Optimising tail risks with limited samples:

Can algorithms engineer effective reductions in variance & model-bias?

Karthyek Murthy
SUTD

(joint work with Anand Deo & Arjun Ramachandra)

Risk Analytics & Optimization

Boiling over

Natural-gas prices are spiking around the world

Unusual weather and supply outages are to blame

The Ever Given Is Moving But Your Supply Chain Will Not

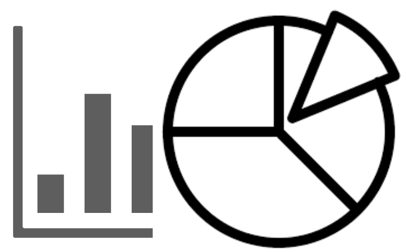
Loading, please wait

The global chip shortage is here for some time

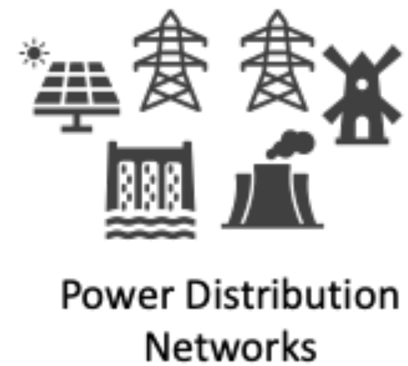
Trading-off risk & return: $$\min_{\theta : R(\theta) \geq r} \rho\left[L(\boldsymbol{X}, \boldsymbol{\theta})\right]$$

Trading-off risk & return: $\min_{\theta : R(\theta) \geq r} \rho\left[L(X, \boldsymbol{\theta})\right]$

Minimize default risk in loan portfolios while meeting target return requirement

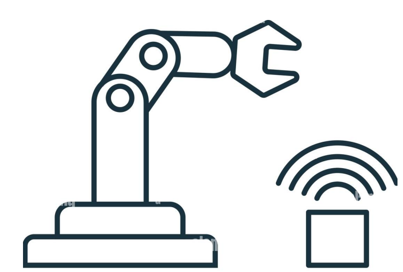


Managing power operations under volatile supply, demand and price risks

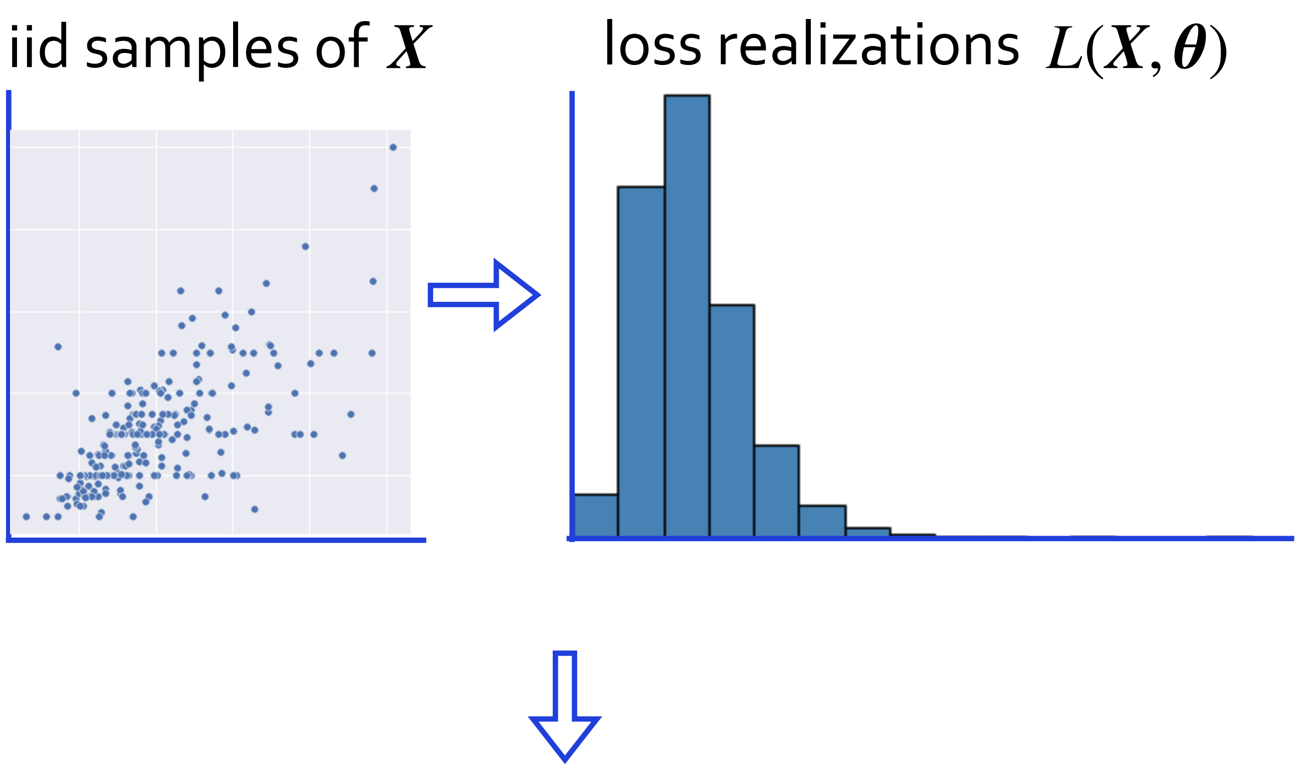


Minimize excess latency under supply and price risks

Minimize safety risk in cyber-physical systems



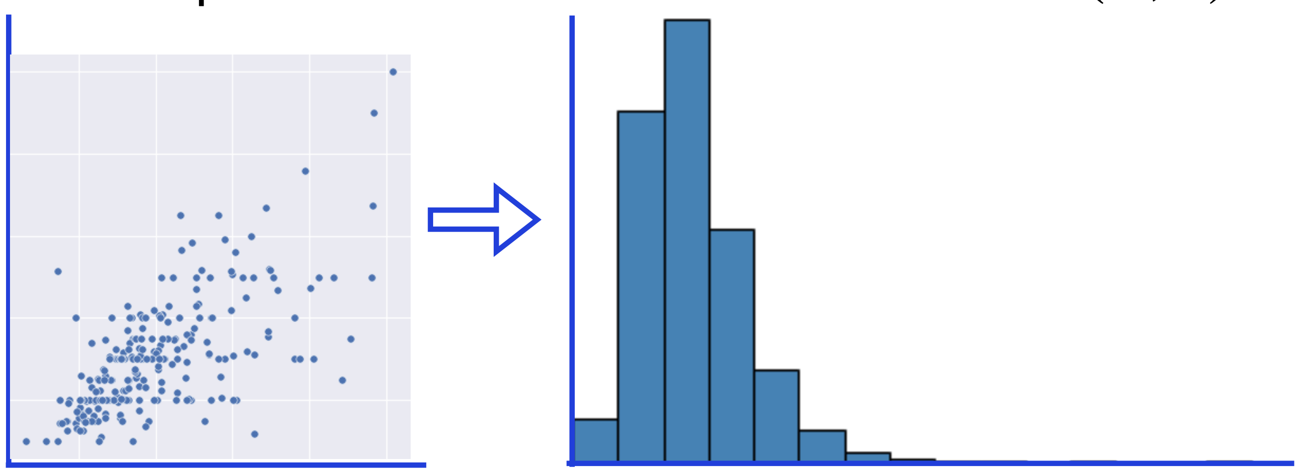
Conventional: Minimize average loss



minimize sample average

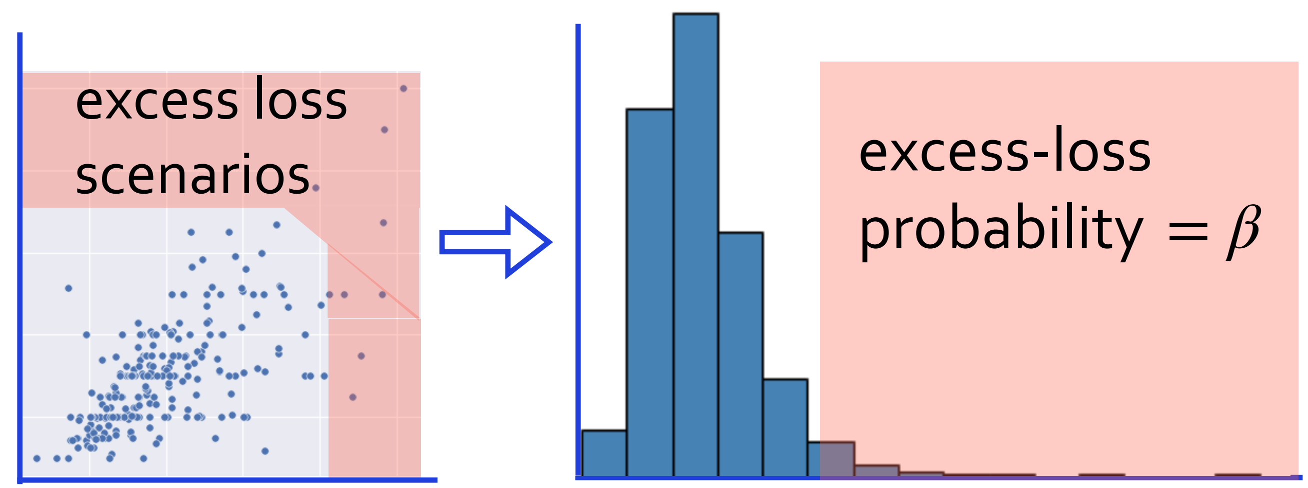
Conventional: Minimize average loss

iid samples of $\boldsymbol{X}$ loss realizations $L(\boldsymbol{X}, \boldsymbol{\theta})$



minimize sample average

Minimizing risk measures



$\textbf{CVaR}_{1-\beta}\left[L(\boldsymbol{X}, \boldsymbol{\theta})\right]$

For minimizing sample average,

$$\text{sample requirement} = \frac{1}{\beta} \times \text{usual requirement}$$

(*Trinidade et al '07*)

Minimizing risk measures

Eg: for tail level $\beta = 1/40$,

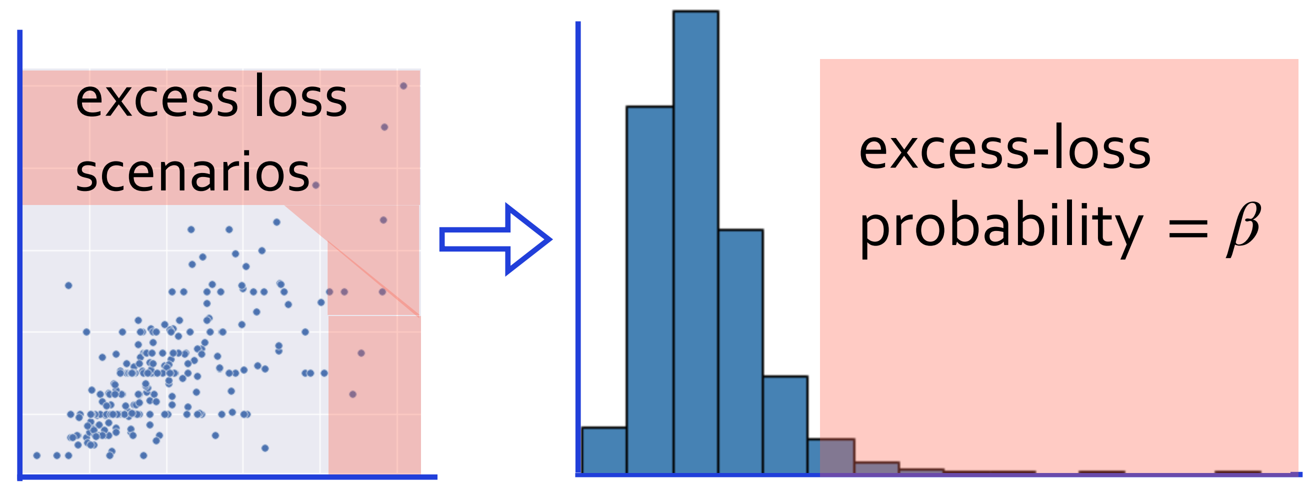
to achieve 10% relative error in optimum portfolio's CVaR for 100 stocks,

need ~14 years of daily returns data

Perils of minimizing sample average with insufficient samples:

Lim, Shanthikumar & Vahn '11

Caccioli, Paap & Condor '18

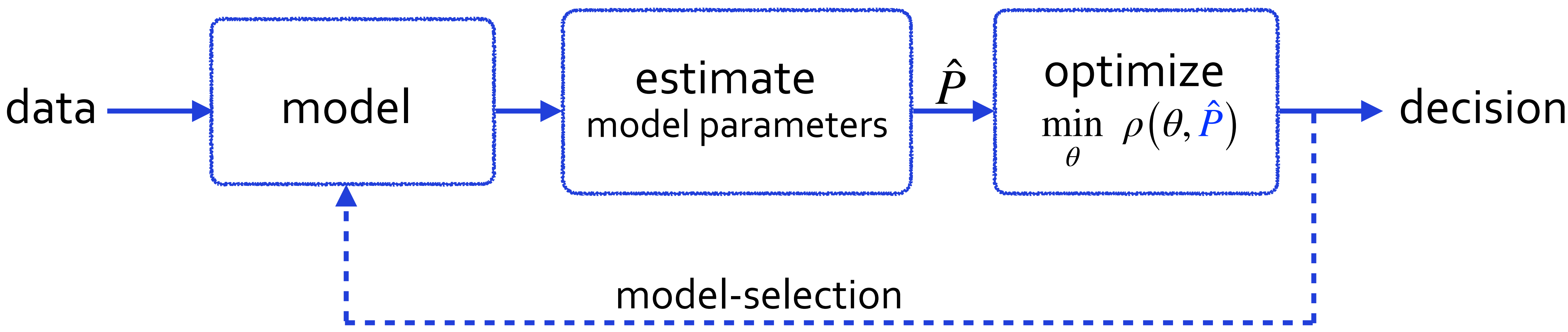


$\mathbf{CVaR}_{1-\beta}\left[L(X,\boldsymbol{\theta})\right]$

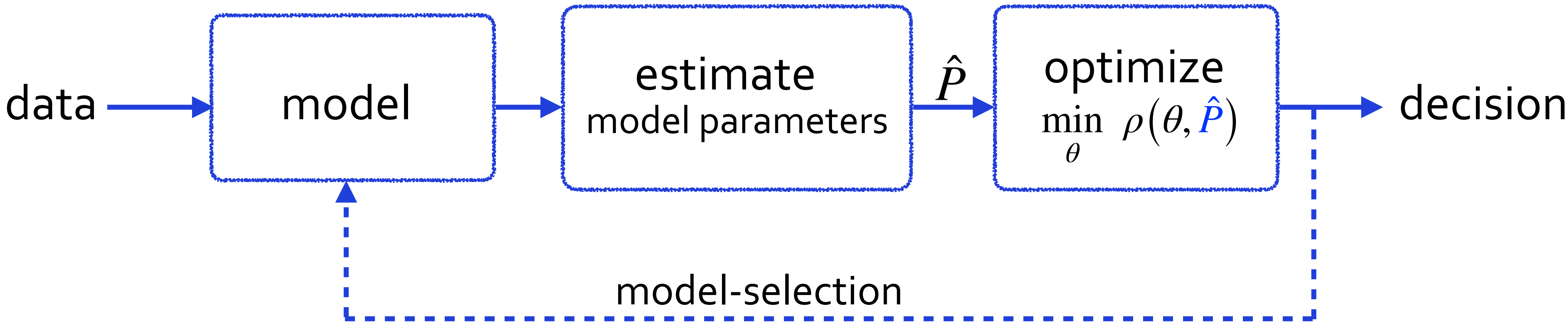
For minimizing sample average,

sample requirement $= \frac{1}{\beta} \times$ usual requirement

Estimate, then optimize



Estimate, then optimize



Copulas

(Gaussian, t, Archimedean,...)

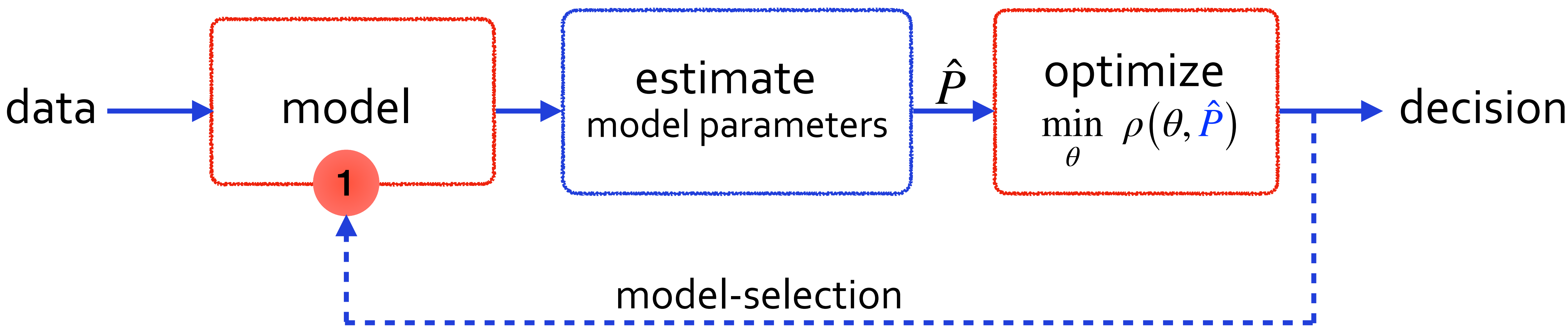
elliptical distributions

multivariate regularly varying

extreme value distributions

quantile regression models

A statistical bottleneck: Incorrect distributional model affects downstream optimization



the need to carefully handle bias created by plugging-in a wrong distributional model

A statistical bottleneck: Incorrect distributional model affects downstream optimization

Handling model-bias:

- Better models
- Expressive model classes
- Inject conservative bias with robust optimization

A statistical bottleneck: Incorrect distributional model affects downstream optimization

Handling model-bias:

- Better models
- Expressive model classes
- Inject conservative bias with robust optimization

Pickands dependence function (Pickands '81)
d-max decreasing neural nets (Hasan et al '22)

A statistical bottleneck: Incorrect distributional model affects downstream optimization

Handling model-bias:

- Better models
- Expressive model classes
- Inject conservative bias with robust optimization

Worst-case CVaR and robust chance constraints

El Ghaoui '03, Calafiore and El Ghaoui `06, Chen et al '10, Zymler '13, Natarajan et al '14, Hanasusanto et al '15, '17, Van Parys et al '15, Van Parys et al '16, Esfahani and Kuhn '18, Lofti & Zenios '18, Duan et al '18, Jiang & Guan '18, Xie '18, Xie & Ahmed '18, Li et al '19, Xie & Ahmed '19, Zhang et al '18, Ji & Lejeune '21, Chen et al'22, Rahimian and Mehrotra '22

A statistical bottleneck: Incorrect distributional model affects downstream optimization

Handling model-bias:

- Better models
- Expressive model classes
- Inject conservative bias with robust optimization

    Convexity constraint (Mottet & Lam '17)
    Orthounimodal shape constraints (Lam et al '21)

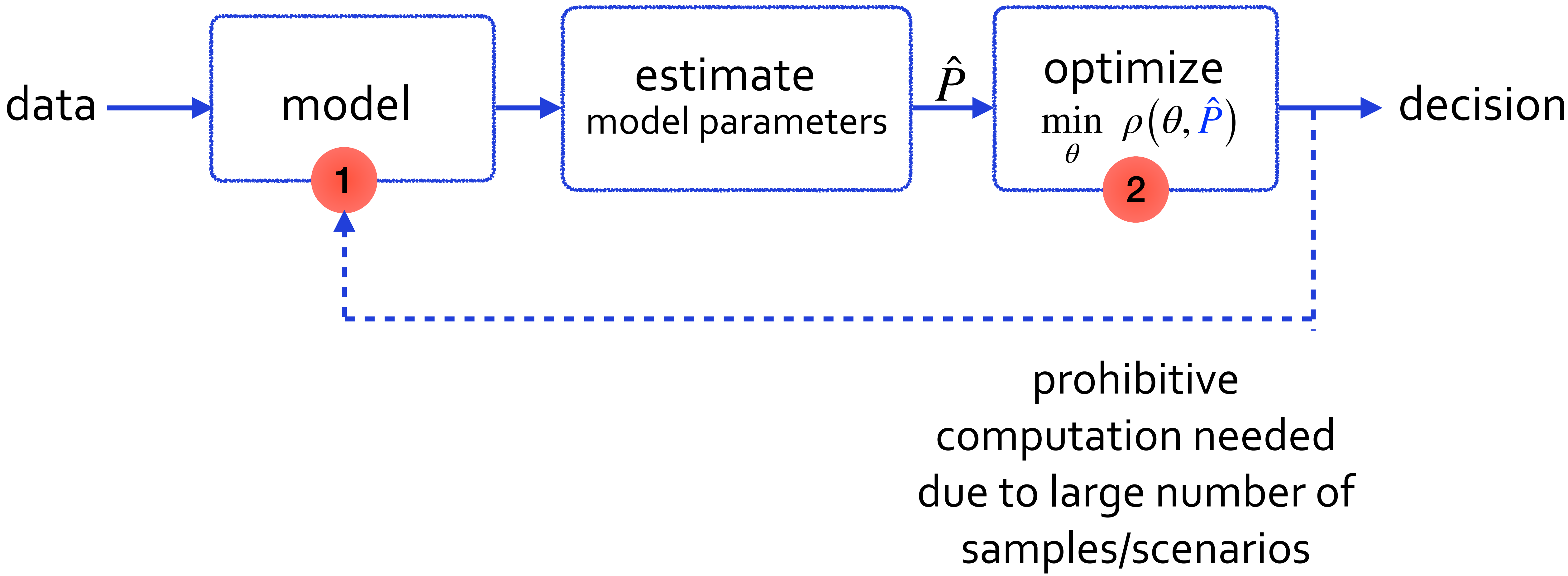
A statistical bottleneck: Incorrect distributional model affects downstream optimization

Handling model-bias:

- Better models
- Expressive model classes
- Inject conservative bias with robust optimization

Given a distributional model, can we have an algorithm to "debias" the objective of its nonparametric model error?

A computational bottleneck: Rarity implies prohibitive no. of scenarios/samples required



A computational bottleneck:
Rarity implies prohibitive no. of scenarios/samples required

Variance reduction techniques

- Importance sampling, stratified sampling, control variates, etc.
- Importance scenario generation
- Problem-driven scenario generation
  Fairbrother et al '19

A computational bottleneck: Rarity implies prohibitive no. of scenarios/samples required

Variance reduction techniques

- Importance sampling, stratified sampling, control variates, etc.
- Importance scenario generation
- Problem-driven scenario generation
  Fairbrother et al '19

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

Prominent hurdles & solution approaches

- Even two random vectors proportional to each other can be “nearly singular” to each other in large dimensions

Nemirovski & Shaprio ‘06

Prominent hurdles & solution approaches

	portfolio VaR	credit risk	sums	pert, dist. n/w	p/w-linear
normal	GHS'00	GL'05 GKS'08		BLN'19	BHLZ '22
extremal	GHS'02	BJZ'08			
iid heavy-tailed			JS'02, AK'06, DLW'07	JKS'07	
log-elliptical			BR, ABJR '11		

Naive Monte Carlo

sample requirement $= \frac{1}{\beta} \times$ usual requirement

⟹

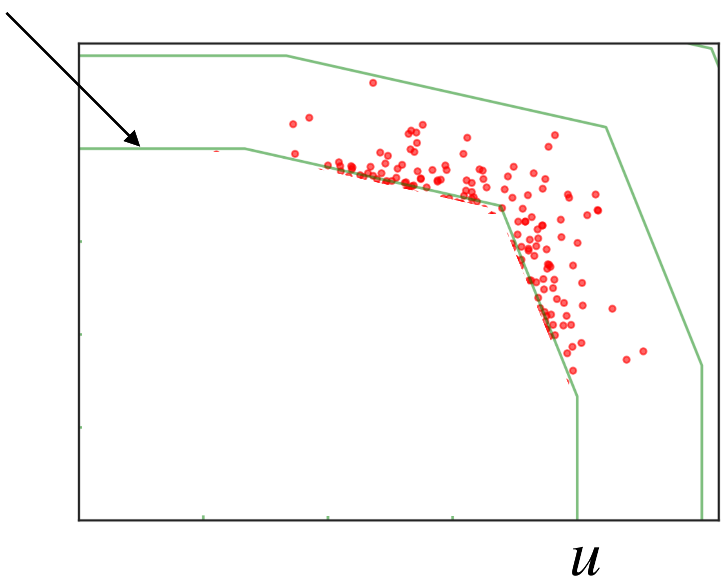
Log-efficient sampler

sample requirement $= \left(\frac{1}{\beta}\right)^{o(1)} \times$ usual

Prominent hurdles & solution approaches

level curves of loss

In red: excess loss samples $X \mid L(X) > u$



X ~ multivariate normal

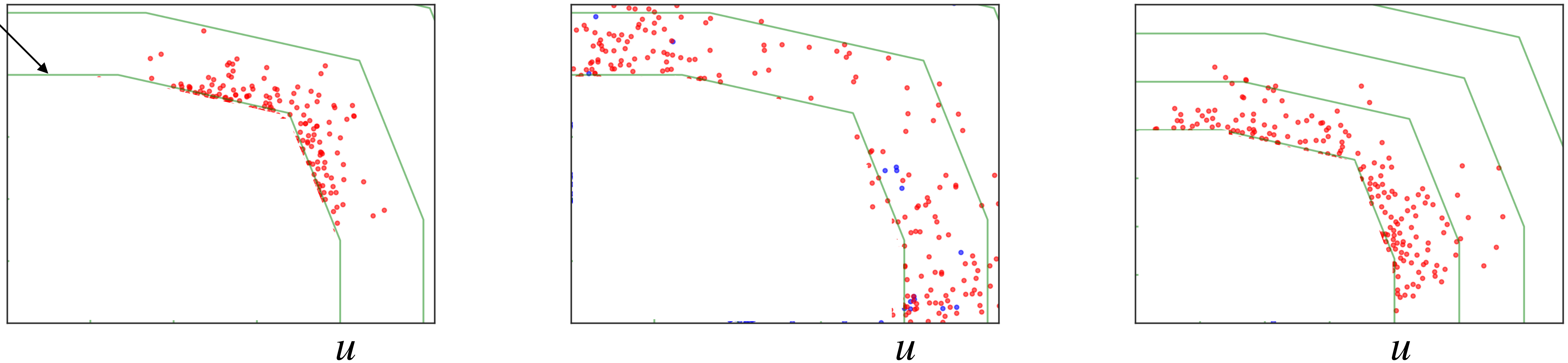
X ~ heavier-tailed Weibull marginals + Gaussian copula

X ~ exponential marginals + Gaussian copula

Prominent hurdles & solution approaches

level curves of loss

In red: excess loss samples $X \mid L(X) > u$

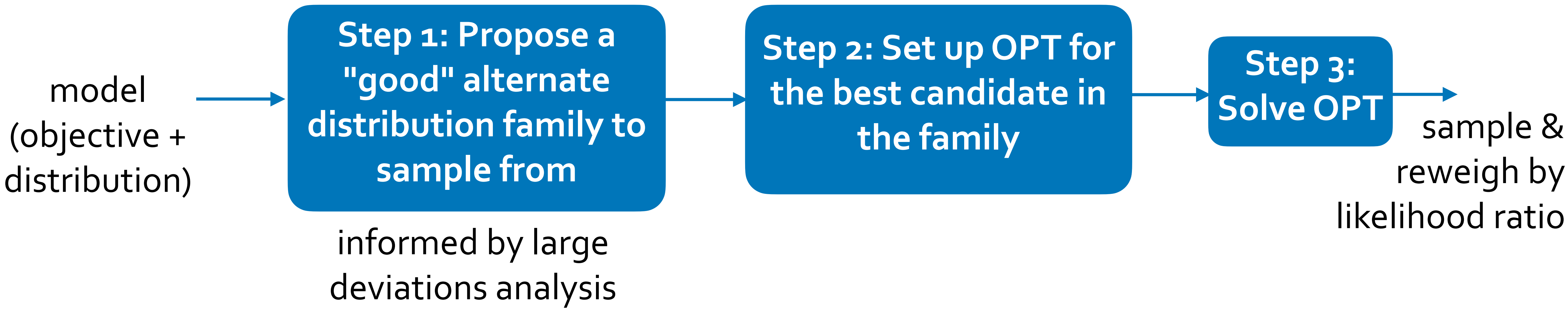


X ~ multivariate normal

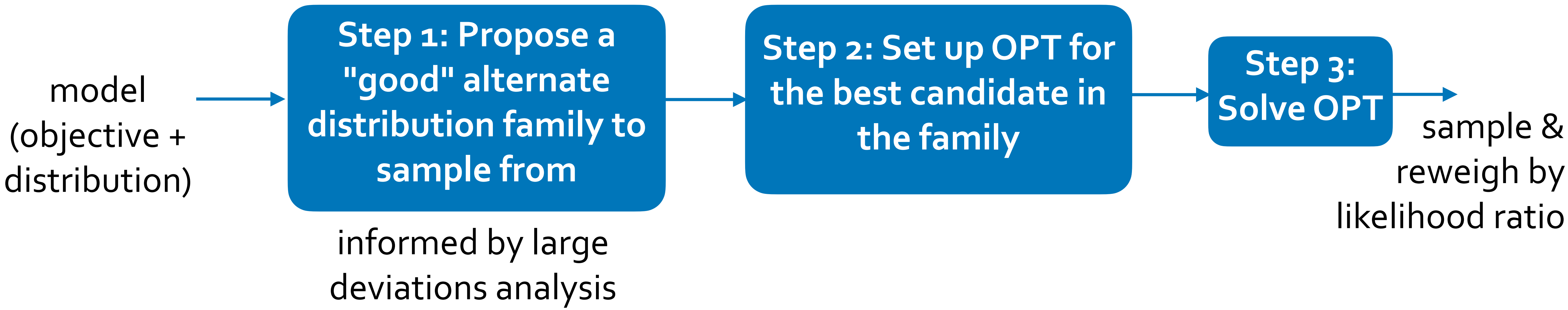
X ~ heavier-tailed Weibull marginals + Gaussian copula

X ~ exponential marginals + Gaussian copula

Prominent hurdles & solution approaches



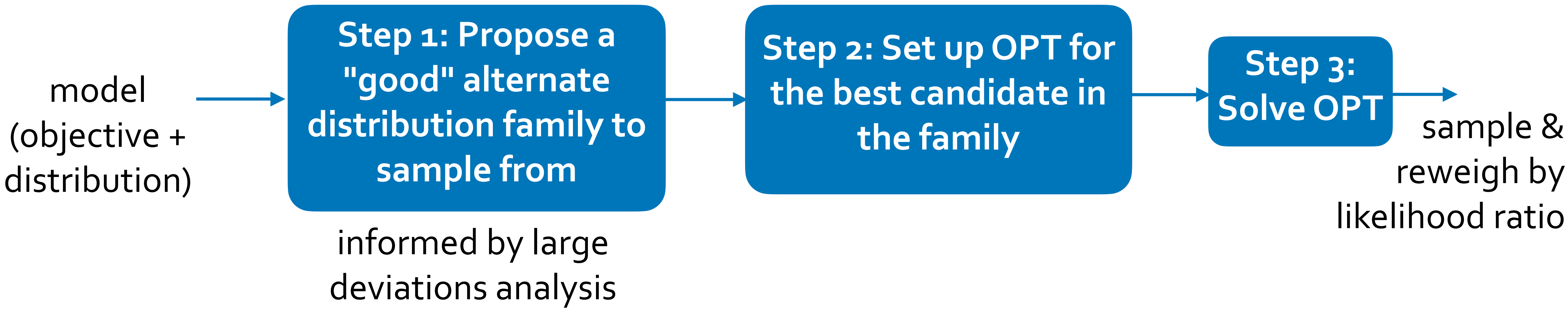
Prominent hurdles & solution approaches

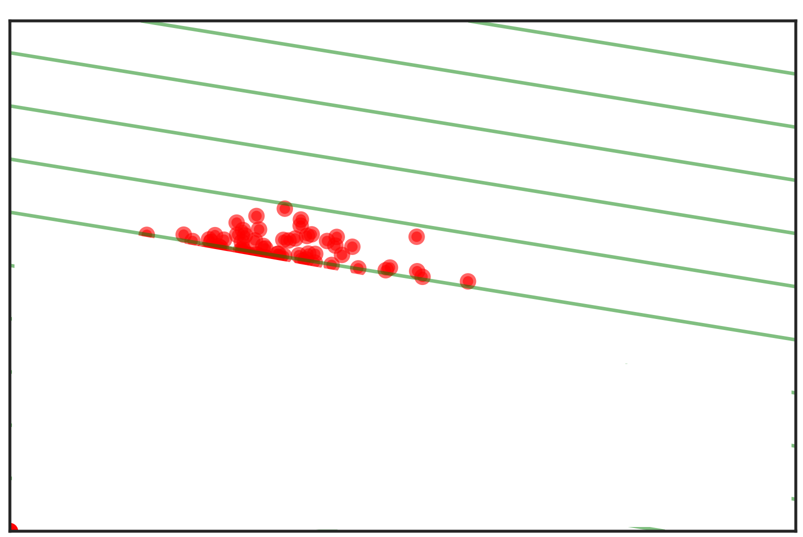


For multivariate normal:

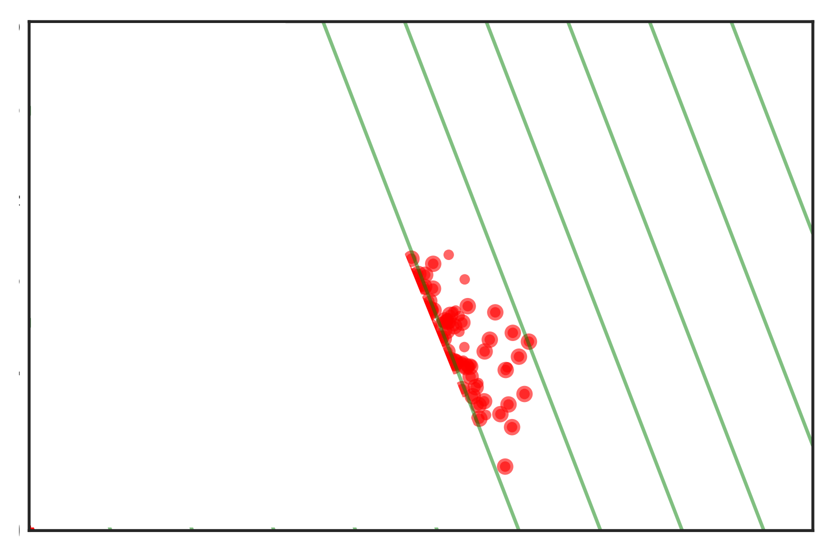
- quadratic program (Glasserman et al '00, '05)
  + combinatorial structure (Glasserman et al'08)
- Mixed-integer program (Bai et al '20)

Prominent hurdles & solution approaches





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

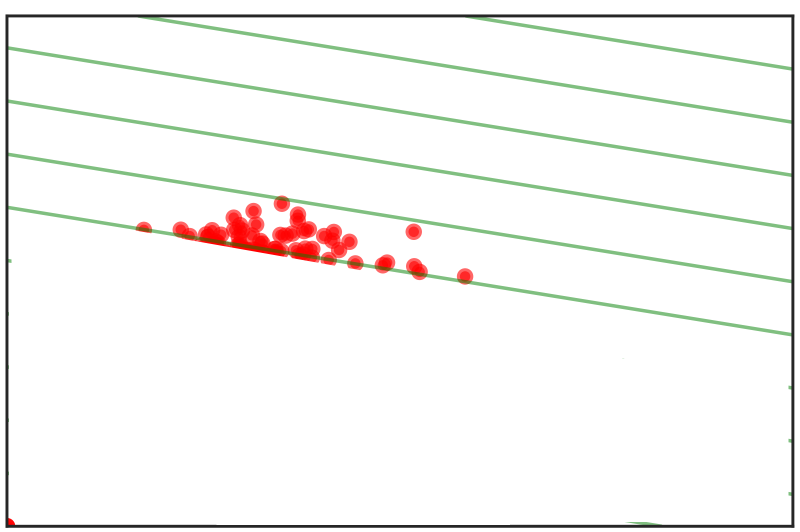


$L(\boldsymbol{X}) = 0.8X_1 + 0.2X_2$

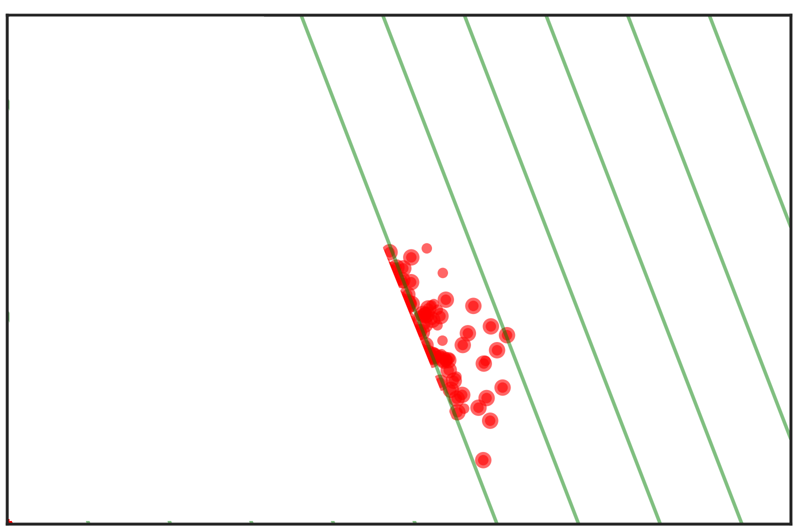
Prominent hurdles & solution approaches

What is a good sampler for one decision choice is often not good for other → a bottleneck in optimization

(Barrera et al '14)



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

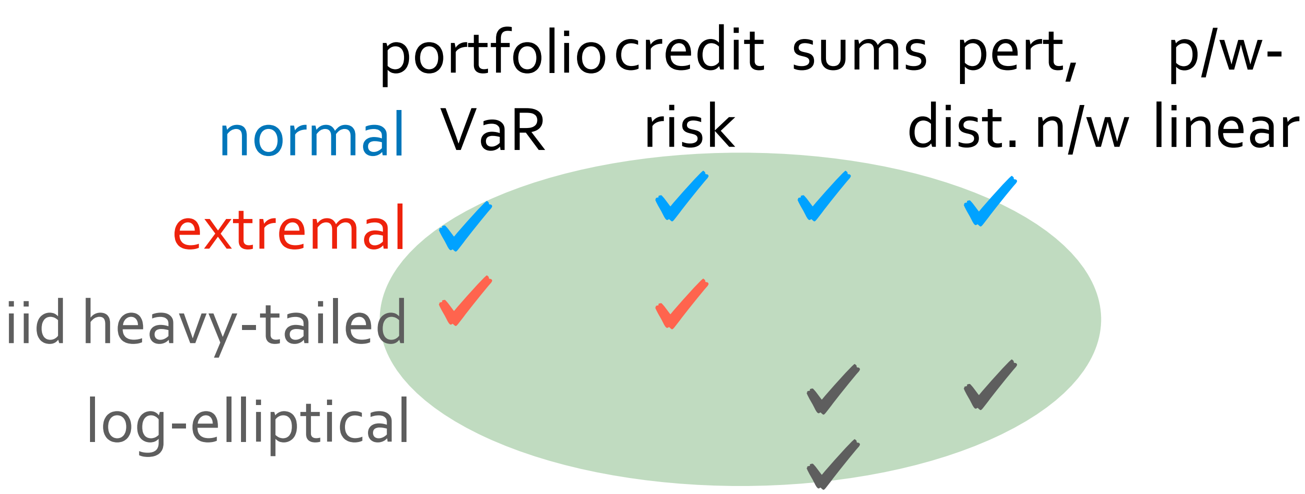


$L(\boldsymbol{X}) = 0.8X_1 + 0.2X_2$

Prominent hurdles & solution approaches

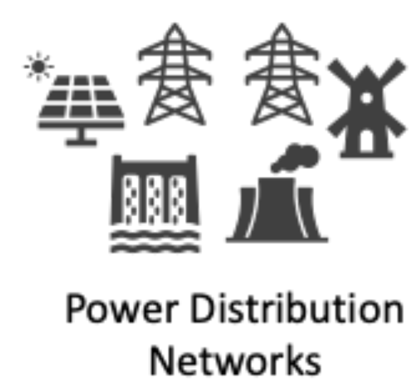
	portfolio VaR	credit risk	sums	pert, dist. n/w	p/w-linear
normal	✓	✓	✓	✓	
extremal	✓	✓			
iid heavy-tailed			✓	✓	
log-elliptical			✓		

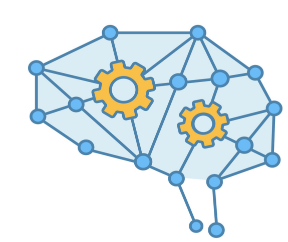
Prominent hurdles & solution approaches



optimization objectives?

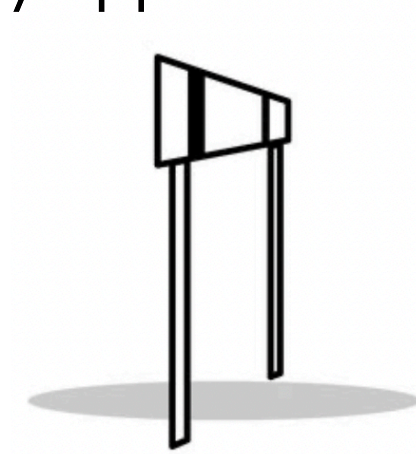
+ ML?





Prominent hurdles & solution approaches

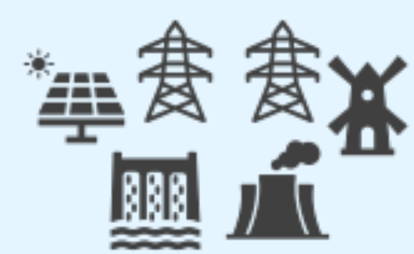
Can we have samplers which are efficient & broadly applicable?



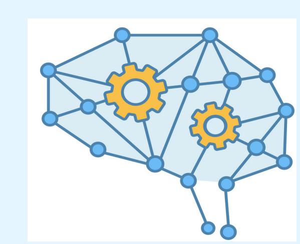
	portfolio VaR	credit risk	sums	pert, dist. n/w	p/w-linear
normal	✓	✓	✓	✓	
extremal	✓	✓			
iid heavy-tailed			✓	✓	
log-elliptical			✓		

optimization objectives?

+ ML?



Power Distribution Networks



Prominent hurdles & solution approaches

Can we have samplers which are efficient & broadly applicable?

How to integrate seamlessly with optimization?

Prominent hurdles & solution approaches

Can we have samplers which are efficient & broadly applicable?

How to integrate seamlessly with optimization?

Can we have an algorithm which adapts its Importance Sampling distribution to the objective at hand?

Two questions in this talk

Q1: Can we have an algorithm which adapts its Importance Sampling distribution to the objective at hand?

(computational bottleneck)

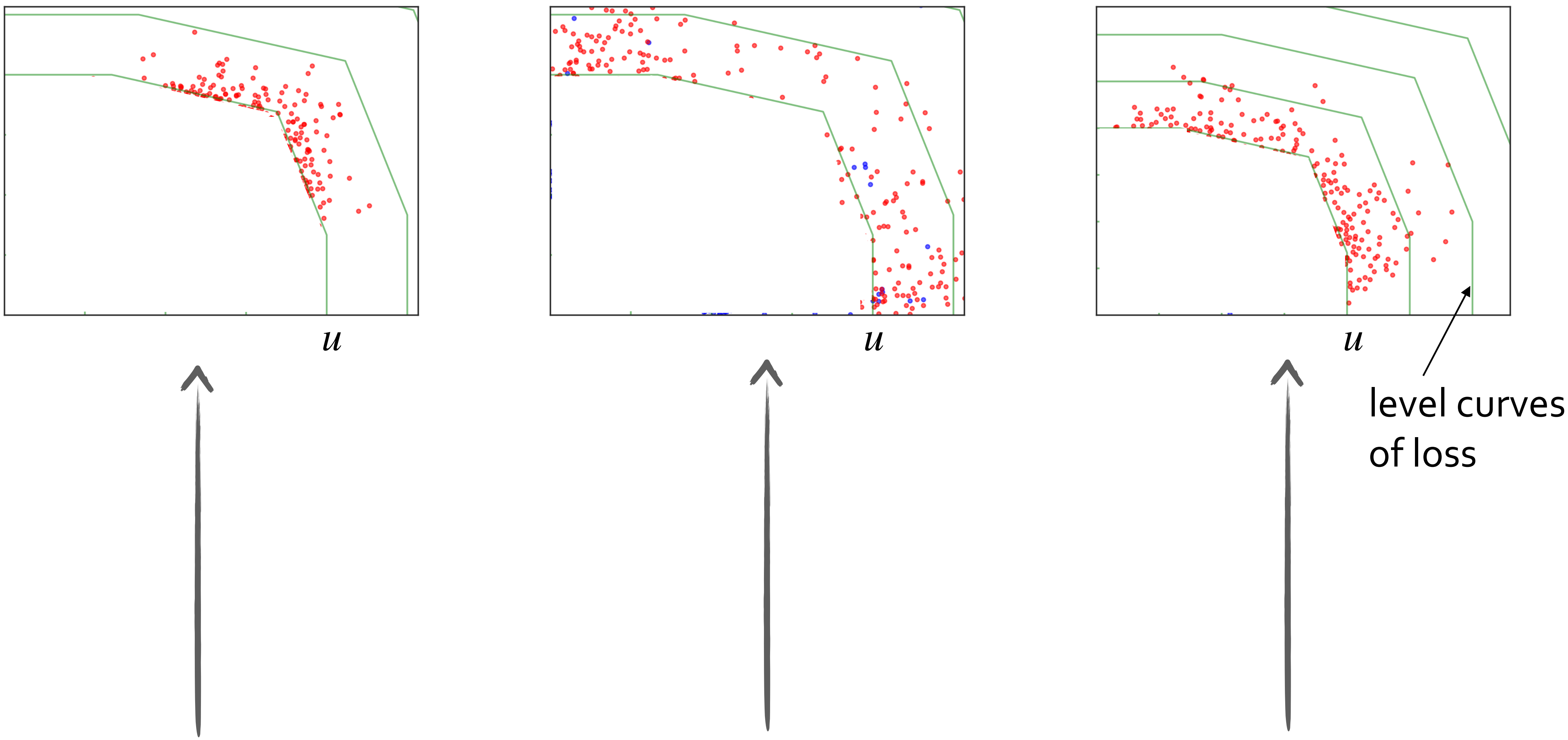
Q2: Given a distributional model, can we have an algorithm to "debias" the objective of its nonparametric model error?

(statistical bottleneck)

A key observation
and its implications for the two bottlenecks

Recall: Why efficient samplers are elusive?

In red: excess loss samples $X \mid L(X) > u$



X ~ multivariate normal

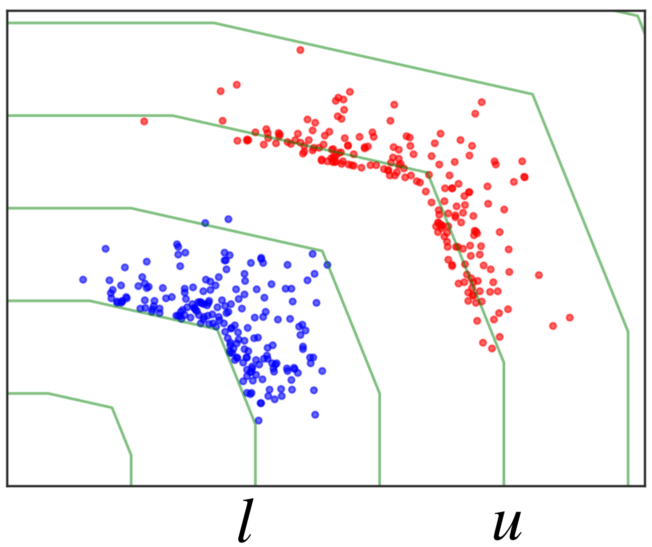
X ~ heavier-tailed Weibull marginals + Gaussian copula

X ~ exponential marginals + Gaussian copula

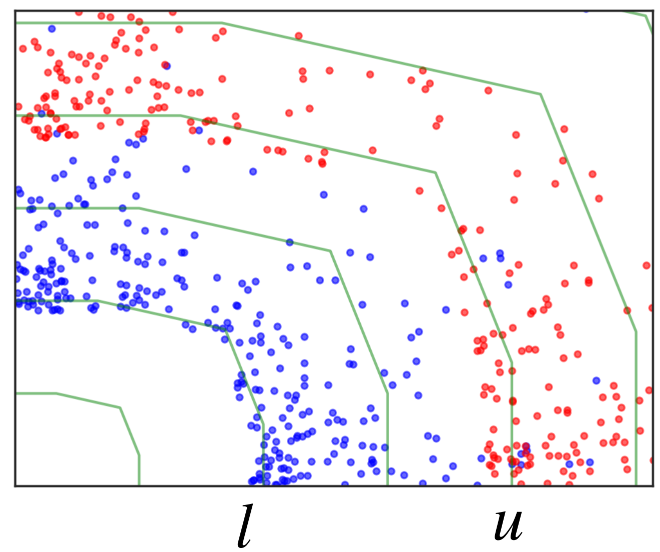
Key observation: Tail events occur in structurally similar ways

In blue: excess loss samples $X \mid L(X) > l$

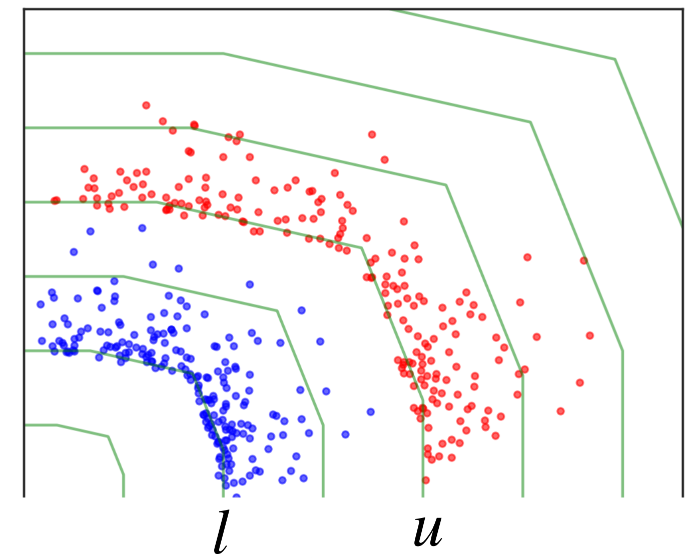
In red: excess loss samples $X \mid L(X) > u$



X ~ multivariate normal

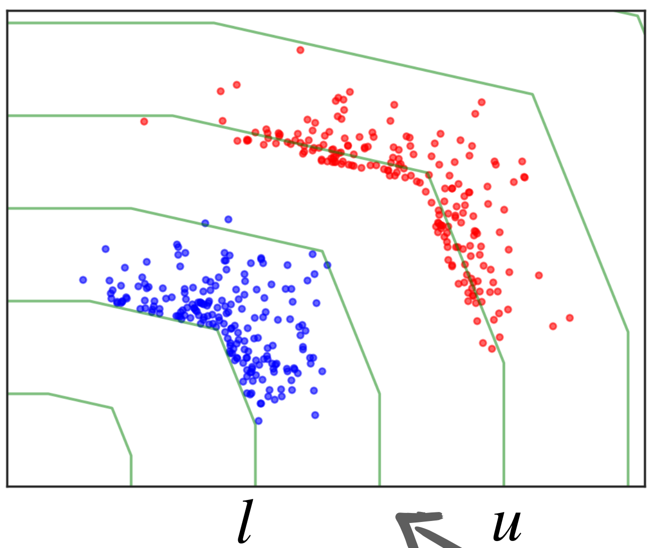


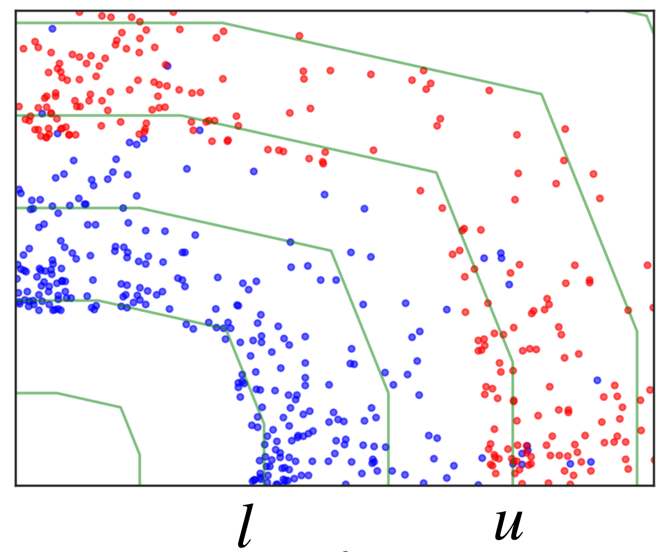
X ~ heavier-tailed Weibull marginals + Gaussian copula

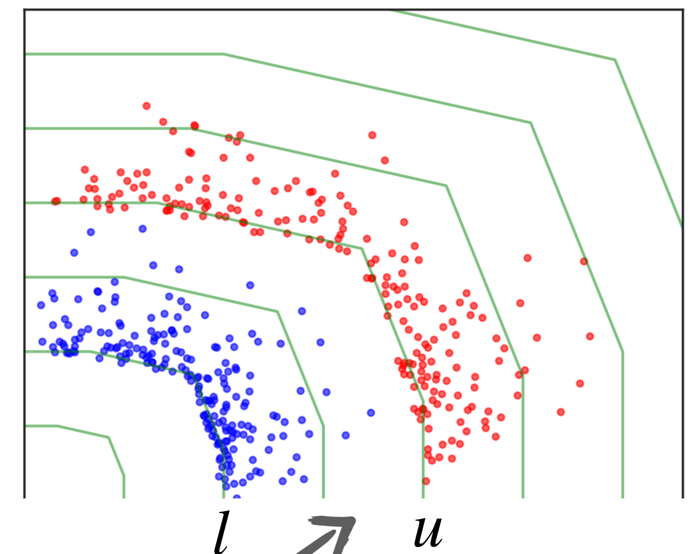


X ~ exponential marginals + Gaussian copula

Key idea: Tail events occur in structurally similar ways







Is there ONE transformation which can induce each of these 'red' profiles when applied to these less-rare 'blue' samples?

X ~ multivariate normal

X ~ heavier-tailed Weibull marginals + Gaussian copula

X ~ exponential marginals + Gaussian copula

Search for a good density $\longrightarrow$ Search for a good transformation

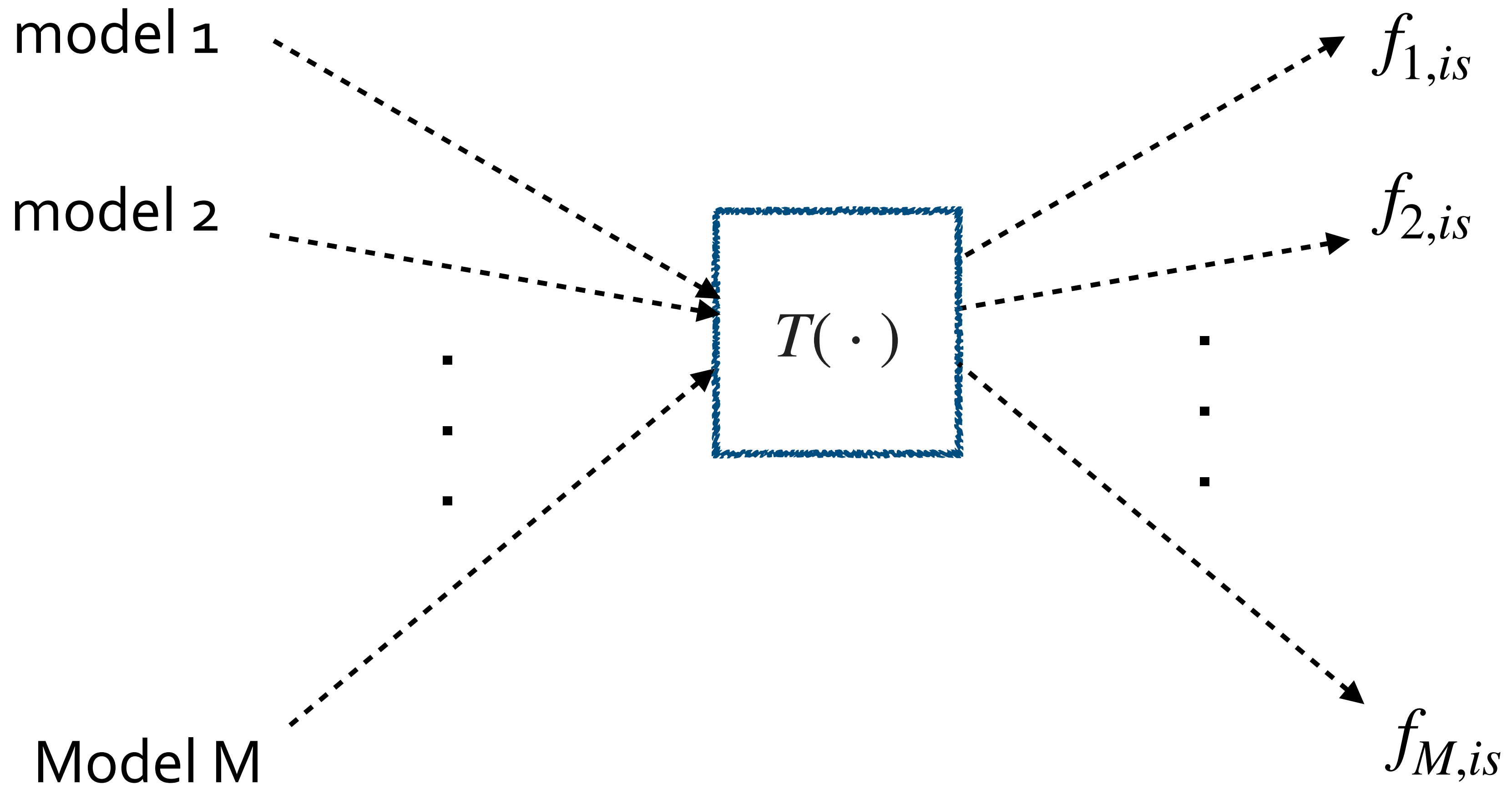
model 1 $\dashrightarrow$ $f_{1,is}$

model 2 $\dashrightarrow$ $f_{2,is}$

$\vdots$

Model M $\dashrightarrow$ $f_{M,is}$

Search for a good density $\longrightarrow$ Search for a good transformation



Resolving the computational bottleneck

Q1: Can we have an algorithm which adapts its Importance Sampling distribution to the objective at hand?

- A fixed elementary transformation of the samples is efficient!
- Suited for a broad variety of risk management models, including those using sophisticated predictors
- Ability to resolve the bottleneck in variance reduction for optimization models with CVaR objectives or chance-constraints

Resolving the statistical bottleneck

Q2: Given a plug-in distributional model, can we have a procedure to "debias" the objective?

- Debiased objective = objective with plug-in + a correction term
- Objective has zero sensitivity to perturbations in plug-in model

Resolving the statistical bottleneck

Q2: Given a plug-in distributional model, can we have a procedure to "debias" the objective?

- Debiased objective = objective with plug-in + a correction term
- Objective has zero sensitivity to perturbations in plug-in model
- If modeller's choice induces a bias = ε_n in the objective, bias in debiased objective is only ε_n^2!
- Convexity retained in the debiased objective

Resolving the statistical bottleneck

Q2: Given a plug-in distributional model, can we have a procedure to "debias" the objective?

- Debiased objective = objective with plug-in + a correction term
- Objective has zero sensitivity to perturbations in plug-in model
- If modeller's choice induces a bias = ε_n in the objective,
  bias in debiased objective is only ε_n^2!
- Convexity retained in the debiased objective

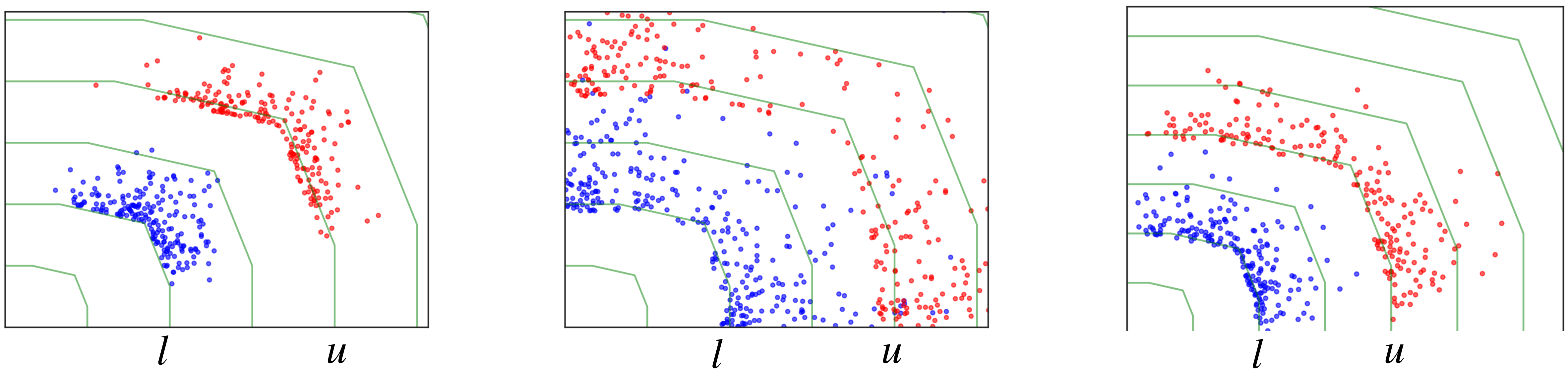
Newey and Stoker, '93
Murphy and van der Vaart '97
Van der Vaart '99
Chernozhukov et al. '16, '17
Foster and Syrgkanis '19
Newey and Ichimura '22

Debiasing in Operations Research literature
Gupta, Huang, Rusmevichientong '21

Resolving the statistical bottleneck

Q2: Given a distributional model, can we have an algorithm to "debias" the objective of its nonparametric model error?

Debiased objective = objective($\hat{P}$) + a correction term



Resolving the statistical bottleneck

Q2: Given a distributional model, can we have an algorithm to "debias" the objective of its nonparametric model error?

Debiased objective = objective($\hat{P}$) + a correction term

- Debiased objective has zero sensitivity to perturbations in plug-in model

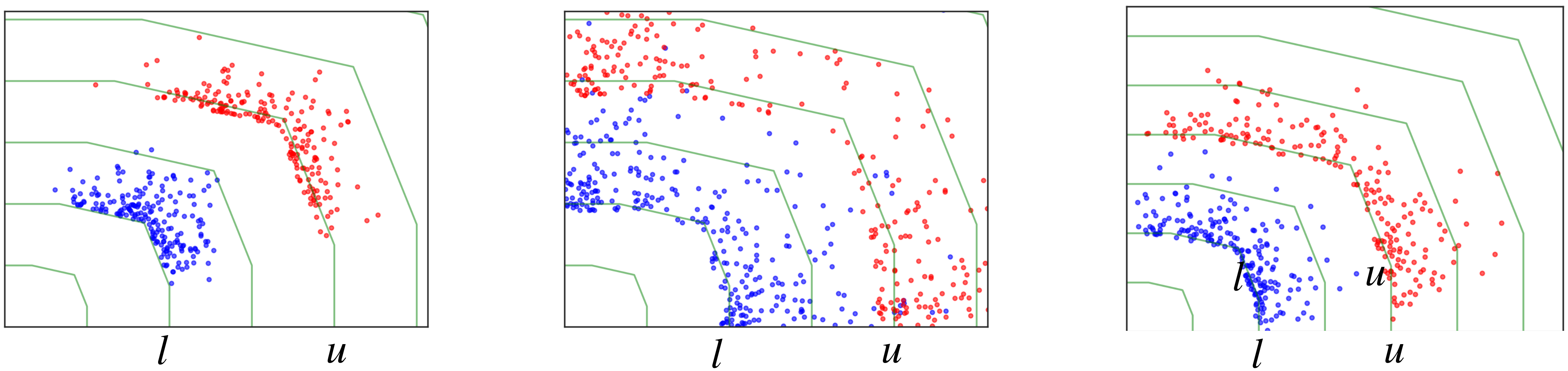
If modeller's choice of $\hat{P}$ induces a bias = $O_p(\varepsilon_n)$ ⇒ **Debiased objective only carries a bias = $O_p(\varepsilon_n^2)$**

- Convexity retained in the debiased objective!

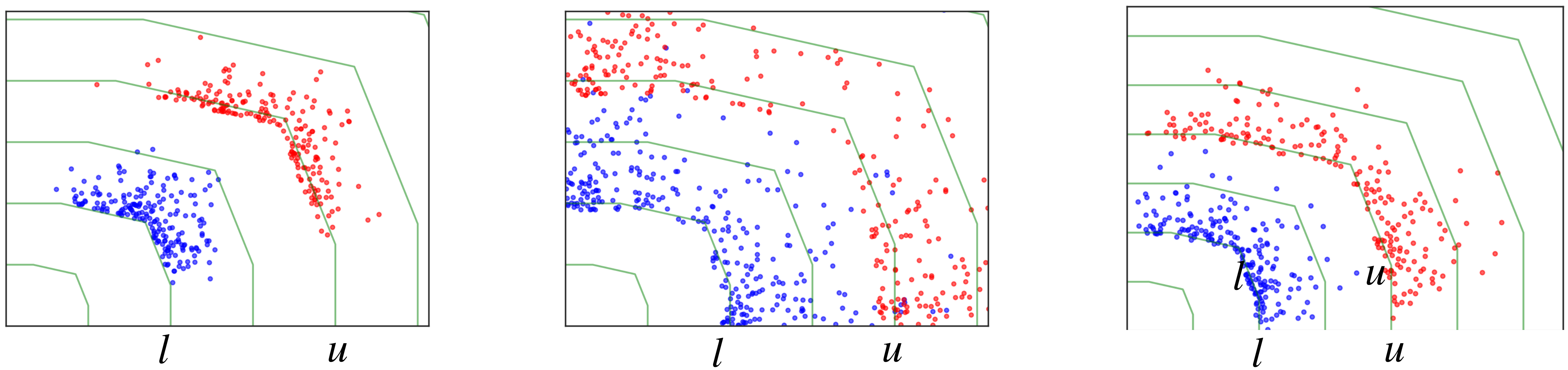
Outline of the talk

- Introduction
- Challenges due to rarity & model-bias
- Why algorithmic approaches have been elusive?
- Key observation & its implications
- Q1: Can a sampler adapt its IS distribution to the problem-at-hand?
- Q2: Can we correct the plug-in model-bias?
- Summary

Tail modeling based for studying self-similarity



Tail modeling based for studying self-similarity



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

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

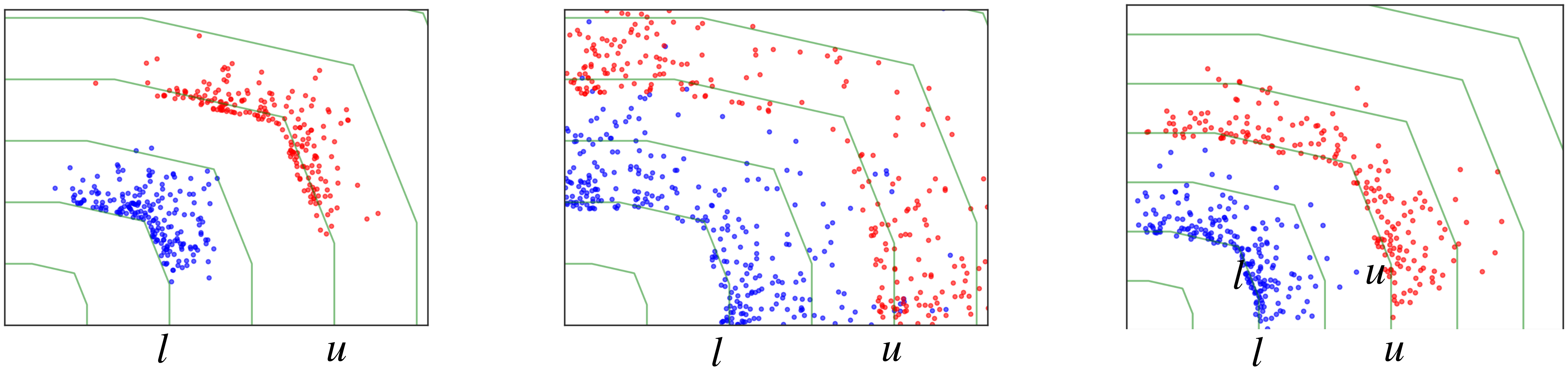
(that is, φ is regularly varying)

Some examples:

elliptical densities,
exponential family,
log-concave densities,
Gaussian copula,
t-copula,
archimedean copula, …

+ light/heavy-tailed marginals

Tail modeling based for studying self-similarity



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

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

(that is, φ is regularly varying)

- Heavy-tailed:

  pdf of $\boldsymbol{X}$ is regularly varying

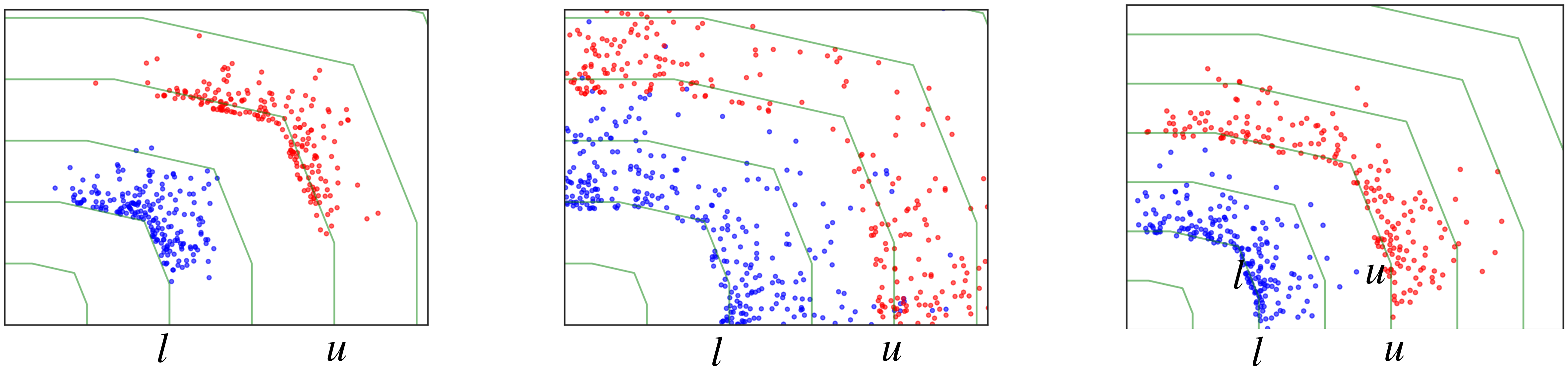
  (See, eg. Resnick '07, '08)

Some examples:

elliptical densities,
exponential family,
log-concave densities,
Gaussian copula,
t-copula,
archimedean copula, ...

+ light/heavy-tailed marginals

Uncovering a large deviations principle



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

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

(that is, φ is regularly varying)

Theorem [Deo and M '21]

X/n satisfies a large deviations principle:

$$P\,(\boldsymbol{X} \in nA) = \exp\{-t_n\varphi^*(A)+o(t_n)\}$$

and the above similarity in conditional excess loss distributions hold

Setup: Assumptions on the loss

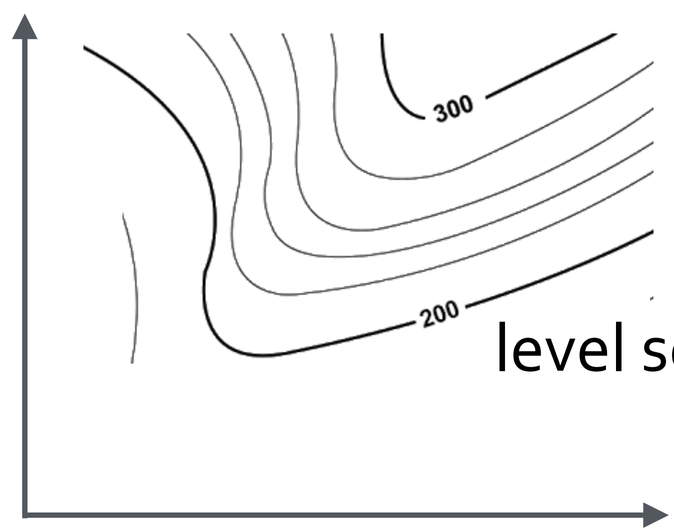
- Asymptotically homogenous loss:

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

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

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

(that is, φ is regularly varying)



Some examples:

LP, MILP, QP objectives with random coefficients,

their optimal values,

⋮

losses written in terms of feature maps/decision rules specified with kernels and ReLU neural networks

Large deviations mechanics:
Intersection of level curves determine the most likely excess loss samples

- Asymptotically homogenous loss:

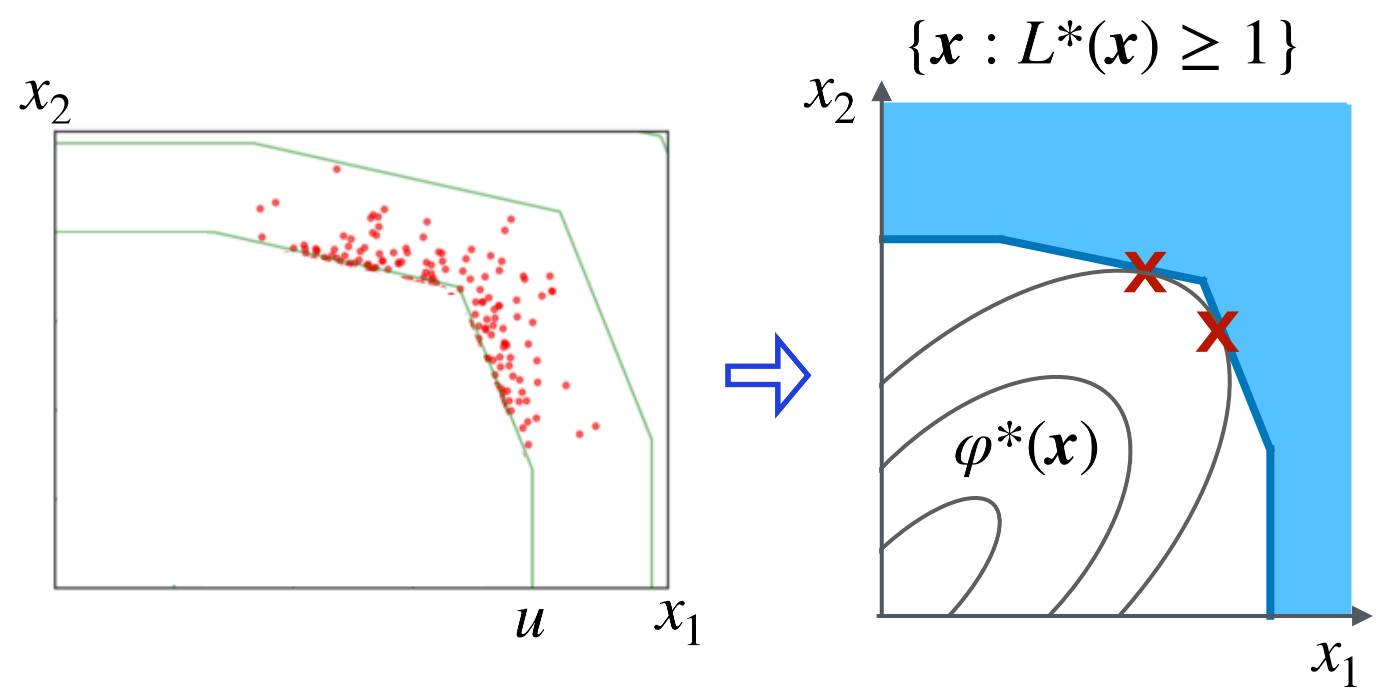
$$\lim_{n\to\infty} \frac{L(n\boldsymbol{x})}{n^{\rho}} = L^*(\boldsymbol{x})$$

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

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

(that is, φ is regularly varying)

eg: + correlated multivariate normal



x $= \arg\min_{\boldsymbol{x}:L^*(\boldsymbol{x})\geq 1} \varphi^*(\boldsymbol{x})$

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

Large deviations mechanics:
Intersection of level curves determine the most likely excess loss samples

- Asymptotically homogenous loss:

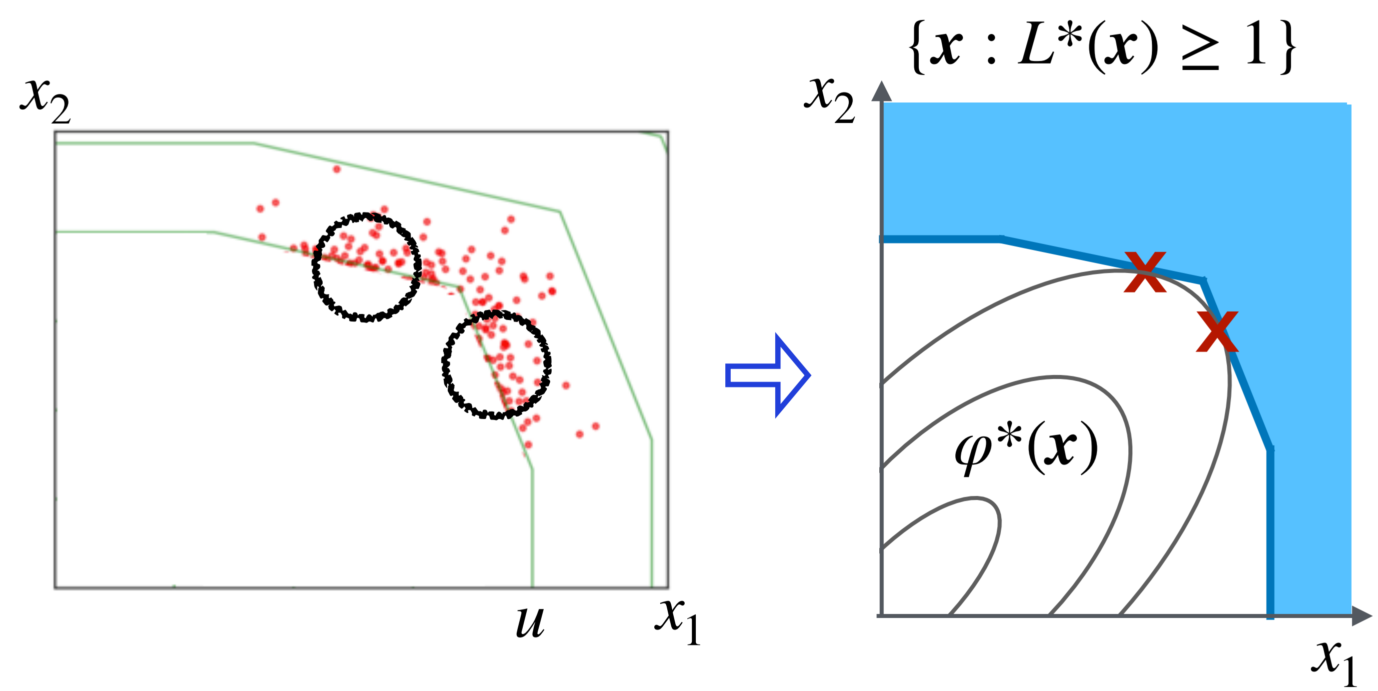
$$\lim_{n\to\infty} \frac{L(n\boldsymbol{x})}{n^\rho} = L^*(\boldsymbol{x})$$

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

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

(that is, φ is regularly varying)

eg: + correlated multivariate normal



$\mathbf{X} = \arg \min_{\boldsymbol{x}:L^*(\boldsymbol{x})\geq 1} \varphi^*(\boldsymbol{x})$

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

Large deviations mechanics: Intersection of level curves determine the most likely excess loss samples

- Asymptotically homogenous loss:

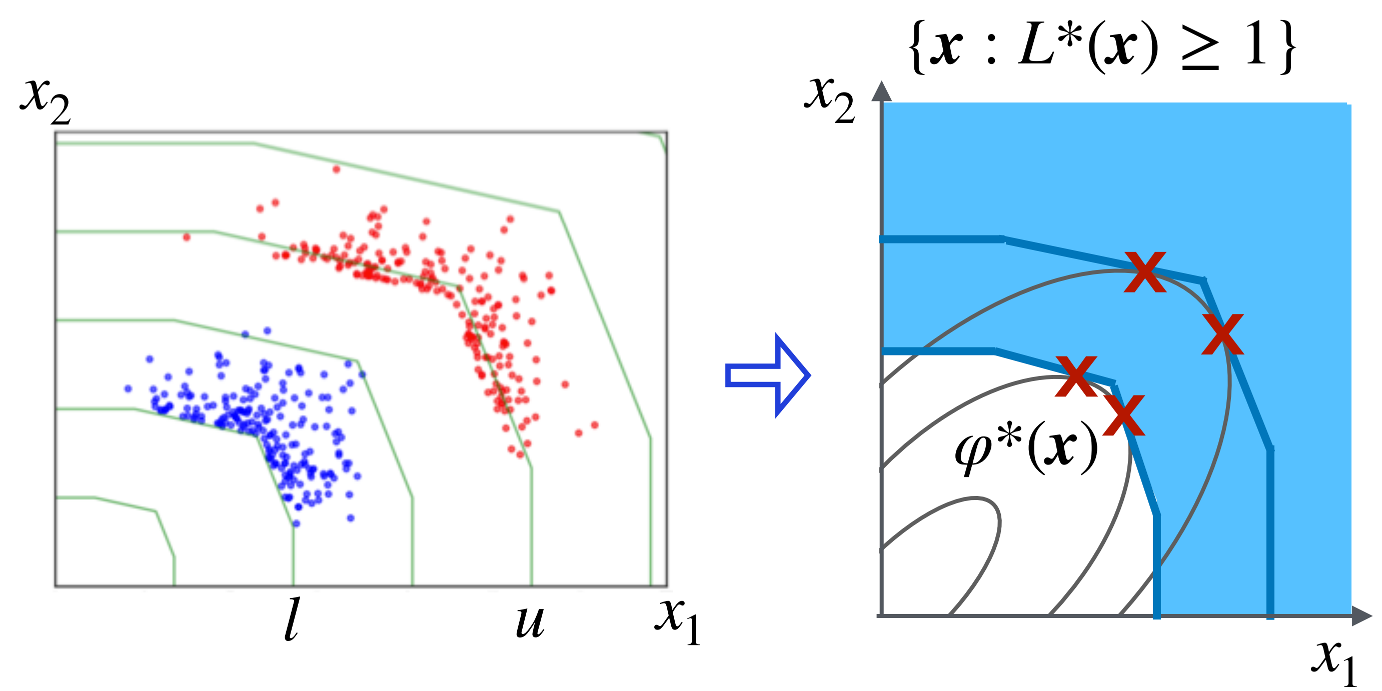
$$\lim_{n\to\infty} \frac{L\left(n\boldsymbol{x}\right)}{n^{\rho}} = L^{*}(\boldsymbol{x})$$

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

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

(that is, φ is regularly varying)

eg: + correlated multivariate normal



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

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

Large deviations mechanics:
Intersection of level curves determine the most likely excess loss samples

- Asymptotically homogenous loss:

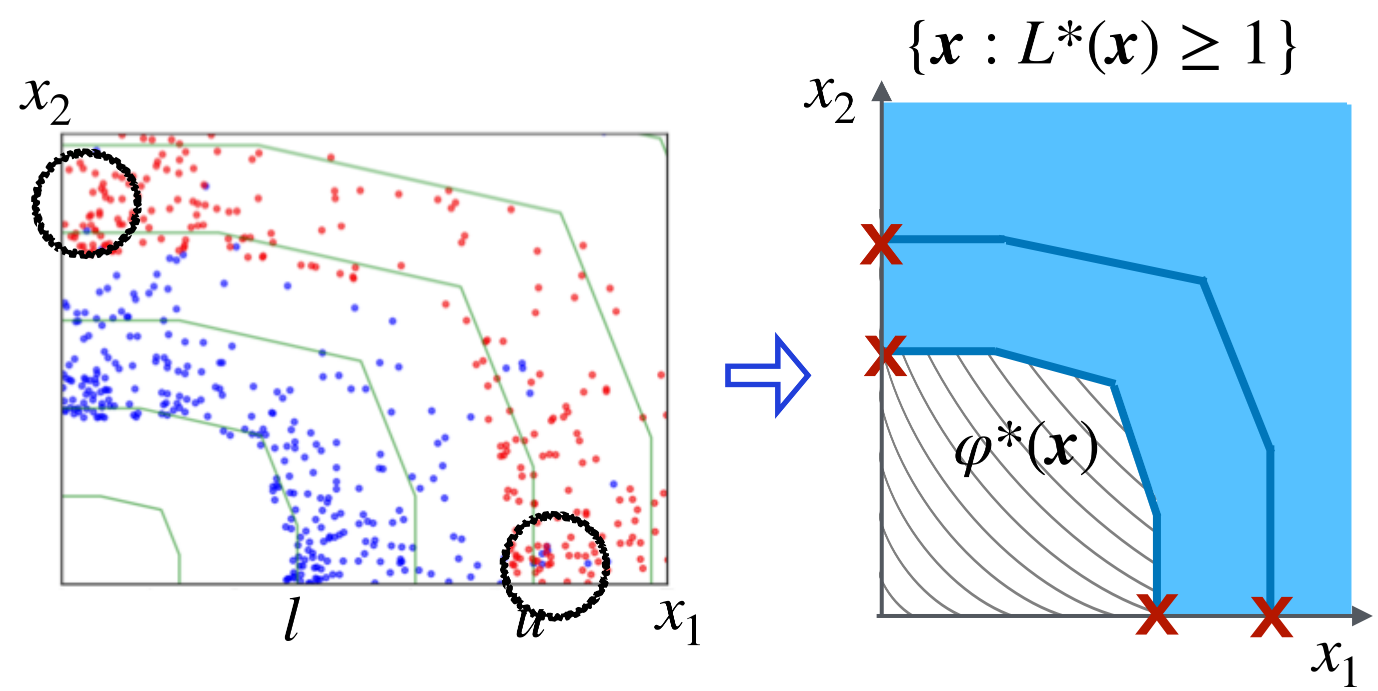
$$\lim_{n\to\infty} \frac{L(n\boldsymbol{x})}{n^{\rho}} = L^{*}(\boldsymbol{x})$$

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

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

(that is, φ is regularly varying)

eg: weibull marginals, gaussian copula



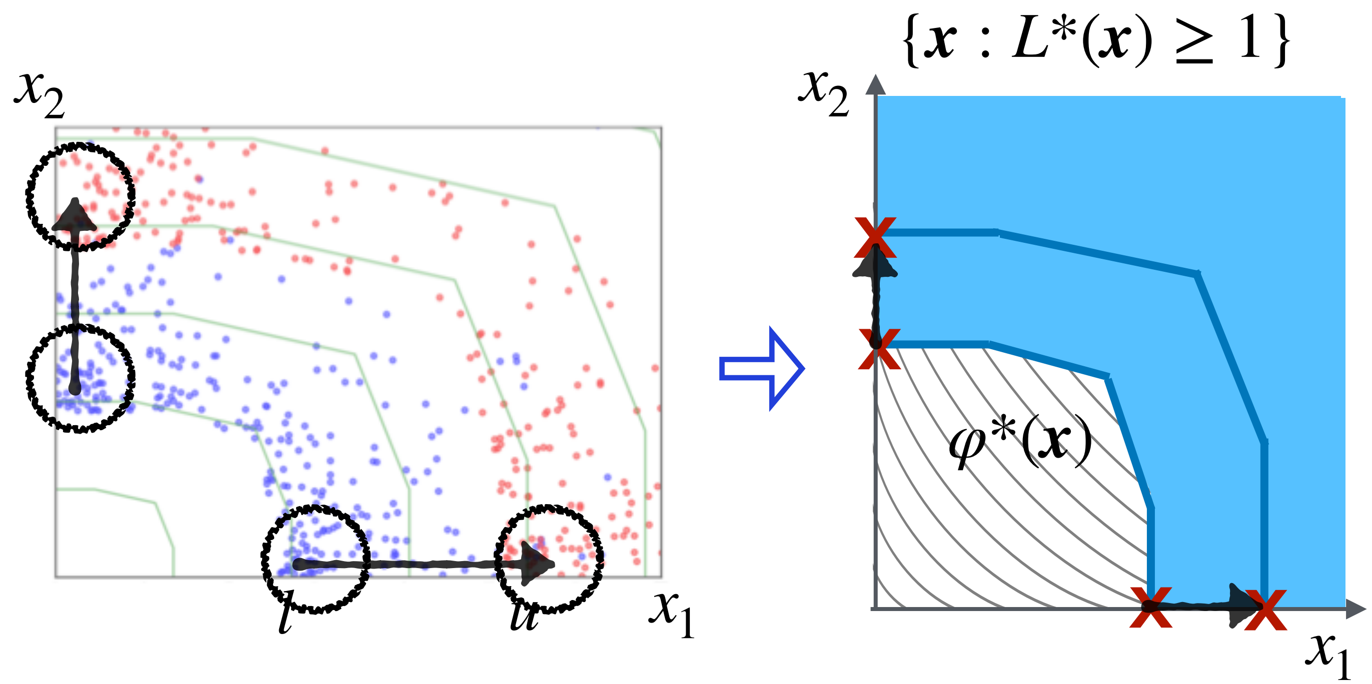
in blue: samples of $\boldsymbol{X} \mid L(\boldsymbol{X}) > l$

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

Back to concentration preserving transformation

Can we find a
rate-point preserving transformation
that is oblivious to the underlying
objective and the distribution?

eg: weibull marginals, gaussian copula



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

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

Concentration-preserving stretching

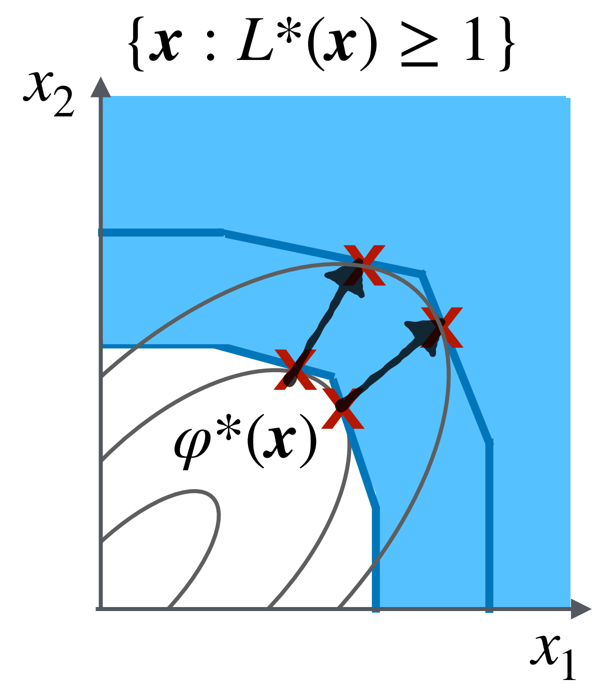
Can we find a
rate-point preserving transformation
that is oblivious to the underlying
objective and the distribution?

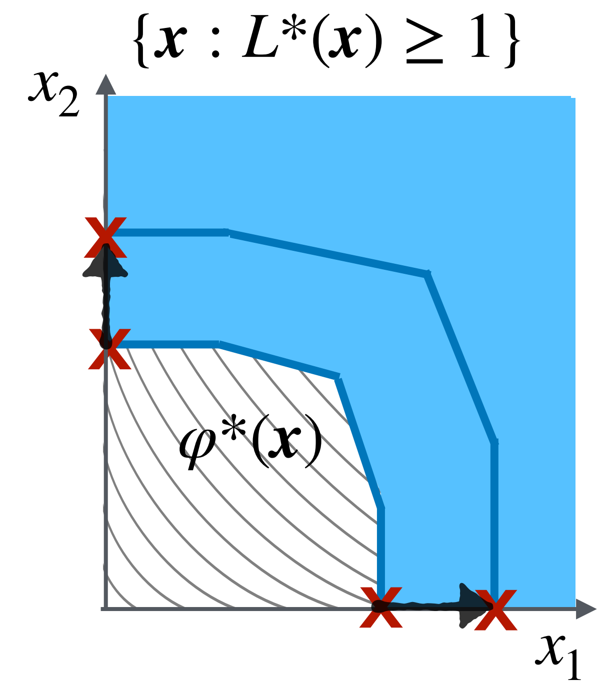
$$\boldsymbol{T}(\boldsymbol{x}) = s^{\kappa(\boldsymbol{x})}\boldsymbol{x}$$

where

$$\kappa(\boldsymbol{x}) = \frac{1}{\rho}\frac{\log|\boldsymbol{x}|}{\log\|\boldsymbol{x}\|_\infty}$$

s = scalar stretch parameter





$$T^*(\boldsymbol{x}) = s^{\alpha_{\min}/\rho} \cdot \boldsymbol{x}$$

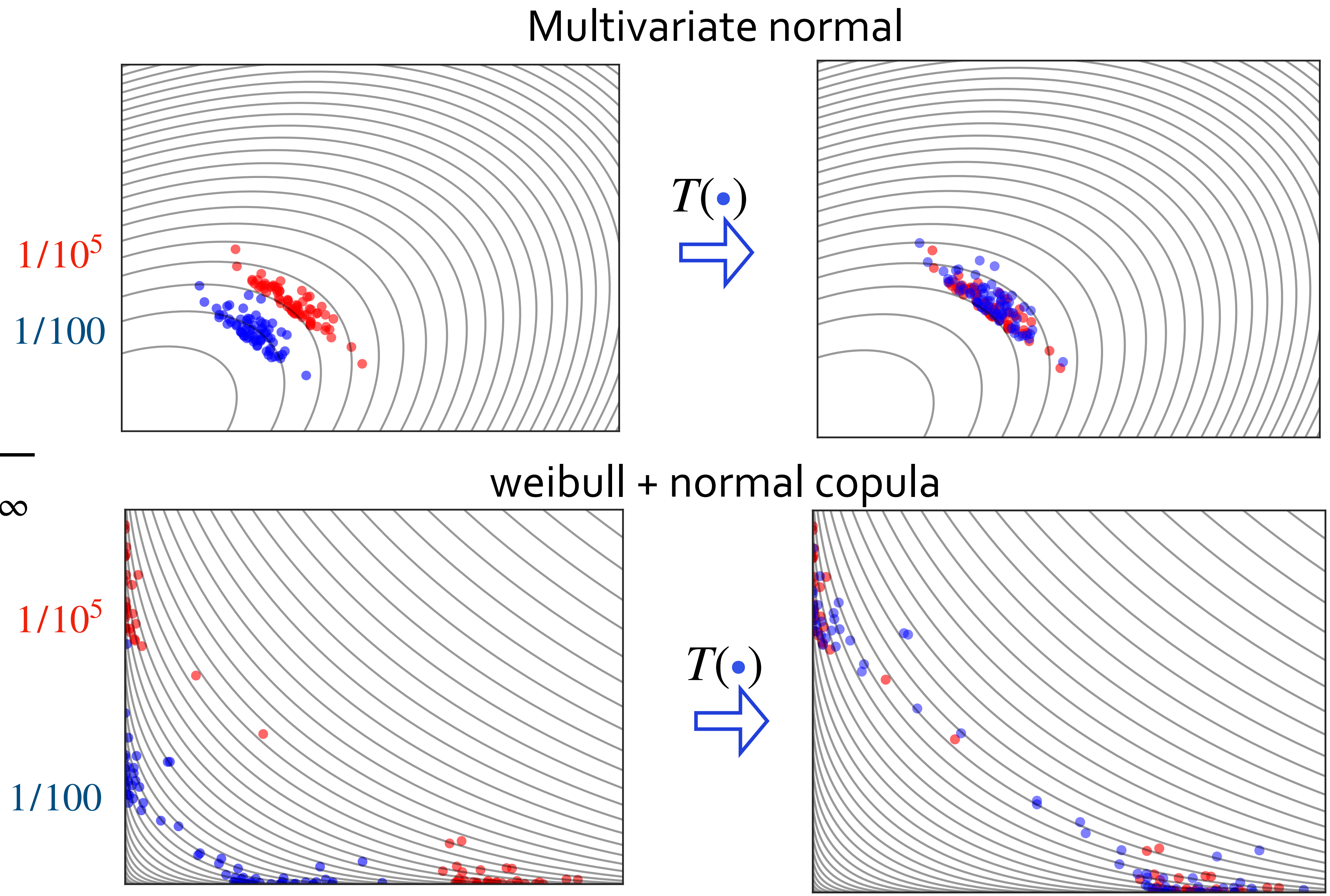
Concentration-preserving stretching, in action

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

where

$$\kappa(\boldsymbol{x}) = \frac{1}{\rho}\frac{\log|\boldsymbol{x}|}{\log\|\boldsymbol{x}\|_\infty}$$

s = 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

Concentration-preserving stretching, in action

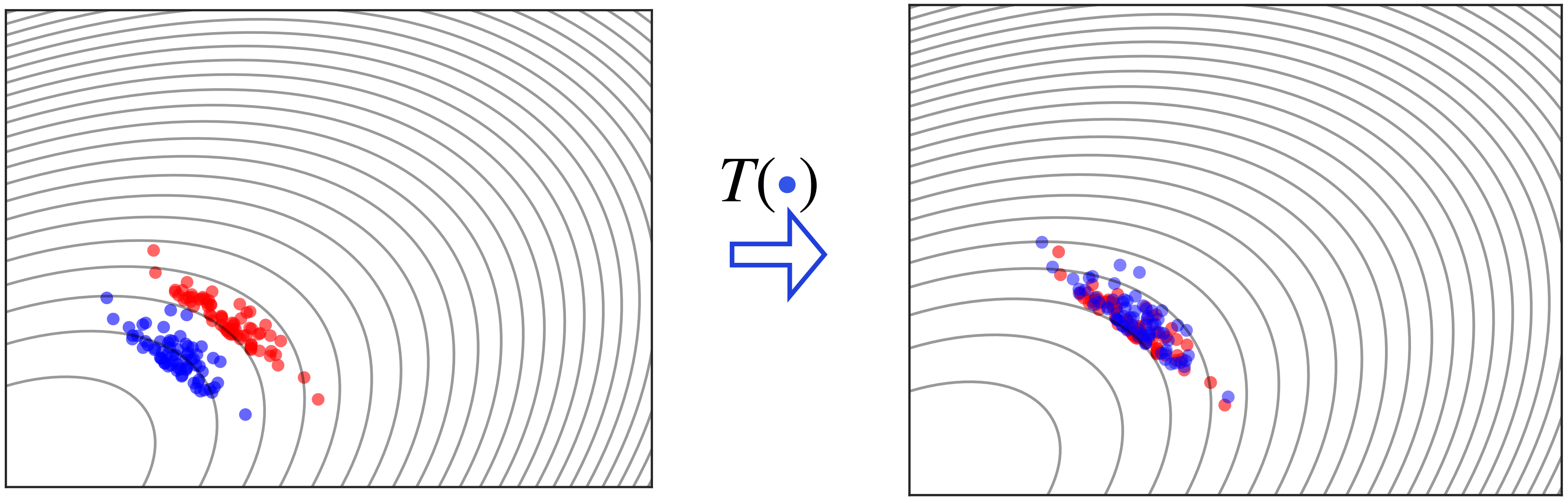
$$\boldsymbol{T}(\boldsymbol{x}) = s^{\kappa(\boldsymbol{x})}\boldsymbol{x}$$

where

$$\kappa(\boldsymbol{x}) = \frac{1}{\rho}\frac{\log|\boldsymbol{x}|}{\log\|\boldsymbol{x}\|_\infty}$$

s = scalar stretch parameter

Multivariate normal



Proposition [Deo & M '21]. In the generality considered,
1) the theoretically optimal sampler and
2) the transformed excess loss samples

concentrate their mass on the same set of points, albeit at different rates

Logarithmic efficiency

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

Theorem.

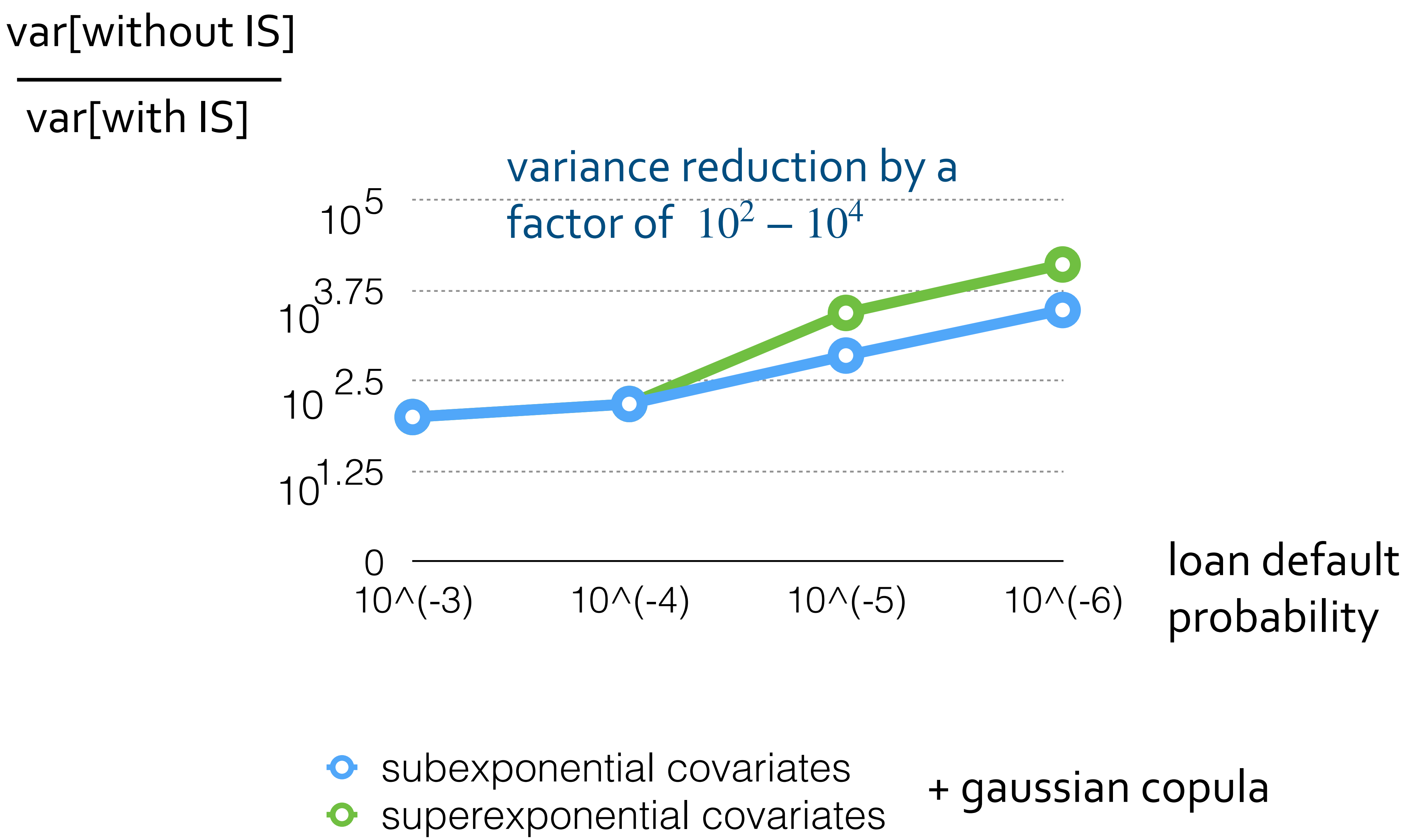
Minimizing $\textbf{CVaR}_{1-\beta}(L_{\boldsymbol{\theta}}(X))$ with the proposed sampler is log-efficient as $\beta \to 0$.

$\Longrightarrow$ # sample required $\propto \begin{cases} \log^k \frac{1}{\beta} & \text{with the proposed sampler} \\ \frac{1}{\beta} & \text{without importance sampling} \end{cases}$

Numerical experiments

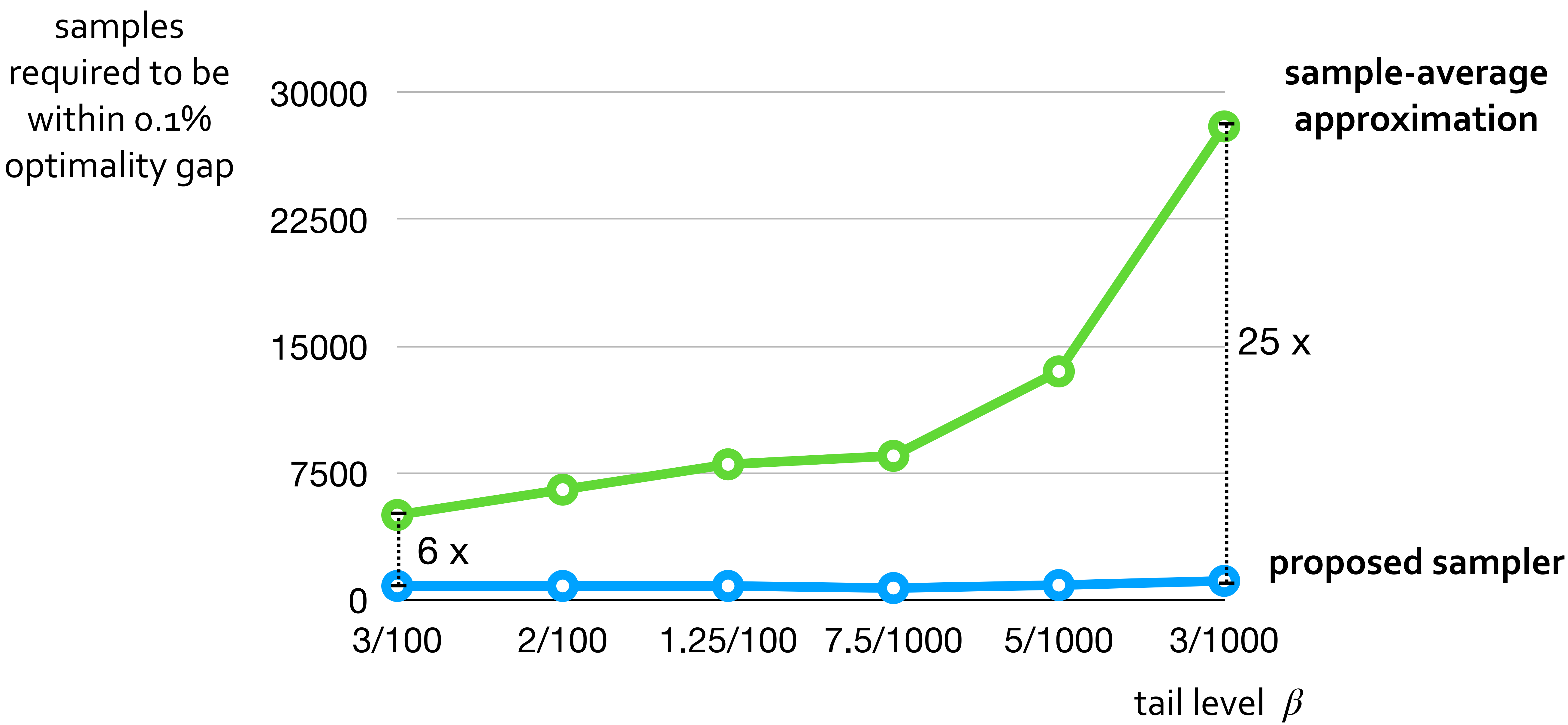
Probability of excess loss in a portfolio with 3000 loans

Default probability modeled by a ReLU network with 1 hidden layer



Minimizing CVaR objective

Illustration of portfolio optimization objective with 15 assets



Outline of the talk

- Introduction
- Challenges due to rarity & model-bias
- Why algorithmic approaches have been elusive?
- Key observation & its implications
- Q1: Can a sampler adapt its IS distribution to the problem-at-hand?
- Q2: Can we correct the plug-in model-bias?
- Summary

Debiased learning: Related literature

Debiasing in Statistics: Old and new

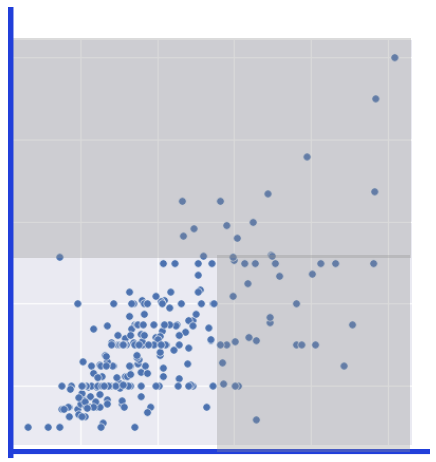
Newey and Stoker, '93
Murphy and van der Vaart '97
Van der Vaart '99
Chernozhukov et al. '16, '17
Foster and Syrgkanis '19
Newey and Ichimura '22

Debiasing in Operations Research literature

Gupta, Huang, Rusmevichientong '21

An overview of the debiased objective

data from unknown P



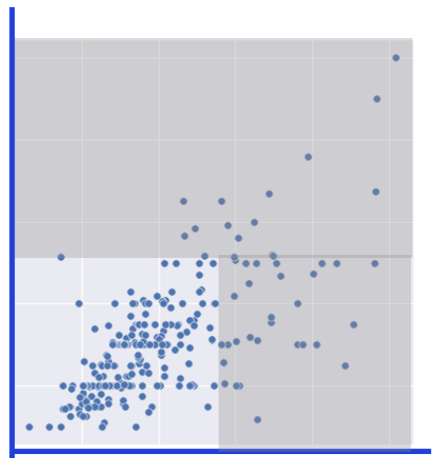
$\hat{\gamma}$ = fraction of samples in grey region

model for grey region → $\hat{P}$ → Optimize with $\hat{P}$ → decision

Debiased objective = objective($\hat{P}$) + a correction term

An overview of the debiased objective

data from unknown P



$\hat{\gamma}$ = fraction of samples in grey region

→ model for grey region → $\hat{P}$ → Optimize with $\hat{P}$ → decision

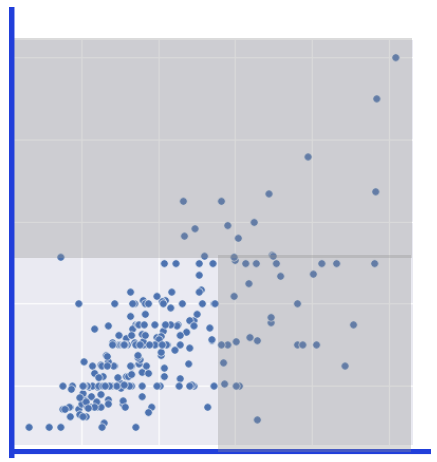
Debiased objective = objective($\hat{P}$) + a correction term

$$\inf_{u,\theta} \left\{ u + \frac{\hat{\gamma}}{\beta} E_{\hat{P}} \left[L(\boldsymbol{X}, \boldsymbol{\theta}) - u \right]^{+} \right\}$$

(Rockafellar & Uryasev '02)

An overview of the debiased objective

data from unknown P



model for grey region → $\hat{P}$ → Optimize with $\hat{P}$ → decision

$\hat{\gamma}$ = fraction of samples in grey region

Debiased objective = objective($\hat{P}$) + a correction term

$$\inf_{u,\theta} \left\{ u + \frac{\hat{\gamma}}{\beta} E_{\hat{P}} \left[L(\mathbf{X}, \boldsymbol{\theta}) - u \right]^+ \right\}$$

(Rockafellar & Uryasev '02)

Call this $E_{\hat{P}}\left[\,\xi\,\right]$

What is the correction term?

$$\xi = \left[\ L(\boldsymbol{X}, \boldsymbol{\theta}) - u\ \right]^{+}$$

error due to model-misspecification

$$= \ E_P\left[\,\xi\,\right] \ - \ E_{\hat{P}}\left[\,\xi\,\right]$$

What is the correction term?

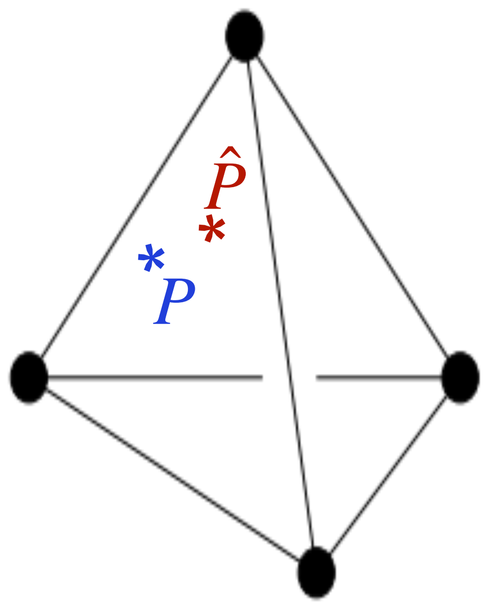
$$\xi = \left[\ L(\boldsymbol{X},\boldsymbol{\theta}) - u\ \right]^{+}$$

error due to model-misspecification

$$= E_P\left[\xi\right] - E_{\hat{P}}\left[\xi\right]$$

$$= E_{\hat{P}}\left[\xi \frac{dP}{d\hat{P}}\right] - E_{\hat{P}}\left[\xi\right]$$

$$= E_{\hat{P}}\left[\xi\left(\frac{dP}{d\hat{P}} - 1\right)\right]$$



$$d\hat{P} = e^{-h}dP$$

What is the correction term?

$$\xi = \left[\, L(\boldsymbol{X}, \boldsymbol{\theta}) - u \,\right]^{+}$$

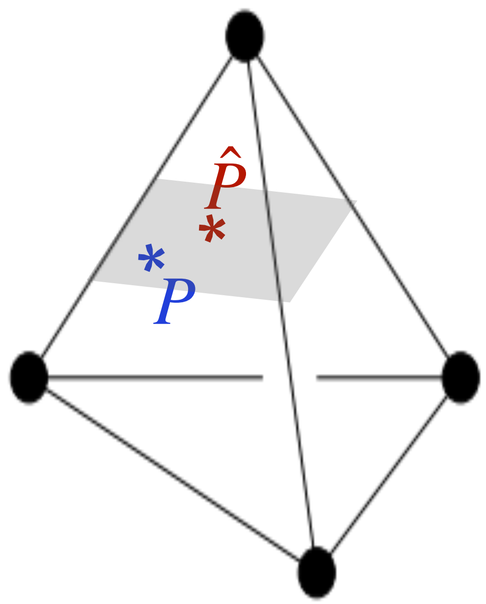
error due to model-misspecification

$$= E_P\left[\xi\right] - E_{\hat{P}}\left[\xi\right]$$

$$= E_{\hat{P}}\left[\xi \frac{dP}{d\hat{P}}\right] - E_{\hat{P}}\left[\xi\right]$$

$$= E_{\hat{P}}\left[\xi\left(\frac{dP}{d\hat{P}} - 1\right)\right]$$

$$\approx E_{\hat{P}}\left[\xi \cdot h\right]$$



$$d\hat{P} = e^{-h} dP$$

$h =$ zero mean,
homogenous function
if $\hat{P}, P$ are tail-similar

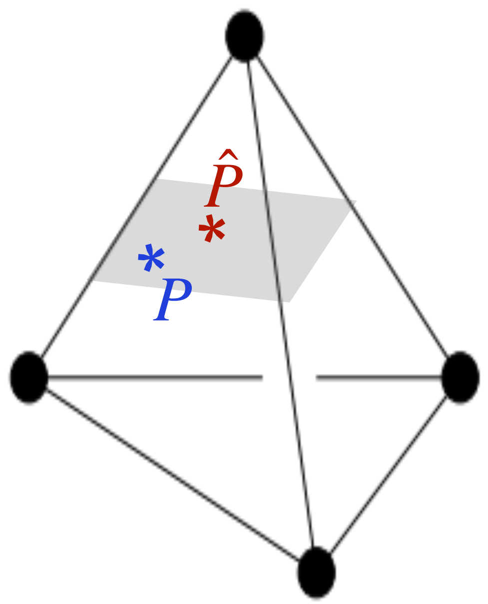
(as log pdf is nearly homogenous)

Restricting to self-similar class lowers variance

$$\xi = \left[\ L(\boldsymbol{X},\boldsymbol{\theta}) - u\ \right]^+$$

error due to model-misspecification

$$= \ E_{\hat{P}}\left[\,\xi \cdot h\,\right]$$



$$d\hat{P} = e^{-h} dP$$

$h =$ zero mean,
homogenous function
if $\hat{P}, P$ are tail-similar

Restricting to self-similar class lowers variance

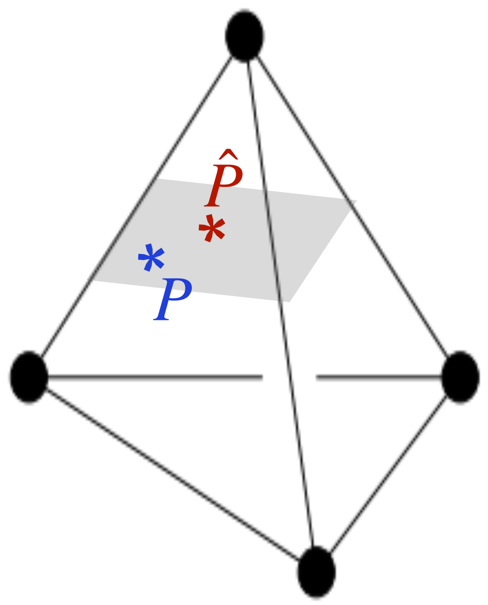
$$\xi = \left[\, L(\boldsymbol{X}, \boldsymbol{\theta}) - u \,\right]^+$$

error due to model-misspecification

$$= \; E_{\hat{P}}\left[\xi \cdot h\right]$$

$$= \; E_{\hat{P}}\left[E_{\hat{P}}\left[\xi \mid \mathscr{F}\right] \cdot h \right]$$

best zero mean
homogenous function
approximating ξ



$$d\hat{P} = e^{-h} dP$$

$h =$ zero mean,
homogenous function
if $\hat{P}, P$ are tail-similar

Restricting to self-similar class lowers variance

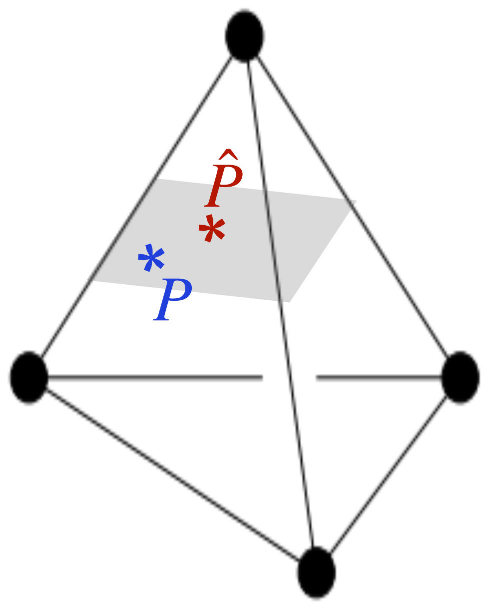
$$\xi = \left[\ L(\boldsymbol{X}, \boldsymbol{\theta}) - u\ \right]^+$$

error due to model-misspecification

$$= \ E_{\hat{P}}\left[\xi \cdot h\right]$$

$$= \ E_{\hat{P}}\Big[E_{\hat{P}}\left[\xi \mid \mathscr{F}\right] \cdot h\Big]$$

best zero mean
homogenous function
approximating ξ



$$d\hat{P} = e^{-h} dP$$

$h =$ zero mean,
homogenous function
if $\hat{P}, P$ are tail-similar

can be understood as gradient at $\hat{P}$
(efficient influence function)

Restricting to self-similar class lowers variance

$$\xi = \left[\ L(\boldsymbol{X}, \boldsymbol{\theta}) - u\ \right]^{+}$$

error due to model-misspecification

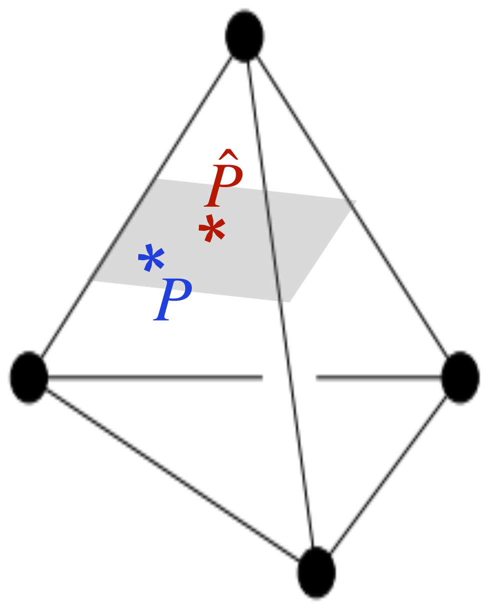
$$= E_{\hat{P}}\left[\xi \cdot h\right]$$

$$= E_{\hat{P}}\Big[E_{\hat{P}}\left[\xi \mid \mathscr{F}\right] \cdot h\Big]$$

$$\approx E_{P}\Big[E_{\hat{P}}\left[\xi \mid \mathscr{F}\right]\Big]$$

= sample mean of $E_{\hat{P}}\left[\xi \mid \mathscr{F}\right]$

+ $n^{-1/2}$ CLT term



$$d\hat{P} = e^{-h} dP$$

$h =$ zero mean,
homogenous function
if $\hat{P}, P$ are tail-similar

Can we correct plug-in model bias?

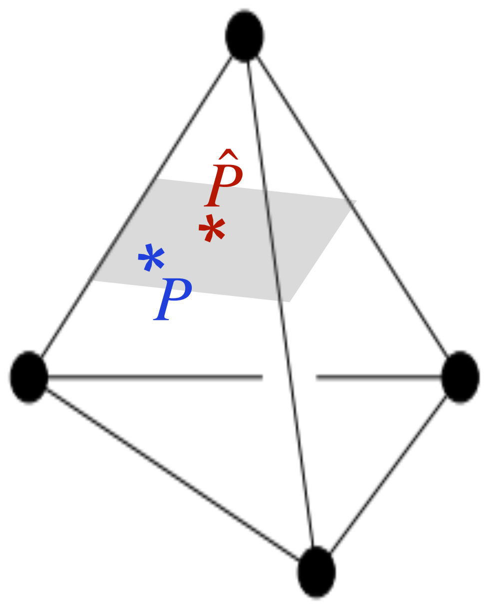
$$\xi = \left[\, L(\boldsymbol{X}, \boldsymbol{\theta}) - u \,\right]^{+}$$

error due to model-misspecification

$$= \text{sample mean of } E_{\hat{P}}\left[\, \xi \mid \mathscr{F} \,\right]$$

$+ \; n^{-1/2}$ CLT term

+ sec. order terms



Can we correct plug-in model bias?

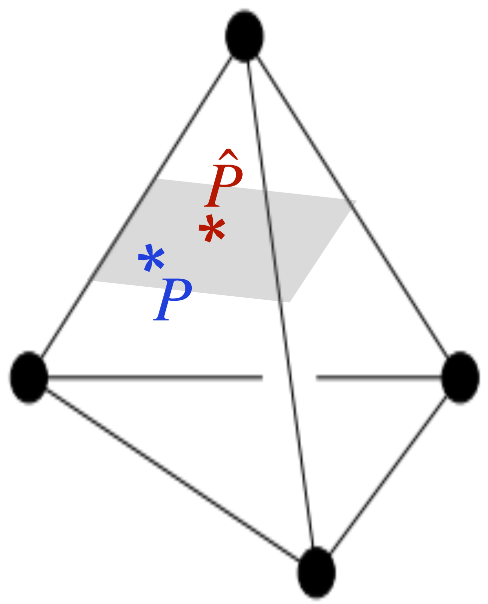
$$\xi = \left[\, L(\boldsymbol{X}, \boldsymbol{\theta}) - u \,\right]^{+}$$

error due to model-misspecification

$$= \text{sample mean of } E_{\hat{P}}\left[\, \xi \mid \mathscr{F} \,\right]$$

$+\, n^{-1/2}$ CLT term

+ sec. order terms



Evaluating $E_{\hat{P}}\left[\, \xi \mid \mathscr{F} \,\right]$ amounts to finding the best approx. to ξ in the span(e_1, e_2) under the plug-in measure

here $\quad e_1(x) = \hat{\varphi}(\boldsymbol{x}) - E_{\hat{P}_{\angle x}}\left[\, \hat{\varphi}(\boldsymbol{X}) \,\right]$

$$e_2(x) = \hat{\varphi}(\boldsymbol{x}) \log \boldsymbol{x} - E_{\hat{P}_{\angle x}}\left[\, \hat{\varphi}(\boldsymbol{X}) \log \boldsymbol{X} \,\right]$$

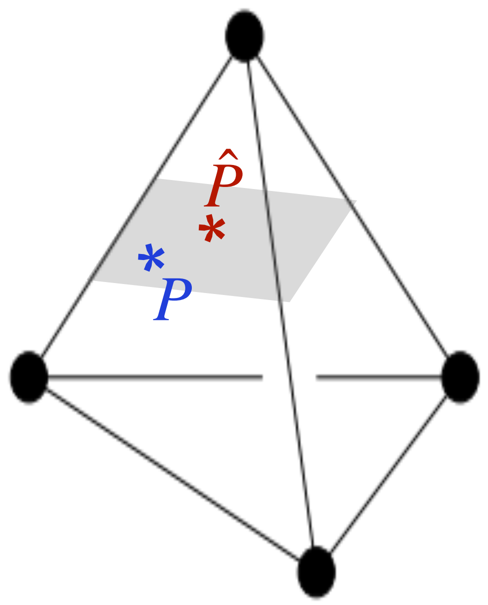
Can we correct plug-in model bias?

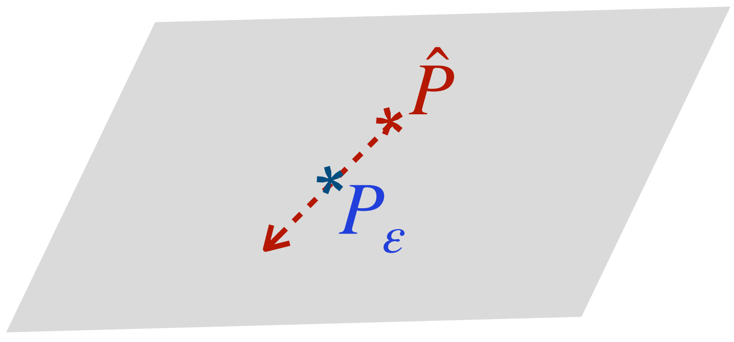
$$\xi = \left[\ L(\boldsymbol{X}, \boldsymbol{\theta}) - u\ \right]^+$$

debiased objective

$$= E_{\hat{P}}\,[\,\xi\,] + \text{ sample mean of } E_{\hat{P}}\left[\,\xi \mid \mathscr{F}\,\right]$$

- Neyman orthogonal: orthogonal to model perturbations





Derivative of the debiased objective w.r.to ε is zero

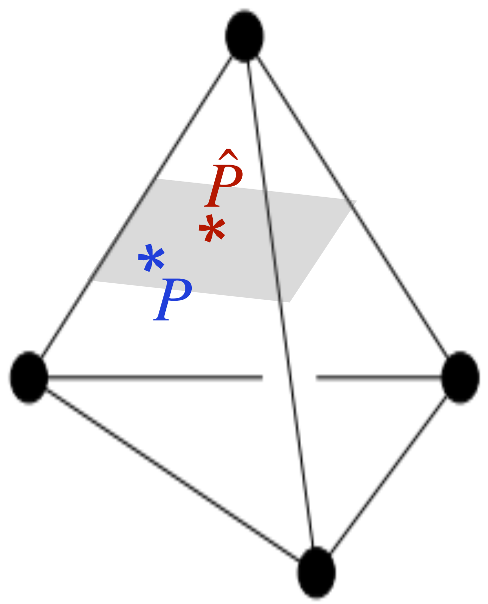
Can we correct plug-in model bias?

$$\xi = \left[\ L(\boldsymbol{X}, \boldsymbol{\theta}) - u\ \right]^+$$

debiased objective

$$= E_{\hat{P}}\,[\,\xi\,] + \text{sample mean of } E_{\hat{P}}\left[\,\xi \mid \mathscr{F}\,\right]$$

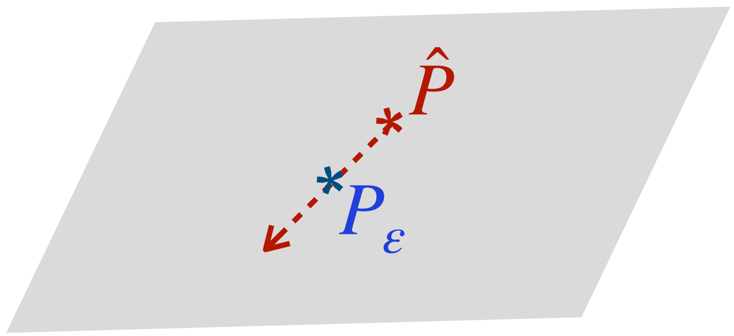
- Neyman orthogonal: orthogonal to model perturbations



Error rates

Plug-in objective: $O_p\left(\textbf{rate}(N)\right)$

Debiased objective: $O_p\left(\textbf{rate}^2(N)\right)$



Derivative of the debiased objective w.r.to ε is zero

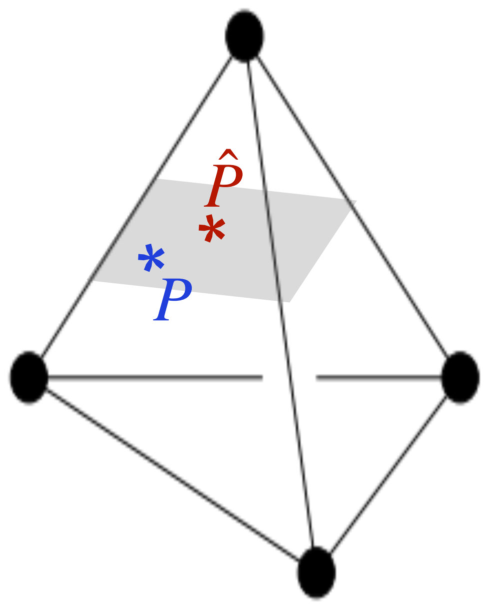
Can we correct plug-in model bias?

$$\xi = \left[\ L(\boldsymbol{X}, \boldsymbol{\theta}) - u\ \right]^+$$

debiased objective

$= E_{\hat{P}}\,[\,\xi\,] +$ sample mean of $E_{\hat{P}}\left[\,\xi \mid \mathscr{F}\,\right]$

- Neyman orthogonal: orthogonal to model perturbations
- Convexity is retained
- Can be understood as a first-order Taylor approximation on the subset of distributions with self-similar tails



Can we correct plug-in model bias?

$$\xi = \left[\ L(\boldsymbol{X}, \boldsymbol{\theta}) - u\ \right]^{+}$$

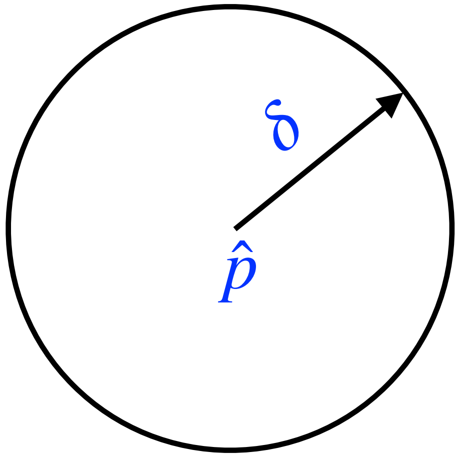
debiased objective

$$= E_{\hat{P}}\ [\ \xi\ ] + \text{sample mean of } E_{\hat{P}}\left[\ \xi \mid \mathscr{F}\ \right]$$

Contrast with RO / DRO

worst-case objective $\displaystyle \sup_{\|\boldsymbol{x}-\boldsymbol{p}\|<\delta} \rho(\boldsymbol{x}) = \rho(\hat{\boldsymbol{p}}) + \delta \, \|\nabla \rho(\hat{\boldsymbol{p}})\|$

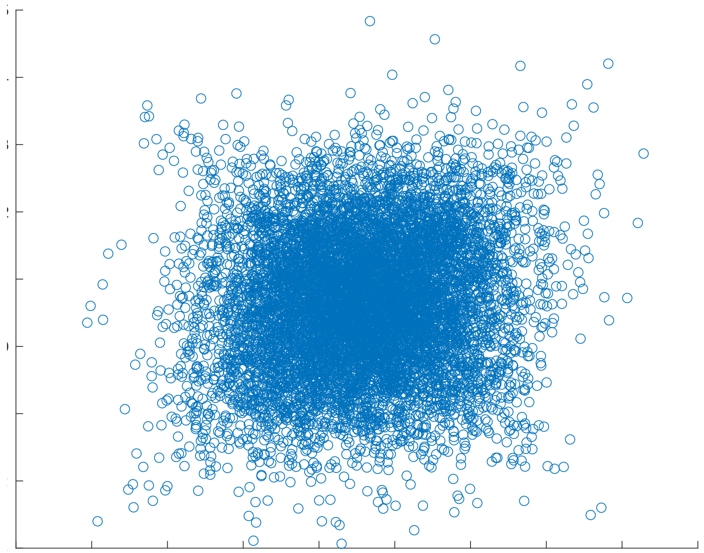
debiased objective $\rho(\boldsymbol{p}) \approx \rho(\hat{\boldsymbol{p}}) + \langle \nabla \rho(\hat{\boldsymbol{p}}), \boldsymbol{p} - \hat{\boldsymbol{p}} \rangle$



(debiasing = a targeted notion of robustness)

Numerical experiments

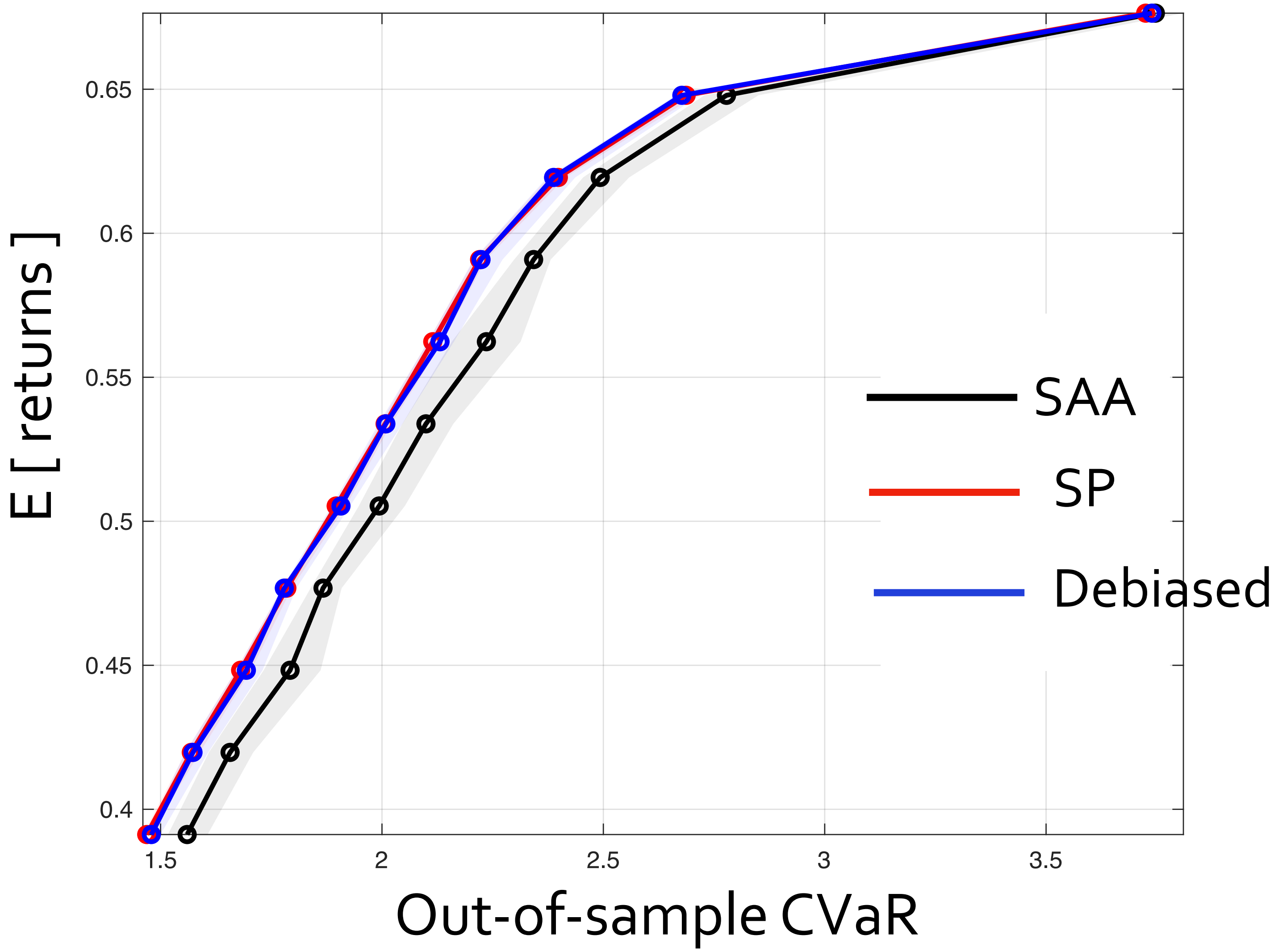
Portfolio optimization with 5 assets, given 1000 return samples



data generated from
70% multivariate normal +
30% t-copula

plug-in = multivariate normal

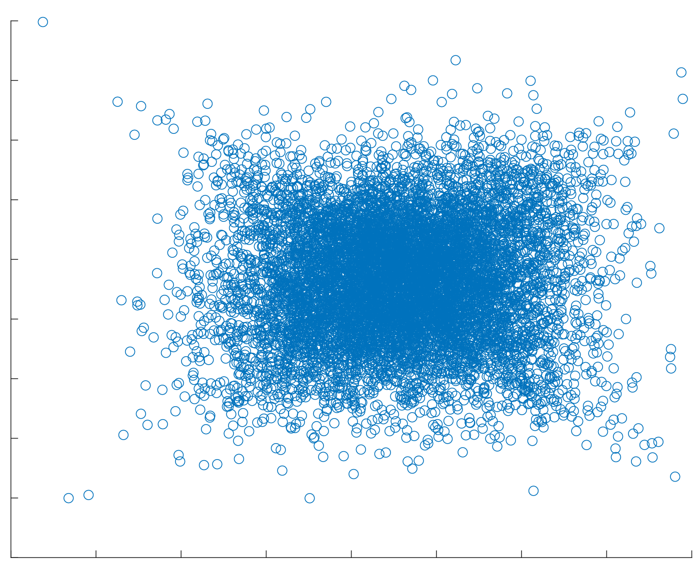
Out-of-sample Pareto frontier



Minimize CVaR at tail level 1 / 300
subject to return requirement

Numerical experiments

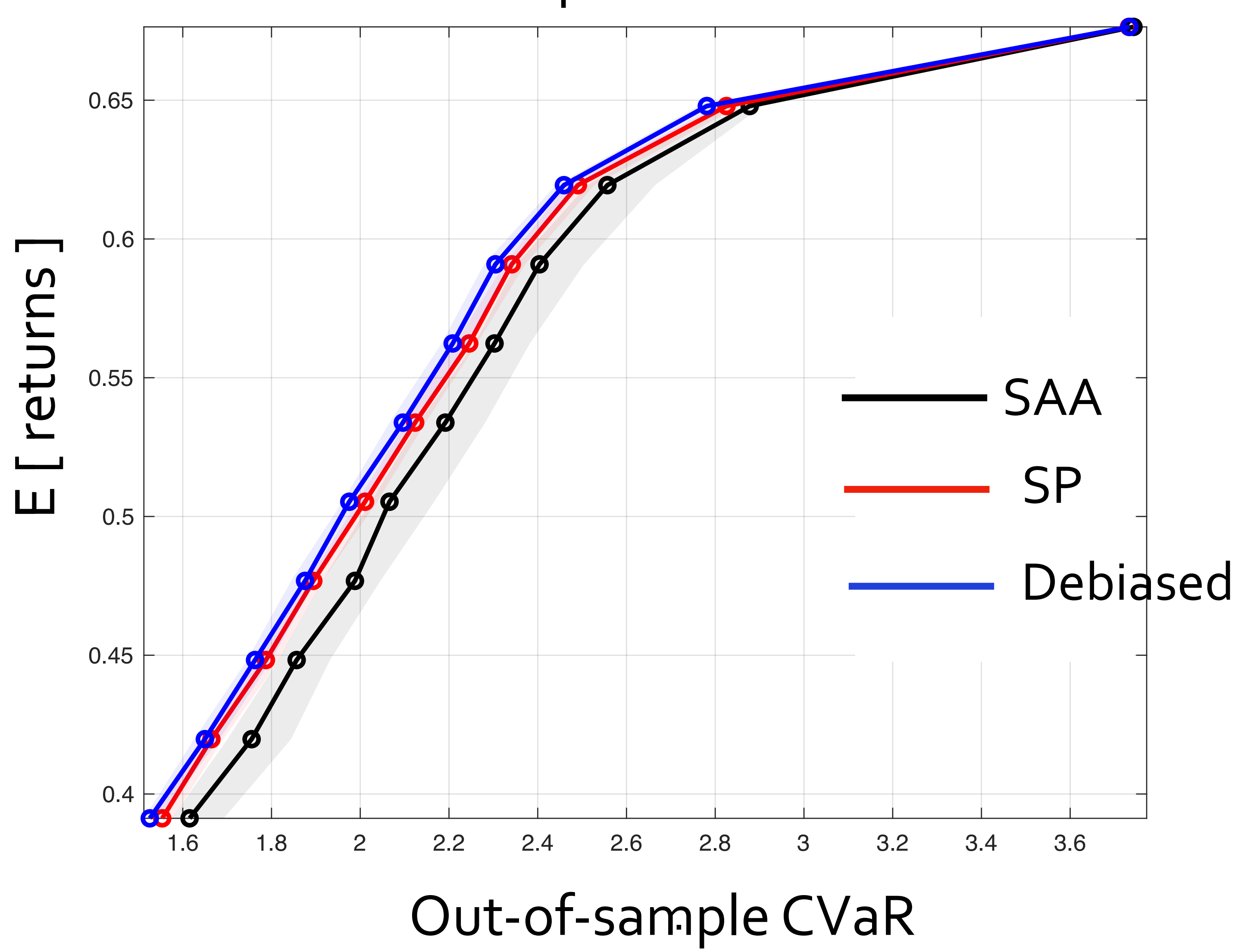
Portfolio optimization with 5 assets, given 1000 return samples



data generated from
50% multivariate normal +
50% t-copula

plug-in = multivariate normal

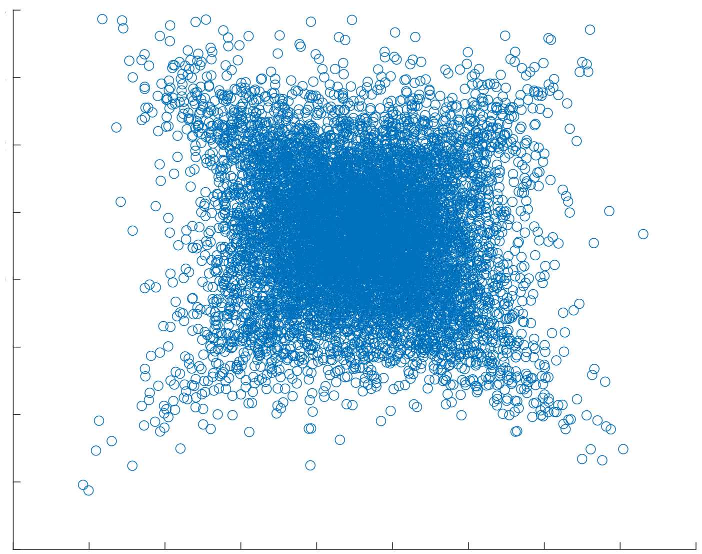
Out-of-sample Pareto frontier



Minimize CVaR at tail level 1 / 300
subject to return requirement

Numerical experiments

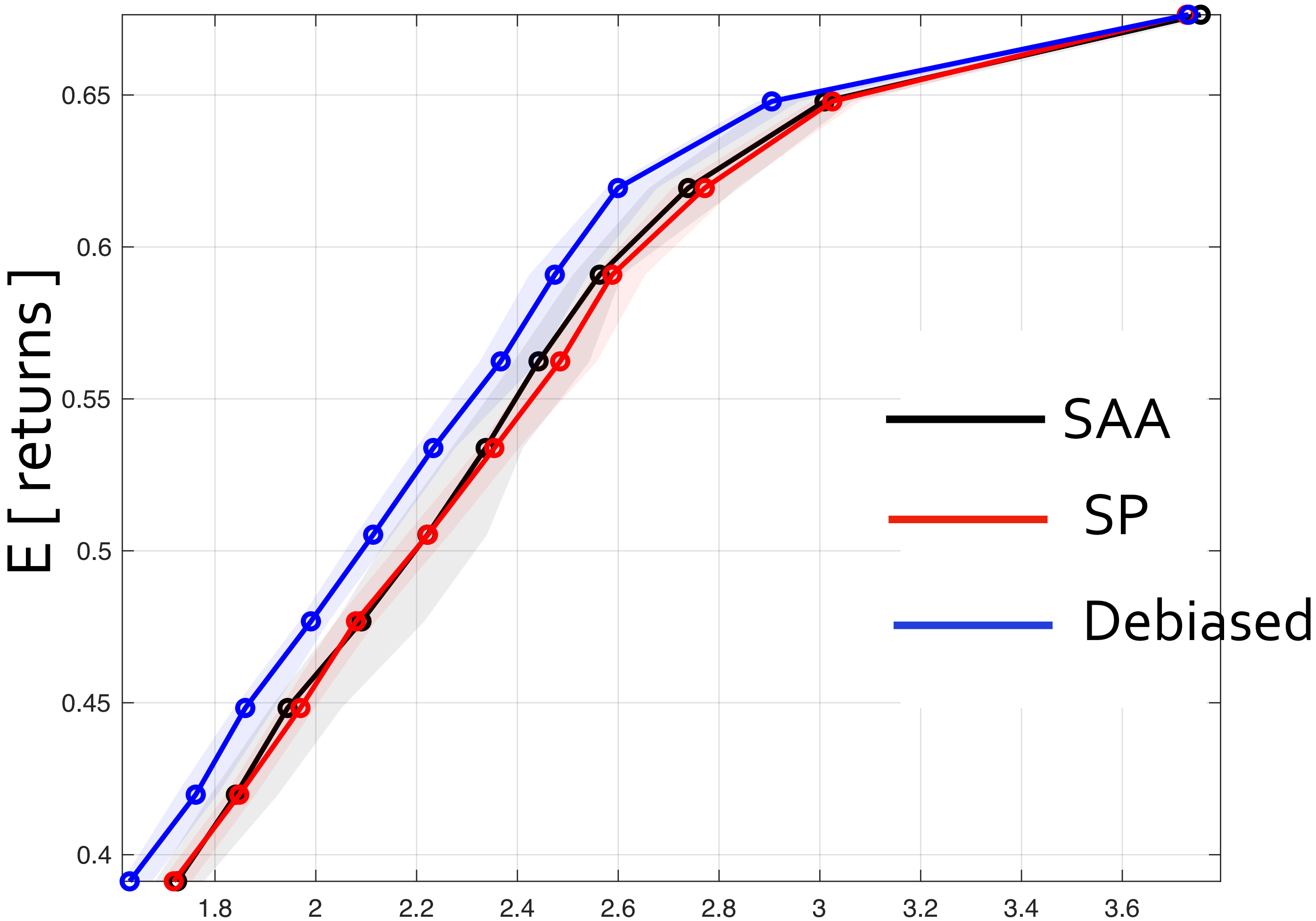
Portfolio optimization with 5 assets, given 1000 return samples



data generated from
20% multivariate normal +
80% t-copula

plug-in = multivariate normal

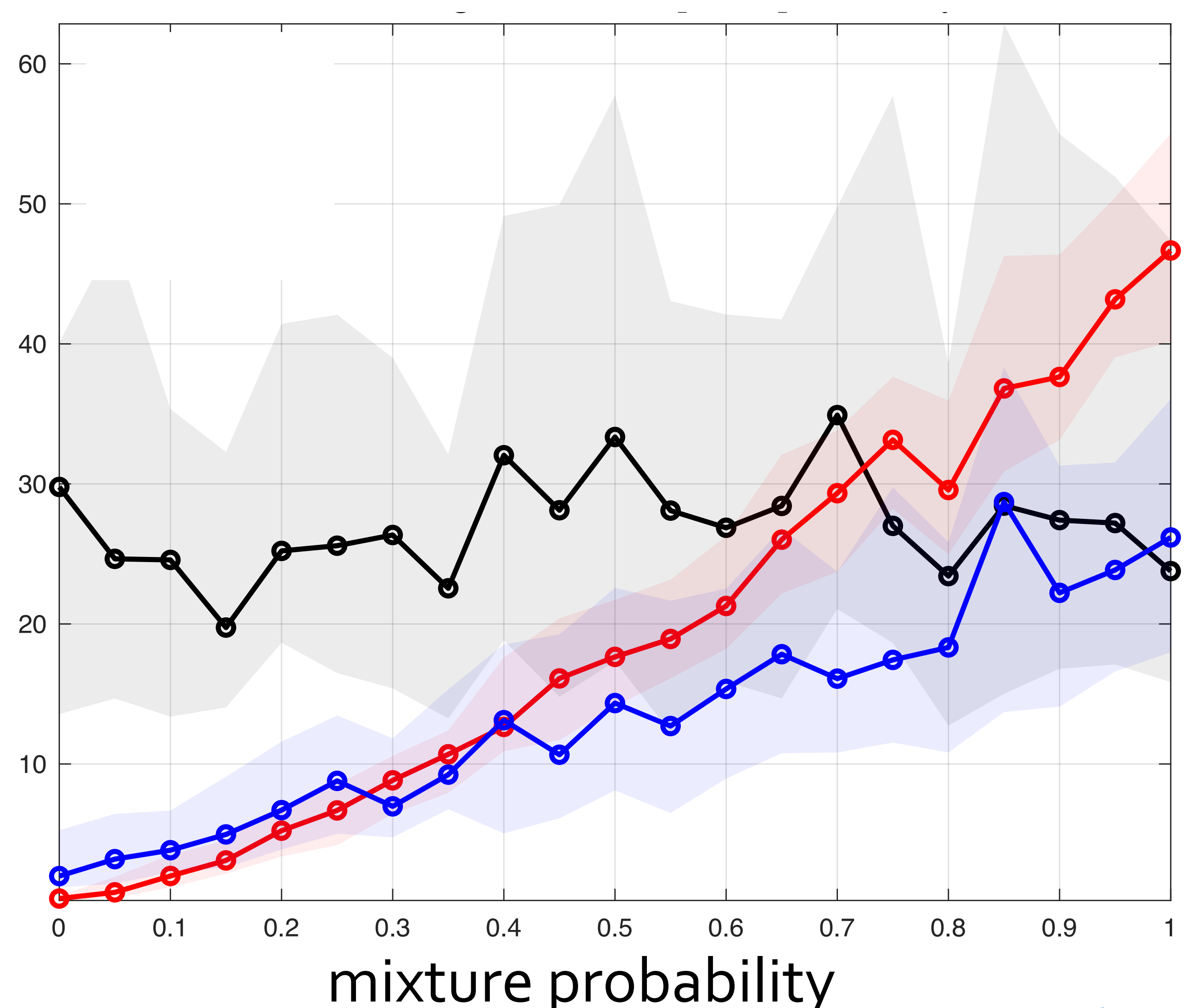
Out-of-sample Pareto frontier

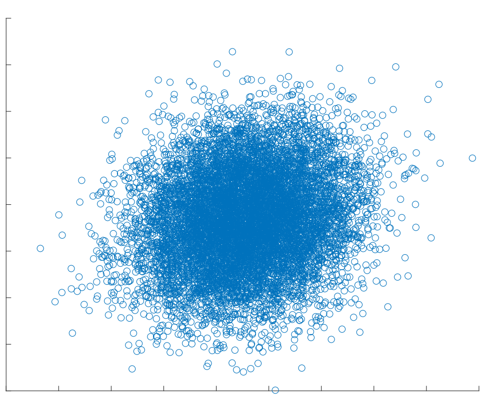


Minimize CVaR at tail level 1 / 300
subject to return requirement

Performance under distribution shift

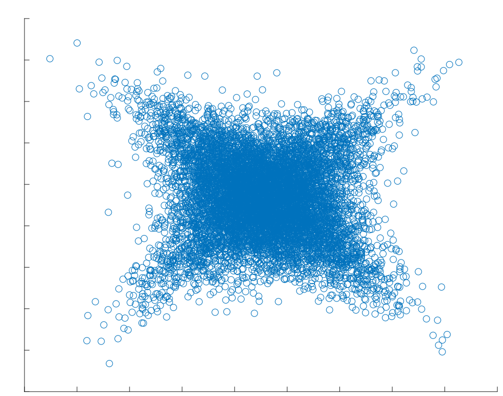
Portfolio optimization with 5 assets, given 750 return samples





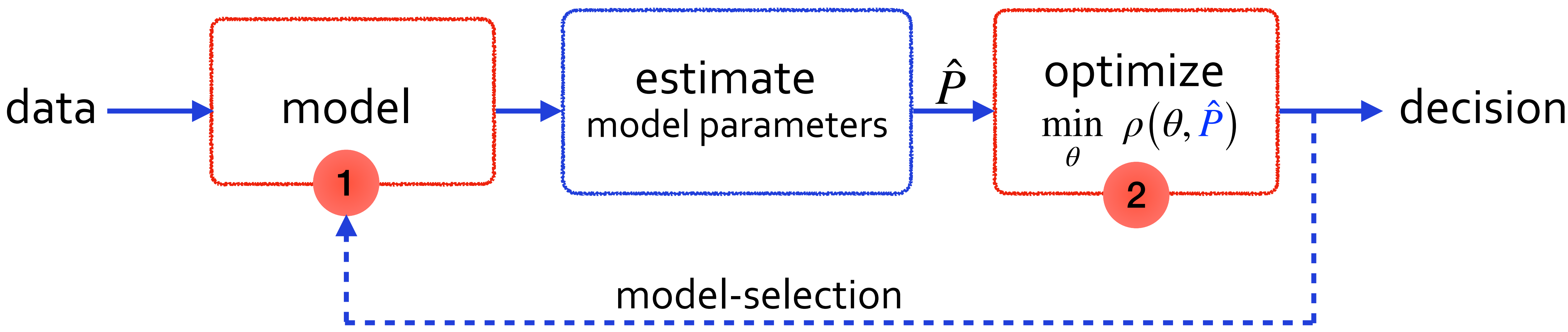
multivariate normal

plug-in = multivariate normal



normal marginals + t-copula

Summary: The two bottlenecks



the need to carefully handle bias created by plugging-in a wrong distributional model

prohibitive computation needed due to large number of samples/scenarios

Summary: Algorithmic variance reduction with self-similar tails

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

Automate, informed by large deviations

search for alternate distribution ⇨ search for a concentration-preserving transformation

Summary: Algorithmic variance reduction with self-similar tails

Applied probability, rare events

Tailored efficient samplers for stylized QRM models

Optimization under uncertainty

minimizing CVaR, chance-constraints with sample-averaging

Summary: Algorithmic bias reduction with self-similar tails

- Debiased objective = objective ($\hat{P}$) + a correction term
- Objective has zero sensitivity to model perturbations
- If modeller's choice induces a bias = ε_n in the objective,
  bias in debiased objective is only ε_n^2!
- Convexity retained in the debiased objective