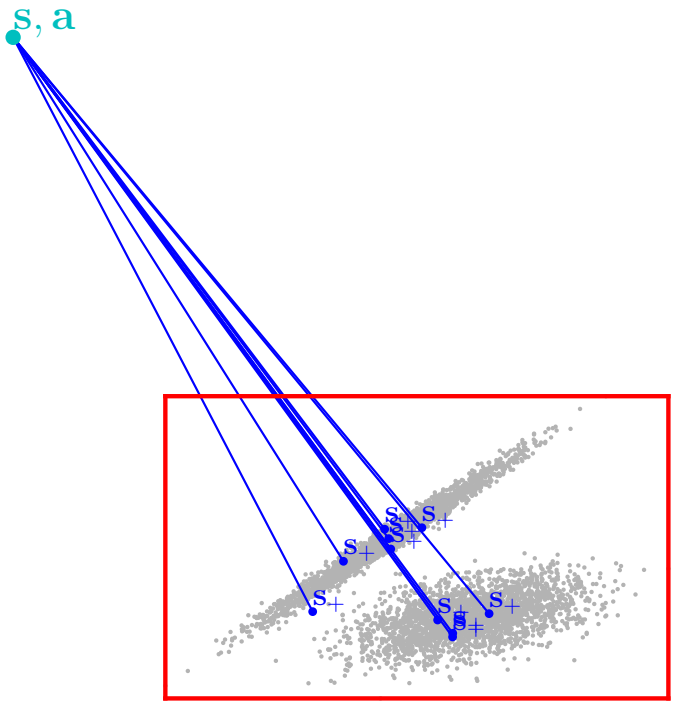
Distribution of one-step forward



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

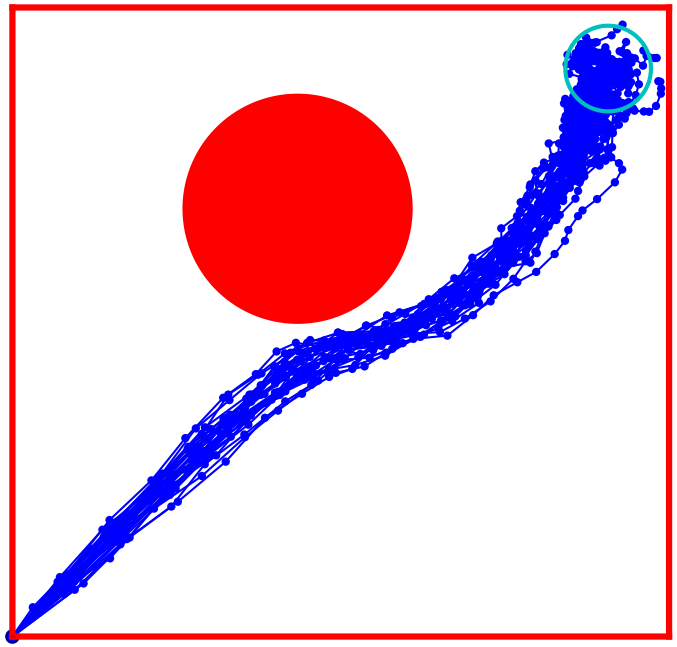
Distribution of one-step forward with "container"



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

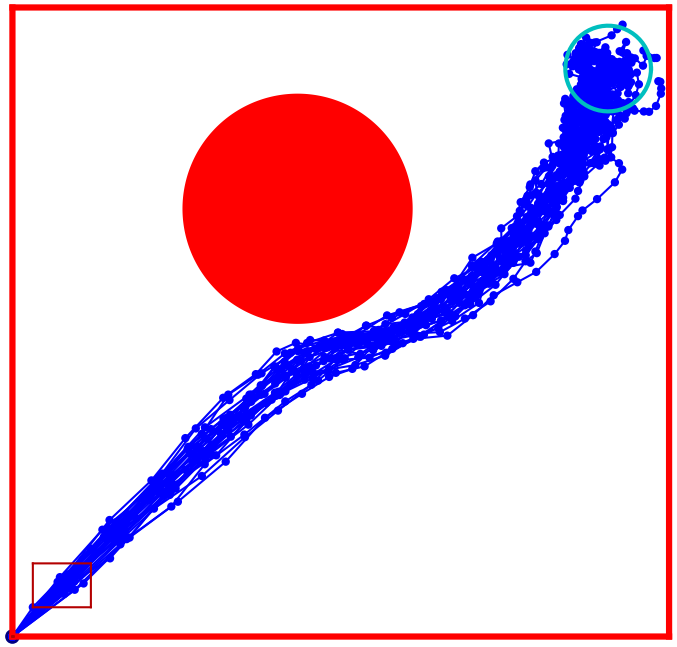
Trajectories



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

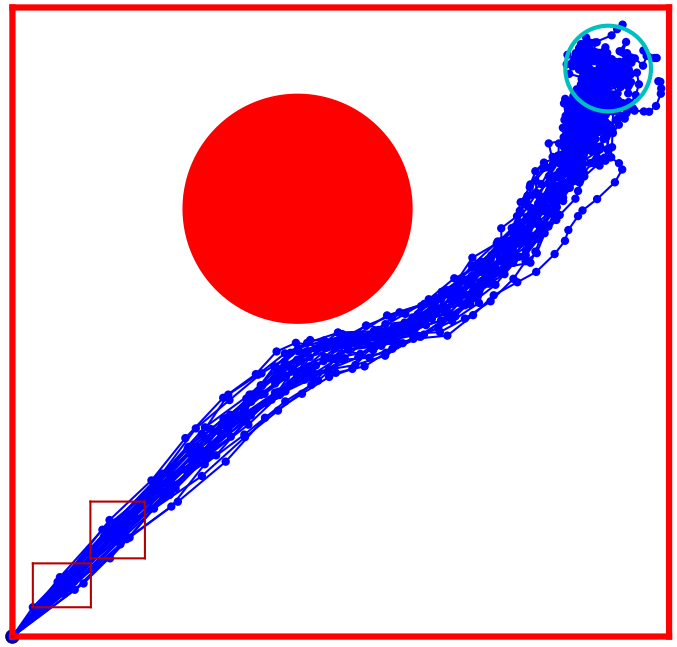
Trajectories with "containers"



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

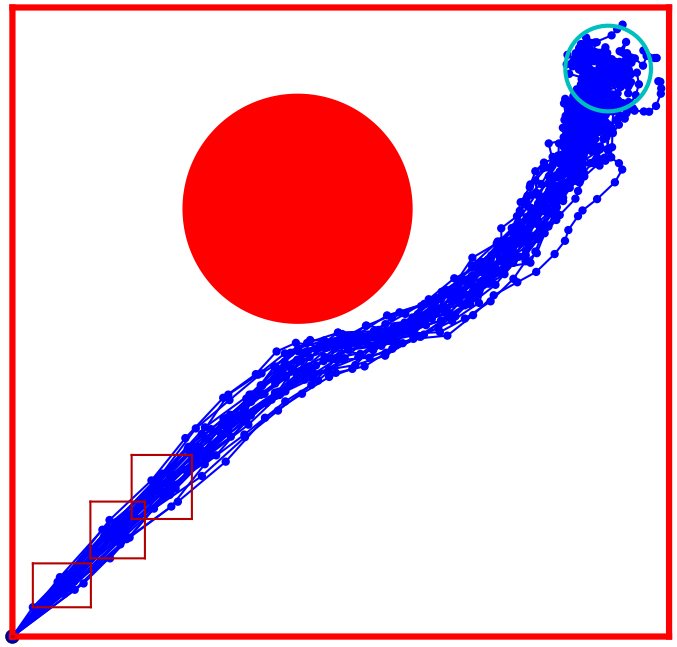
Trajectories with "containers"



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

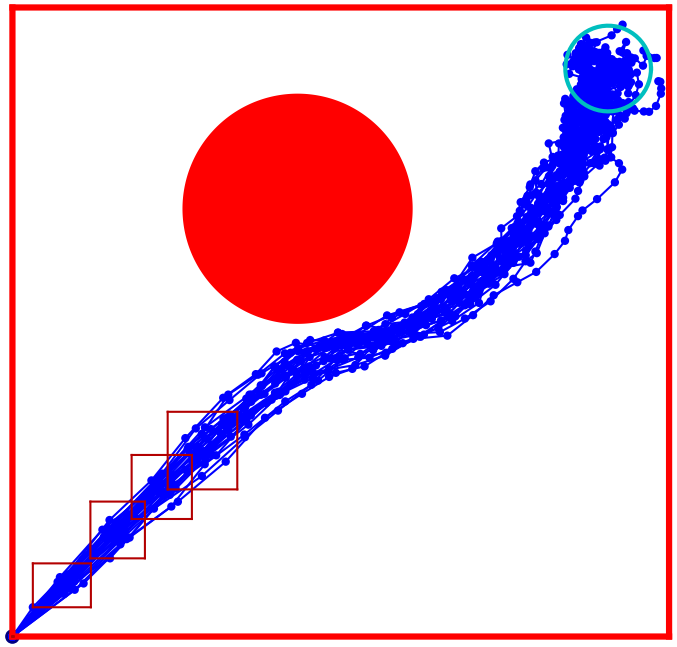
Trajectories with "containers"



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

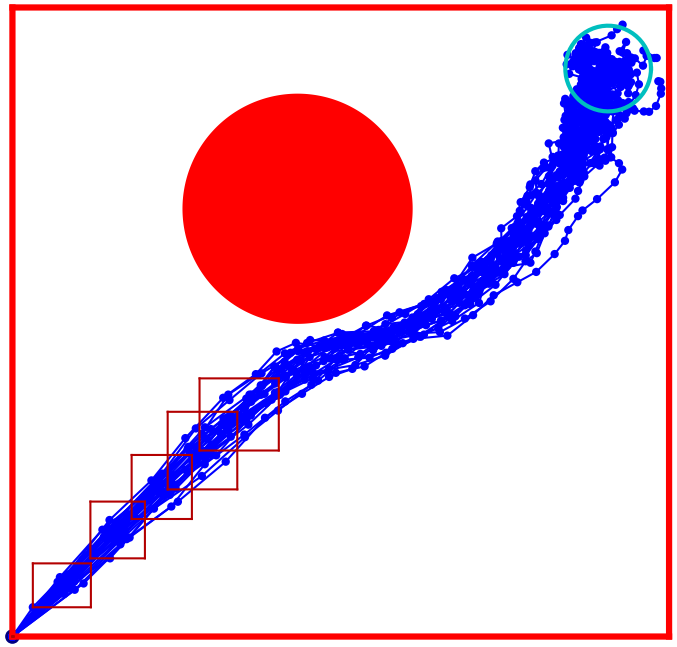
Trajectories with "containers"



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

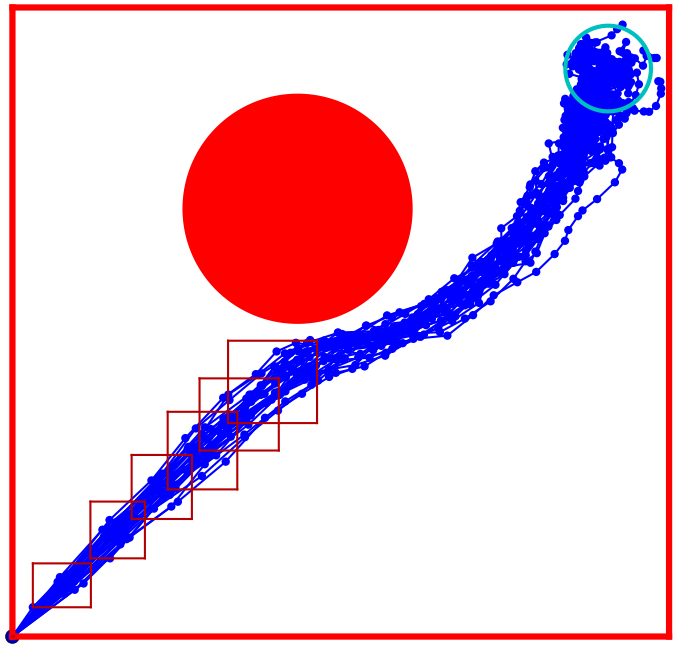
Trajectories with "containers"



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

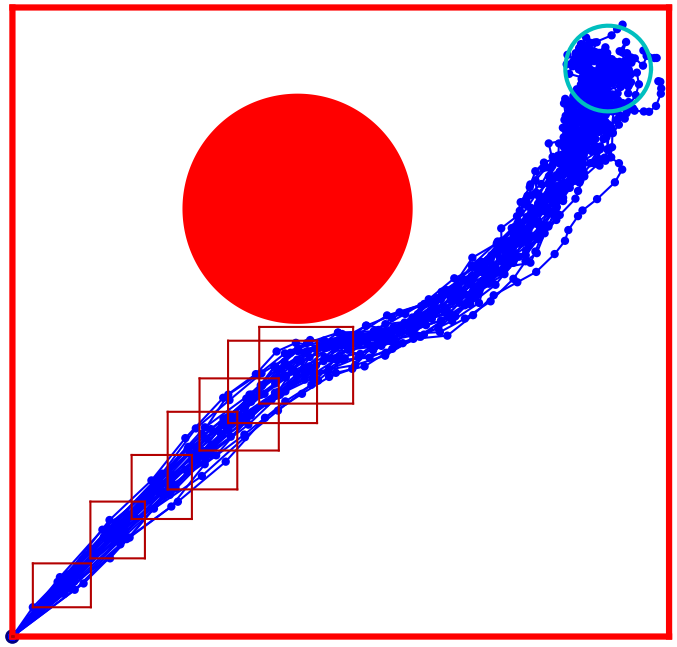
Trajectories with "containers"



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

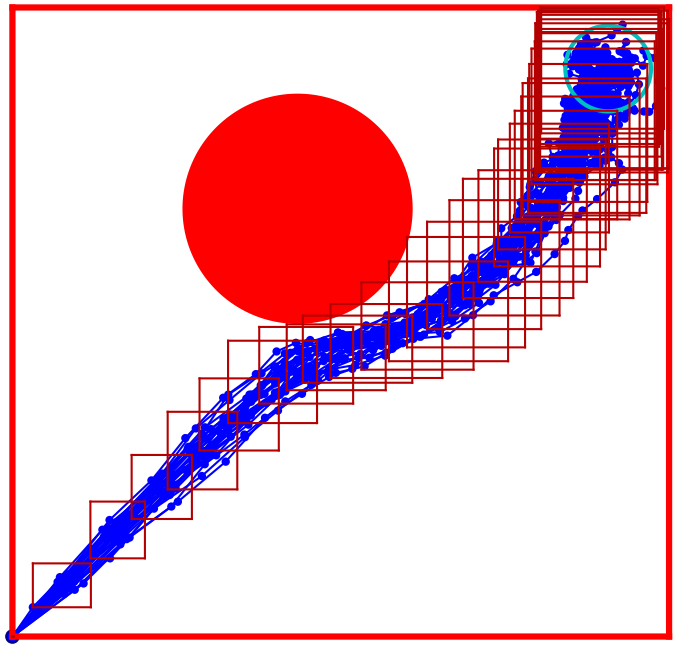
Trajectories with "containers"



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

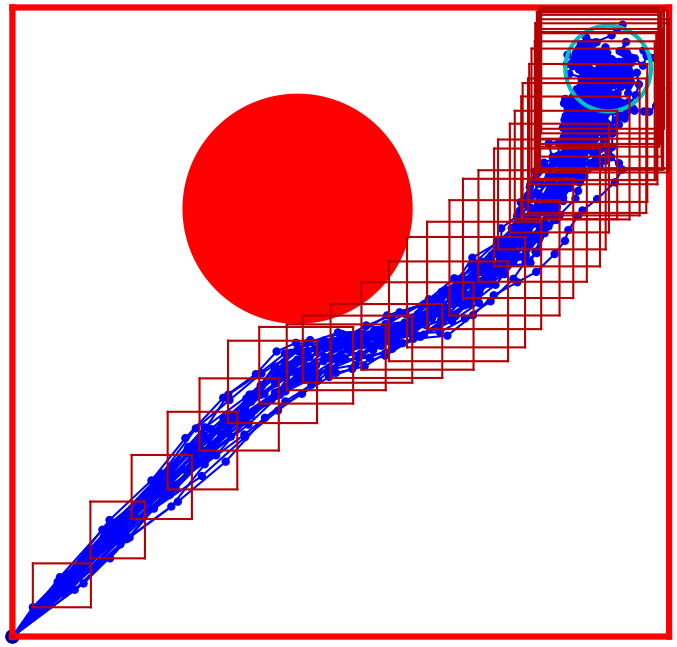
Trajectories with "containers"



Pessimistic Models for Decision Making

- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

Trajectories with "containers"



Remarks

- The propagation of the "containers" in the model predictions can be expensive / difficult
- Pessimistic propagations are usually needed $\rightarrow$ pessimistic over pessimistic
- Policy based on worst-case perspective makes the decisions highly conservative
- Often labelled "Robust" decision making

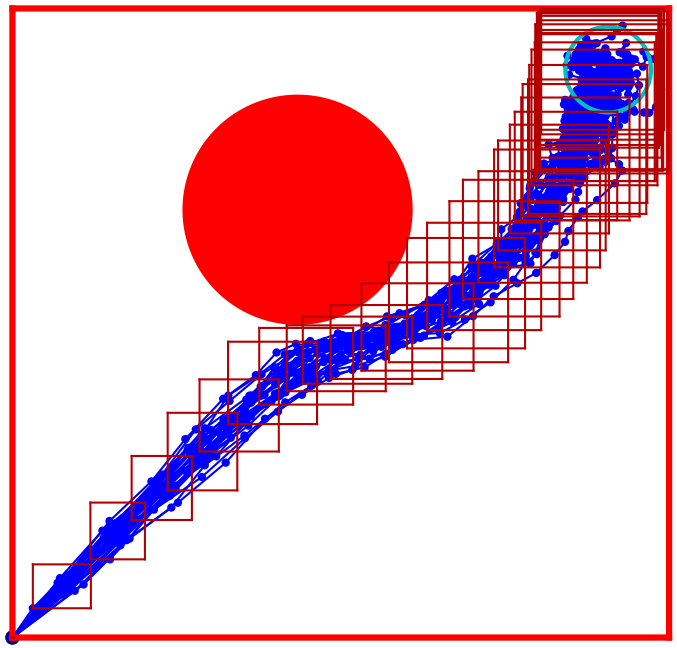
Pessimistic Models for Decision Making

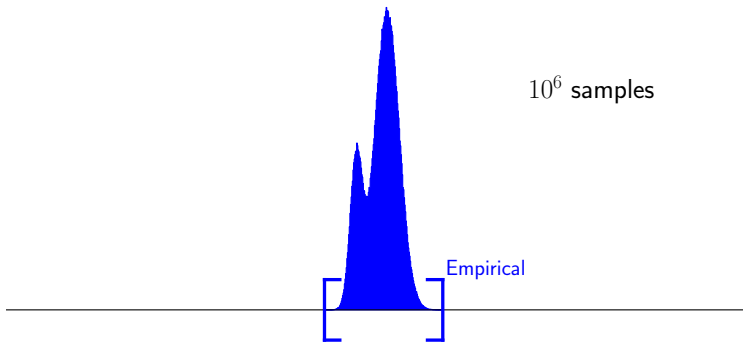
- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

Epistemic problem: build container...

- from data, but it may expand with more data and safety is "asymptotic", as data $\to \infty$

Trajectories with "containers"





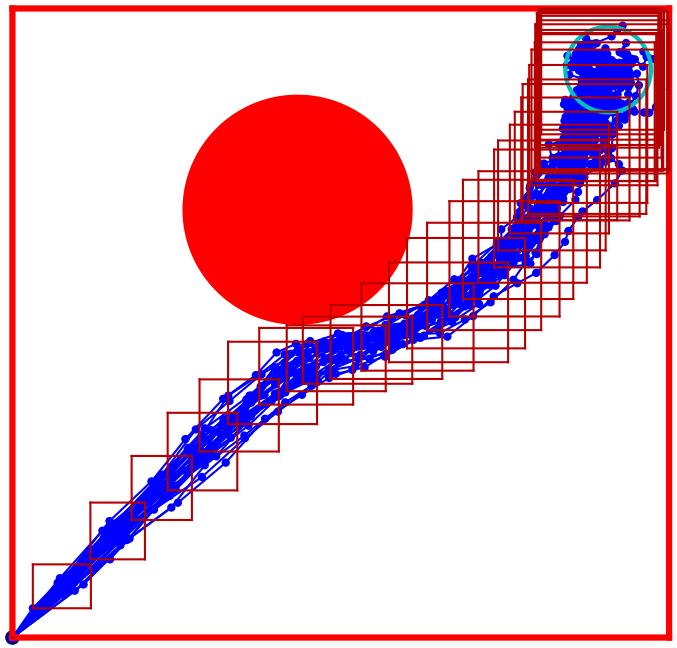
Pessimistic Models for Decision Making

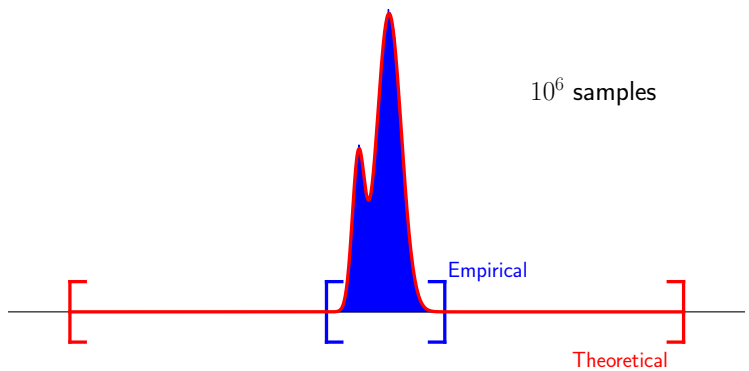
- Model must "contain" the uncertainty
- "Container" (set) should be simple for computational reasons
- Trajectories predicted by pessimistic model will "cover" the real world
- Decision policy wants to be safe w.r.t. the "containers"

Epistemic problem: build container...

- from data, but it may expand with more data and safety is "asymptotic", as data $\to \infty$
- "theoretically" (from knowledge), but can be impractically pessimistic

Trajectories with "containers"



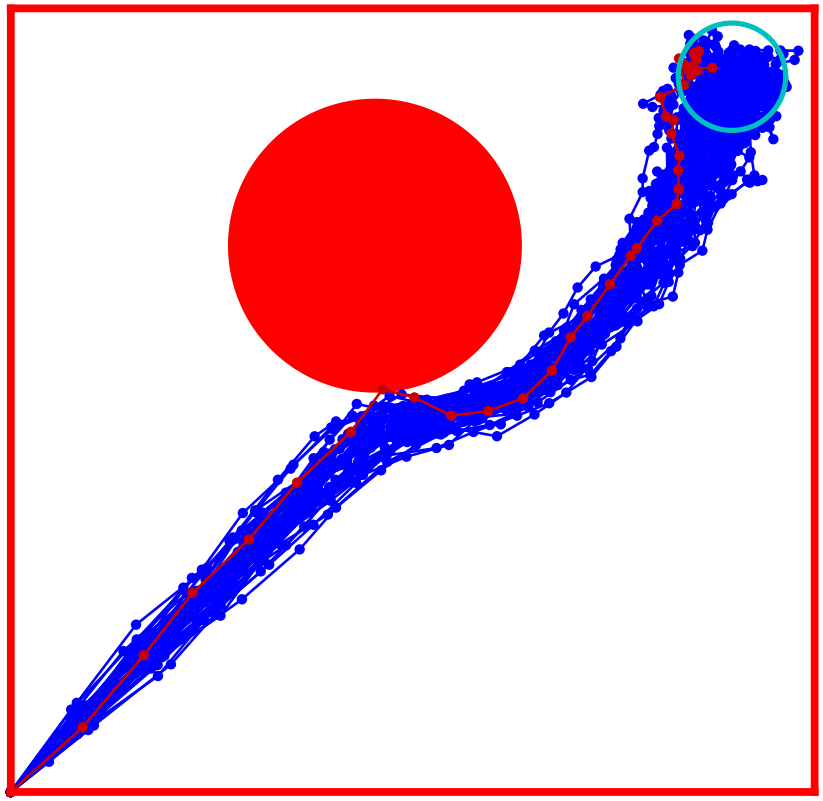


MDPs with probabilistic safety

In words

$$\pi^\star = \arg\max_{\pi} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

$$\text{s.t.} \quad \text{Probability of no violation} \geq c$$



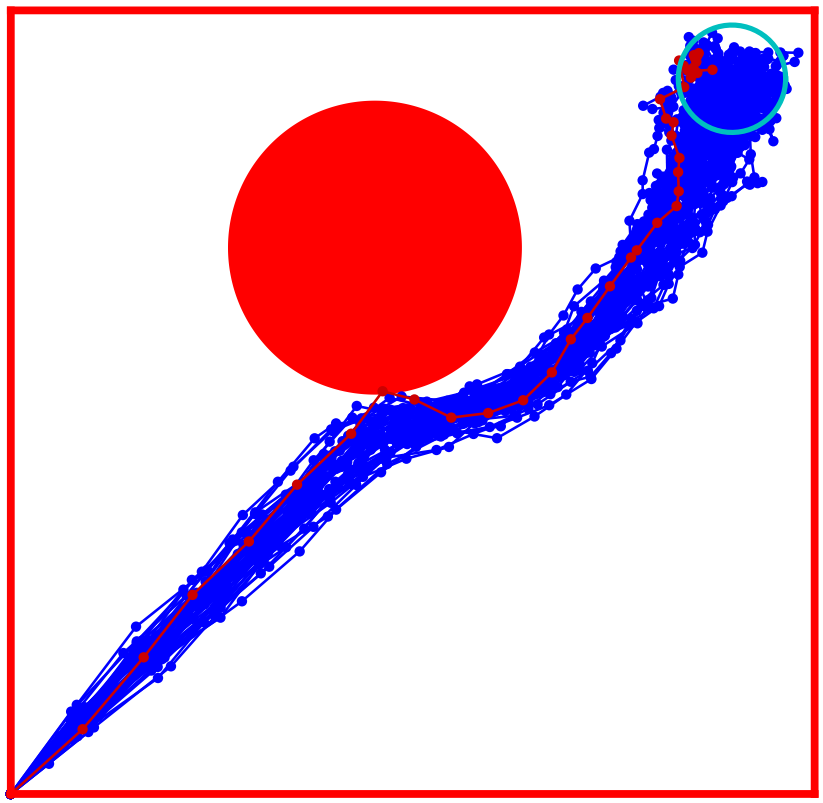
MDPs with probabilistic safety

In words

$$\pi^\star = \arg\max_{\pi} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$
$$\text{s.t.} \quad \text{Probability of no violation} \geq c$$

Formally

$$\pi^\star = \arg\max_{\pi} \quad \mathrm{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{s}_k, \mathbf{a}_k\right)\right]$$
$$\text{s.t.} \quad \mathrm{P}\left[\mathbf{s}_{0,\ldots\infty} \in \mathbb{S}\right] \geq c$$



MDPs with probabilistic safety

In words

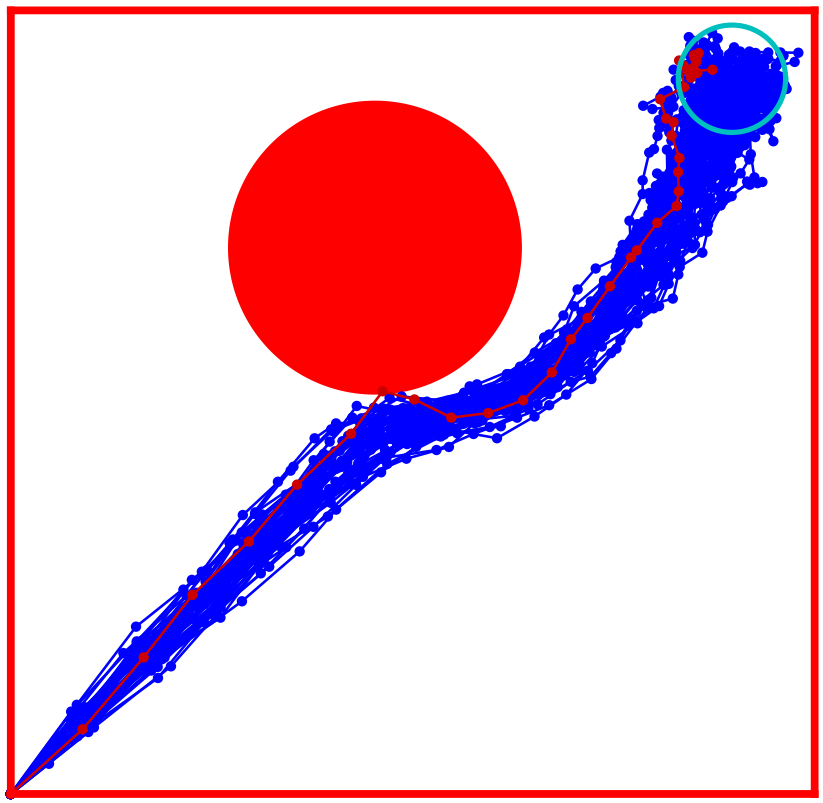
$$\pi^\star = \arg\max_{\pi} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$
$$\text{s.t.} \quad \text{Probability of no violation} \geq c$$

Formally

$$\pi^\star = \arg\max_{\pi} \quad \mathrm{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{s}_k, \mathbf{a}_k\right)\right]$$
$$\text{s.t.} \quad \mathrm{P}\left[\mathbf{s}_{0,\ldots\infty} \in \mathbb{S}\right] \geq c$$

Remarks:

- If we can tolerate $c > 0$ (small) it can make a big gain in performance
- Aligned with industrial / practical standards on "large series"
- Problem needs a "termination" (time or goal reached)
- Building decisions can be difficult from a computational point of view



Risk-sensitive MDPs

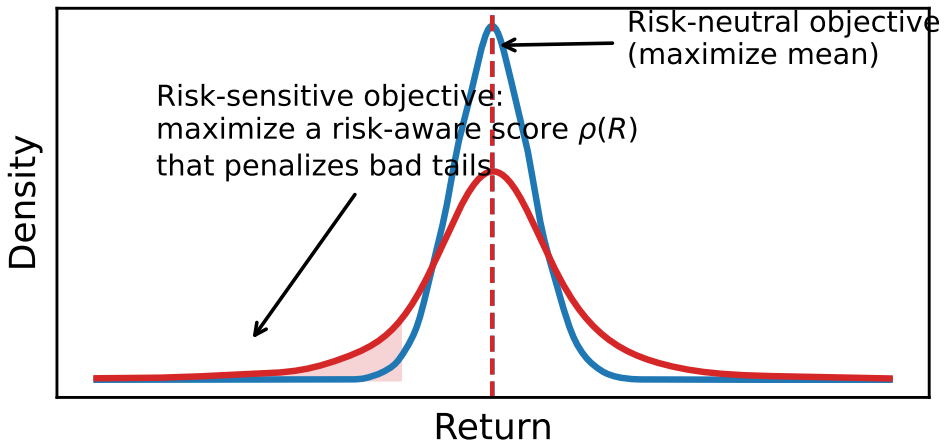
In words

- Instead of constraints, the utility function itself is altered to contain a risk measure of the return.
- Intuition: a "tail-risk filter" even if outcomes are good most of the time, severe tails influence decisions.

Risk-sensitive MDPs

In words

- Instead of constraints, the utility function itself is altered to contain a risk measure of the return.
- Intuition: a "tail-risk filter" even if outcomes are good most of the time, severe tails influence decisions.



Risk-sensitive MDPs

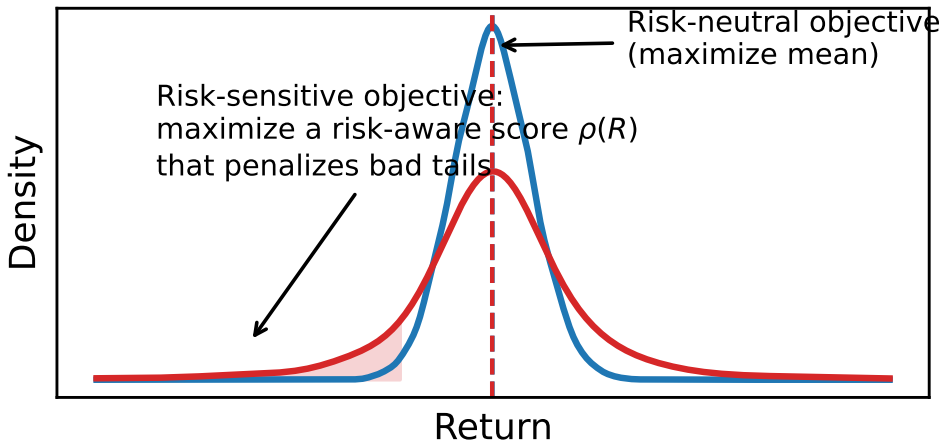
In words

- Instead of constraints, the utility function itself is altered to contain a risk measure of the return.
- Intuition: a "tail-risk filter" even if outcomes are good most of the time, severe tails influence decisions.

Formally

$$\pi^* = \arg\max_{\pi} \ \rho\Big(\sum_{t=0}^{\infty} \gamma^t r_t\Big)$$

ρ: VaR, CVaR, Entropic Risk.



Risk-sensitive MDPs

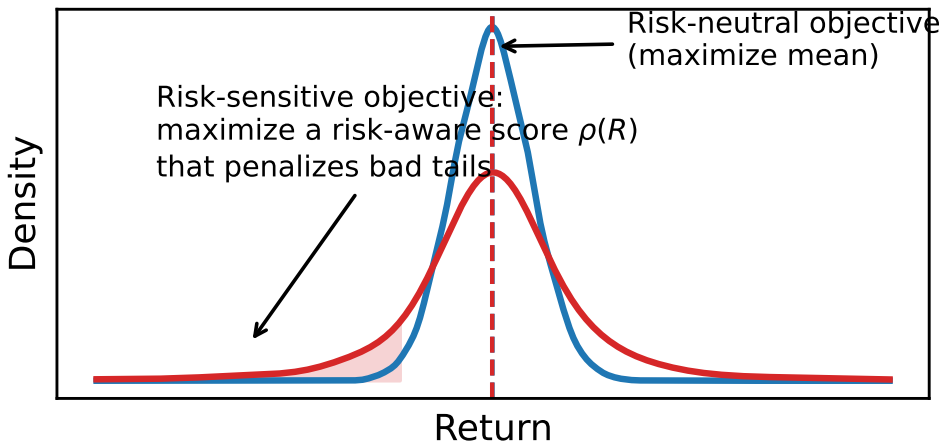
In words

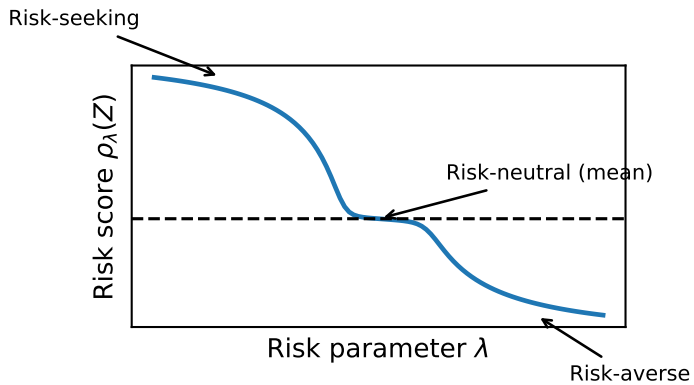
- Instead of constraints, the utility function itself is altered to contain a risk measure of the return.
- Intuition: a "tail-risk filter" even if outcomes are good most of the time, severe tails influence decisions.

Formally

$$\pi^* = \arg\max_{\pi} \ \rho\Big(\sum_{t=0}^{\infty} \gamma^t r_t\Big)$$

ρ: VaR, CVaR, Entropic Risk.





Risk-sensitive MDPs

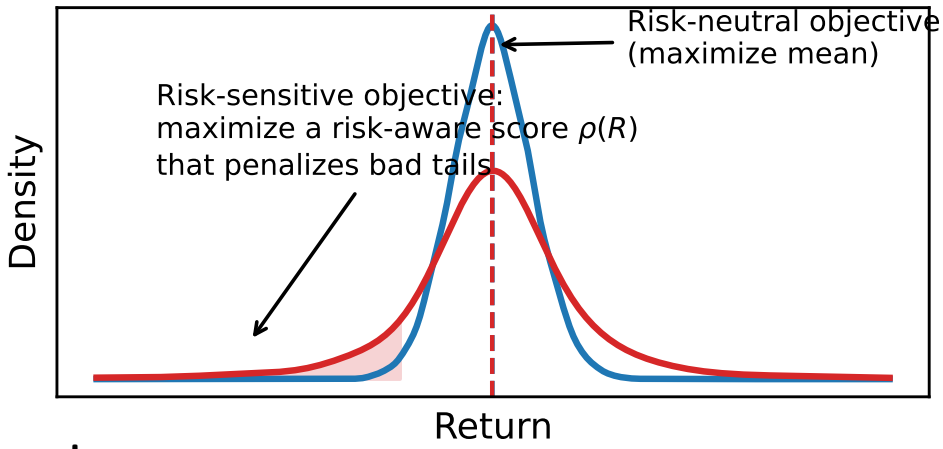
In words

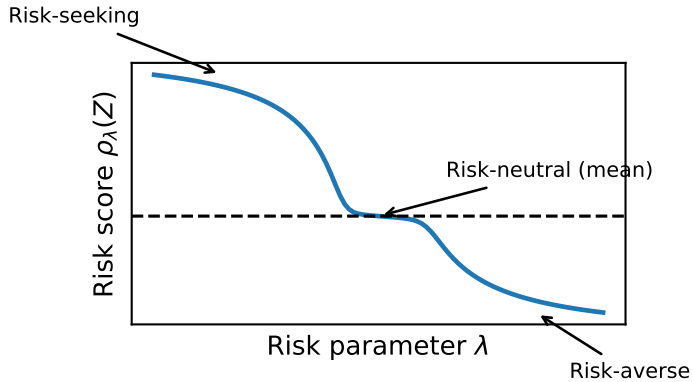
- Instead of constraints, the utility function itself is altered to contain a risk measure of the return.
- Intuition: a "tail-risk filter" even if outcomes are good most of the time, severe tails influence decisions.

Formally

$$\pi^* = \arg\max_{\pi} \rho\Big(\sum_{t=0}^{\infty} \gamma^t r_t\Big)$$

ρ: VaR, CVaR, Entropic Risk.



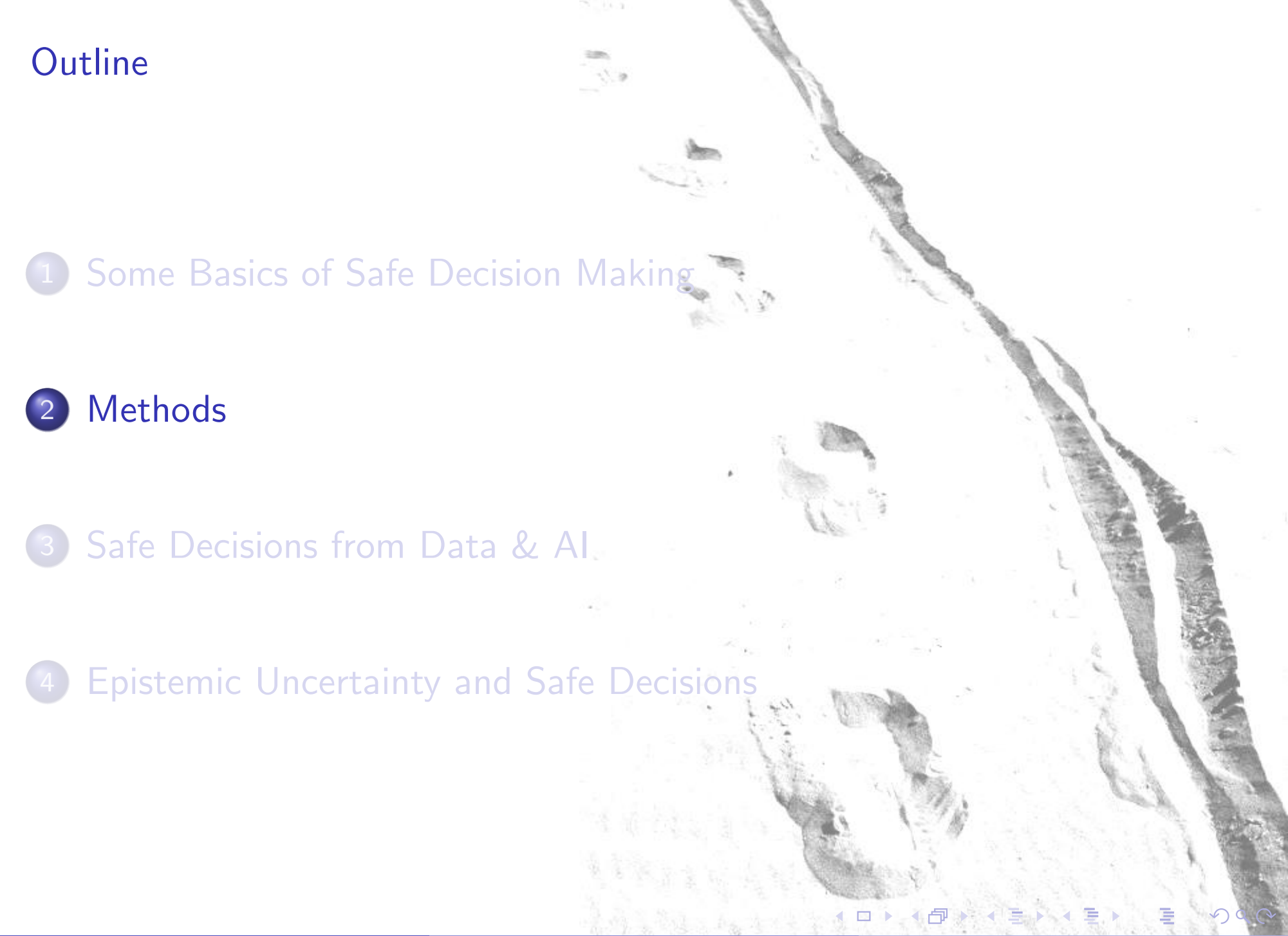


Remarks:

- Views risk from through the lense of utility (not undesirable events).
- Risk-sensitive criteria do *not* guarantee a bound on violation.
- Cares only about the average severity of bad outcomes in the tail and not about the rest.

Outline

1. Some Basics of Safe Decision Making
2. Methods
3. Safe Decisions from Data & AI
4. Epistemic Uncertainty and Safe Decisions



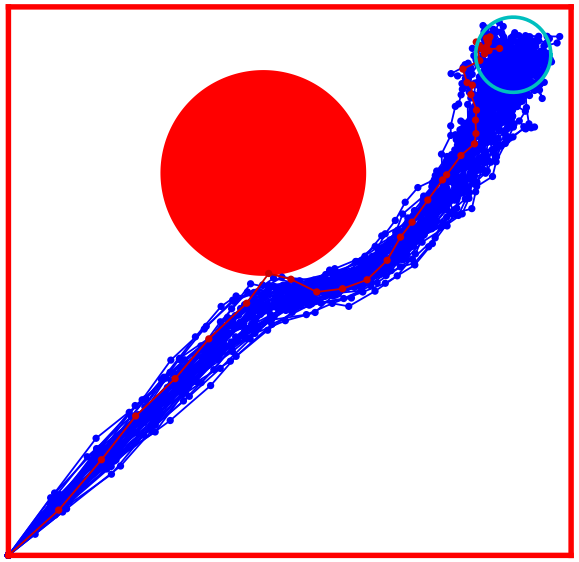
Robust Repeated Planning

- Introduction: Robust MPC, Scenario Trees, MC, robust multi-stage stochastic programing, MPPI?
- Difficulties: guarantees for nonlinear systems, persistent safety (recursive feasibility)

TODOTODOTODO

Dynamic Programming & RL with Probabilistic Safety

$$\boldsymbol{\pi}^\star = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{s}_k, \mathbf{a}_k\right)\right]$$
$$\text{s.t.} \quad \mathrm{P}\left[\mathbf{s}_{0,\ldots\infty} \in \mathbb{S}\right] \geq c$$



Expected value form

$$\boldsymbol{\pi}^\star = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{k=0}^{\infty} \gamma^k L\left(\mathbf{s}_k, \mathbf{a}_k\right)\right]$$
$$\text{s.t.} \quad \mathrm{E}\left[\begin{array}{ll} 1 & \text{if no violation occurred} \\ 0 & \text{otherwise} \end{array}\right] \geq c$$

- Expected value form enables classical techniques (ref. 1st lecture), i.e. DP and RL
- Difficulties: estimate expected values (sample based) when $\mathrm{P}\left[\mathbf{s}_{0,\ldots\infty} \in \mathbb{S}\right]$ is close to 1. Illustrate?? SG will try on 11.9

Outline

1. Some Basics of Safe Decision Making
2. Methods
3. Safe Decisions from Data & AI
4. Epistemic Uncertainty and Safe Decisions

Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbb{S}(\mathbf{s})$

Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathrm{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

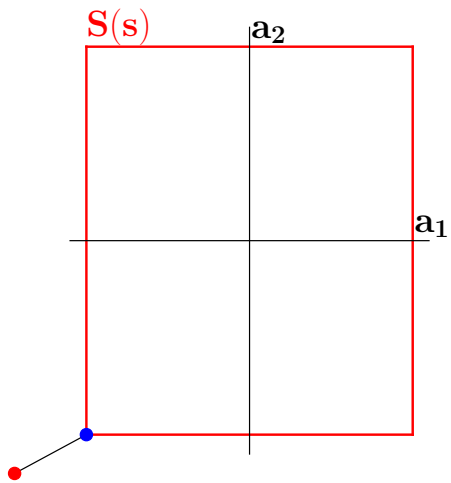
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



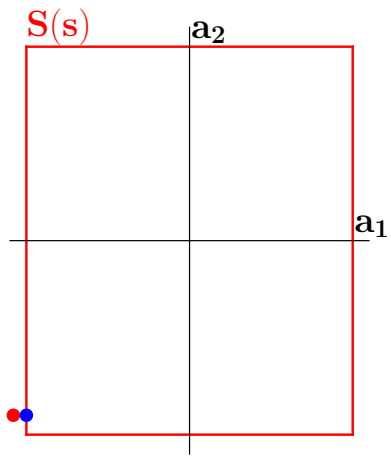
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



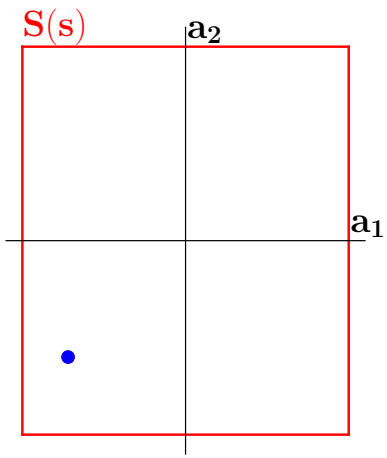
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



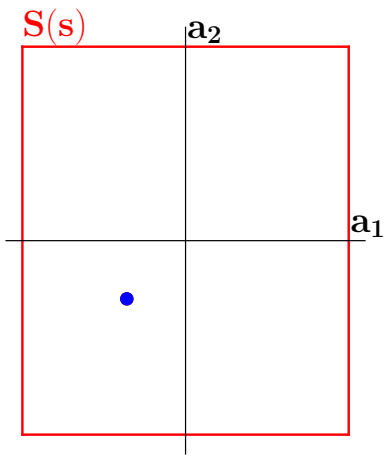
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



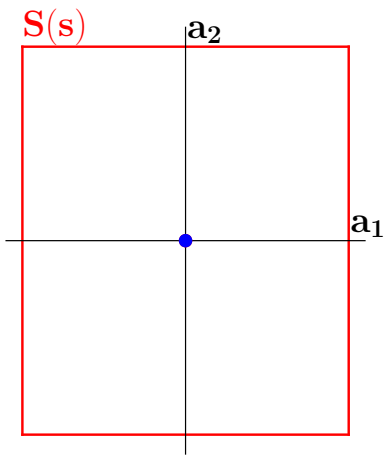
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathrm{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathrm{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



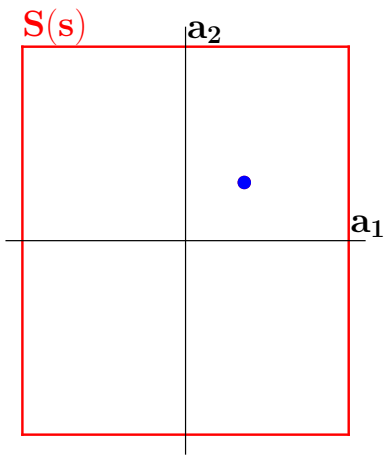
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



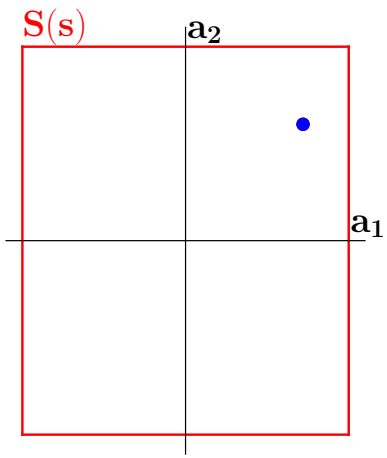
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



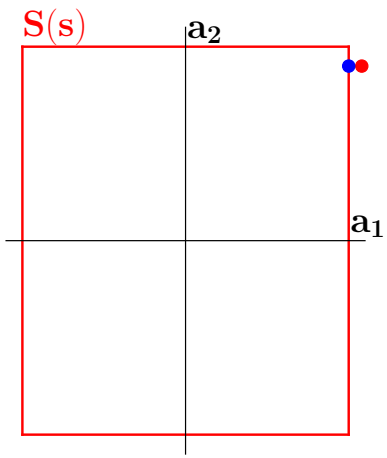
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



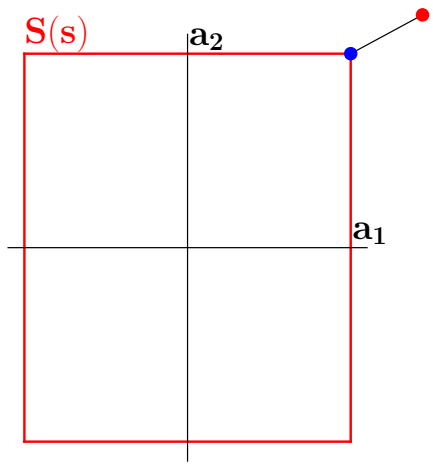
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



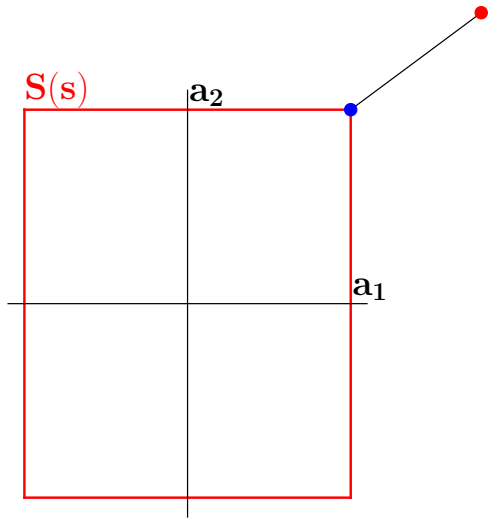
Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$



Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

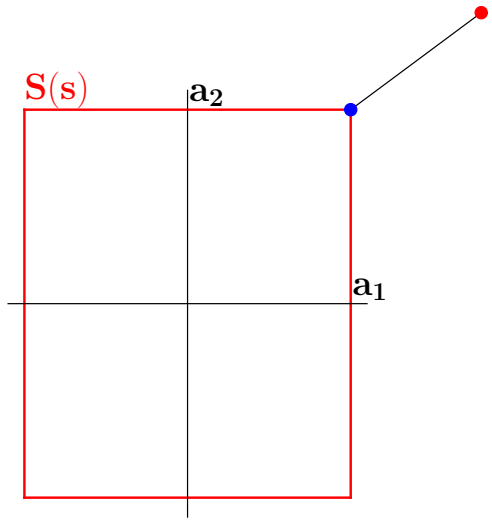
$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathbf{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$

Remark: "oracle" can be very hard to build, or not...

Not losing the king



Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathbf{S}(\mathbf{s})$
- We can learn an "unsafe" policy

$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathbf{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

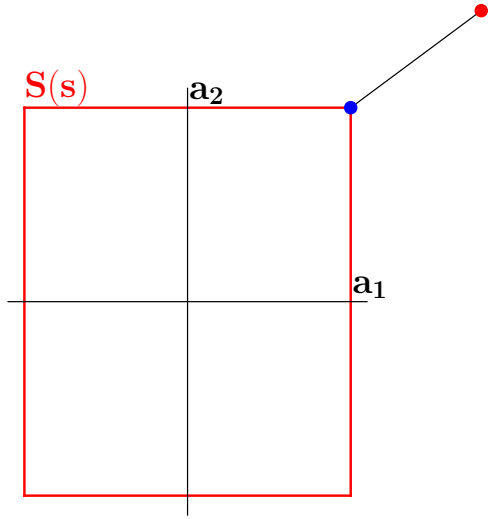
that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathbf{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$

Remark: "oracle" can be very hard to build, or not...

Not losing the king

Able to stop within the visible distance



Safety Filters & Reinforcement Learning

- For any state $\mathbf{s}$, an "oracle" tells you which actions will not jeopardize safety in the long run$\rightarrow$"Safe set" $\mathrm{S}(\mathbf{s})$
- We can learn an "unsafe" policy

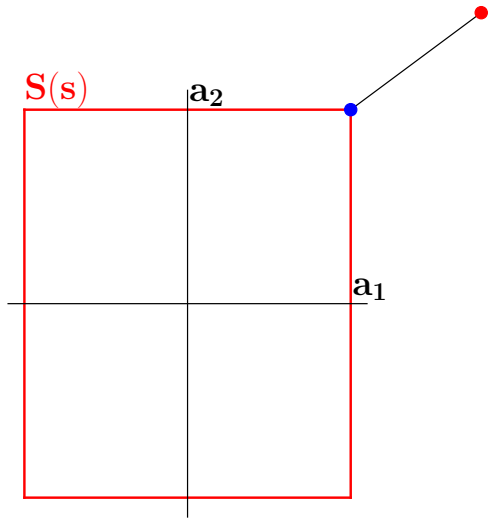
$$\boldsymbol{\pi}^{\star} = \arg\max_{\boldsymbol{\pi}} \quad \mathrm{E}\left[\sum_{\text{time}} \text{Utility}\right]$$

that focuses only on utility

- For state $\mathbf{s}$, take the decisions using: $\text{Action} = \text{Projection}_{\mathrm{S}(\mathbf{s})}\left[\boldsymbol{\pi}^{\star}(\mathbf{s})\right]$

Remarks:

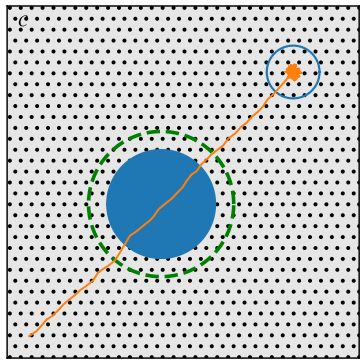
- Often close to practice, e.g., flight envelop protection in modern airplane, semi-autonomous driving in cars, etc
- Oracle $\Rightarrow$ knowledge-based & conservative!



Control Barrier Functions (CBFs)

In words

- CBFs provide a formal, model-based way to build the "oracle" safe set.
- Use a barrier function $h(x)$ so that staying safe means $h(x) \geq 0$.
- At each step, if the proposed action would leave the safe set, a small QP minimally adjusts it to be safe.



Unsafe trajectory: violates $h(x) \geq 0$.

Control Barrier Functions (CBFs)

In words

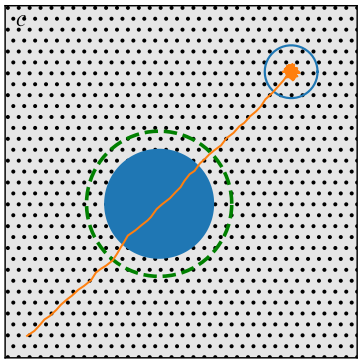
- CBFs provide a formal, model-based way to build the "oracle" safe set.
- Use a barrier function $h(x)$ so that staying safe means $h(x) \geq 0$.
- At each step, if the proposed action would leave the safe set, a small QP minimally adjusts it to be safe.

Formally

$$\dot{x} = f(x) + g(x)u, \qquad \mathcal{C} = \{x : h(x) \geq 0\}$$

$$u^\star(x) = \arg\min_u \tfrac{1}{2}\|u - u_{\text{des}}(x)\|_2^2 \quad \text{s.t.}$$

$$\underbrace{L_f h(x) + L_g h(x)\, u + \alpha(h(x)) \geq 0}_{\text{CBF constraint}}$$



Unsafe trajectory: violates $h(x) \geq 0$.

Control Barrier Functions (CBFs)

In words

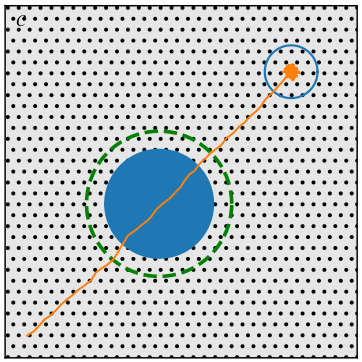
- CBFs provide a formal, model-based way to build the "oracle" safe set.
- Use a barrier function $h(x)$ so that staying safe means $h(x) \geq 0$.
- At each step, if the proposed action would leave the safe set, a small QP minimally adjusts it to be safe.

Formally

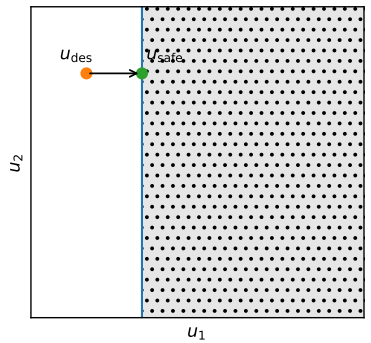
$$\dot{x} = f(x) + g(x)u, \qquad \mathcal{C} = \{x : h(x) \geq 0\}$$

$$u^\star(x) = \arg\min_u \tfrac{1}{2}\|u - u_{\text{des}}(x)\|_2^2 \quad \text{s.t.}$$

$$\underbrace{L_f h(x) + L_g h(x)\, u + \alpha(h(x)) \geq 0}_{\text{CBF constraint}}$$



Unsafe trajectory: violates $h(x) \geq 0$.



CBF filter: project u_{des} to feasible u_{safe}.

Control Barrier Functions (CBFs)

In words

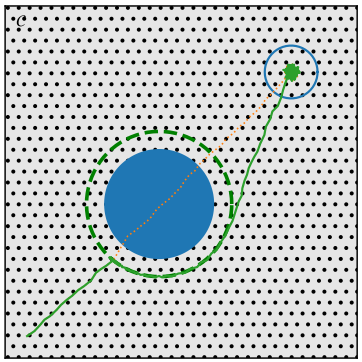
- CBFs provide a formal, model-based way to build the "oracle" safe set.
- Use a barrier function $h(x)$ so that staying safe means $h(x) \geq 0$.
- At each step, if the proposed action would leave the safe set, a small QP minimally adjusts it to be safe.

Formally

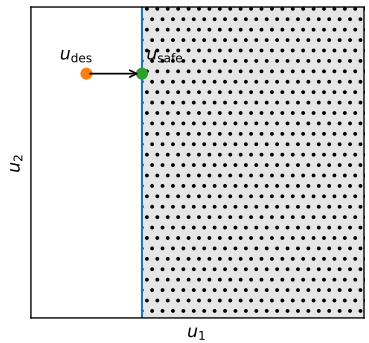
$$\dot{x} = f(x) + g(x)u, \qquad \mathcal{C} = \{x : h(x) \geq 0\}$$

$$u^\star(x) = \arg\min_u \tfrac{1}{2}\|u - u_{\text{des}}(x)\|_2^2 \quad \text{s.t.}$$

$$\underbrace{L_f h(x) + L_g h(x)\, u + \alpha(h(x)) \geq 0}_{\text{CBF constraint}}$$



CBF-filtered trajectory.



CBF filter: project u_{des} to feasible u_{safe}.

- The barrier function and safe set $\mathcal{C}$ are typically **constructed from domain knowledge** (physics, rules, safety envelopes).
- Requires a (possibly simplified, conservative) **system model**: **robust** approach

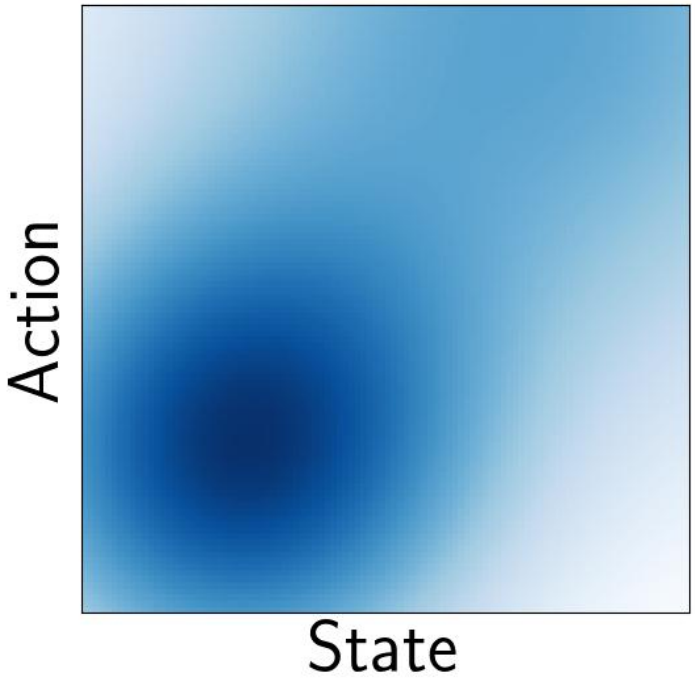
CVaR

Akhil can do on 11.9

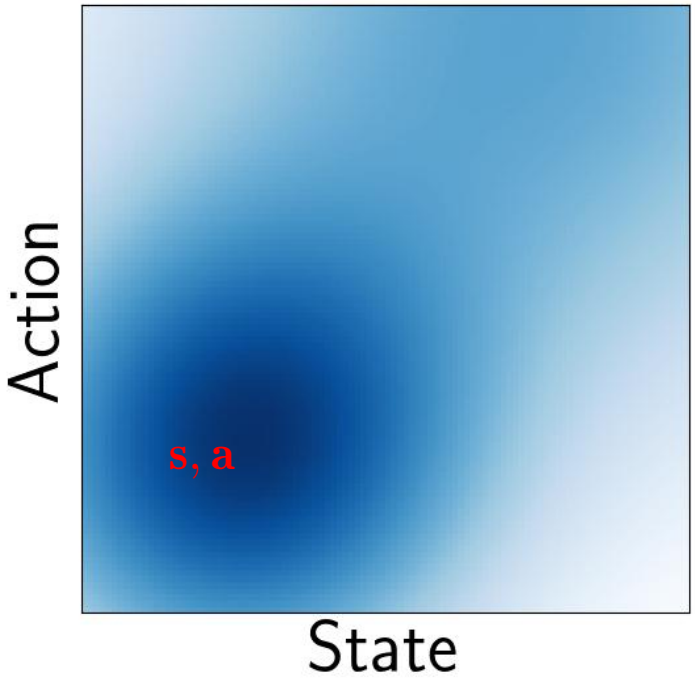
Outline

1. Some Basics of Safe Decision Making
2. Methods
3. Safe Decisions from Data & AI
4. Epistemic Uncertainty and Safe Decisions

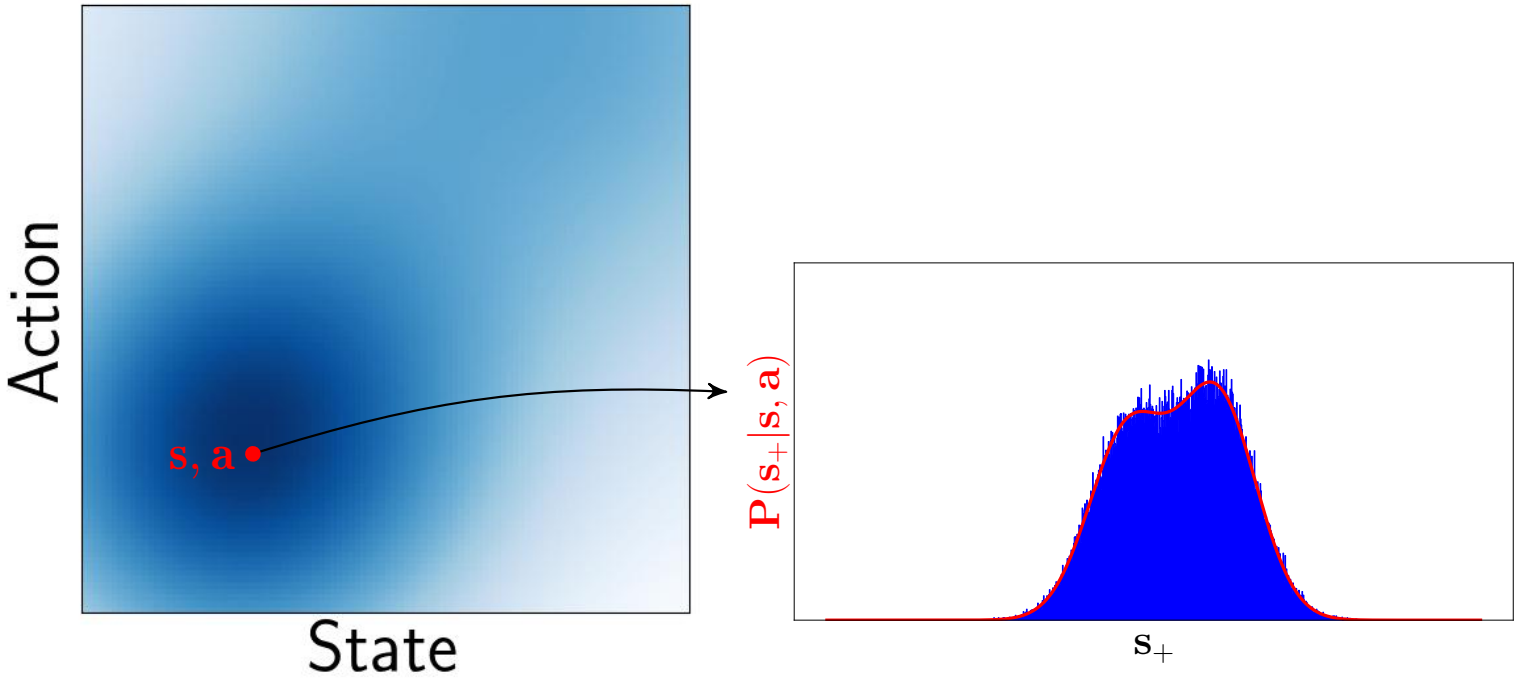
Data & Epistemic Uncertainty



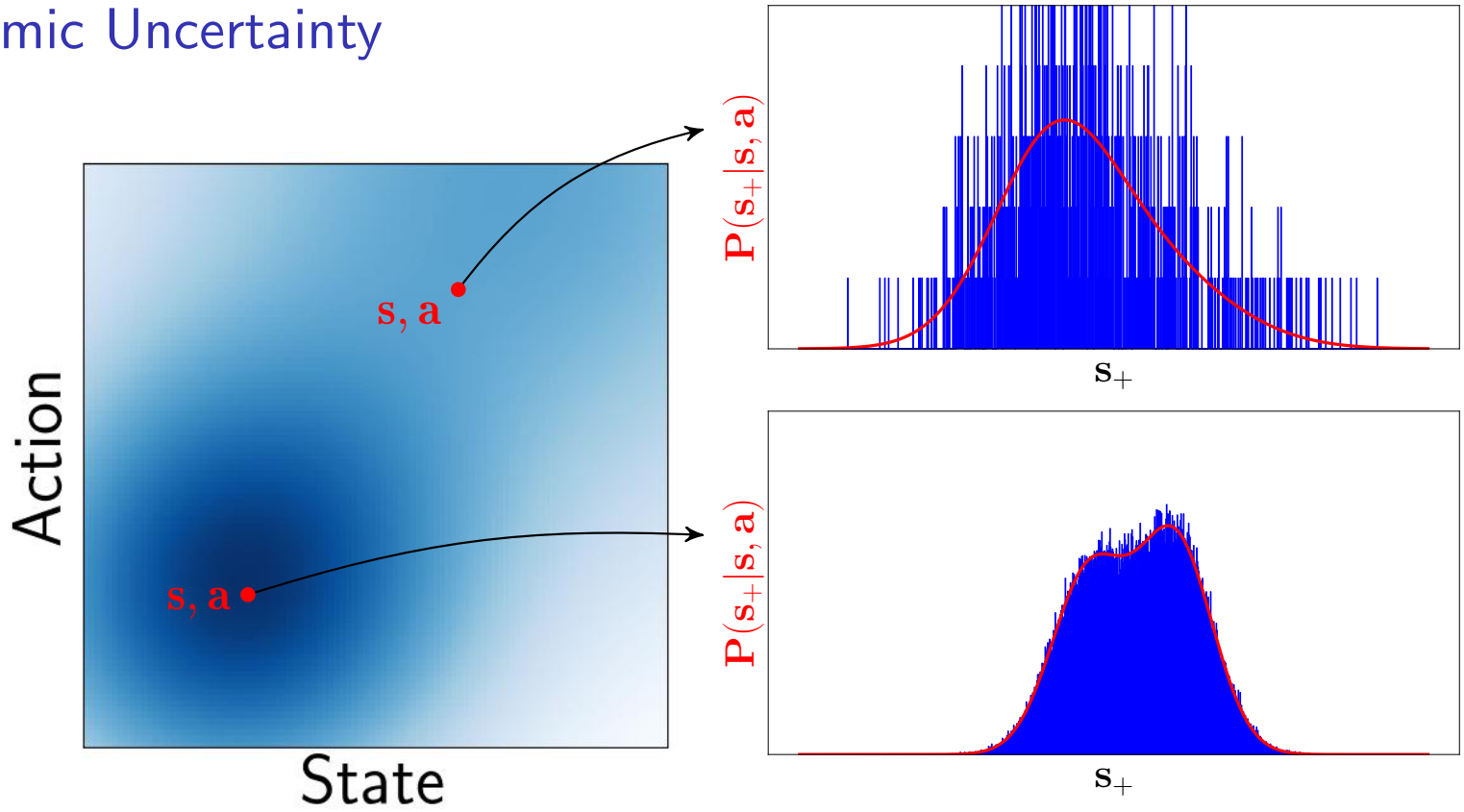
Data & Epistemic Uncertainty



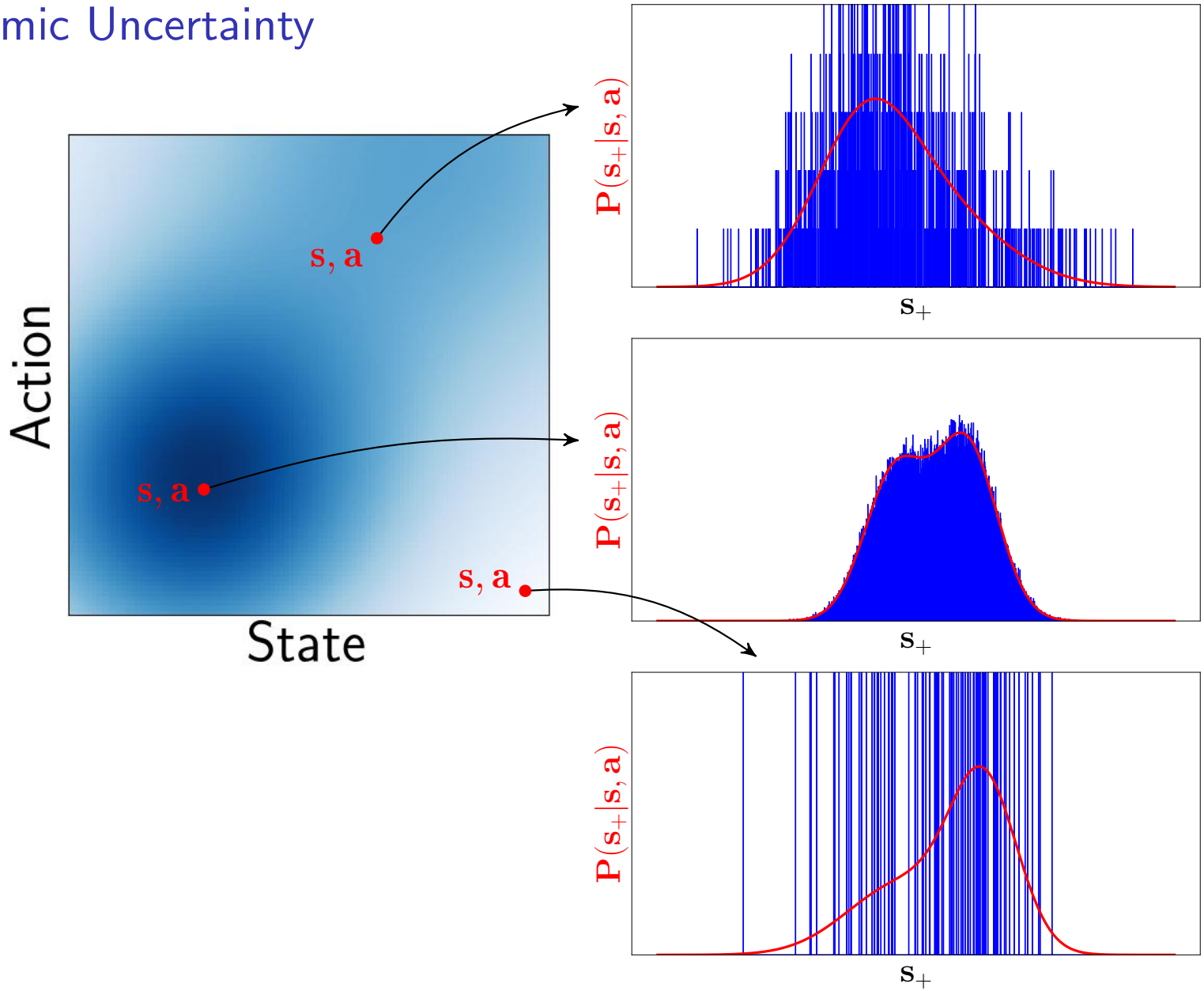
Data & Epistemic Uncertainty



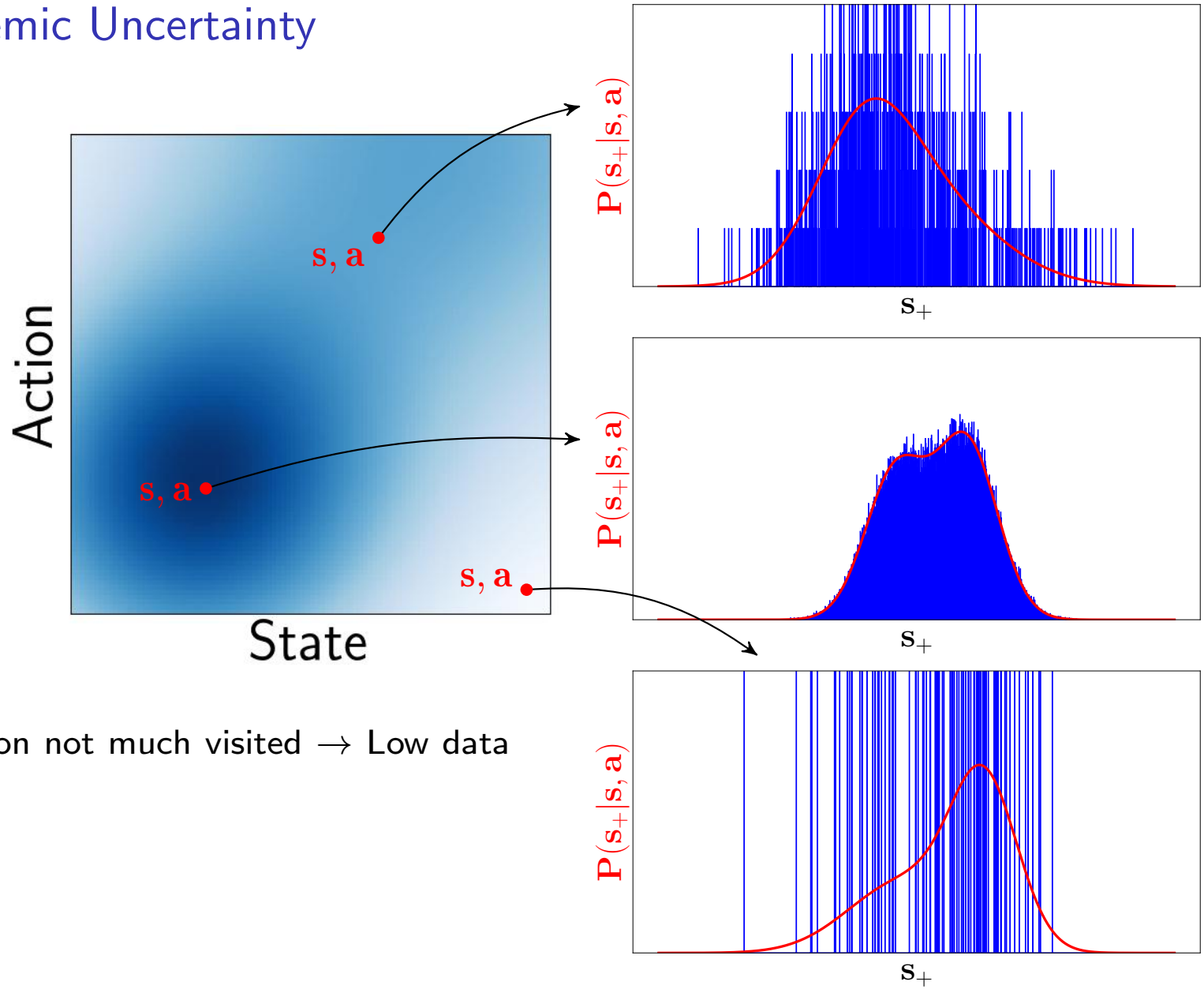
Data & Epistemic Uncertainty



Data & Epistemic Uncertainty



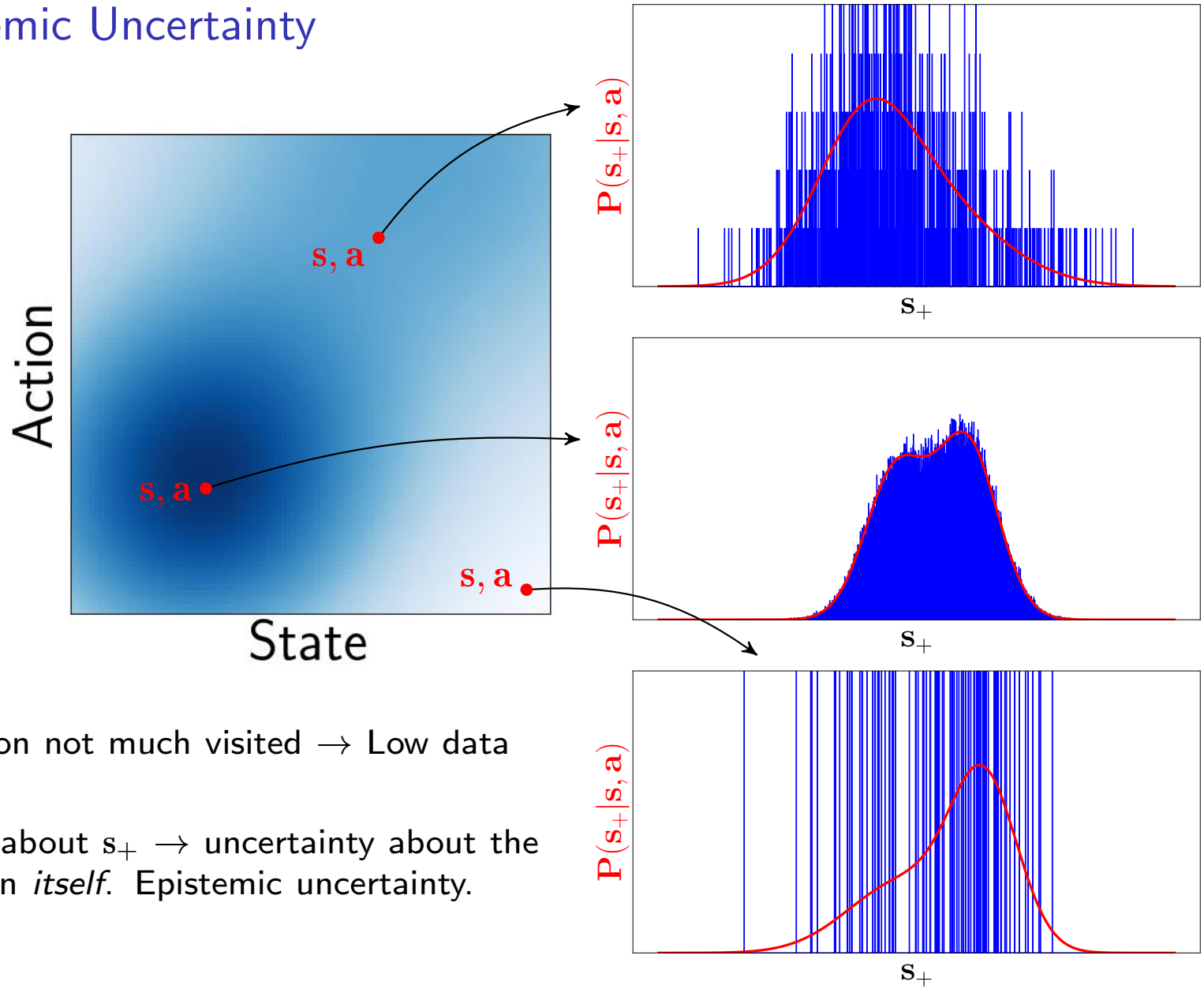
Data & Epistemic Uncertainty



Remarks

- State-action not much visited $\rightarrow$ Low data about s_+

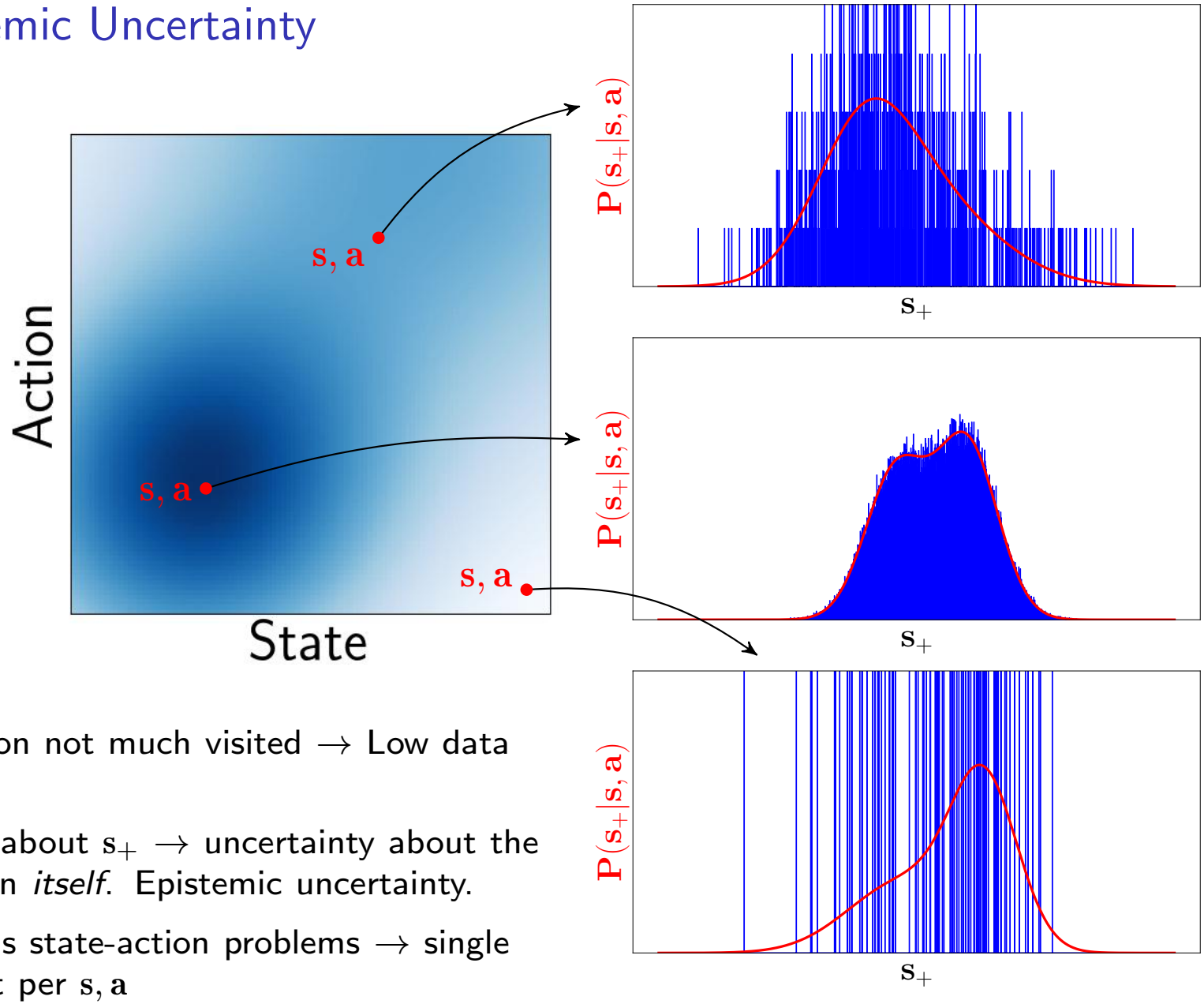
Data & Epistemic Uncertainty



Remarks

- State-action not much visited $\rightarrow$ Low data about s_+
- Low data about s_+ $\rightarrow$ uncertainty about the distribution *itself*. Epistemic uncertainty.

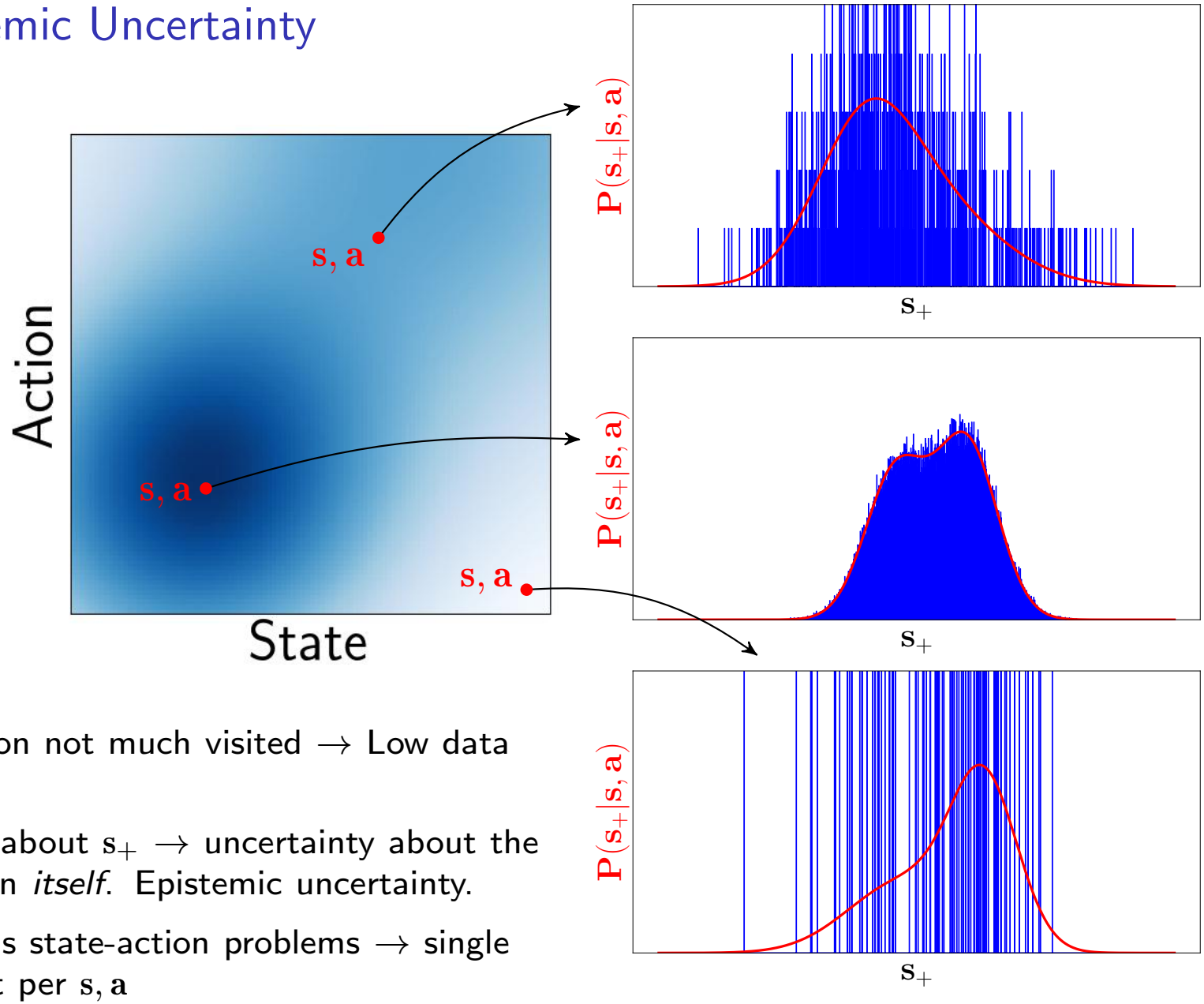
Data & Epistemic Uncertainty



Remarks

- State-action not much visited $\rightarrow$ Low data about $\mathbf{s}_+$
- Low data about $\mathbf{s}_+$ $\rightarrow$ uncertainty about the distribution *itself*. Epistemic uncertainty.
- Continuous state-action problems $\rightarrow$ single data point per $\mathbf{s}, \mathbf{a}$

Data & Epistemic Uncertainty



Remarks

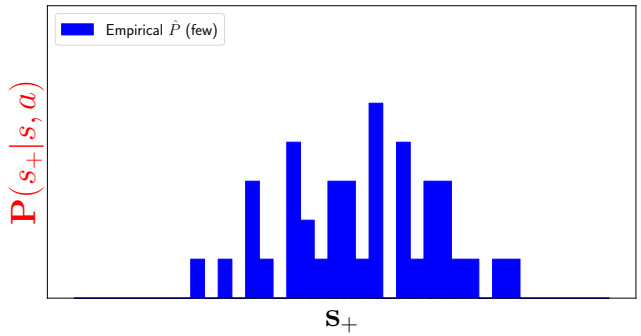
- State-action not much visited $\rightarrow$ Low data about $\mathbf{s}_+$
- Low data about $\mathbf{s}_+$ $\rightarrow$ uncertainty about the distribution *itself*. Epistemic uncertainty.
- Continuous state-action problems $\rightarrow$ single data point per $\mathbf{s}, \mathbf{a}$
- Rely on regularity across $\mathbf{s}, \mathbf{a}$!!

Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.

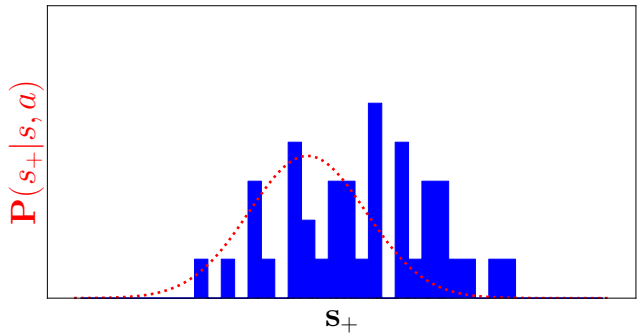
Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.



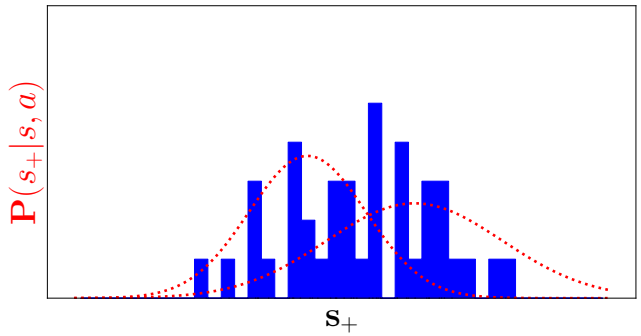
Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.



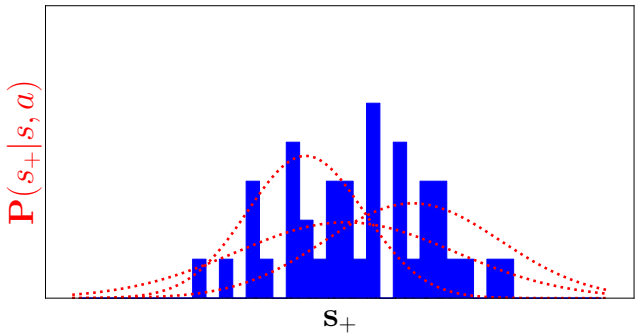
Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.



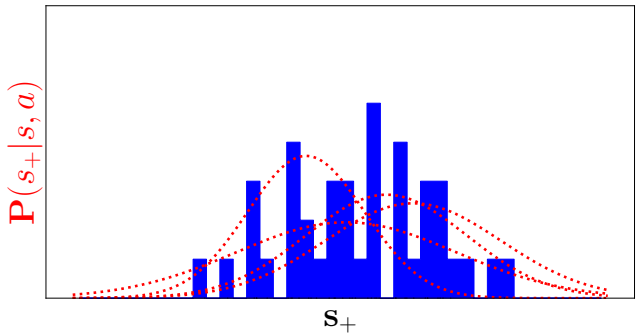
Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.
- Instead of a single density, use a distribution of densities.



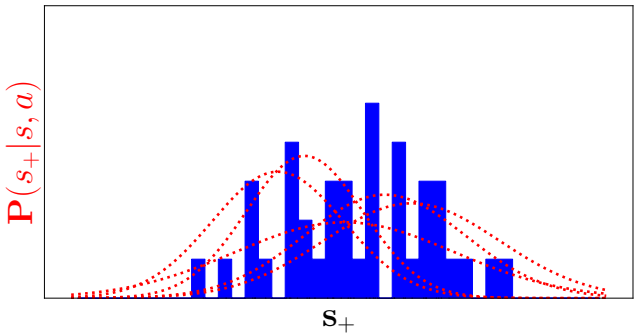
Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.
- Instead of a single density, use a distribution of densities.



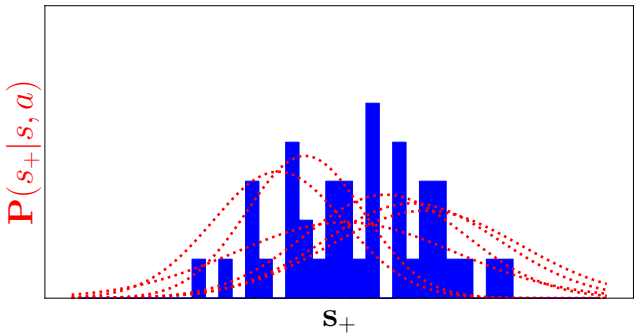
Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.
- Instead of a single density, use a distribution of densities.



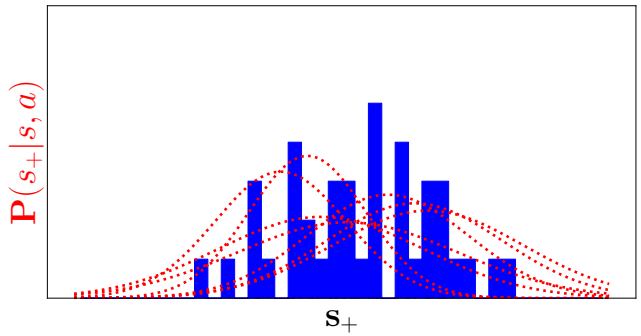
Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.
- Instead of a single density, use a distribution of densities.



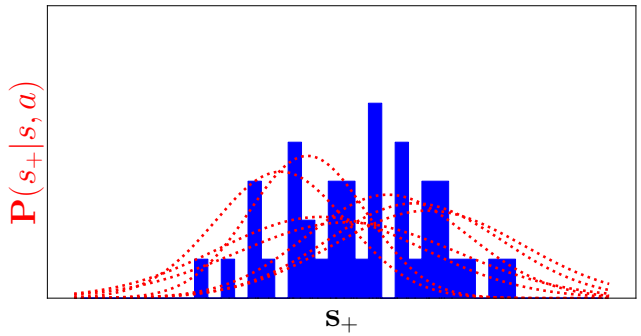
Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.
- Instead of a single density, use a distribution of densities.



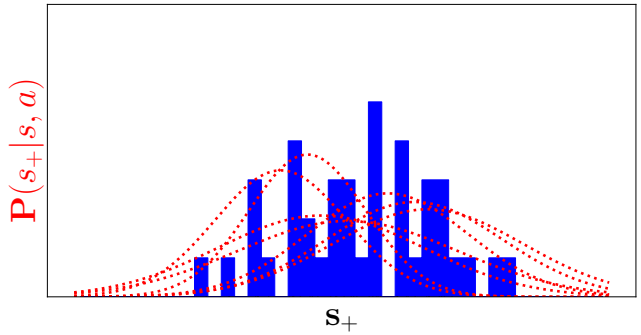
Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.
- Instead of a single density, use a distribution of densities.



Distributional View on MDPs

- True dynamics are unknown; only samples or estimates available.
- Instead of a single density, use a distribution of densities.



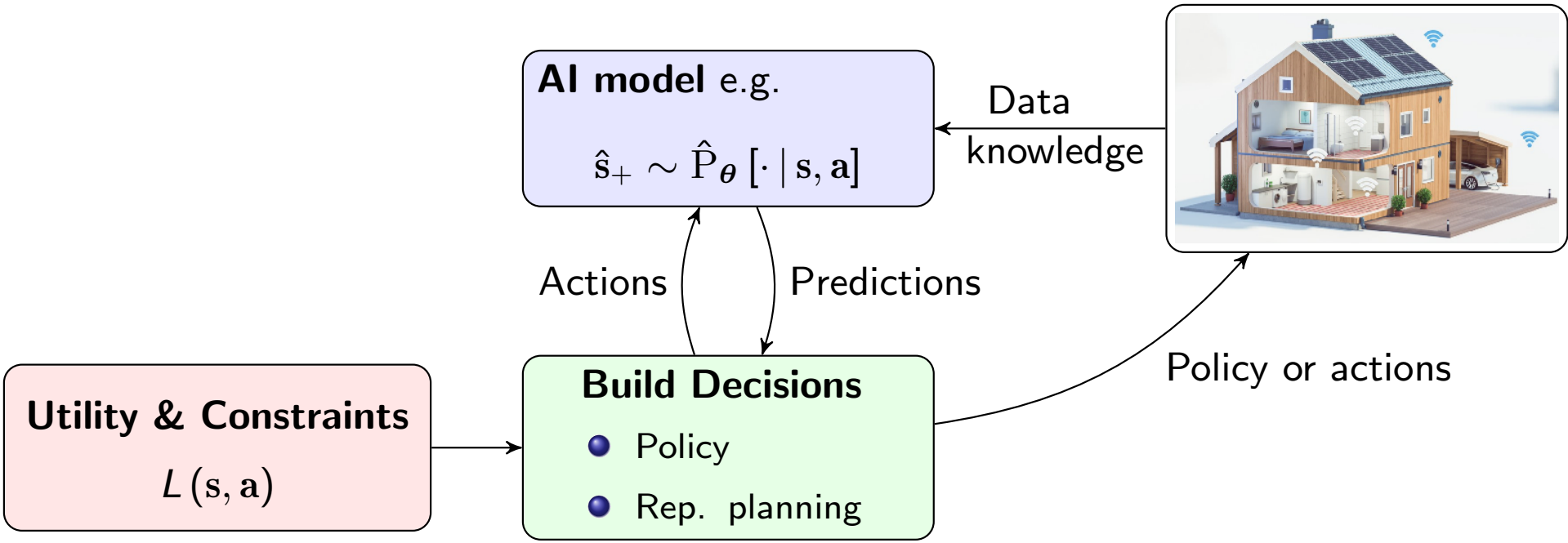
Remarks:

- Accounts for **epistemic uncertainty** from limited data.
- Distributional robust MDP: less pessimistic than robust MDP.
- Probabilistic Safety: confidence level based on model undertainty??
- Distributional view is applied to risk-sensitive MDP via a distribution of returns (e.g distributional RL).
- connection to bayesian vs freequentist view??

Out-Of-Distribution (OOD) "Guard" Methods

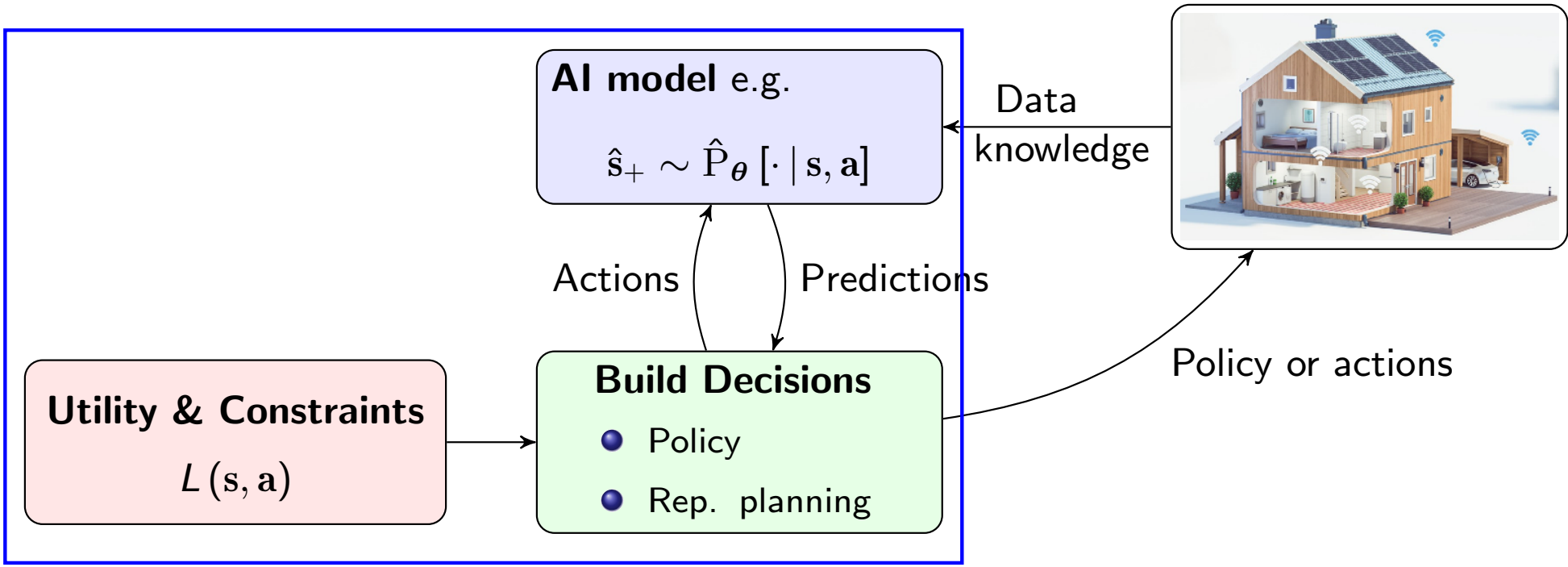
SG can do on 11.9

RL over Decision Making



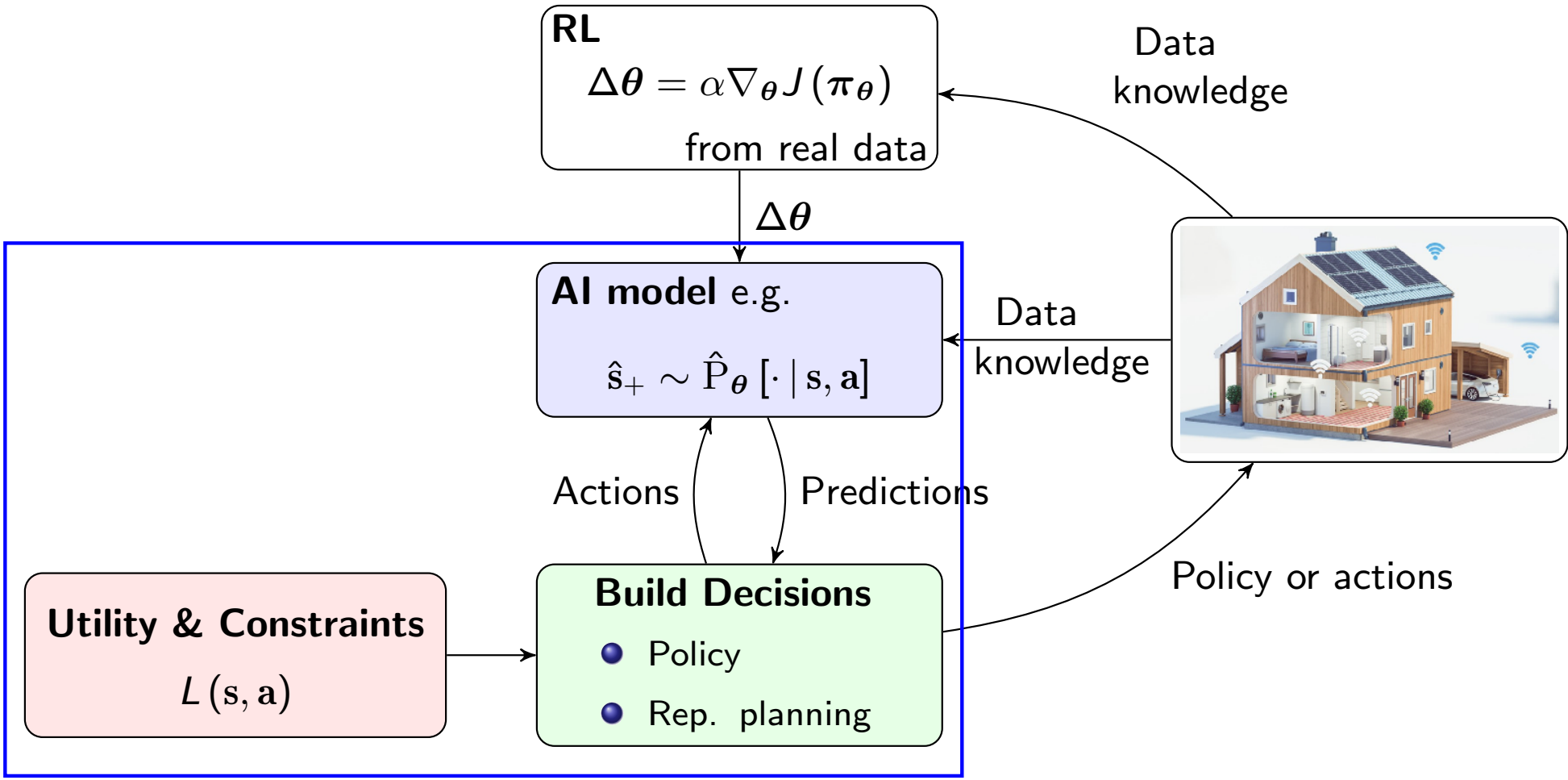
- The model is not $\hat{\mathrm{P}}_{\boldsymbol{\theta}}$, it is the entire "decision box"
- Utility & Constraints to build decisions from the predictions are part of the model
- RL can tune the whole decision box, from real data + true Utility & Constraints

RL over Decision Making



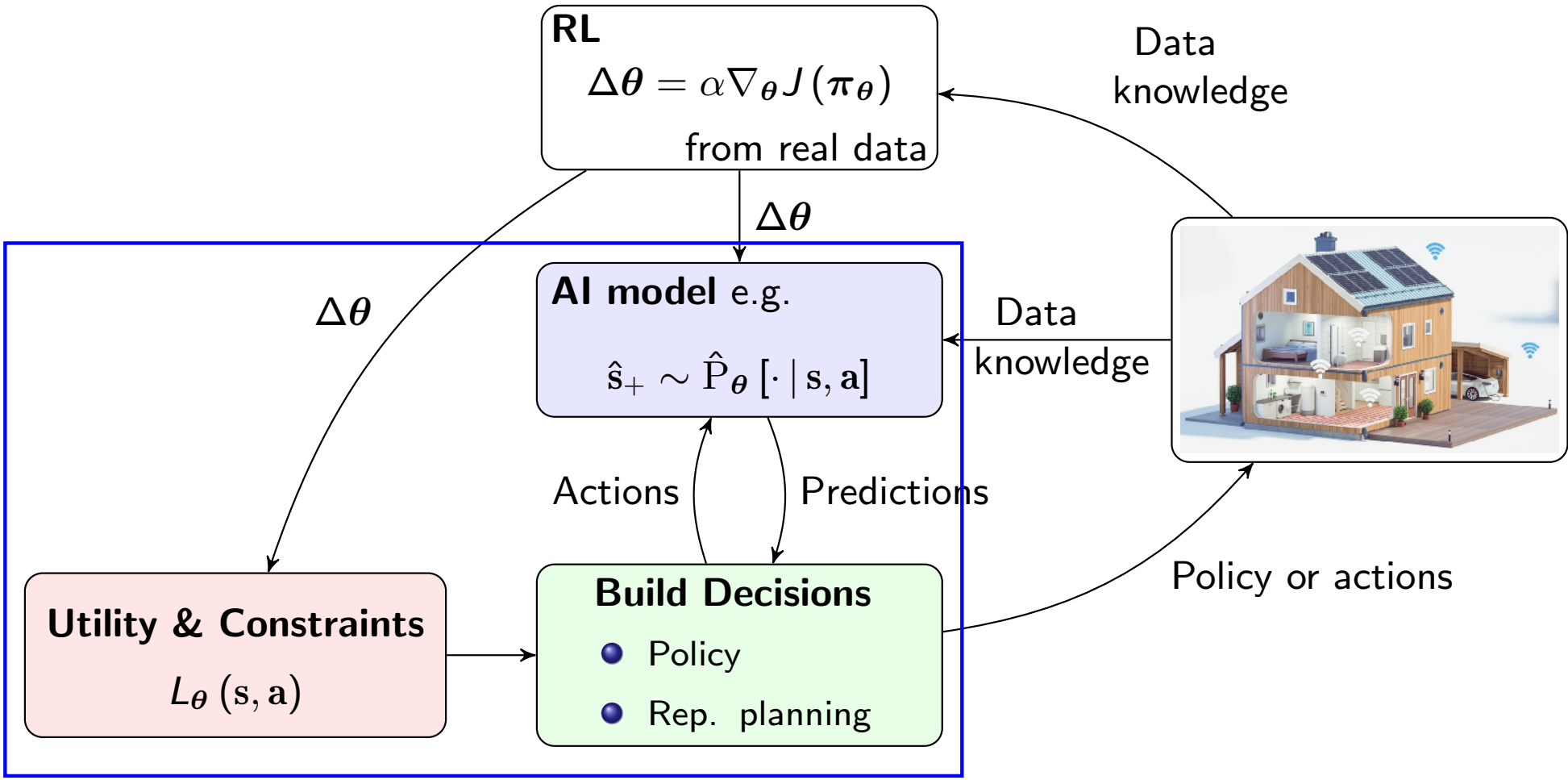
- The model is not $\hat{\mathrm{P}}_{\boldsymbol{\theta}}$, it is the entire "decision box"
- Utility & Constraints to build decisions from the predictions are part of the model
- RL can tune the whole decision box, from real data + true Utility & Constraints

RL over Decision Making



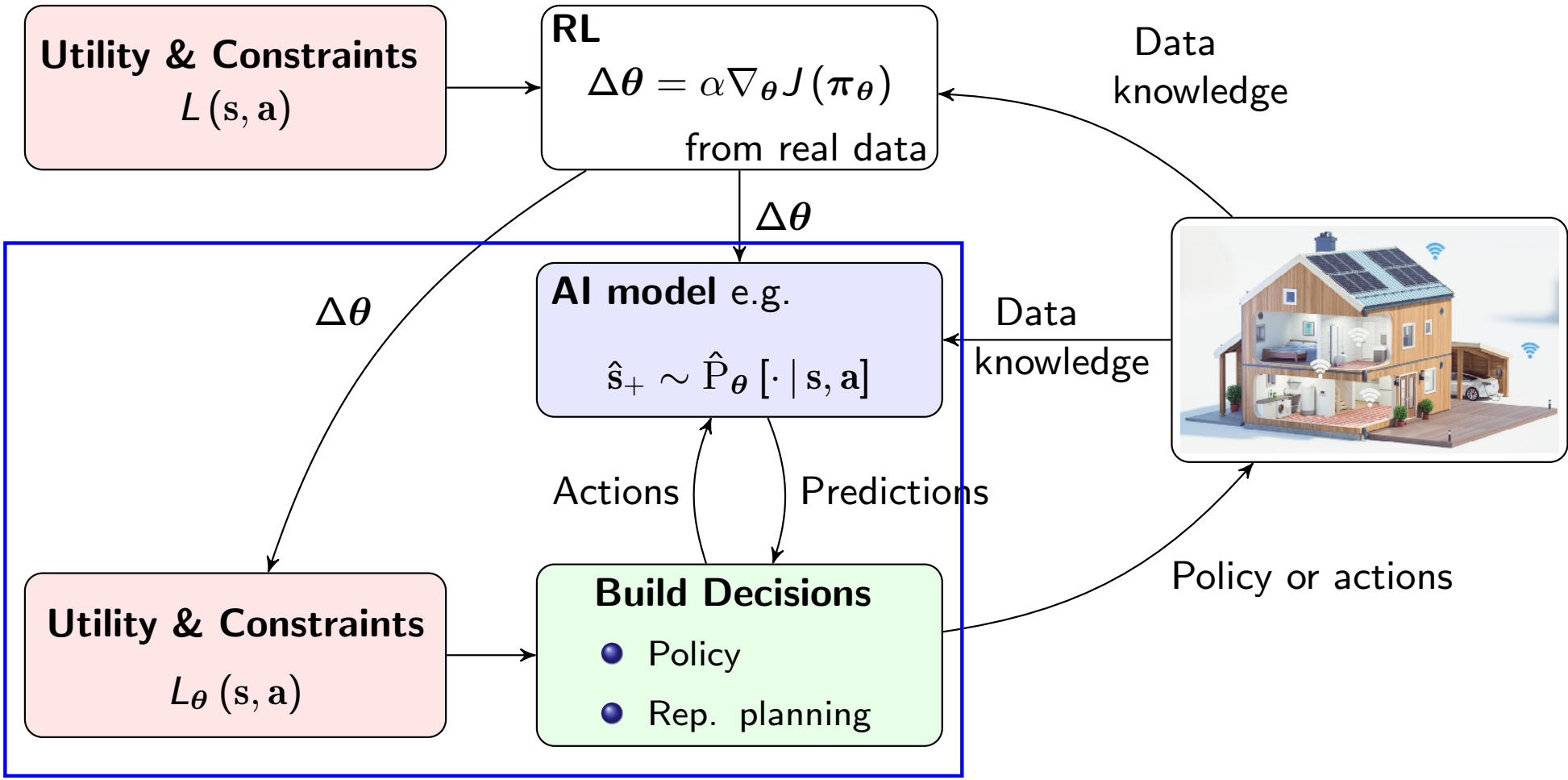
- The model is not $\hat{\mathrm{P}}_{\boldsymbol{\theta}}$, it is the entire "decision box"
- Utility & Constraints to build decisions from the predictions are part of the model
- RL can tune the whole decision box, from real data + true Utility & Constraints

RL over Decision Making



- The model is not $\hat{\mathrm{P}}_{\boldsymbol{\theta}}$, it is the entire "decision box"
- Utility & Constraints to build decisions from the predictions are part of the model
- RL can tune the whole decision box, from real data + true Utility & Constraints

RL over Decision Making



- The model is not $\hat{\mathrm{P}}_{\boldsymbol{\theta}}$, it is the entire "decision box"
- Utility & Constraints to build decisions from the predictions are part of the model
- RL can tune the whole decision box, from real data + true Utility & Constraints