Efficient Importance Scenario Generation for Optimisation with Rare Events

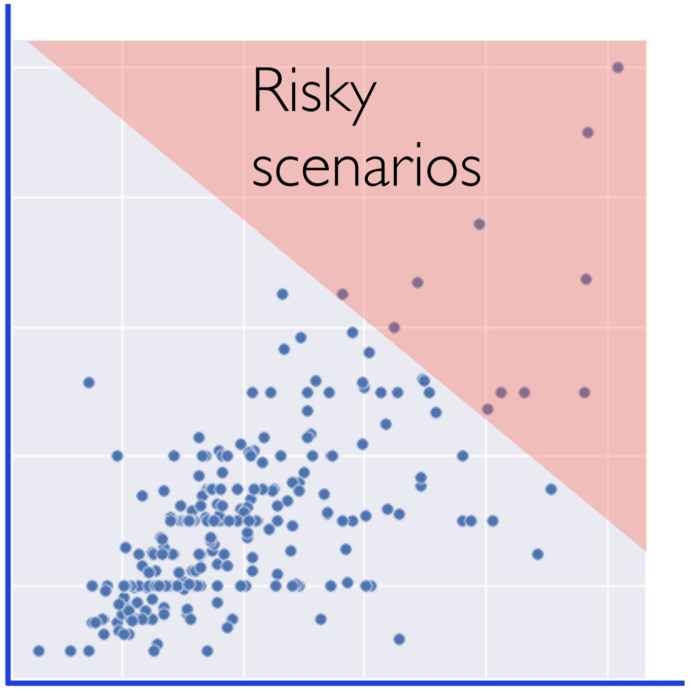
Based on work with Karthyek Murthy

Anand Deo - Indian Institute of Management Bangalore

Optimisation formulations that incorporate extreme risks:

- Tail Risk Measure
- Depends on the most risky β fraction of scenarios

$$\inf_{\theta \in \Theta} \rho(\theta, P^*)$$



$P^* \rightarrow$ Data generating distribution

$\rho \rightarrow$ Tail Risk Functional

$\theta \rightarrow$ Decision

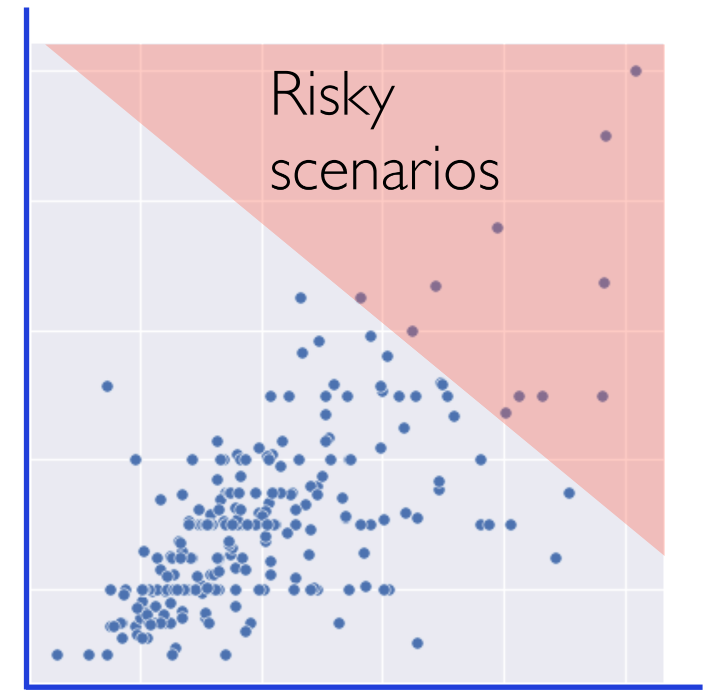
$\Theta \rightarrow$ Service constraint (avg. performance)

Optimisation formulations that incorporate extreme risks:

- Tail Risk Measure
- Depends on the most risky β fraction of scenarios

$\beta \sim 10^{-3}$

$$\inf_{\theta \in \Theta} \; \rho(\theta, P^*)$$



$P^* \rightarrow$ Data generating distribution

$\rho \rightarrow$ Tail Risk Functional

$\theta \rightarrow$ Decision

$\Theta \rightarrow$ Service constraint (avg. performance)

Optimisation formulations that incorporate extreme risks:

1) Risk Minimisation

$$\min_{\theta \in \Theta} \text{CVaR}_{1-\beta}[L(X, \theta)]$$

Θ is a set of service requirements

Portfolio Optimisation

$$L(X, \theta) = -\theta^\intercal \boldsymbol{X}$$

$$\Theta = \{\theta : \mu^\intercal \theta \geq r\}$$

$\mu \rightarrow$ mean vector of returns

$r \rightarrow$ Target return

Optimisation formulations that incorporate extreme risks:

1) Risk Minimisation

$$\min_{\theta \in \Theta} \text{CVaR}_{1-\beta}[L(X,\theta)]$$

Θ is a set of service requirements

Portfolio Optimisation

$$L(X,\theta) = -\theta^\intercal X$$

$$\Theta = \{\theta : \mu^\intercal \theta \geq r\}$$

$\mu \rightarrow$ mean vector of returns

$r \rightarrow$ Target return

2) Mean-CVaR Optimisation

$$\min_{\theta \in \Theta} \lambda E[L(\boldsymbol{X},\theta)] + (1-\lambda)\text{CVaR}_{1-\beta}[L(\boldsymbol{X},\theta)]$$

$\lambda \in [0,1] \rightarrow$ Risk appetite.

Two stage problem

$$L(\boldsymbol{X},\theta) = \boldsymbol{c}^\intercal \theta + Q(\boldsymbol{X},\theta)$$

where

$$Q(\boldsymbol{x},\theta) = \inf_{y \in S(\theta)} y^\intercal \boldsymbol{x}$$

Optimisation formulations that incorporate extreme risks:

1) Risk Minimisation

$$\min_{\theta \in \Theta} \text{CVaR}_{1-\beta}[L(X,\theta)]$$

Θ is a set of service requirements

Portfolio Optimisation

$$L(X,\theta) = -\theta^\intercal X$$

$$\Theta = \{\theta : \mu^\intercal \theta \geq r\}$$

$\mu \rightarrow$ mean vector of returns

$r \rightarrow$ Target return

2) Mean-CVaR Optimisation

$$\min_{\theta \in \Theta} \lambda E[L(X,\theta)] + (1-\lambda)\text{CVaR}_{1-\beta}[L(X,\theta)]$$

$\lambda \in [0,1] \rightarrow$ Risk appetite.

Two stage problem

$$L(X,\theta) = c^\intercal \theta + Q(X,\theta)$$

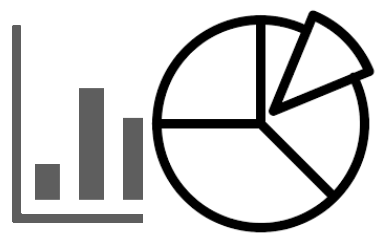
where

$$Q(x,\theta) = \inf_{y \in S(\theta)} y^\intercal x$$

Find a risk minimising decision: $\min_{\theta:S(\theta)\geq r} \textbf{CVaR}_{1-\beta}[L(X,\theta)]$

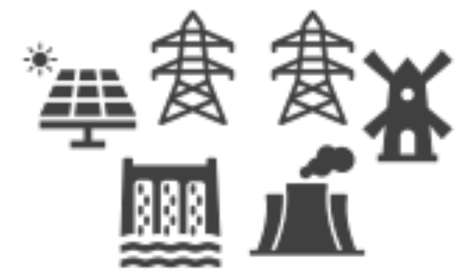
Minimise the tail risk of a loan portfolio while meeting target return guarantees

Rockafellar and Uryasev 2002, Krohkhmal et al. 2002,…



Managing power operations subject to line failures, supply and demand fluctuations

Bienstock et al. 2014, Summers et al. 2015,…

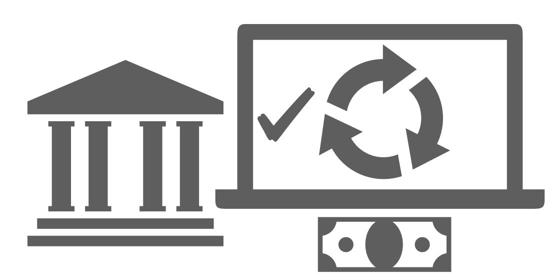


Manage supply and price risk in service operations

Blanchet et al. 2019

Finding classifiers which don't penalise minority subpopulations

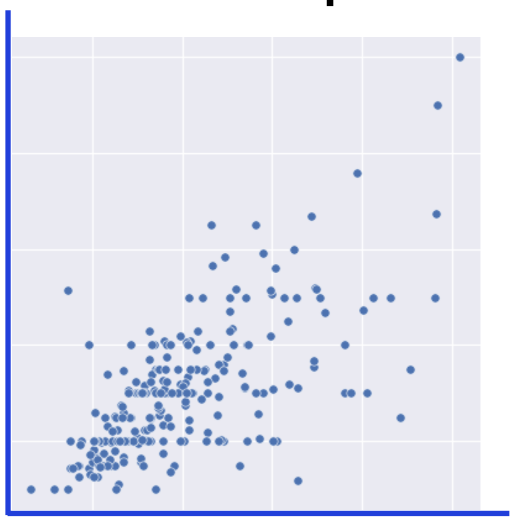
Williamson and Menon, 2019…

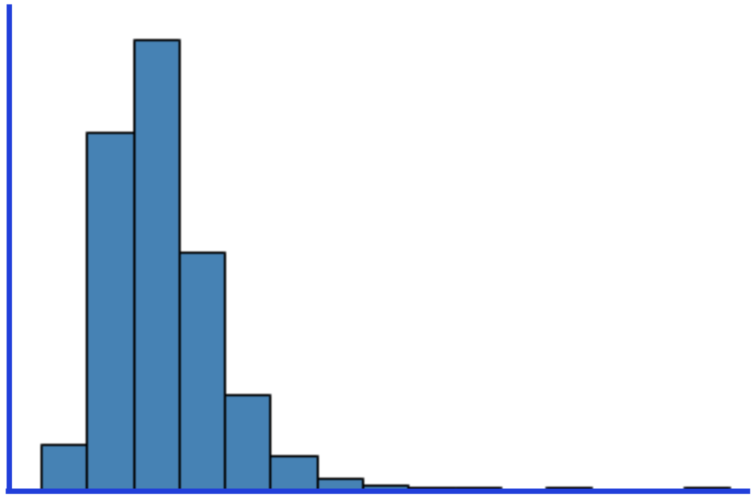


Conventional Stochastic Optimisation

$$\min_{\theta \in \Theta} E[L(X, \theta)]$$

i.i.d. samples $\Longrightarrow$ Loss realisations





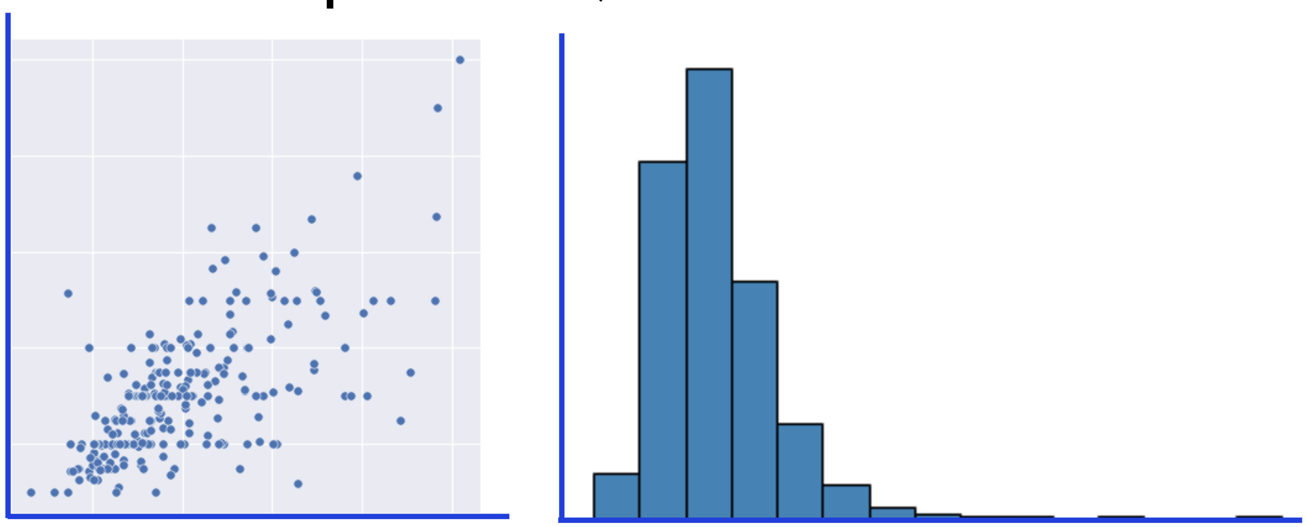
Minimise Sample Averages

$$\min_{\theta \in \Theta} \frac{1}{n} \sum_{i \leq n} L(X_i, \theta)$$

Conventional Stochastic Optimisation

$$\min_{\theta \in \Theta} E[L(X, \theta)]$$

i.i.d. samples $\Longrightarrow$ Loss realisations

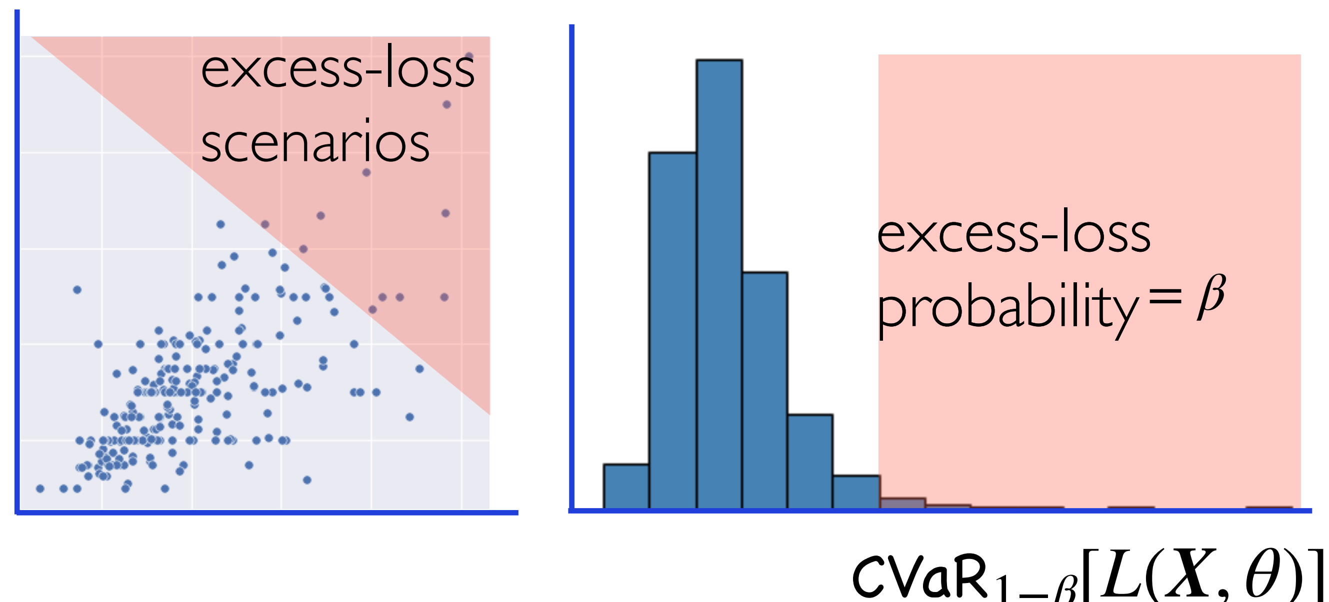


Minimise Sample Averages

$$\min_{\theta \in \Theta} \frac{1}{n} \sum_{i \leq n} L(X_i, \theta)$$

Optimisation with tail risk measures

$$\min_{\theta \in \Theta} \mathrm{CVaR}_{1-\beta}[L(X, \theta)]$$



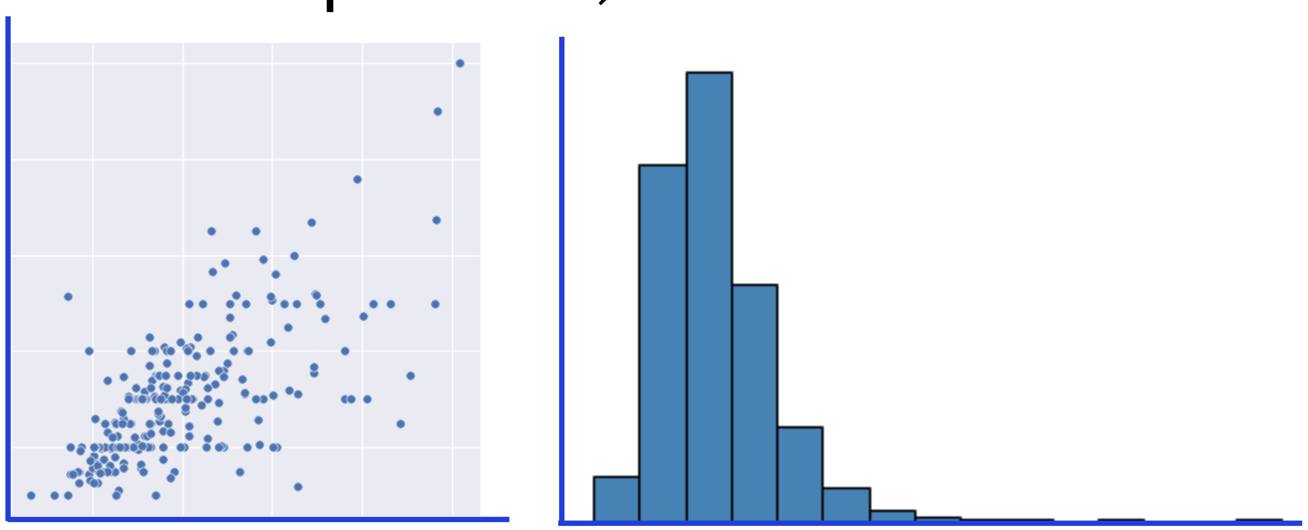
Curse of Rarity

Lack of representative tail samples
(Lim et al. 2011, Caccioli et al. 2018)

Conventional Stochastic Optimisation

$$\min_{\theta \in \Theta} E[L(X, \theta)]$$

i.i.d. samples $\Longrightarrow$ Loss realisations

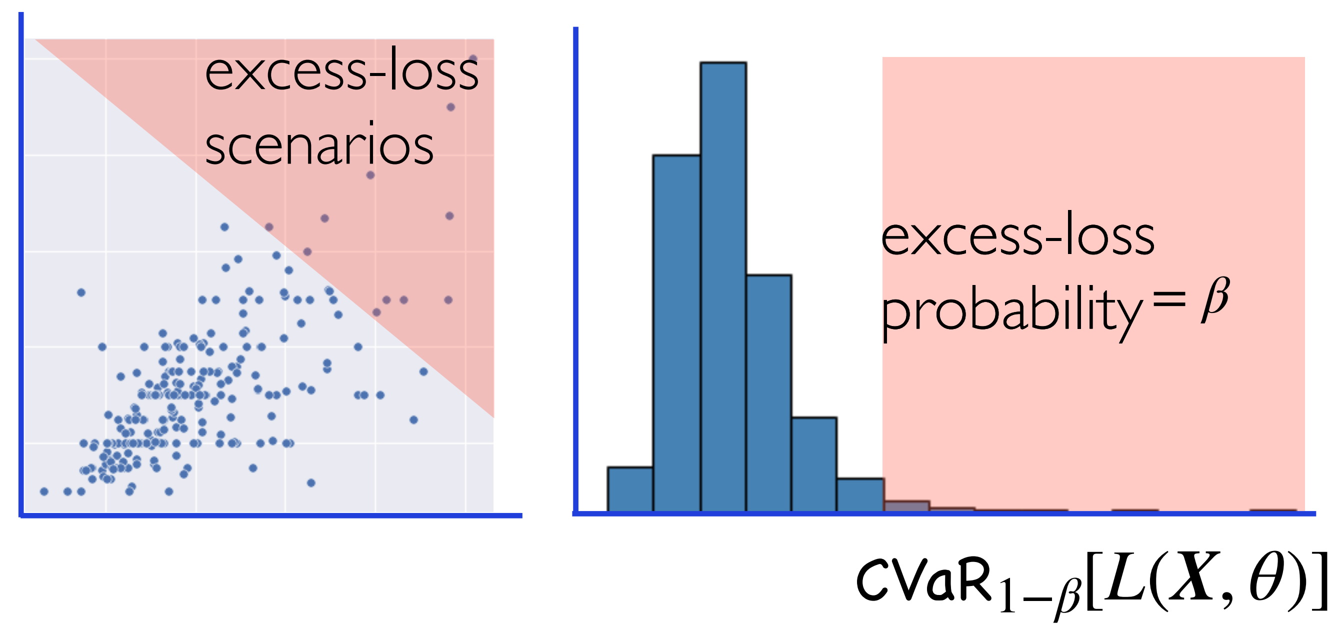


Minimise Sample Averages

$$\min_{\theta \in \Theta} \frac{1}{n} \sum_{i \leq n} L(X_i, \theta)$$

Optimisation with tail risk measures

$$\min_{\theta \in \Theta} \mathrm{CVaR}_{1-\beta}[L(X, \theta)]$$



Curse of Rarity

~ 30 years of data required for CVaR optimisation at $\beta = 0.1$

Today's Talk

CVaR minimisation: $$\min_{u \geq 0,\ \theta \in \Theta} \left(u + \beta^{-1} E[L(X,\theta) - u]^{+} \right)$$

Assumption: **The distribution of X is known**

Reduction in Sample Requirement

Adaptive Importance Sampling Paradigm

Algorithm is scalable

Today's Talk

CVaR minimisation: $$\min_{u \geq 0,\ \theta \in \Theta} \left(u + \beta^{-1} E[L(X,\theta) - u]^+ \right)$$

Assumption: **The distribution of X is known**

Reduction in Sample Requirement

Adaptive Importance Sampling Paradigm

Algorithm is scalable

Requires only the tuning of a single hyper parameter

Achieves the "gold-standard" in terms of reduction of sample requirement

Today's Talk

CVaR minimisation: $\min_{u \geq 0,\ \theta \in \Theta} \left(u + \beta^{-1} E[L(X, \theta) - u]^{+} \right)$

Assumption: **The distribution of X is known**

`Data driven setting: Work in progress`

Reduction in Sample Requirement

Requires only the tuning of a single hyper parameter

Adaptive Importance Sampling Paradigm

Achieves the "gold-standard" in terms of reduction of sample requirement

Algorithm is scalable

Today's Talk

CVaR minimisation: $\min_{u \geq 0,\ \theta \in \Theta} \left(u + \beta^{-1} E[L(X,\theta) - u]^{+} \right)$

Assumption: **The distribution of X is known**

Data driven setting: Work in progress

Reduction in Sample Requirement

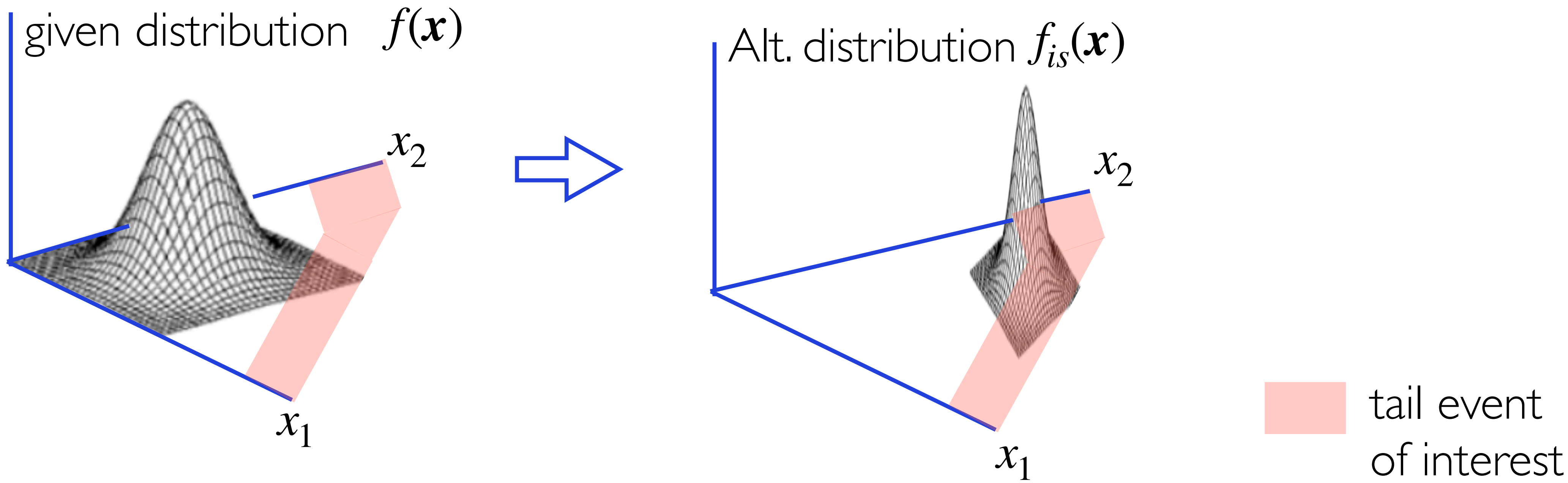
Adaptive Importance Sampling Paradigm

Algorithm is scalable

Requires only the tuning of a single hyper parameter

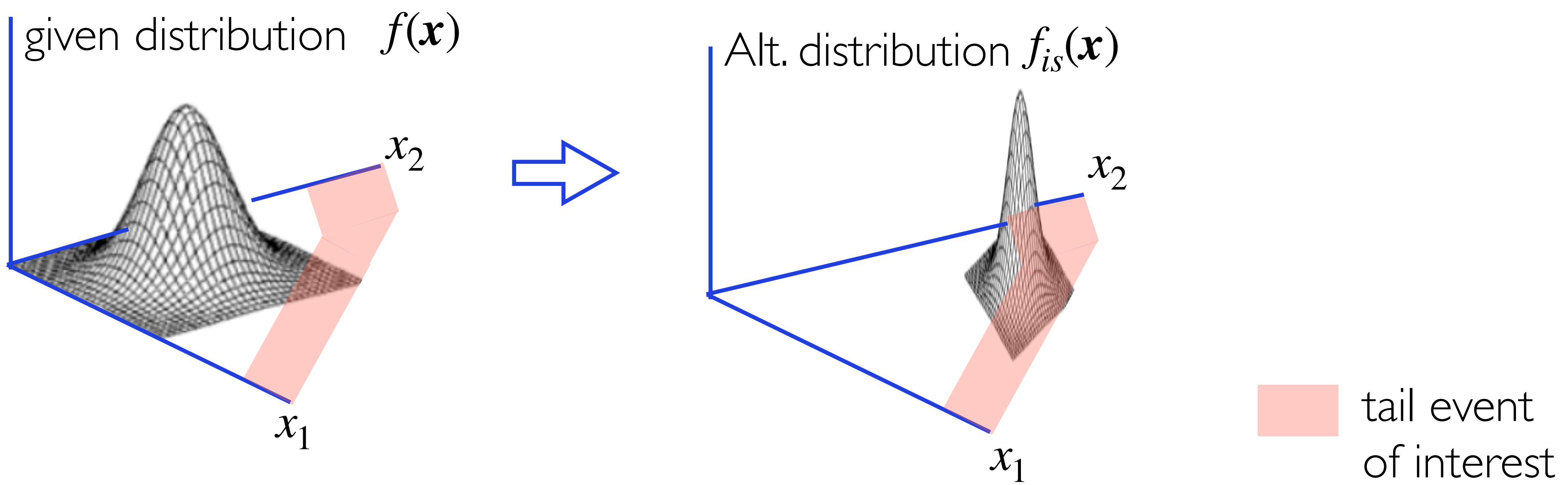
Achieves the "gold-standard" in terms of reduction of sample requirement

Importance Sampling



CVaR minimisation: $\min_{u \geq 0,\ \theta \in \Theta} \left(u + \beta^{-1} E[L(X, \theta) - u]^{+} \right)$

Importance Sampling

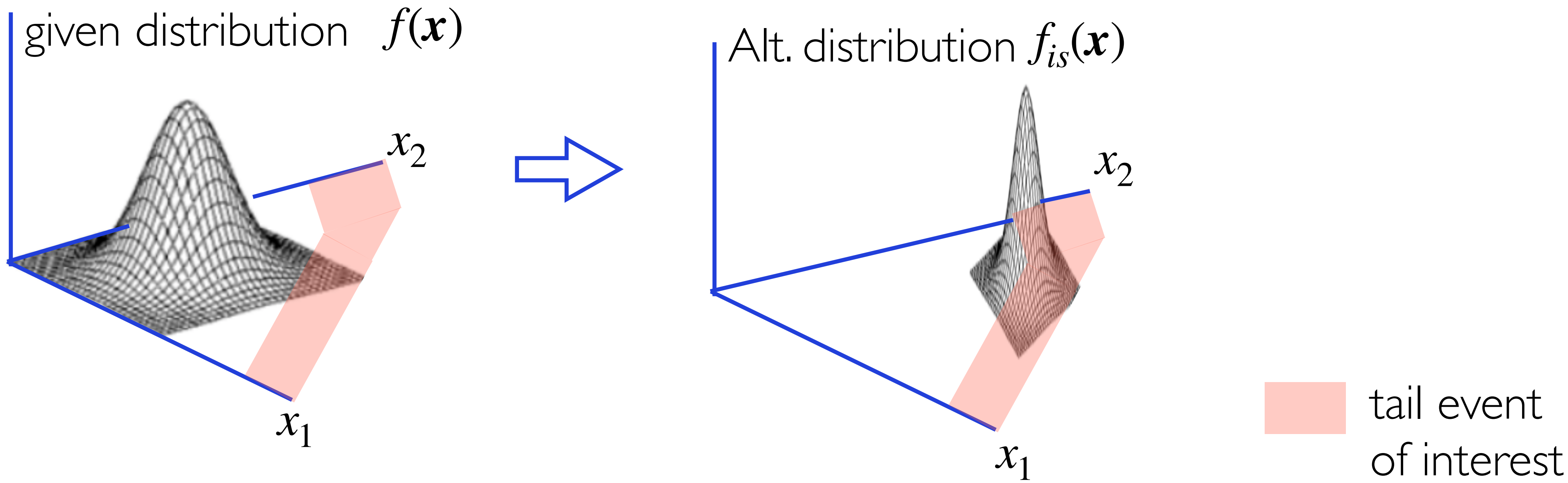


+

Reweigh objective to eliminate bias

CVaR minimisation: $\min_{u \geq 0,\ \theta \in \Theta} \left(u + \beta^{-1} E \left([L(Z,\theta) - u]^{+} \mathscr{L} \right) \right)$

Importance Sampling

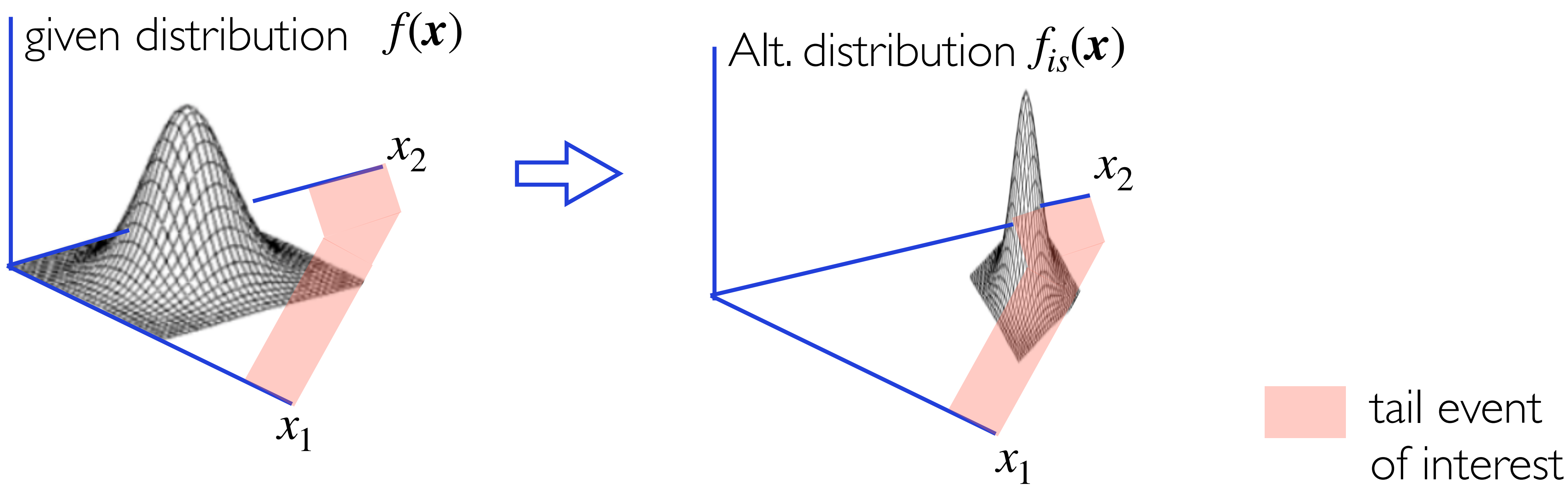


+

Reweigh objective to eliminate bias

CVaR minimisation: $$\min_{u \geq 0,\ \theta \in \Theta} \left(u + \beta^{-1} E \left([L(Z,\theta) - u]^{+} \mathscr{L} \right) \right)$$

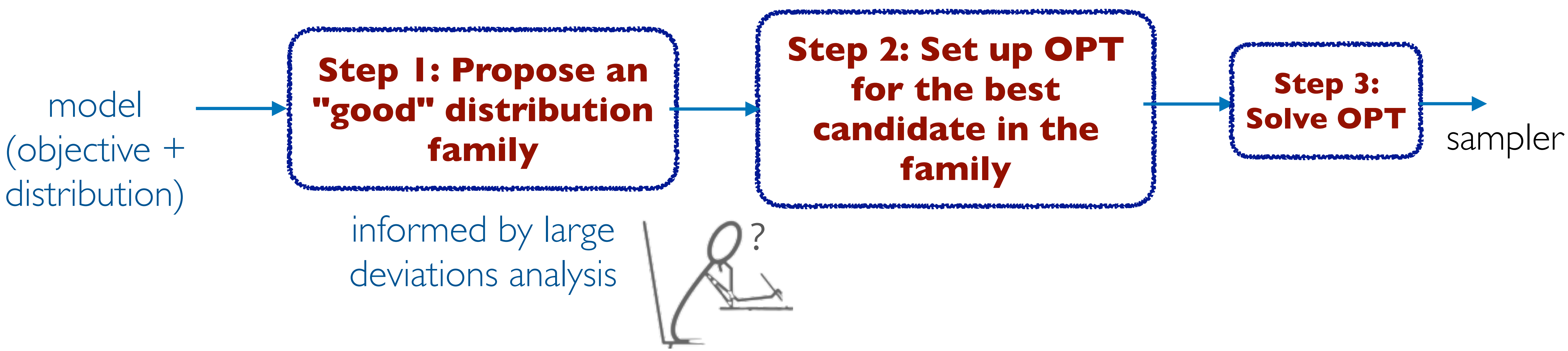
Importance Sampling



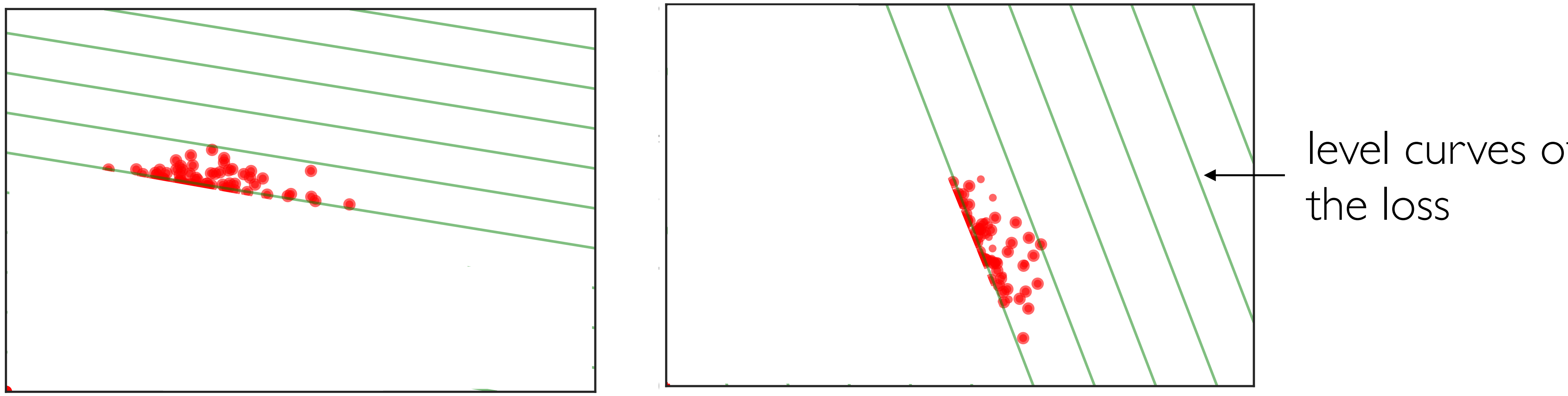
model (objective + distribution) → **Step 1: Propose an "good" distribution family** → **Step 2: Set up OPT for the best candidate in the family** → **Step 3: Solve OPT** → sampler

informed by large deviations analysis

Traditional Work Flow



CVaR minimisation: $\min_{\theta} \text{CVaR}_{1-\beta}[L(\boldsymbol{X}, \theta)]$

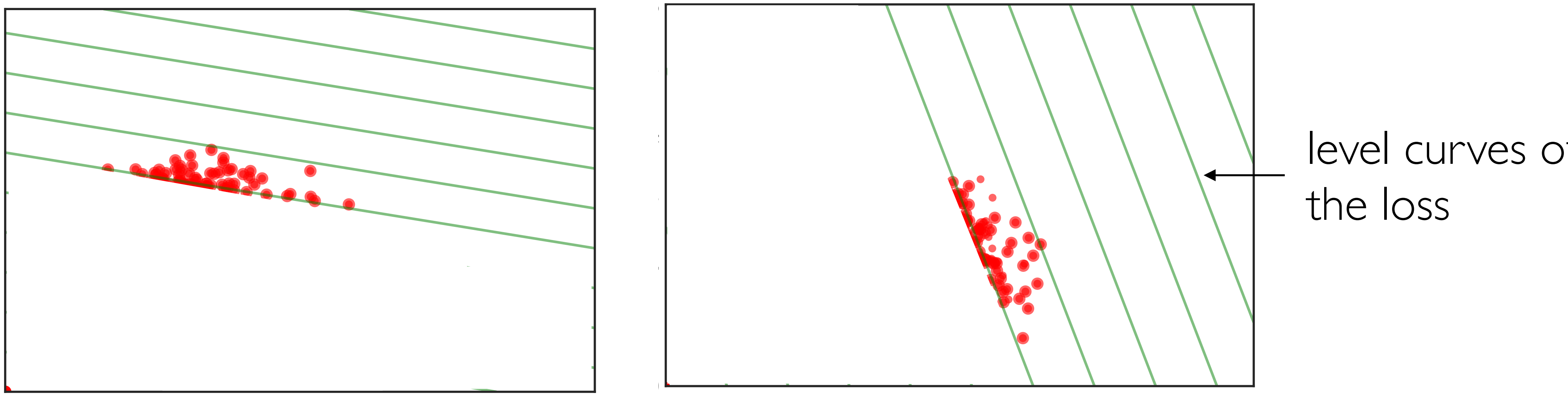


$L(\boldsymbol{X}, \theta) = 0.2X_1 + 0.8X_2$ $\qquad$ $L(\boldsymbol{X}, \theta) = 0.8X_1 + 0.2X_2$

model (objective + distribution) → **Step 1: Propose an "good" distribution family** → **Step 2: Set up OPT for the best candidate in the family** → **Step 3: Solve OPT** → sampler

informed by large deviations analysis

CVaR minimisation: $\min_{\theta} \text{CVaR}_{1-\beta}[L(\boldsymbol{X}, \theta)]$



$L(\boldsymbol{X}, \theta) = 0.2X_1 + 0.8X_2$ $\qquad$ $L(\boldsymbol{X}, \theta) = 0.8X_1 + 0.2X_2$

What is a good sampler for one **decision** is often not good for other ⇨ a bottleneck in design of IS algorithms

model (objective + distribution) → **Step 1: Propose an "good" distribution family** → **Step 2: Set up OPT for the best candidate in the family** → **Step 3: Solve OPT** → sampler

informed by large deviations analysis

CVaR minimisation: $\min_{\theta} \text{CVaR}_{1-\beta}[L(\boldsymbol{X}, \theta)]$

A brief literature review

Variance Reduction Techniques

Optimization under uncertainty

minimizing CVaR,
chance-constraints
with sample-averaging

Dantzig & Glynn '90
Dantzig & Infanger '93
Rubinstein & Shapiro '93
Shapiro & Homem-de-Mello '98
Nemirovski & Shapiro '06
Barrera et al '14
Kozmik & Morton '14
Parpas et al '15
Birge '12, Homem-de-Mello & Bayraskan '15 (reviews)
Blanchet , Zhang & Zwart '20
He, Jiang , Lam & Fu, '21

Image credit: Frits Ahlefeldt

A brief literature review

Variance Reduction Techniques

Optimization under uncertainty

minimizing CVaR,
chance-constraints
with sample-averaging

- Importance Sampling for multistage linear programs

Dantzig & Glynn '90
Dantzig & Infanger '93
Rubinstein & Shapiro '93
Shapiro & Homem-de-Mello '98
Nemirovski & Shapiro '06
Barrera et al '14
Kozmik & Morton '14
Parpas et al '15
Birge '12, Homem-de-Mello & Bayraskan '15 (reviews)
Blanchet , Zhang & Zwart '20
He, Jiang , Lam & Fu, '21

Image credit: Frits Ahlefeldt

A brief literature review

Variance Reduction Techniques

Optimization under uncertainty

minimizing CVaR, chance-constraints with sample-averaging

- Variance reduction techniques for chance constrained optimisation.
- Use *exponential twisting*
- Applied to a communication networks problem.

Dantzig & Glynn '90
Dantzig & Infanger '93
Rubinstein & Shapiro '93
Shapiro & Homem-de-Mello '98
Nemirovski & Shapiro '06
Barrera et al '14
Kozmik & Morton '14
Parpas et al '15
Birge '12, Homem-de-Mello & Bayraskan '15 (reviews)
Blanchet , Zhang & Zwart '20
He, Jiang , Lam & Fu, '21

Image credit: Frits Ahlefeldt

A brief literature review

Variance Reduction Techniques

Optimization under uncertainty

minimizing CVaR, chance-constraints with sample-averaging

- MCMC for multistage linear programs

Dantzig & Glynn '90
Dantzig & Infanger '93
Rubinstein & Shapiro '93
Shapiro & Homem-de-Mello '98
Nemirovski & Shapiro '06
Barrera et al '14
Kozmik & Morton '14
Parpas et al '15
Birge '12, Homem-de-Mello & Bayraskan '15 (reviews)
Blanchet , Zhang & Zwart '20
He, Jiang , Lam & Fu, '21

Image credit: Frits Ahlefeldt

Overcoming circularity: Adaptive IS

Variance Reduction Techniques

Optimization under uncertainty
minimizing CVaR, chance-constraints with sample-averaging

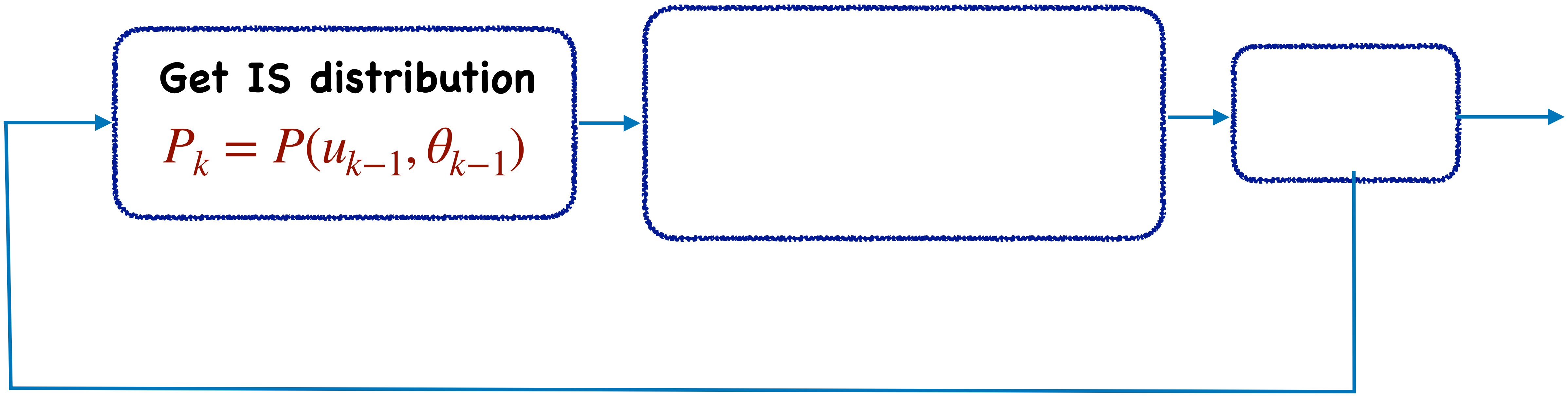
- Develop an adaptive IS algorithm for stochastic optimisation.
- "Tuning" the IS distribution leads to a significantly improved performance

Dantzig & Glynn '90
Dantzig & Infanger '93
Rubinstein & Shapiro '93
Shapiro & Homem-de-Mello '98
Nemirovski & Shapiro '06
Barrera et al '14
Kozmik & Morton '14
Parpas et al '15
Birge '12, Homem-de-Mello & Bayraskan '15 (reviews)
Blanchet , Zhang & Zwart '20
He, Jiang , Lam & Fu, '21

Image credit: Frits Ahlefeldt

Adaptive IS-SAA

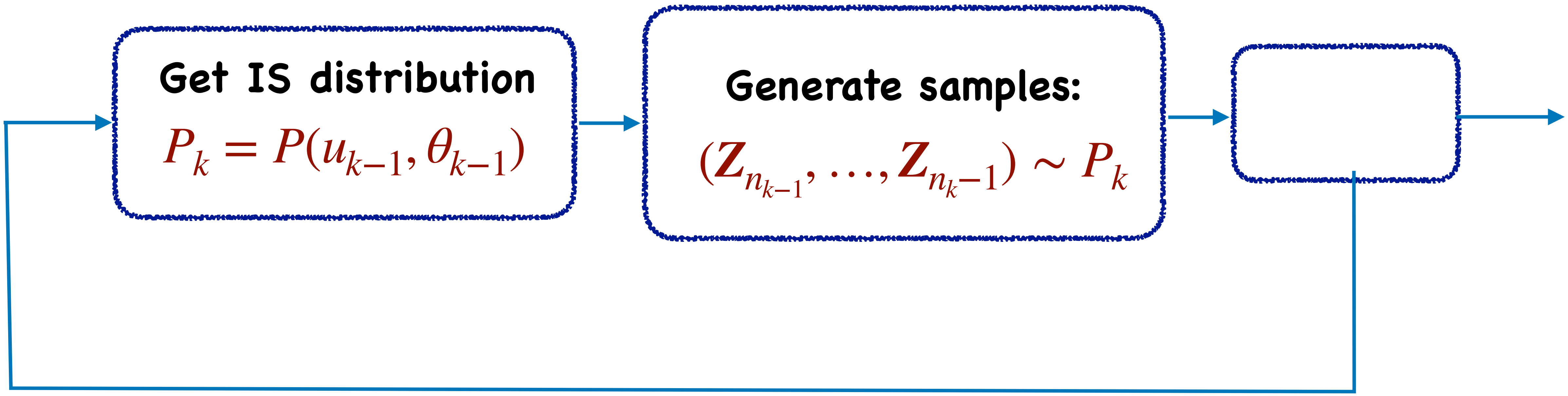
INPUT: Family of distributions $P(\cdot)$ such that $P(u,\theta)$ is a "good" IS distribution for decision (u,θ)



CVaR minimisation: $\min_{u \geq 0,\ \theta \in \Theta} \left(u + \beta^{-1} E[L(X,\theta) - u]^+ \right)$

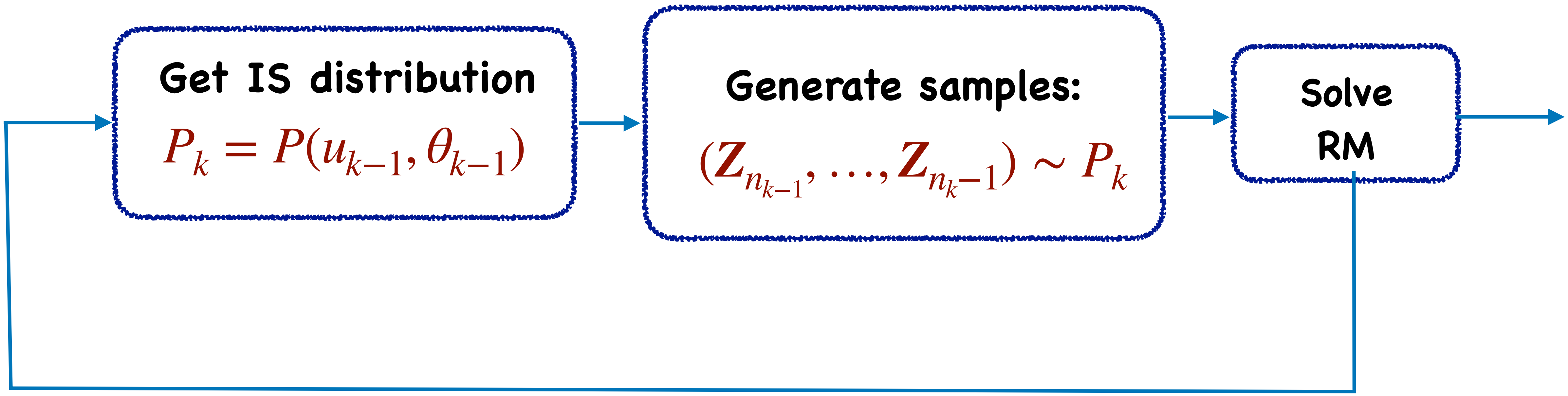
Adaptive IS-SAA

INPUT: Family of distributions $P(\cdot)$ such that $P(u, \theta)$ is a "good" IS distribution for decision (u, θ)



Adaptive IS-SAA

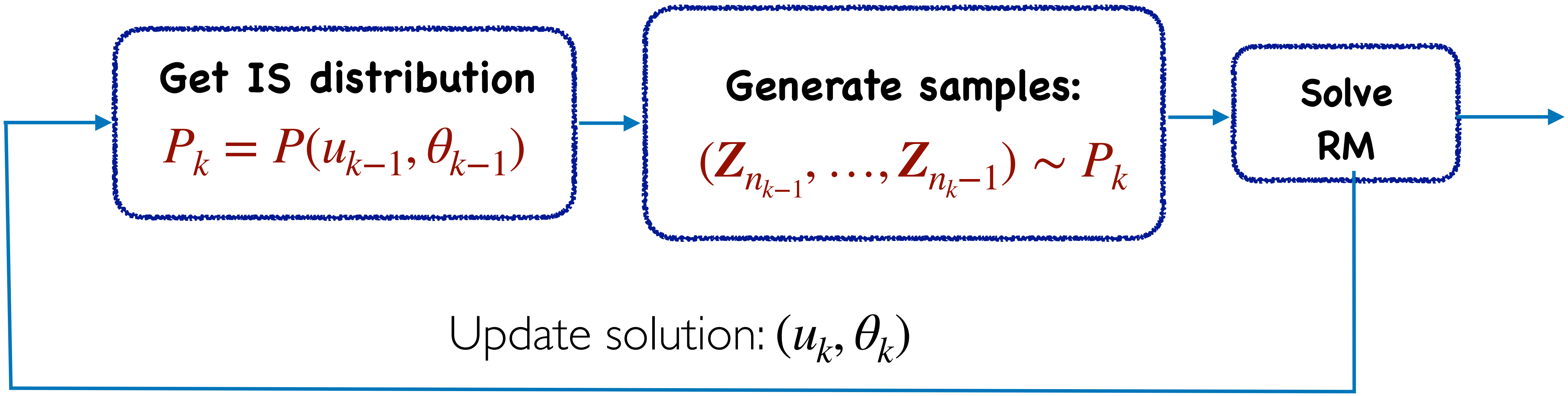
INPUT: Family of distributions $P(\cdot)$ such that $P(u, \theta)$ is a "good" IS distribution for decision (u, θ)



$$\textbf{RM} = \min_{(u,\theta)} \left\{ u + \frac{1}{n_k \beta} \sum_{i=1}^{n_k} (L(Z_i, \theta) - u)^+ \mathscr{L}_i \right\}$$

Adaptive IS-SAA

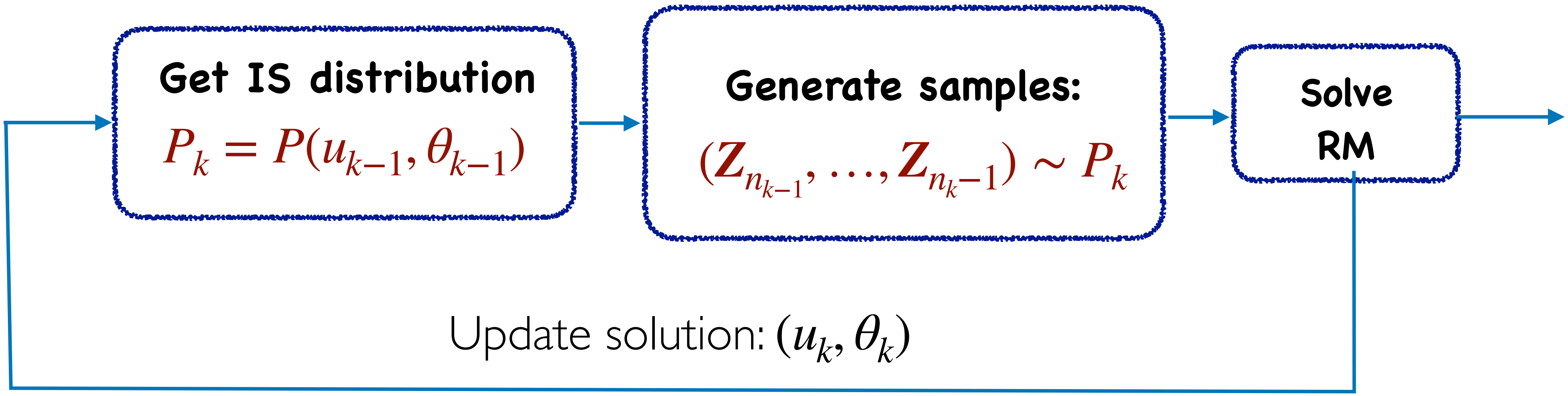
INPUT: Family of distributions $P(\cdot)$ such that $P(u, \theta)$ is a "good" IS distribution for decision (u, θ)



$$\textbf{RM} = \min_{(u,\theta)} \left\{ u + \frac{1}{n_k \beta} \sum_{i=1}^{n_k} (L(Z_i, \theta) - u)^+ \mathscr{L}_i \right\}$$

Adaptive IS-SAA

INPUT: Family of distributions $P(\cdot)$ such that $P(u, \theta)$ is a "good" IS distribution for decision (u, θ)

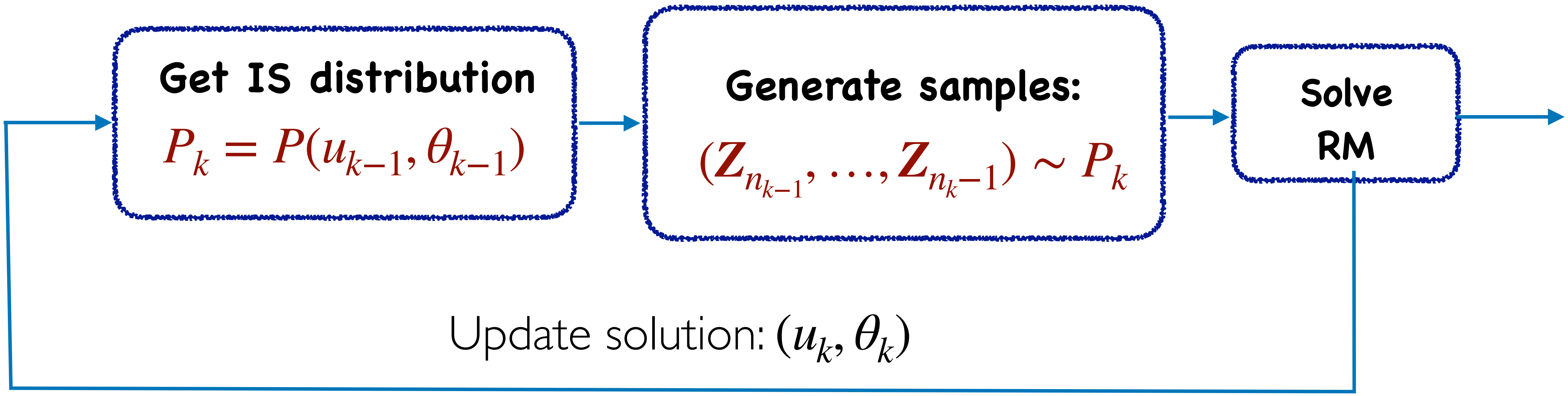


Why adaptive IS works

1) Iterates (u_k, θ_k) converge to optimal solution

Adaptive IS-SAA

INPUT: Family of distributions $P(\cdot)$ such that $P(u, \theta)$ is a "good" IS distribution for decision (u, θ)



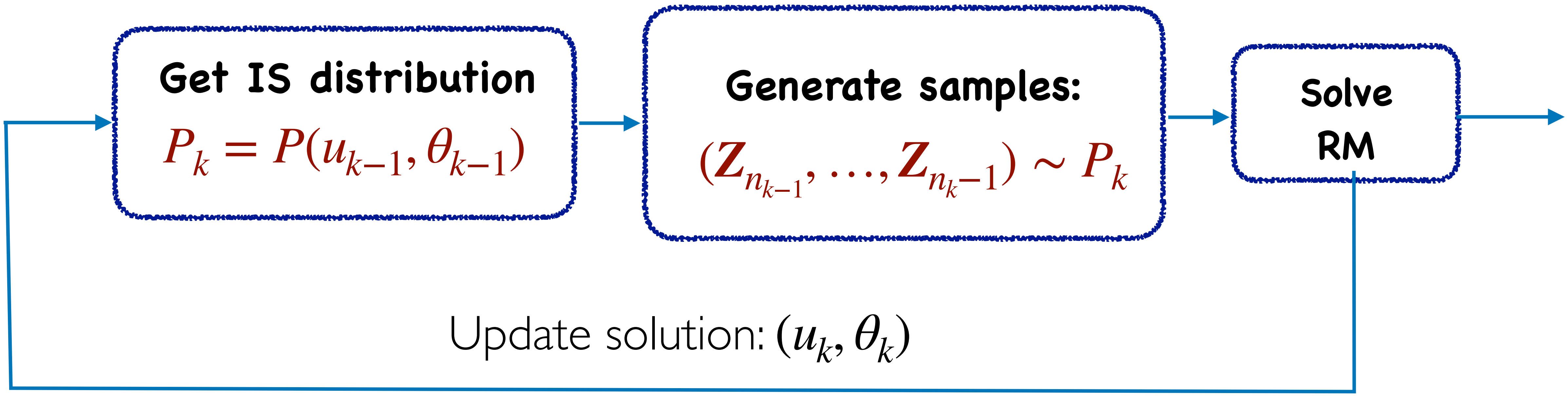
Why adaptive IS works

1) Iterates (u_k, θ_k) converge to optimal solution

2) In turn, this causes the IS distribution to be "more suitable" for the optimal decision

Adaptive IS-SAA

INPUT: Family of distributions $P(\cdot)$ such that $P(u, \theta)$ is a "good" IS distribution for decision (u, θ)



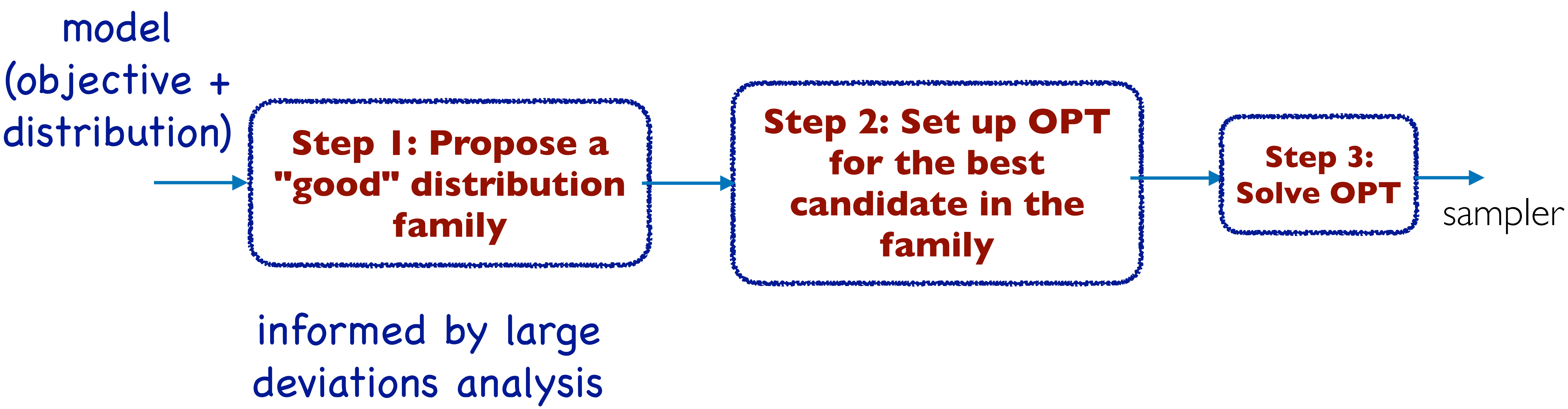
Known Results

1) He et al. (2021) derive a CLT for the solution iterates

2) With a good choice of $P(\cdot)$, significant variance reduction obtained

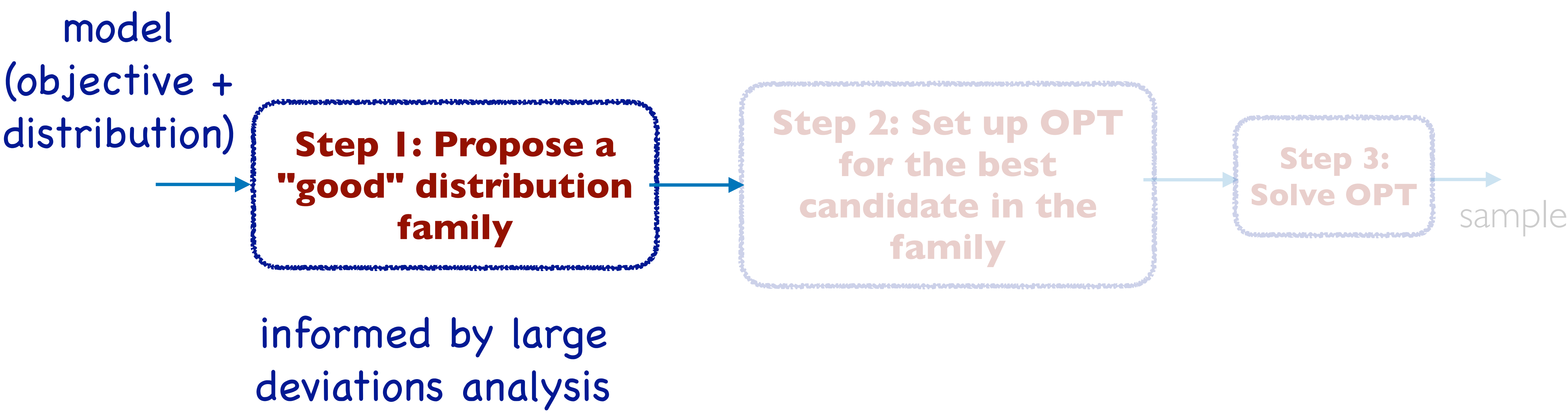
Challenges in implementing IS-SAA

INPUT: Family of distributions $P(\cdot)$ such that $P(u,\theta)$ is a "good" IS distribution for decision (u,θ)



Challenges in implementing IS-SAA

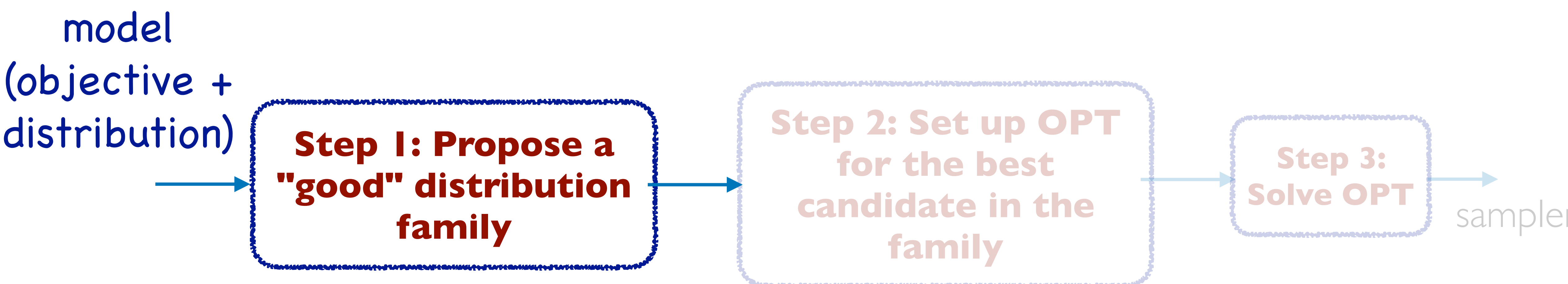
INPUT: Family of distributions $P(\cdot)$ such that $P(u,\theta)$ is a "good" IS distribution for decision (u,θ)



Difficulty 1: What is a "good" family of IS distributions??

Challenges in implementing IS-SAA

INPUT: Family of distributions $P(\cdot)$ such that $P(u, \theta)$ is a "good" IS distribution for decision (u, θ)



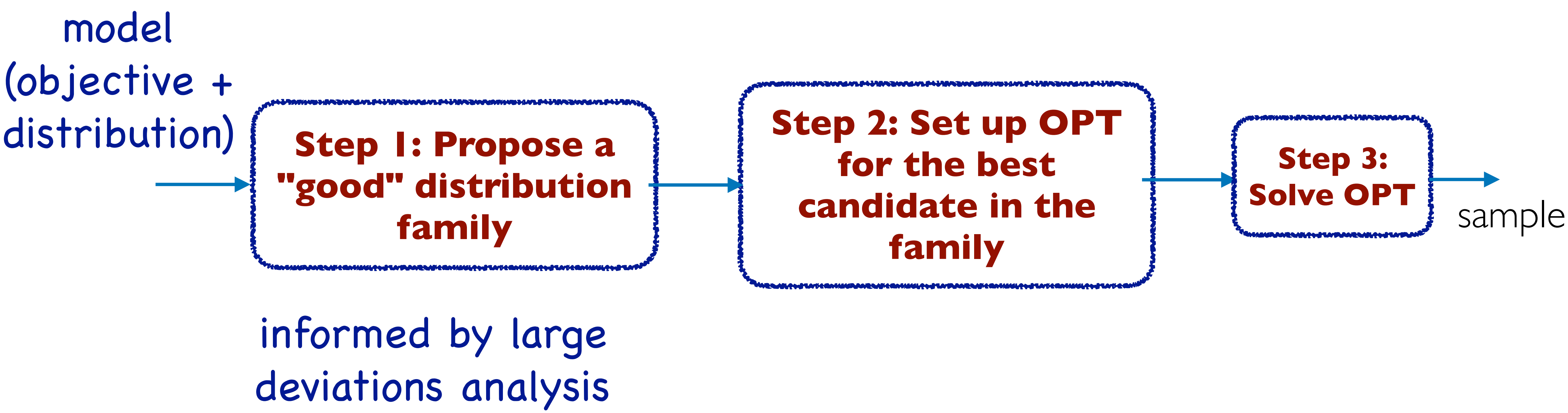
informed by large deviations analysis

Difficulty 1: What is a "good" family of IS distributions??

Partial resolutions: Exponential/Hazard Rate Twisting (Glasserman et al. 2002, Juneja et al. 2008, He et al. 2021…)

Challenges in implementing IS-SAA

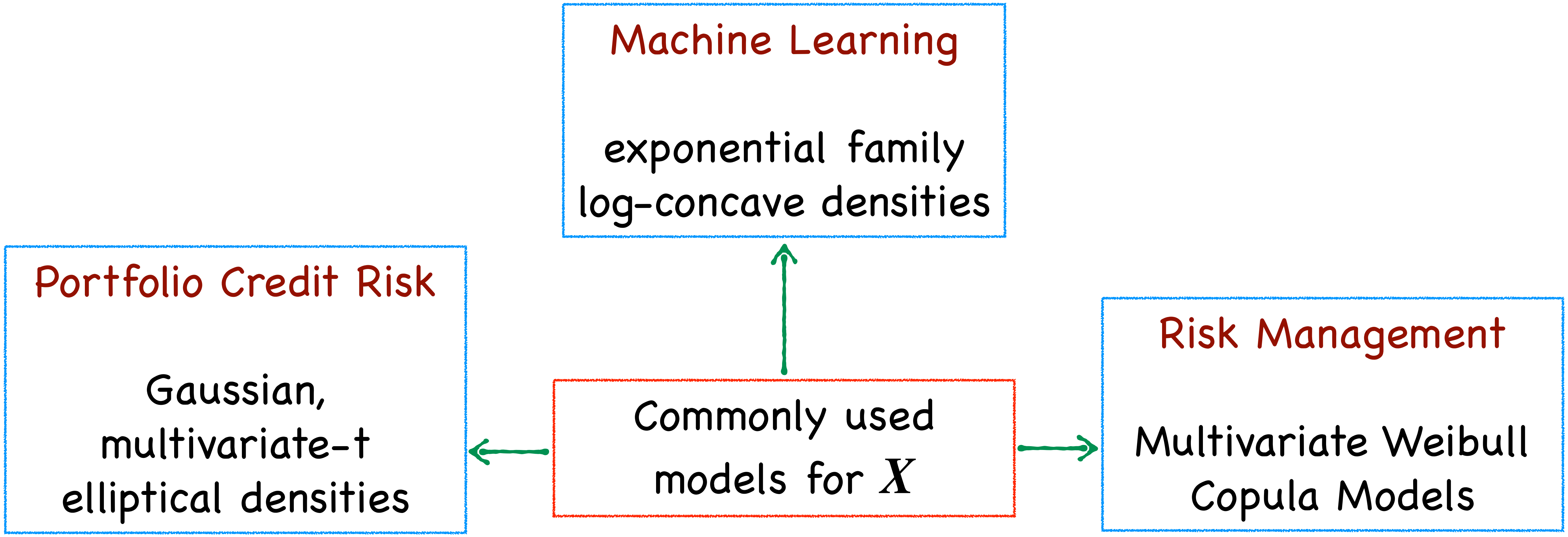
INPUT: Family of distributions $P(\cdot)$ such that $P(u,\theta)$ is a "good" IS distribution for decision (u,θ)



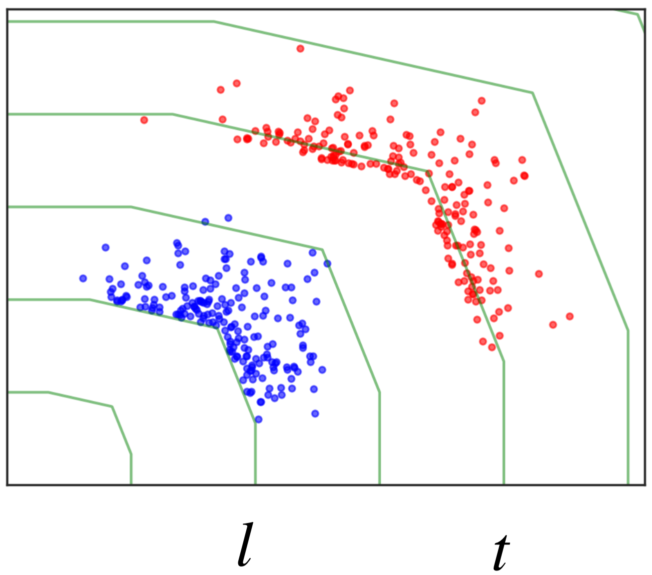
Difficulty 2: How to solve "OPT"??

Can be non-convex (or) combinatorially very challenging

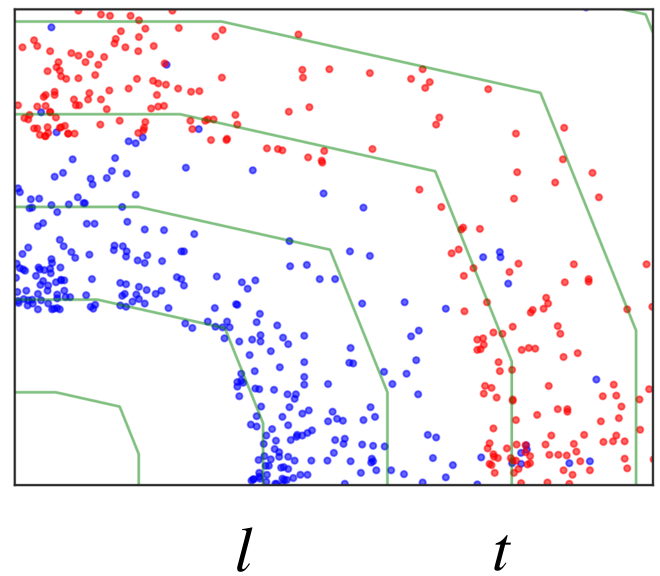
Our Solution: Tail events occur in structurally similar ways



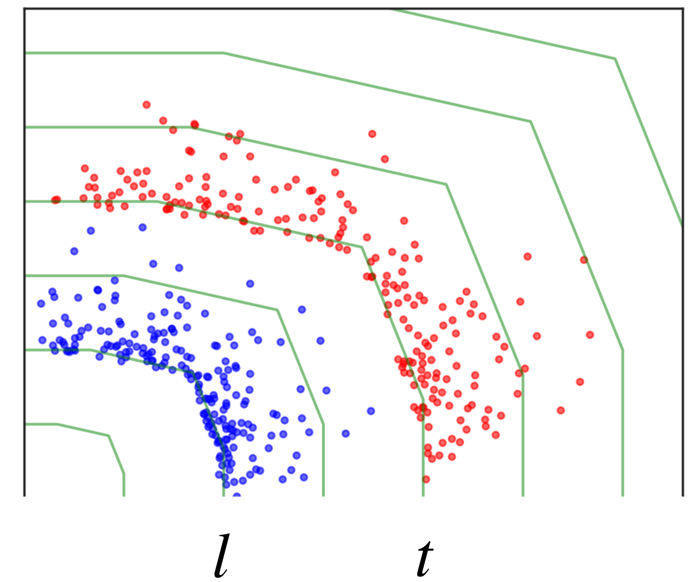
Our Solution: Tail events occur in structurally similar ways



Elliptical Density



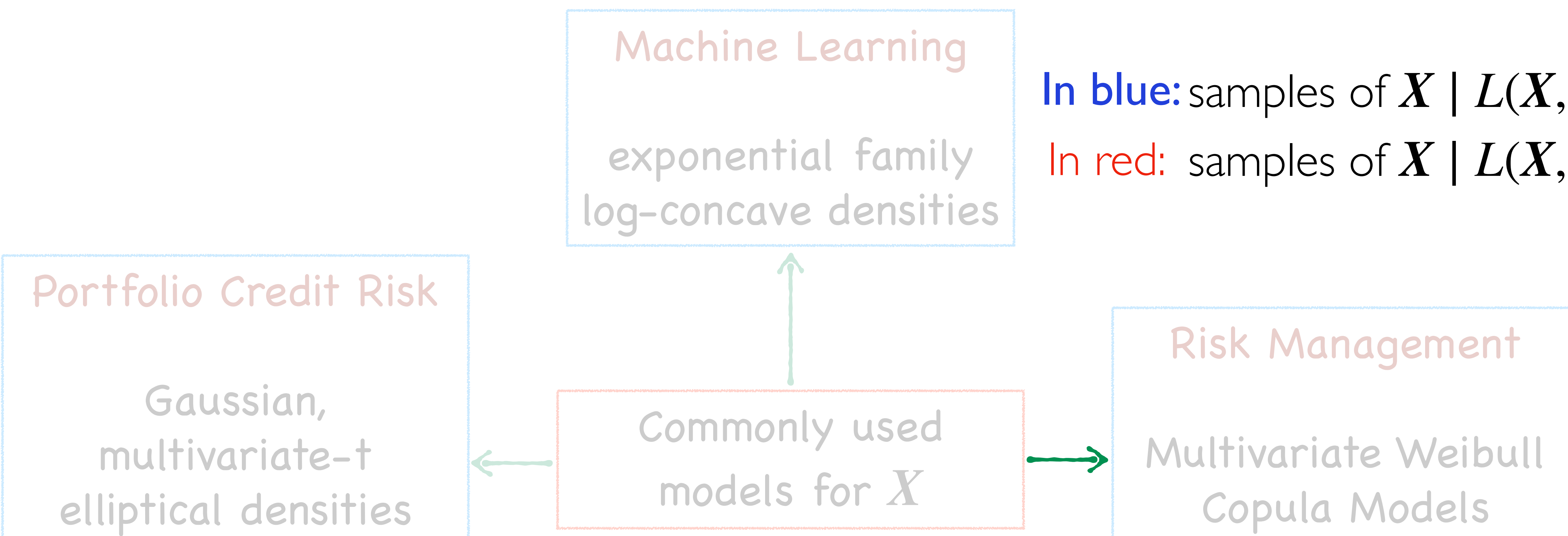
Heavy tailed Weibull



Exponential Family

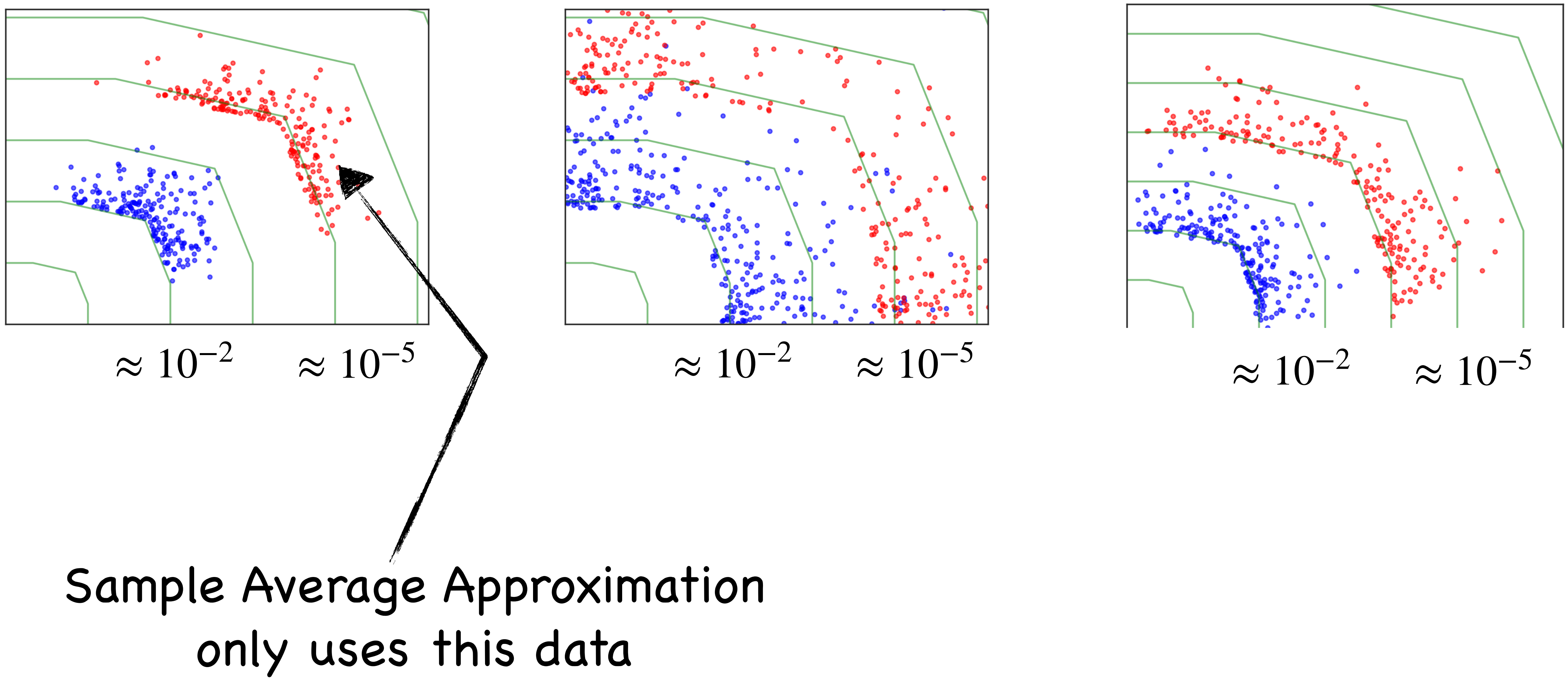
In blue: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > l$

In red: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > t$



Tail events occur in structurally similar ways

Objective: $\min_{\theta \in \Theta} \mathsf{CVaR}_{1-\beta}[L(\boldsymbol{X}, \theta)]$

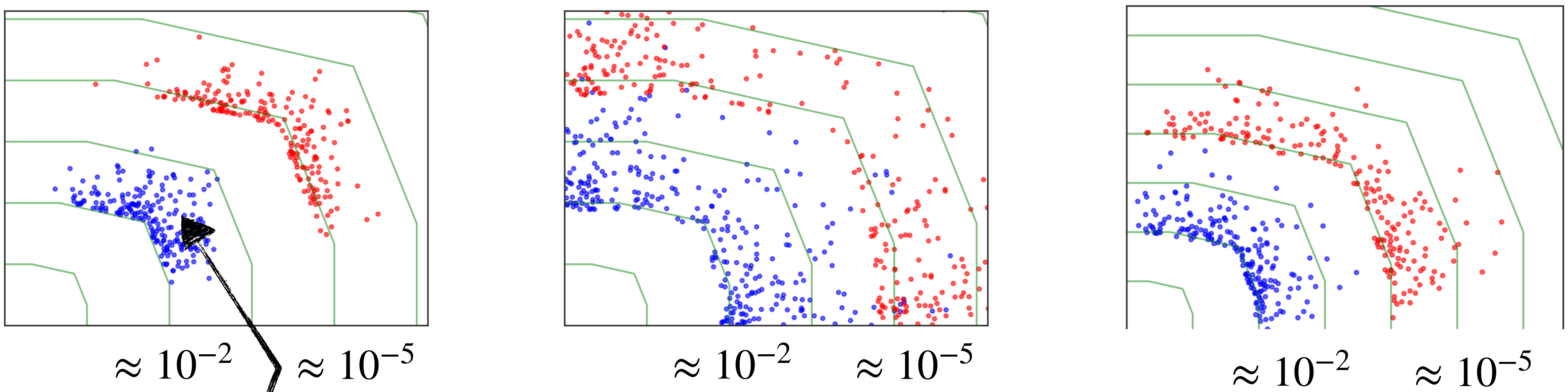


In blue: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > l$

In red: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > t$

Tail events occur in structurally similar ways

Objective: $\min_{\theta \in \Theta} \text{CVaR}_{1-\beta}[L(\boldsymbol{X}, \theta)]$



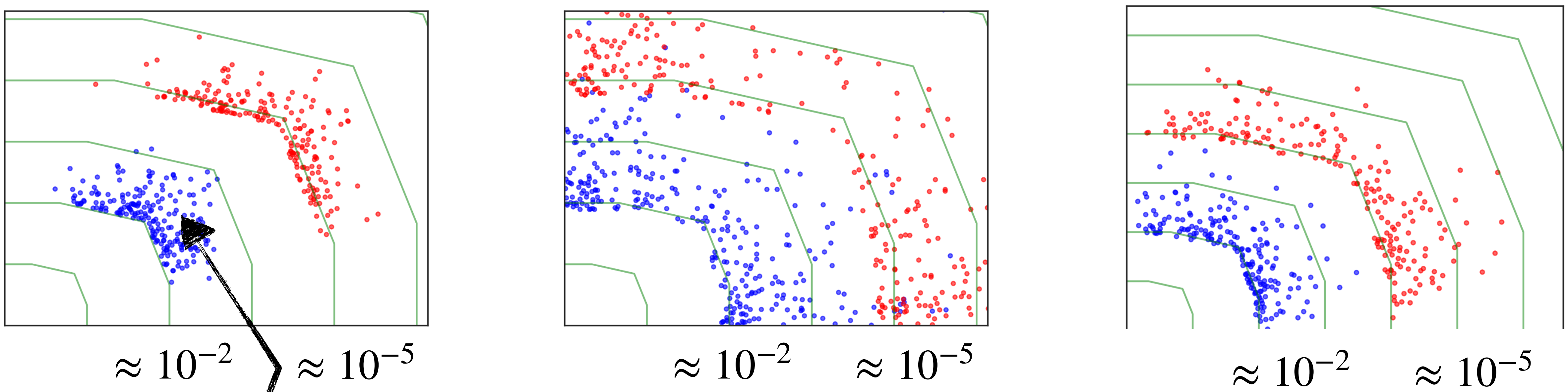
Can leverage information in this data

In blue: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > l$

In red: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > t$

Tail events occur in structurally similar ways

Objective: $\min_{\theta \in \Theta} \text{CVaR}_{1-\beta}[L(\boldsymbol{X}, \theta)]$



Can leverage information in this data

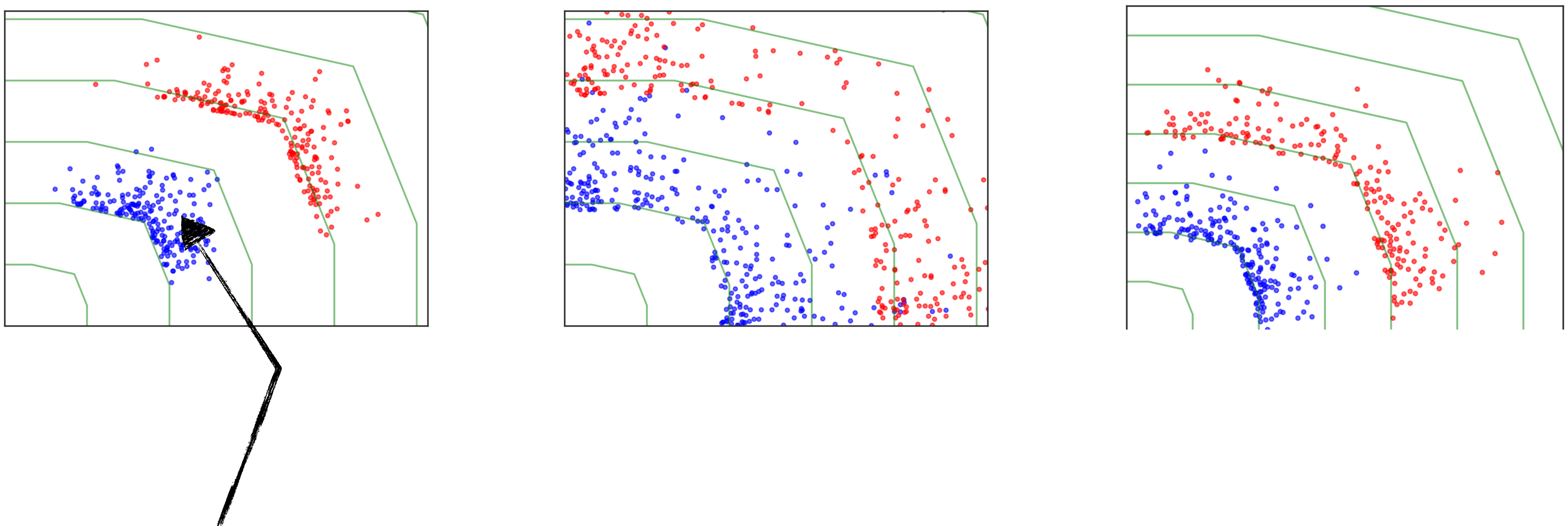
Blue samples seen $\approx 10^3$ times more frequently in data than red ones

In blue: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > l$

In red: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > t$

Tail events occur in structurally similar ways

Objective: $\min_{\theta \in \Theta} \text{CVaR}_{1-\beta}[L(\boldsymbol{X}, \theta)]$



Can leverage information in this data

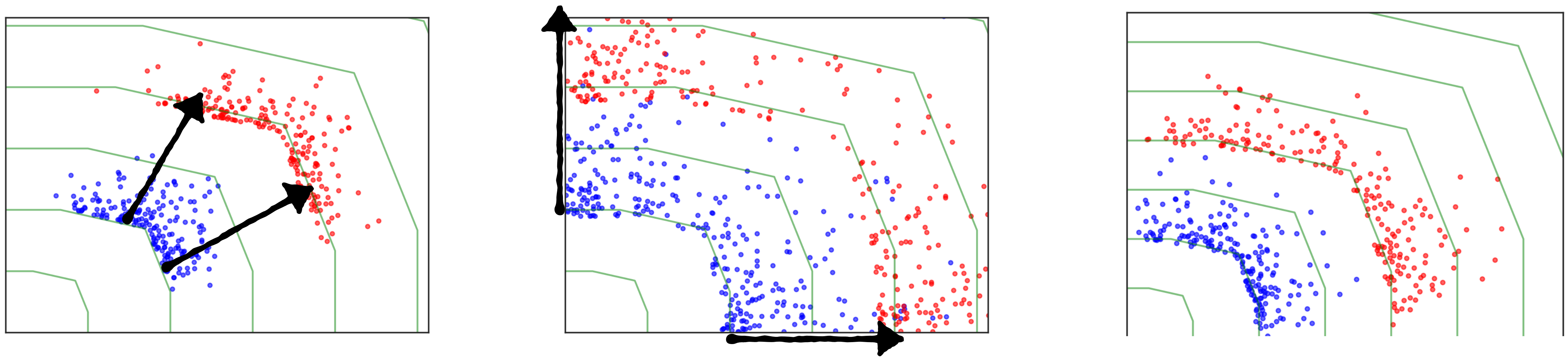
Theorem [D and Murthy 21]: Red and blue points follow a large deviations principle with the same rate function

In blue: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > l$

In red: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > t$

Tail events occur in structurally similar ways

Objective: $\min_{\theta \in \Theta} \text{CVaR}_{1-\beta}[L(\boldsymbol{X}, \theta)]$



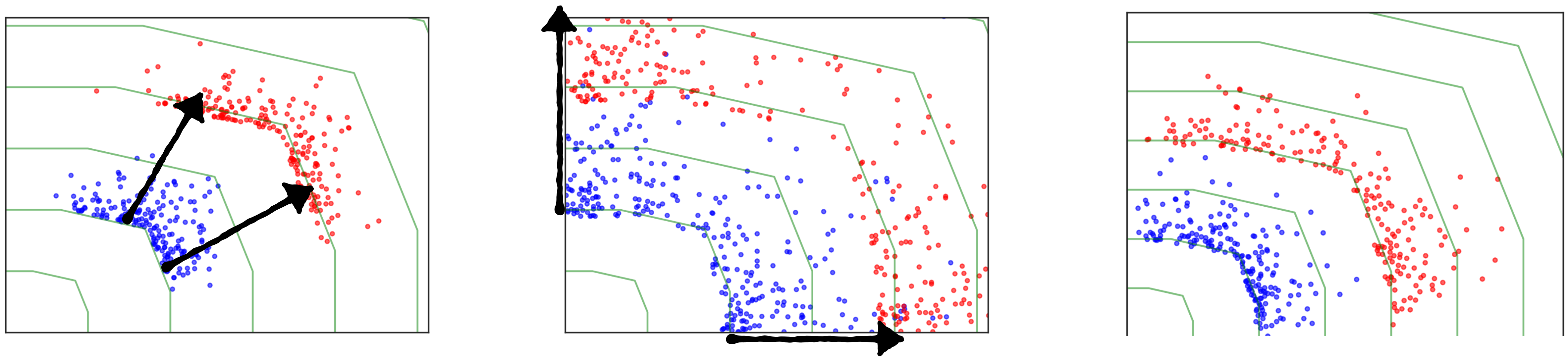
Scaling blue samples appropriately provides representation of the red samples!

In blue: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > l$

In red: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}, \theta) > t$

Self-Similarity → Efficient Samplers

Objective: $\min_{\theta \in \Theta} \mathrm{CVaR}_{1-\beta}[L(X, \theta)]$



Scaling blue samples appropriately provides representation of the red samples!

Questions

1. How to identify this scaling?
2. How to use it in adaptive IS?

In blue: samples of $X \mid L(X, \theta) > l$

In red: samples of $X \mid L(X, \theta) > t$

Self-Structuring transformation

$$\boldsymbol{T}_r(\boldsymbol{x}) = r^{\kappa(\boldsymbol{x})}\boldsymbol{x}$$

where

$$\kappa(\boldsymbol{x}) = \frac{\log|\boldsymbol{x}|}{\rho \log ||\boldsymbol{x}||_\infty}$$

r = scalar stretch parameter

Questions

1. How to identify this scaling? ✓
2. How to use it in adaptive IS?

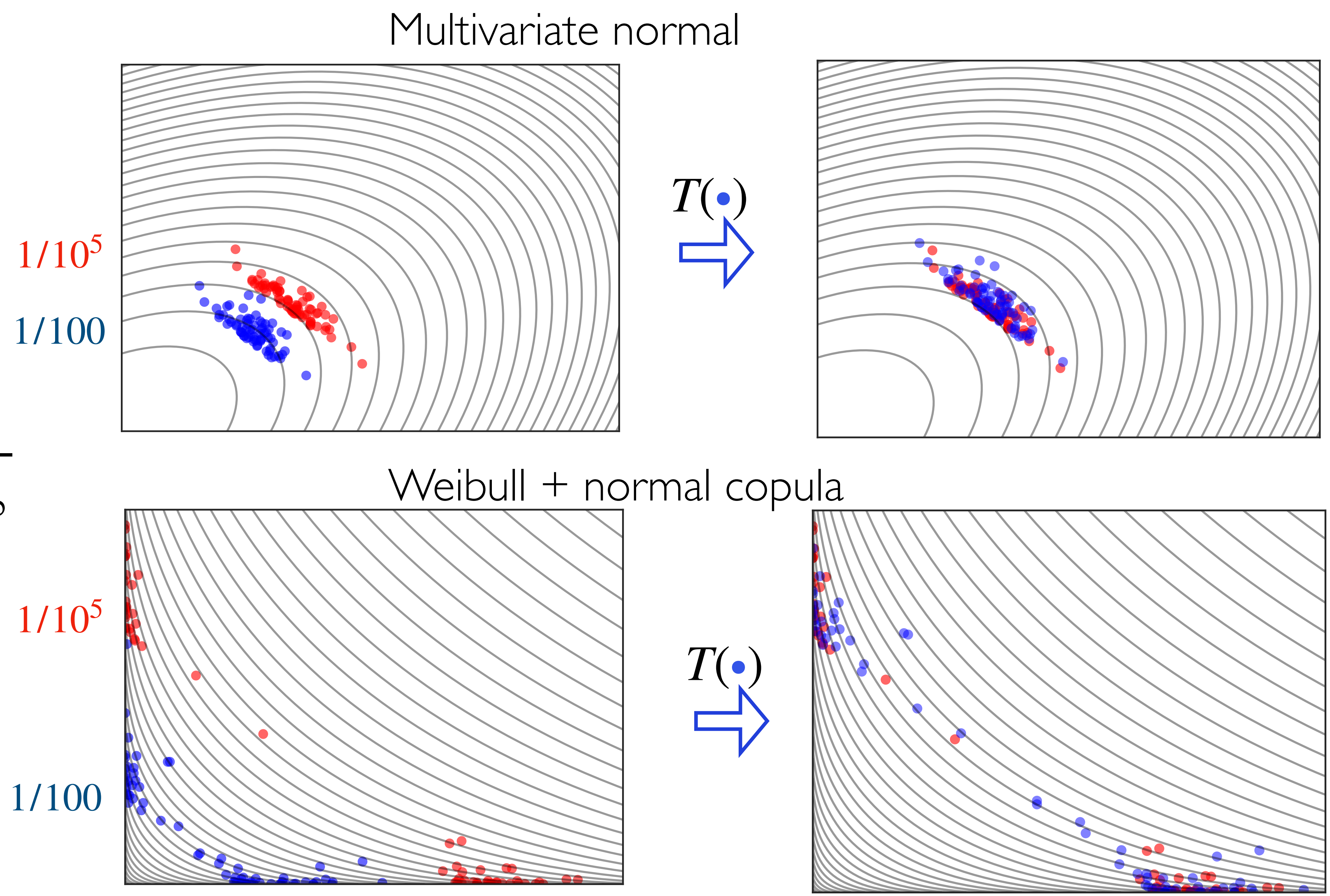
Self-Structuring transformation

$$T_r(\boldsymbol{x}) = r^{\kappa(\boldsymbol{x})}\boldsymbol{x}$$

where

$$\kappa(\boldsymbol{x}) = \frac{\log|\boldsymbol{x}|}{\rho \log \|\boldsymbol{x}\|_\infty}$$

r = scalar stretch parameter



in blue: excess loss samples at 1/100 risk level

in red: excess loss samples at 1/100,000 risk level

in blue: transported excess loss samples

Conditions under which $T_r(\cdot)$ is self structuring

Asymptotically homogenous loss:

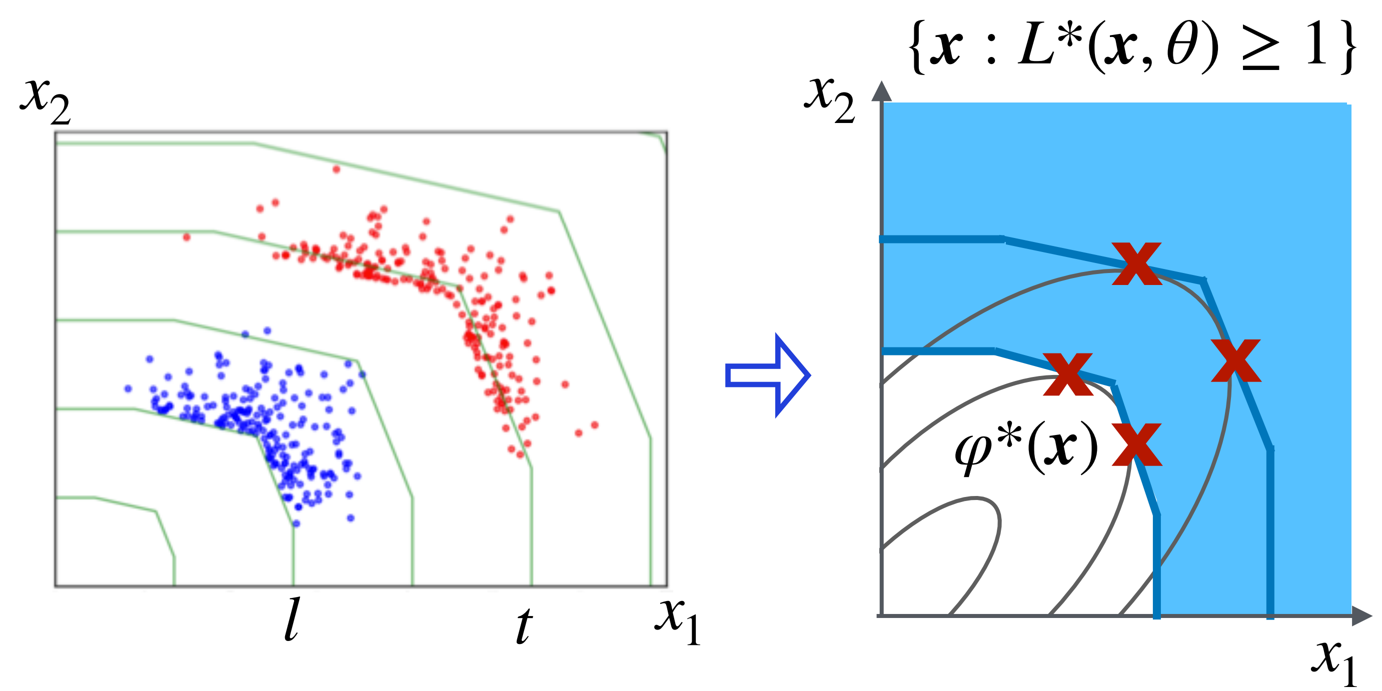
$$\lim_{n\to\infty} \frac{L(n\boldsymbol{x},\theta)}{n^{\rho}} = L^*(\boldsymbol{x},\theta)$$

pdf of ***X*** **:** $f_{\mathbf{X}}(\boldsymbol{x}) = \exp(-\varphi(\boldsymbol{x}))$

$$\lim_{n\to\infty} \frac{\varphi(n\boldsymbol{x})}{\varphi(n\mathbf{1})} = \varphi^*(\boldsymbol{x})$$ → Light Tails

$$\lim_{n\to\infty} \frac{f(n\boldsymbol{x})}{f(n\mathbf{1})} = f^*(\boldsymbol{x})$$ → Heavy tails

eg: + correlated multivariate normal



in blue: samples of $\boldsymbol{X} \mid L(\boldsymbol{X},\theta) > t$

in red: samples of $\boldsymbol{X} \mid L(\boldsymbol{X},\theta) > u$

x: Rate point

Conditions under which $T_r(\cdot)$ is self structuring

Asymptotically homogenous loss:

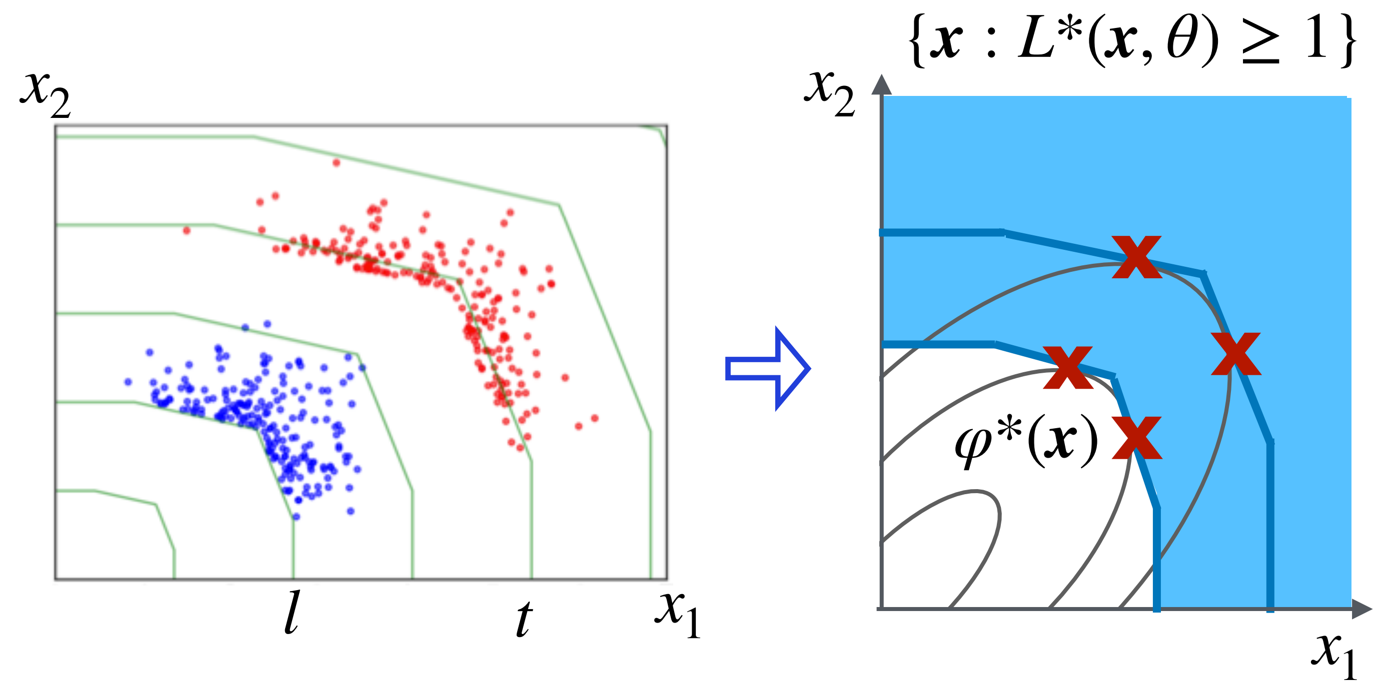
$$\lim_{n\to\infty} \frac{L(n\boldsymbol{x},\theta)}{n^\rho} = L^*(\boldsymbol{x},\theta)$$

pdf of $\boldsymbol{X}$: $f_{\mathbf{X}}(\boldsymbol{x}) = \exp(-\varphi(\boldsymbol{x}))$

$$\lim_{n\to\infty} \frac{\varphi(n\boldsymbol{x})}{\varphi(n\mathbf{1})} = \varphi^*(\boldsymbol{x}) \rightarrow \text{Light Tails}$$

$$\lim_{n\to\infty} \frac{f(n\boldsymbol{x})}{f(n\mathbf{1})} = f^*(\boldsymbol{x}) \rightarrow \text{Heavy tails}$$

eg: + correlated multivariate normal



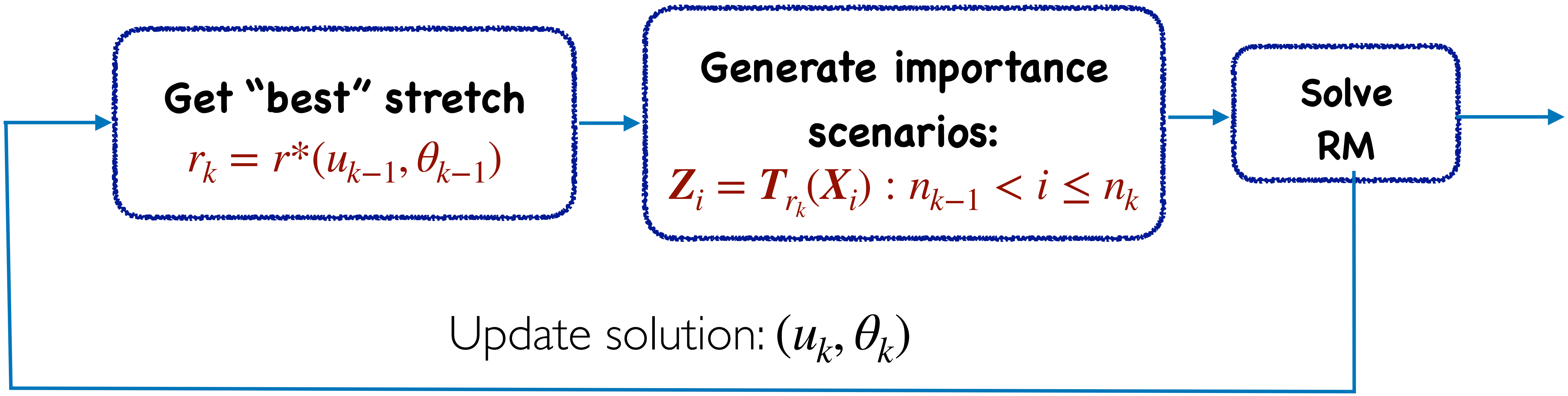
in blue: samples of $\boldsymbol{X} \mid L(\boldsymbol{X},\theta) > t$

in red: samples of $\boldsymbol{X} \mid L(\boldsymbol{X},\theta) > u$

Proposition [D & Murthy '21] (informal version): The theoretically "optimal" sampler and the transformed excess loss samples concentrate their mass on the same set of points

Incorporating Self-Structuring Transformations into adaptive IS [D and Murthy 22']

INPUT: Family of self-structuring transformations:
$\{\boldsymbol{T}_r(\cdot) : r \in [1,\infty)\}$, samples $\boldsymbol{X}_1, \ldots$



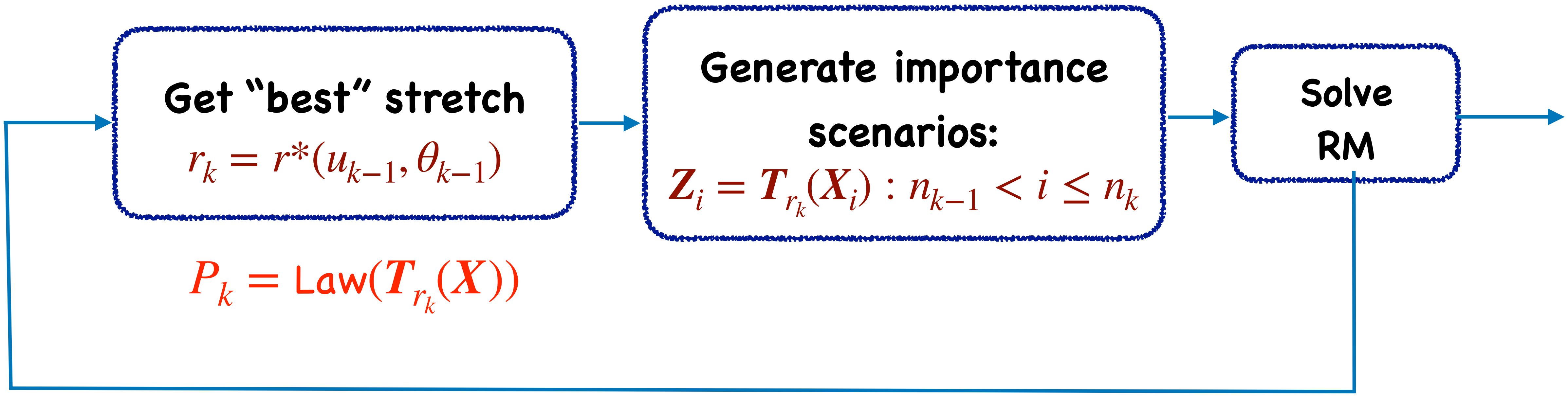
Questions

1. How to identify this scaling?
2. How to use it in adaptive IS?

Incorporating Self-Structuring Transformations into adaptive IS [D and Murthy 22']

INPUT: Family of self-structuring transformations:

$\{T_r(\cdot) : r \in [1,\infty)\}$, samples $X_1, \ldots$

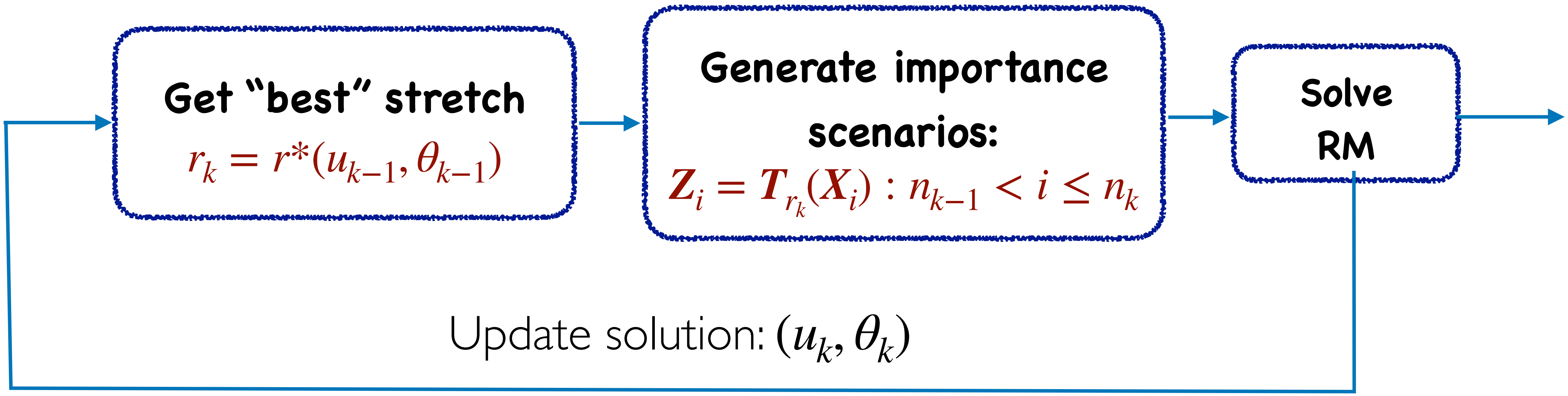


Questions

1. How to identify this scaling?
2. How to use it in adaptive IS?

Incorporating Self-Structuring Transformations into adaptive IS [D and Murthy 22']

INPUT: Family of self-structuring transformations:
$\{\boldsymbol{T}_r(\cdot) : r \in [1,\infty)\}$, samples $\boldsymbol{X}_1, \ldots$

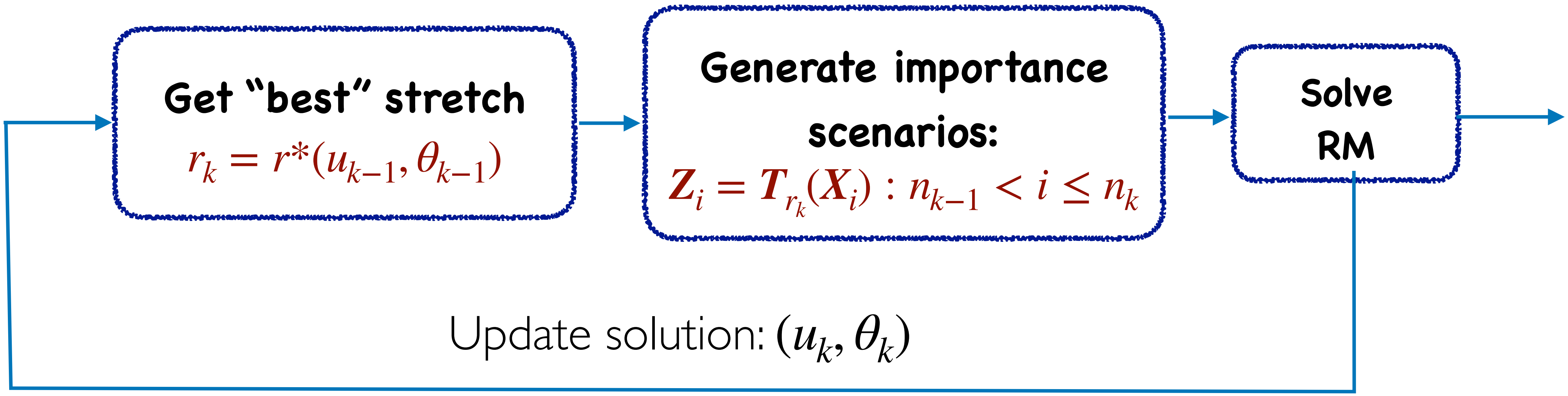


$$\mathbf{RM} = \min_{(u,\theta)} \left\{ u + \frac{1}{n_k \beta} \sum_{i=1}^{n_k} (L(Z_i, \theta) - u)^+ \mathscr{L}_i \right\}$$

Incorporating Self-Structuring Transformations into adaptive IS [D and Murthy 22']

INPUT: Family of self-structuring transformations:

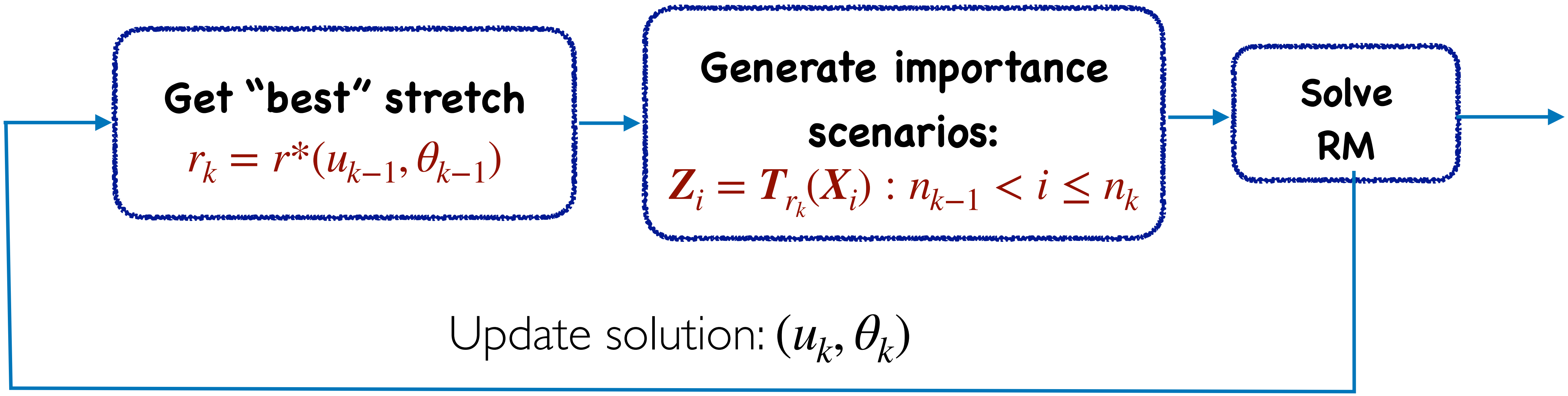
$\{\boldsymbol{T}_r(\cdot) : r \in [1,\infty)\}$, samples $\boldsymbol{X}_1, \ldots$



$$r^*(u,\theta) \in \arg\min_r \left\{ E\left(L(\boldsymbol{T}_r(\boldsymbol{X}), \theta) - u \right)^+ \mathscr{L}_r \right)^2 \right\}$$

Incorporating Self-Structuring Transformations into adaptive IS [D and Murthy 22']

INPUT: Family of self-structuring transformations:
$\{\boldsymbol{T}_r(\cdot) : r \in [1,\infty)\}$, samples $\boldsymbol{X}_1, \ldots$



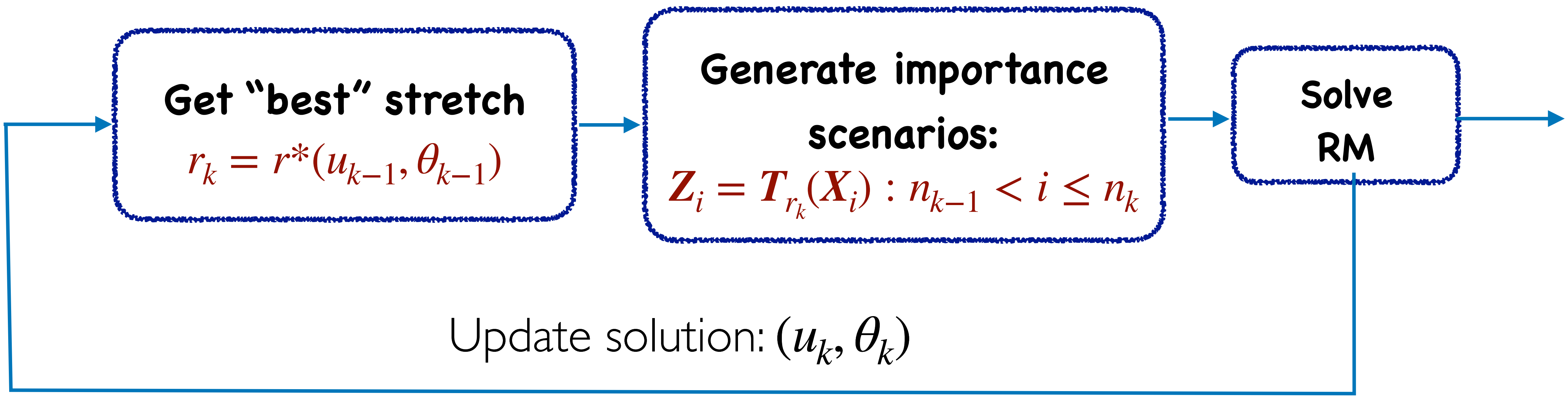
$$r^*(u, \theta) \in \arg\min_r \left\{ E\left(L(\boldsymbol{T}_r(\boldsymbol{X}), \theta) - u)^+ \mathscr{L}_r\right)^2 \right\}$$

Replace expectation by a sample average to identify r_k...

Incorporating Self-Structuring Transformations into adaptive IS [D and Murthy 22']

INPUT: Family of self-structuring transformations:

$\{\boldsymbol{T}_r(\cdot) : r \in [1,\infty)\}$, samples $\boldsymbol{X}_1, \ldots$



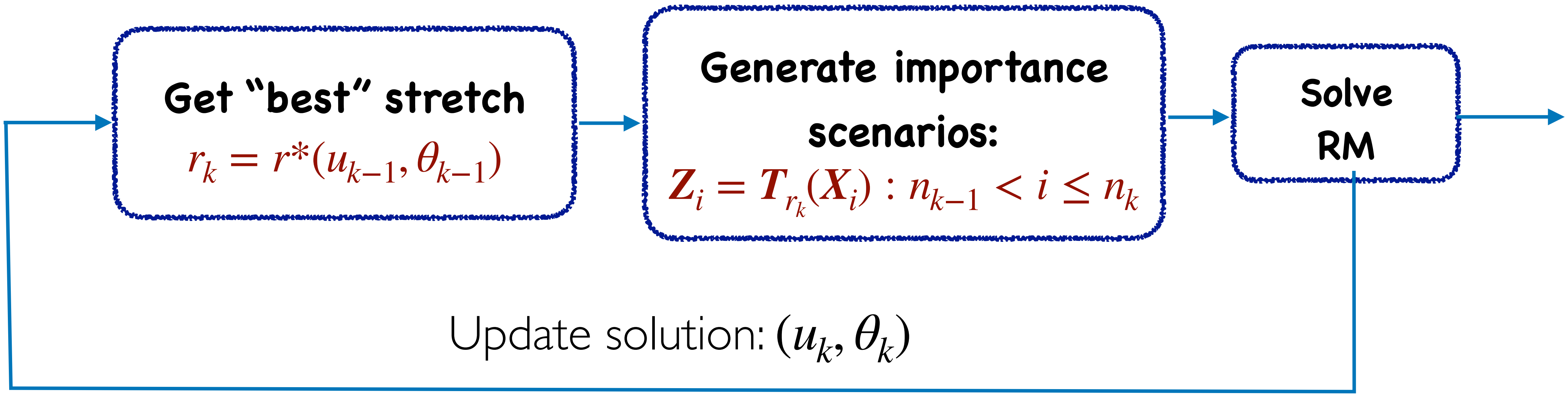
How does this address traditional difficulties?

1. Only need access to samples of $\boldsymbol{X}$
2. r_k can be identified easily

Incorporating Self-Structuring Transformations into adaptive IS [D and Murthy 22']

INPUT: Family of self-structuring transformations:

$\{T_r(\cdot) : r \in [1,\infty)\}$, samples $X_1, \ldots$



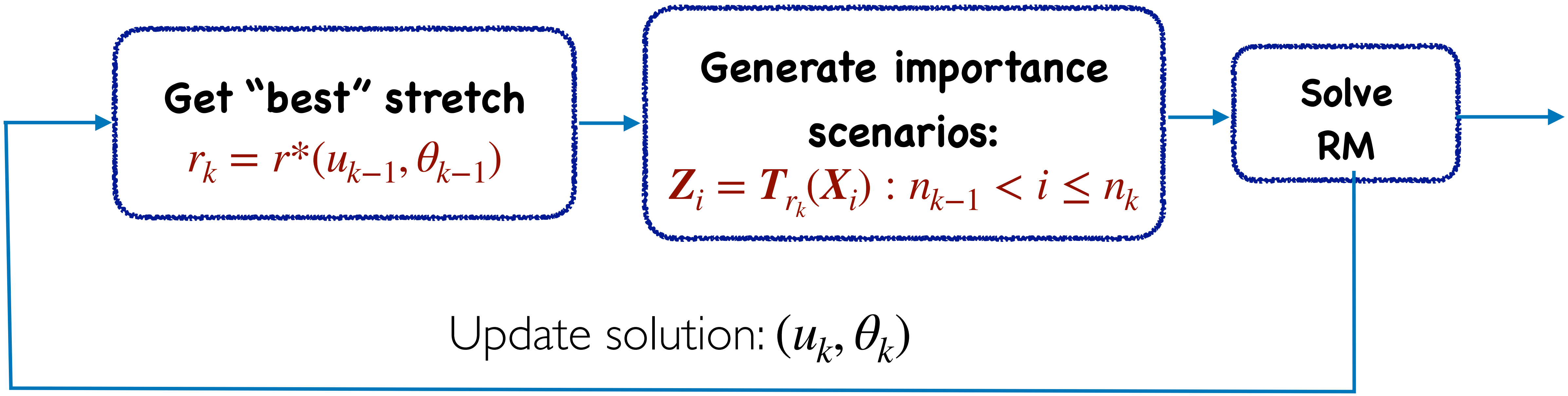
Theorem [D and Murthy, 2023]: Let $n_{\mathtt{ALG}}$ be number of samples required by $\mathtt{ALG}$ to solve the CVaR minimisation problem to within an error of ε. Then, for any $\delta > 0$,

$$\frac{n_{\mathtt{SAA}}}{n_{\mathtt{is}}} \geq \frac{c}{\beta^{1-\delta}} \text{ for all } \beta < \beta_0$$

Incorporating Self-Structuring Transformations into adaptive IS [D and Murthy 22']

INPUT: Family of self-structuring transformations:

$\{T_r(\cdot) : r \in [1,\infty)\}$, samples $X_1, \ldots$



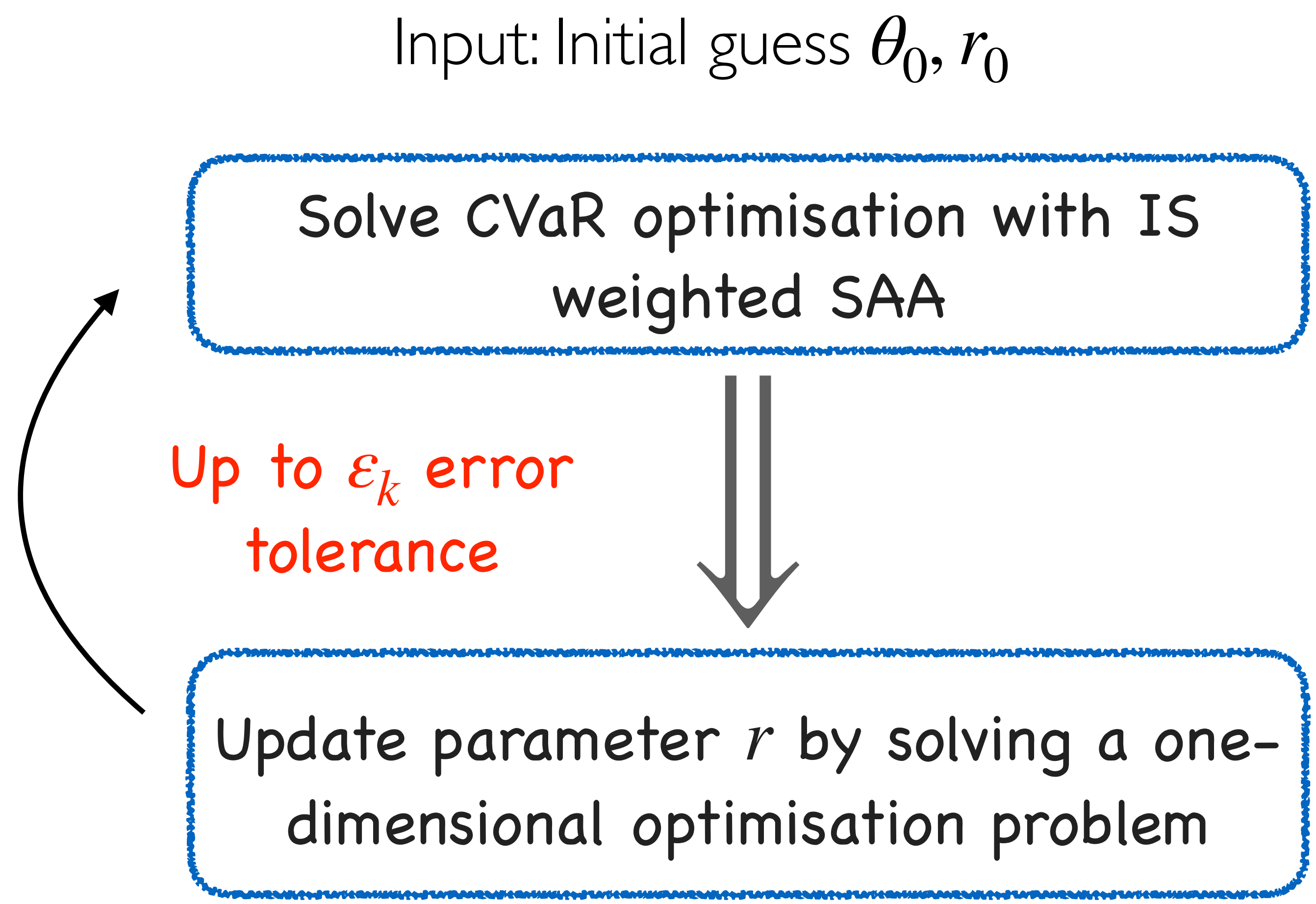
Theorem [D and Murthy, 2023]: Let $n_{\mathtt{ALG}}$ be number of samples required by $\mathtt{ALG}$ to solve the CVaR minimisation problem to within an error of ε. Then, for any $\delta > 0$,

$$\frac{n_{\mathtt{SAA}}}{n_{\mathtt{is}}} \geq \frac{c}{\beta^{1-\delta}}$$ ~~for all $\beta < \beta_0$~~

$$n_{\mathtt{SAA}} \sim \beta^{-1}, \; n_{\mathtt{is}} \sim (-\log \beta)^k$$

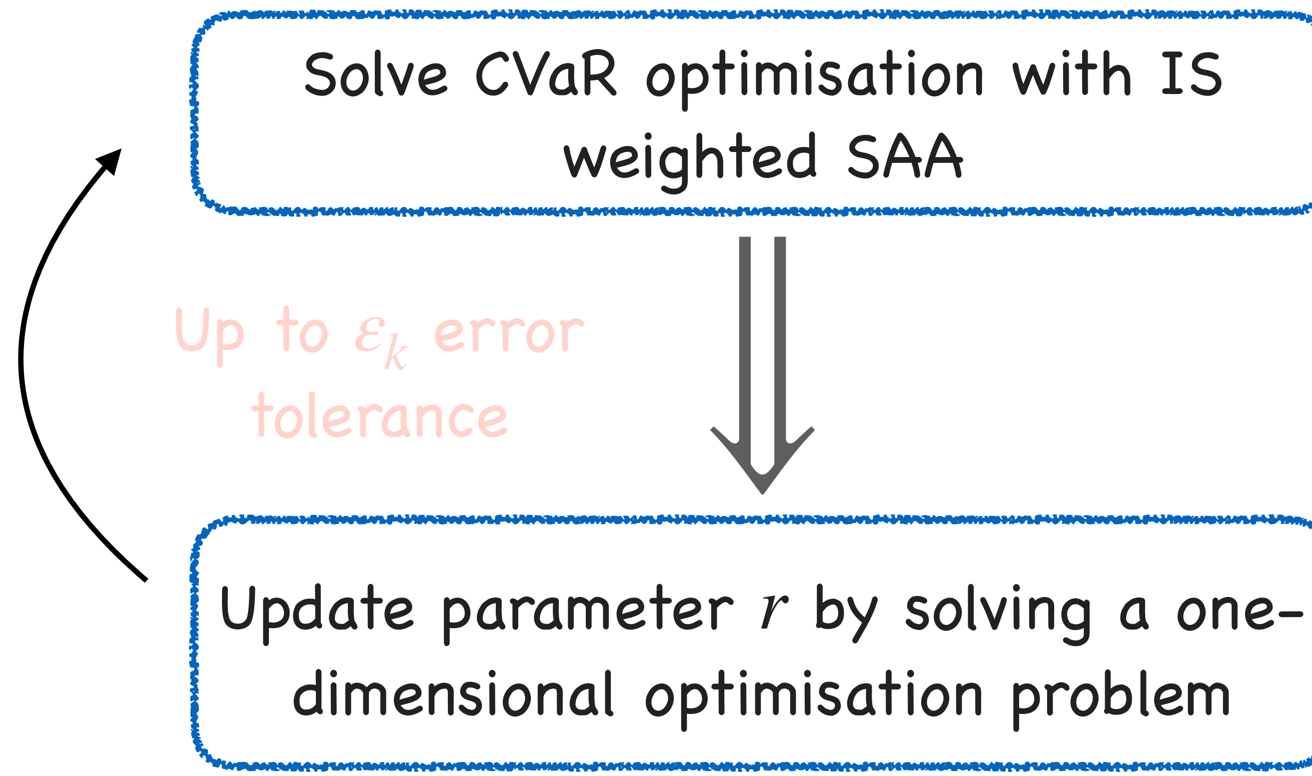
Numerical Exploration

$$T(x) = r^{\kappa(x)} x$$

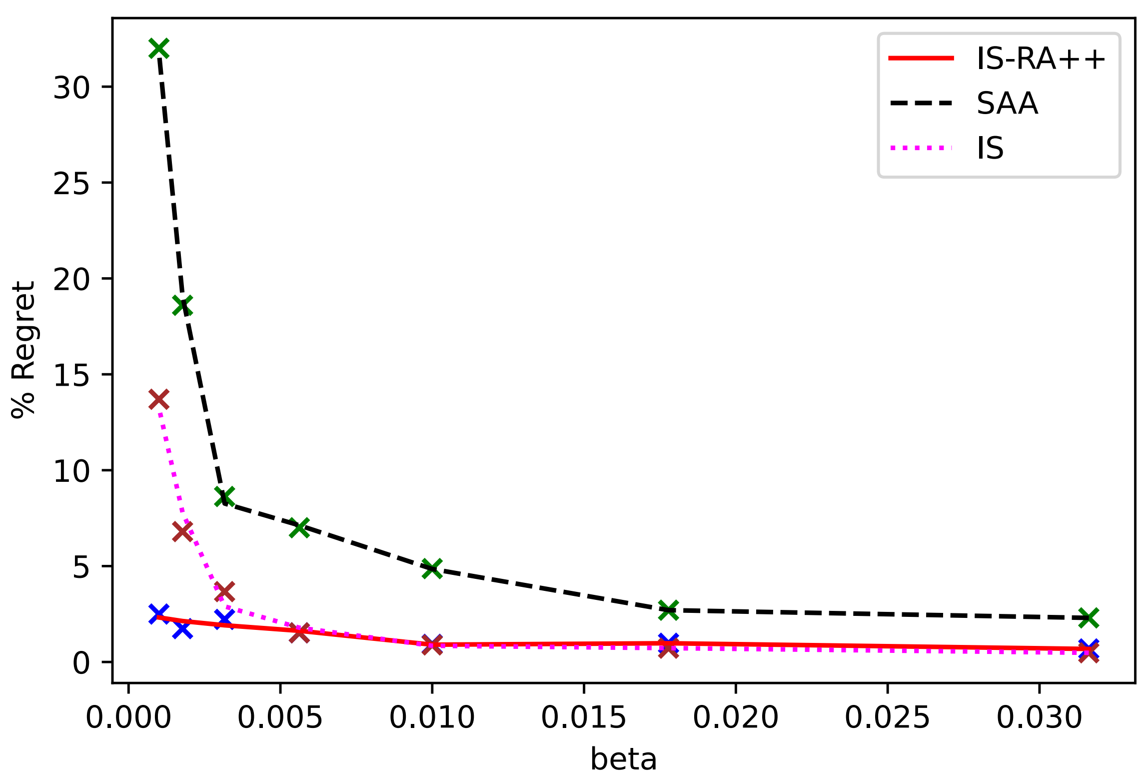


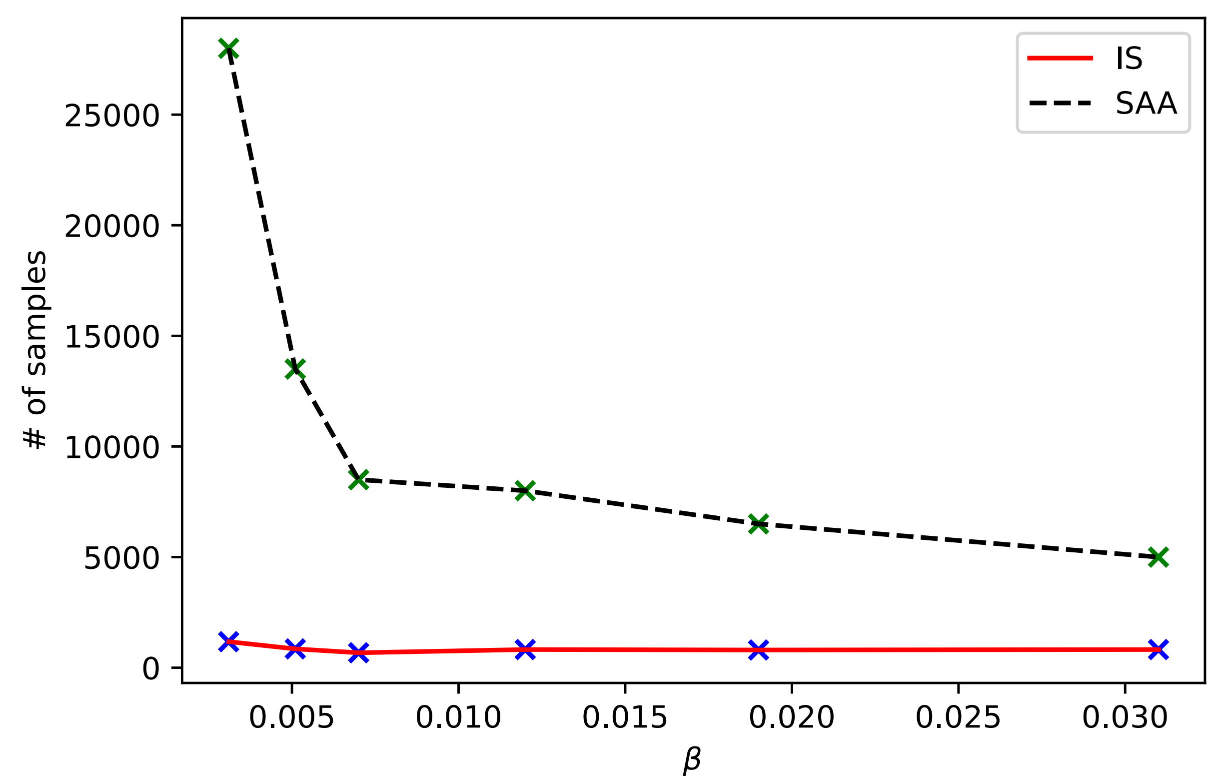
Numerical Exploration

Input: Initial guess θ_0, r_0



$$\boldsymbol{T}(\boldsymbol{x}) = r^{\kappa(\boldsymbol{x})}\boldsymbol{x}$$





CVaR Optimisation with 15-asset portfolio

Summary

Rare event simulation

Optimization under uncertainty

minimizing CVaR,
chance-constraints
with sample-averaging

IS based on self-structuring maps

- Importance Sampling & scenario generation in stochastic programming:

Dantzig & Glynn '90
Dantzig & Infanger '93
Rubinstein & Shapiro '93
Shapiro & Homem-de-Mello '98
Nemirovski & Shapiro '06
Barrera et al '14
Kozmik & Morton '14
Parpas et al '15
Birge '12, Homem-de-Mello & Bayraskan '15 (reviews)
Blanchet , Zhang & Zwart '20
He, Jiang , Lam & Fu, '21

Image credit: Frits Ahlefeldt

(All papers co-authored with Karthyek Murthy)

1. Importance Sampling for minimising tail risks: A tutorial (submitted to WSC 2023)

2. Combining Retrospective Approximation with Importance Sampling for Optimising Conditional Value at Risk (WSC 2022)

3. Achieving Efficiency in Simulation of Distribution Tails with Black Box Importance Samplers (under Revision at Operations Research)

Link to webpage

anand.deo@iimb.ac.in