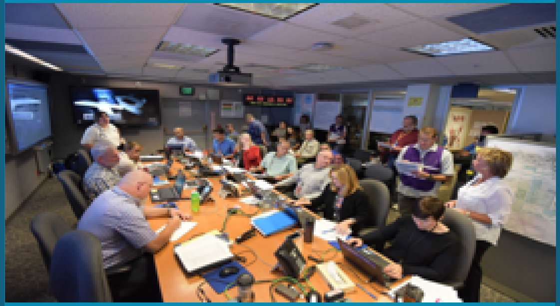
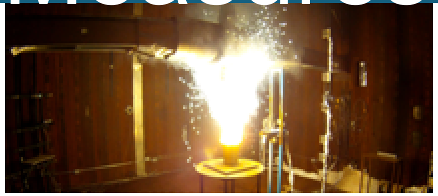
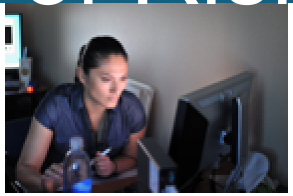
SAND2021-11922C

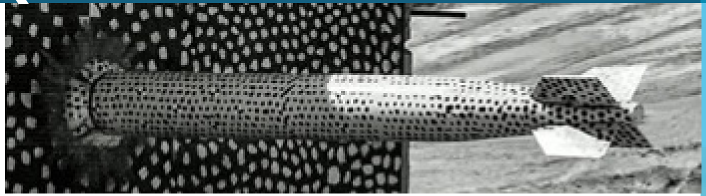
Sandia National Laboratories

Surrogate Modeling for Efficiently, Accurately and Conservatively Estimating Measures of Risk









John Jakeman Sandia National Laboratories

Drew Kouri (SNL), Gabriel Huerta (SNL)

MMLDT-CSET Conference September 27 2021

Unclassified Unlimited Release

U.S. DEPARTMENT OF ENERGY NNSA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

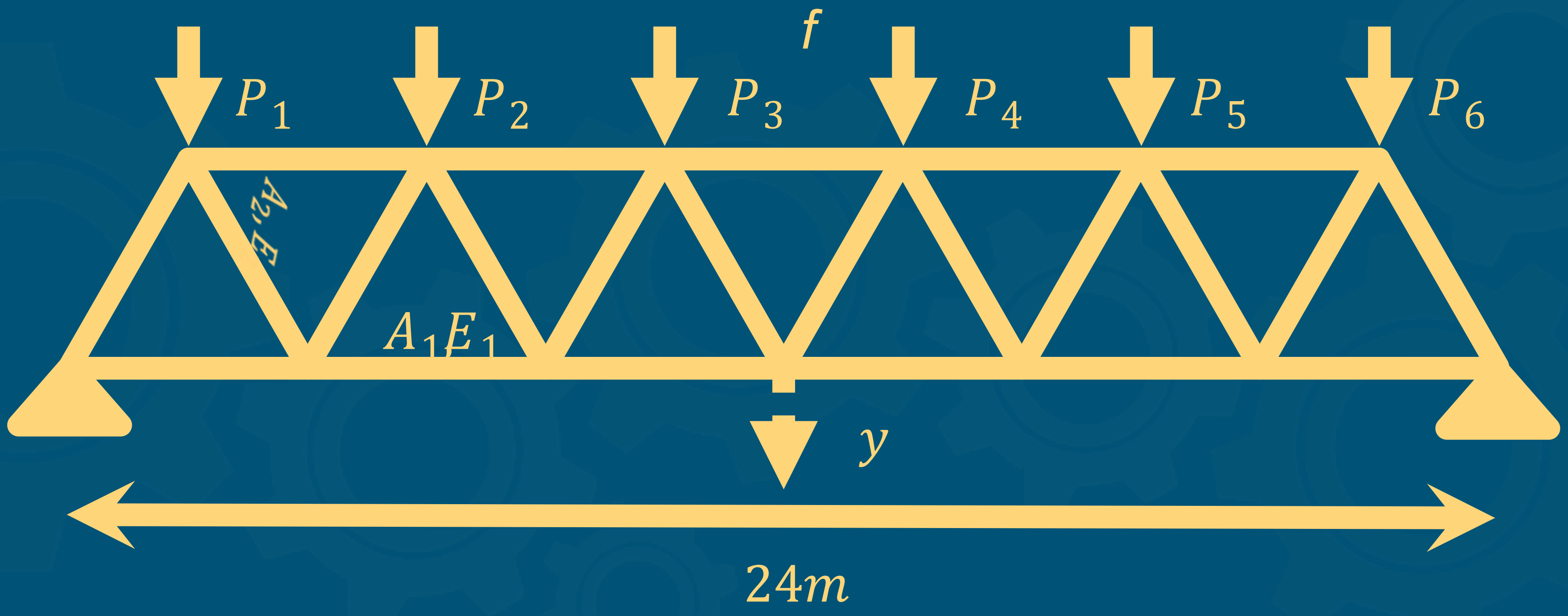
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Digital Twin

- Numerical model consistent with current asset state
- Used to make predictions of future state

Another Digital Twin

Given uncertainty in future loads and material properties predict displacement y using model



$$Y = f(X)$$

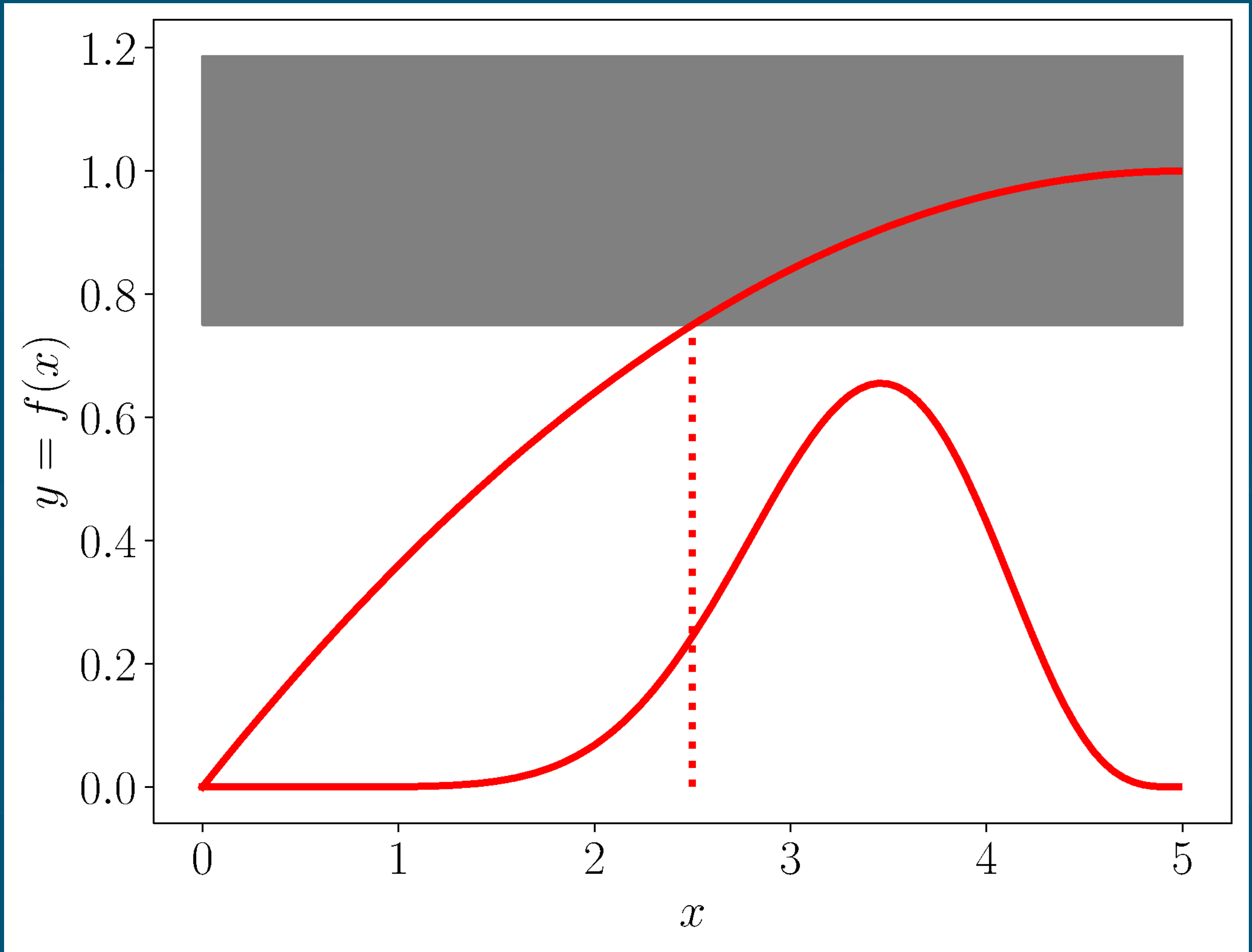
$$X = (A_i, E_i, P_j)$$

Measures of Risk

Risk measures quantify is $Y \leq C$ when uncertainty may occasionally lead to $Y > C$

Common risk measure is probability of failure

$$\mathcal{R}(Y) = \mathbb{P}(Y > C)$$



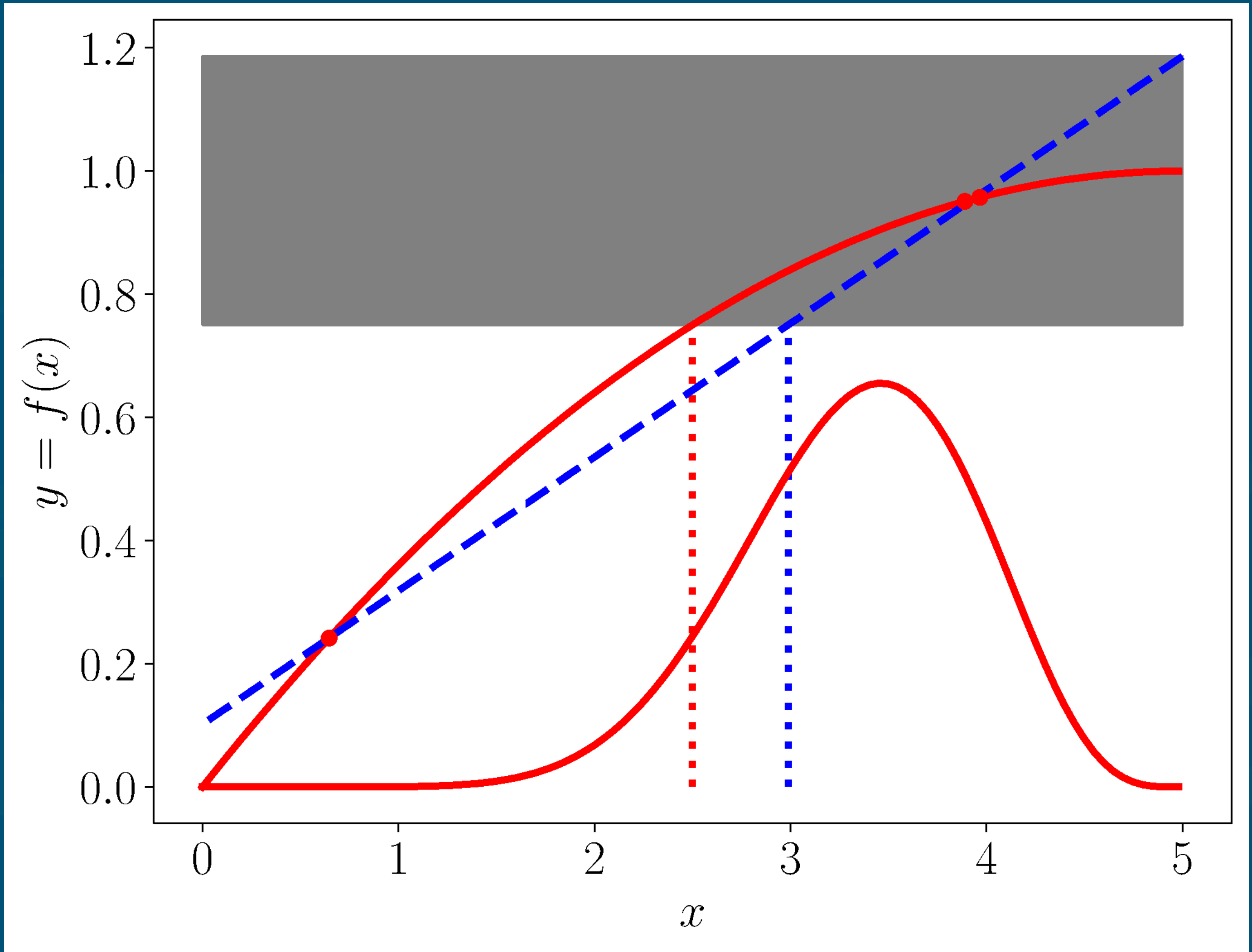
Computing Probability Of Failure with Surrogates

Evaluating risk using high-fidelity model is intractable

Use surrogates instead

Surrogates can under-estimate probability of failure

Surrogates that conservatively estimate (do not under-estimate) risk are needed

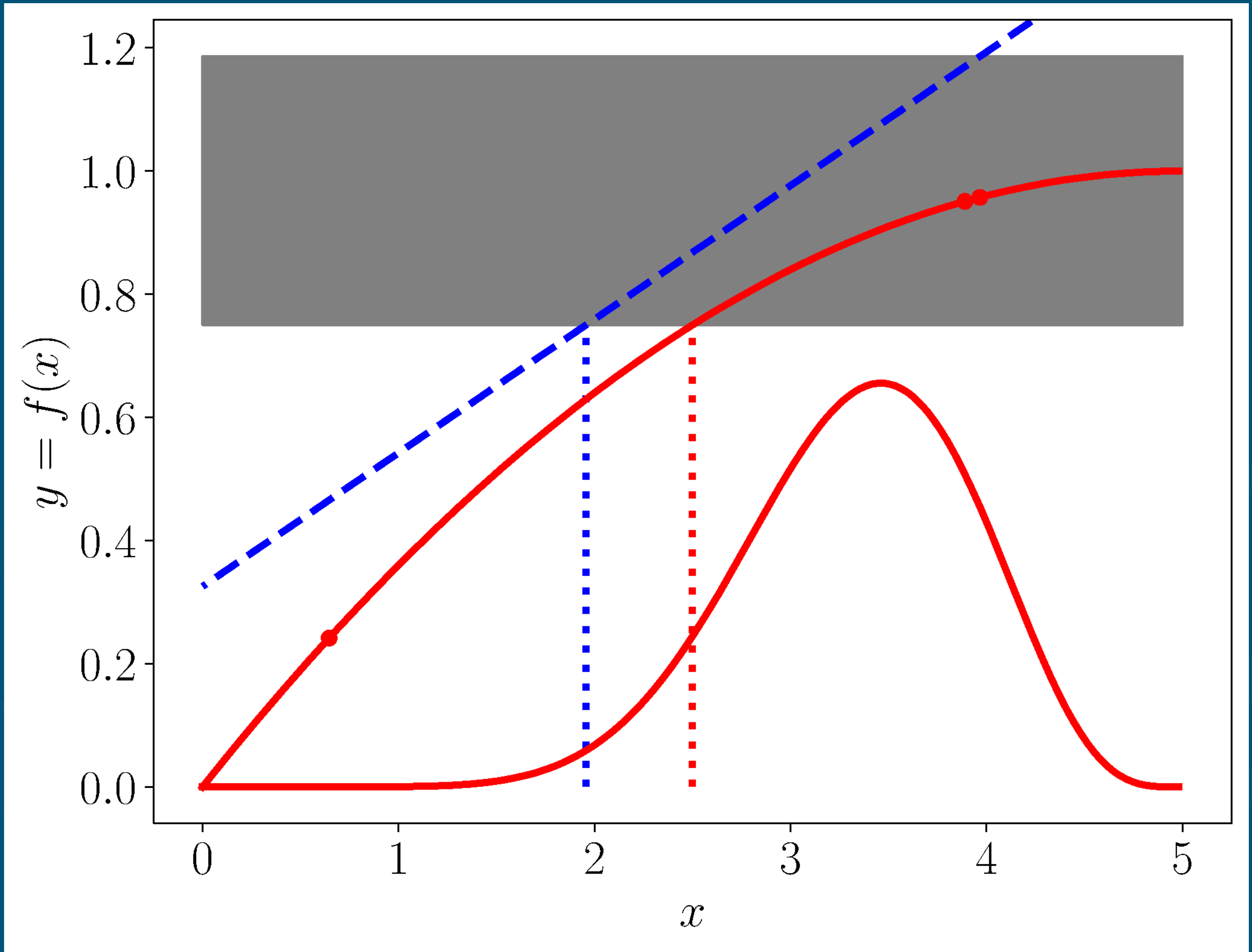


Conservatively Estimating Risk Measures

This talk will present methods for building surrogates that:

Conservatively estimate a **SINGLE** risk measure

Conservatively estimate a **SET** of risk measures



Different Risk Measures

- Mean:

$$\mathcal{R}(Y) = \mathbb{E}[Y]$$

- Mean-plus-standard deviation:

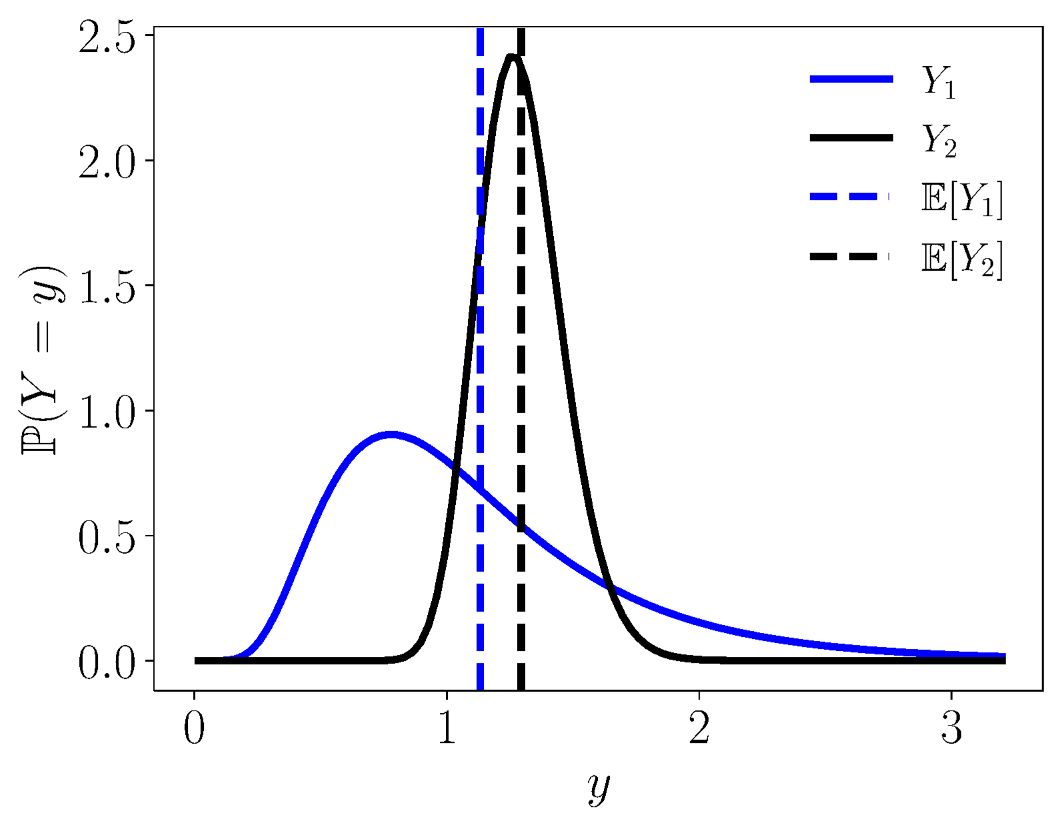
$$\mathcal{R}(Y) = \mathbb{E}[Y] + \lambda \mathbb{V}[Y]^{\frac{1}{2}}$$

- Upper-quantile:

$$\mathcal{R}(Y) = q_p[Y] \coloneqq \inf \{y \mid F_Y(y) \geq p\}$$

- Average-value-at-risk (AVaR):

$$\mathcal{R}(Y) = \mathrm{AVaR}_p[Y] \coloneqq \mathbb{E}\big[\max\big(0, Y - q_p(Y)\big)\big]$$



Different Risk Measures

- Mean:

$$\mathcal{R}(Y) = \mathbb{E}[Y]$$

- Mean-plus-standard deviation:

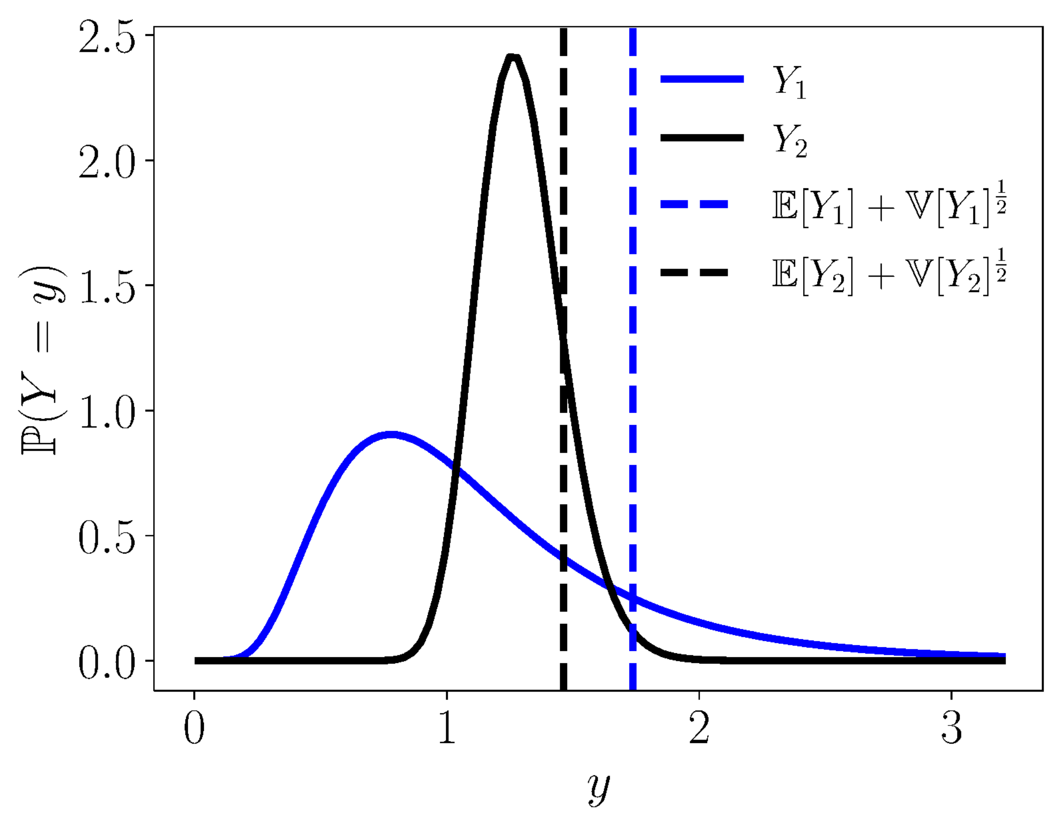
$$\mathcal{R}(Y) = \mathbb{E}[Y] + \lambda \mathbb{V}[Y]^{\frac{1}{2}}$$

- Upper-quantile:

$$\mathcal{R}(Y) = q_p[Y] \coloneqq \inf\{y \mid F_Y(y) \geq p\}$$

- Average-value-at-risk (AVaR):

$$\mathcal{R}(Y) = \mathrm{AVaR}_p[Y] \coloneqq \mathbb{E}\big[\max\big(0, Y - q_p(Y)\big)\big]$$



Different Risk Measures

- Mean:

$$\mathcal{R}(Y) = \mathbb{E}[Y]$$

- Mean-plus-standard deviation:

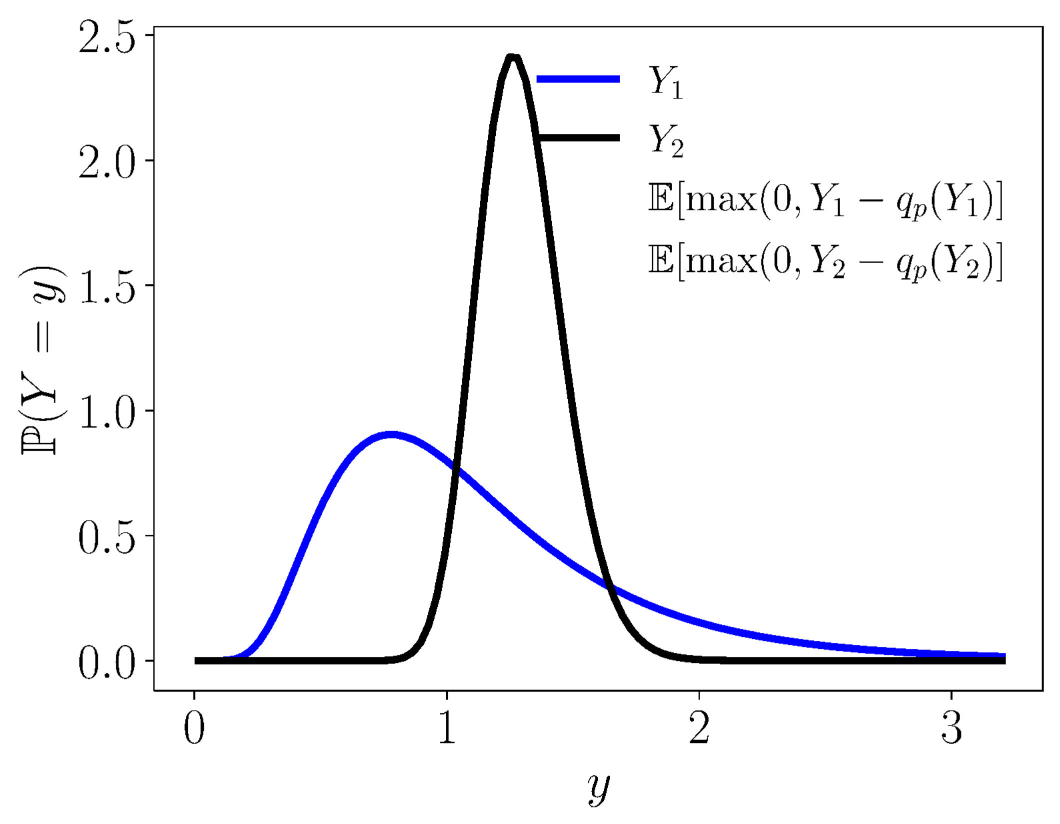
$$\mathcal{R}(Y) = \mathbb{E}[Y] + \lambda \mathbb{V}[Y]^{\frac{1}{2}}$$

- Upper-quantile:

$$\mathcal{R}(Y) = q_p[Y] := \inf \{y \mid F_Y(y) \geq p\}$$

- Average-value-at-risk (AVaR):

$$\mathcal{R}(Y) = \text{AVaR}_p[Y] := \frac{1}{1-p} \mathbb{E}[\max(0, Y - q_p(Y))]$$



Eliciting Risk Measures: Utility/Regret

Risk measures are subjective and must be tailored to beliefs of decision makers

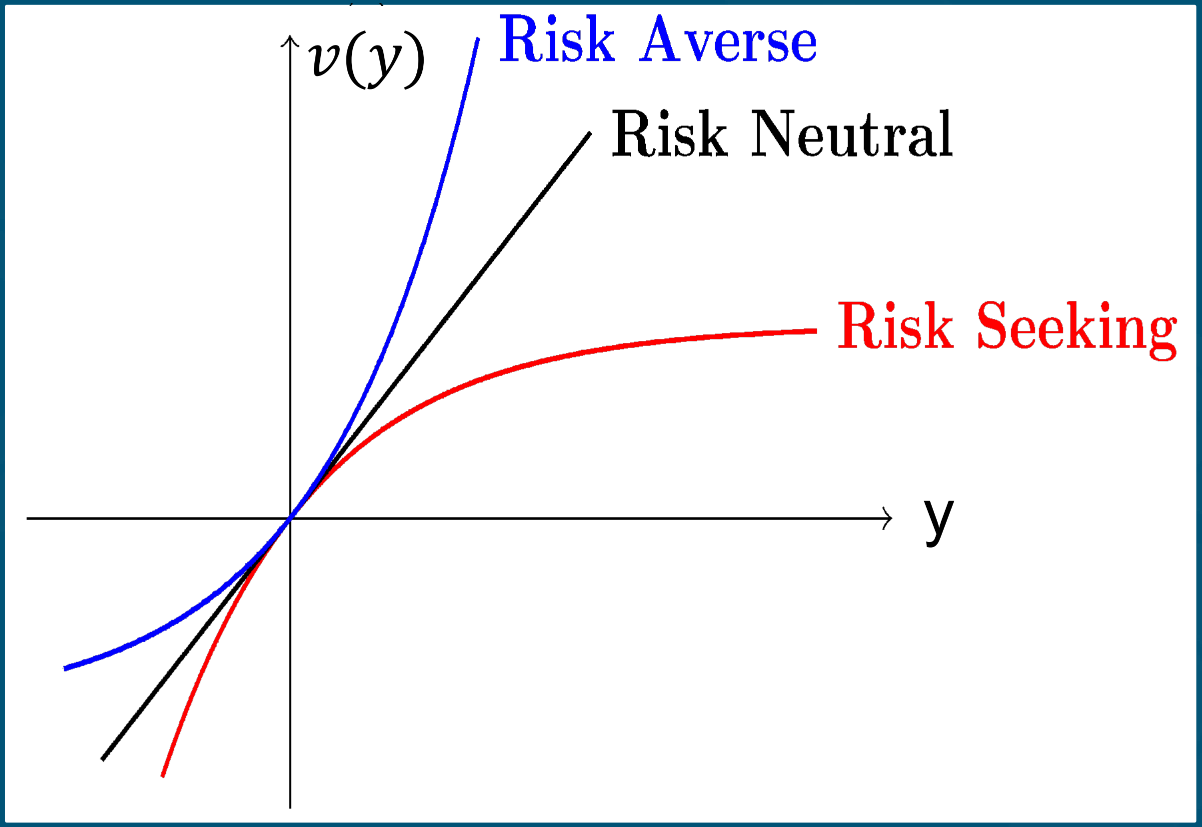
First elicit regret function that quantifies displeasure with the outcomes y

Then formulate a regret measure that quantifies anticipated displeasure

$$\mathcal{V}(Y) = \mathbb{E}[v(Y)]$$

We want to avoid large values so we focus on regret instead of utility

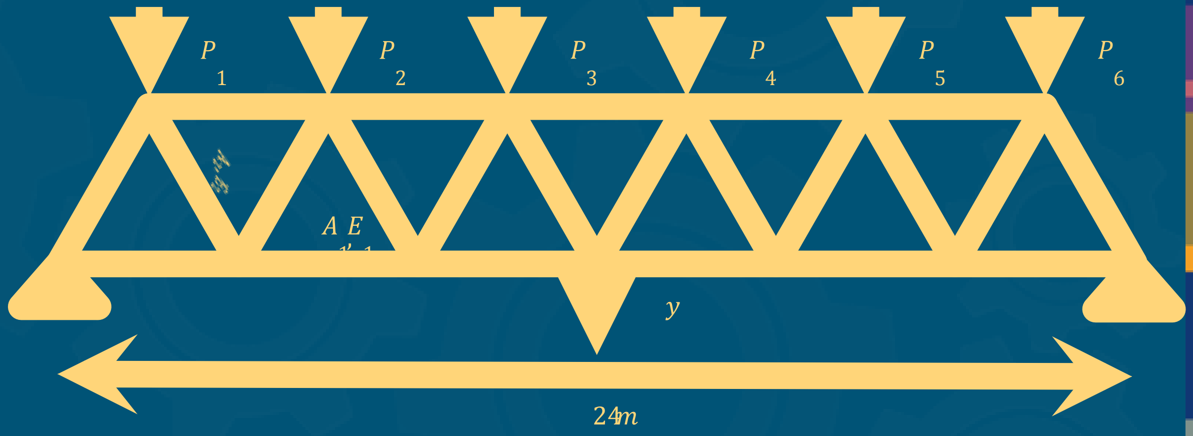
$$\mathcal{V}(Y) = -\mathcal{U}(-Y)$$



Eliciting Risk Measures:
optimized uncertainty equivalent Measures

Regret measures can be used to construct optimized uncertainty equivalent risk measures that encode stakeholder beliefs

$$\mathcal{R}(Y) = \inf_{d \in \mathbb{R}} \{ d + \mathcal{V}(Y - d) \}$$



d:	additional capacity added today
$\mathcal{V}(Y-d)$:	quantifies displeasure in future capacity shortfall
$d+\mathcal{V}(Y-d)$:	anticipated total (current + future) displeasure
$\mathcal{R}(Y)$:	smallest possible anticipated future shortfall

Surrogate Loss Functions Matter

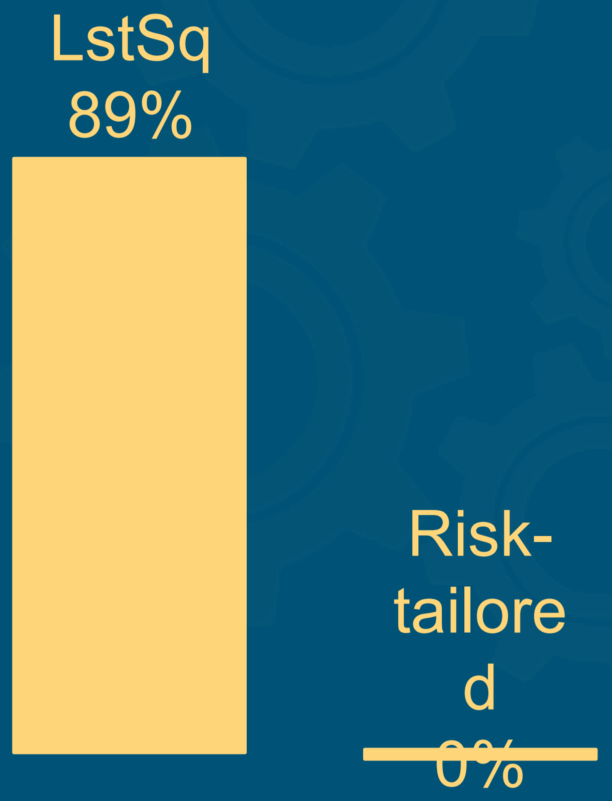
Typically least squares loss function is used to train surrogates

$$\mathcal{E}(Z) = \mathbb{E}[Z^2], \qquad Z = \hat{Y} - Y$$

This approach frequently under-estimates risk

We can limit underestimation by tailoring loss function to risk measure

Frequency of under-estimating risk



Tailoring Loss Functions To Estimate Risk Measures

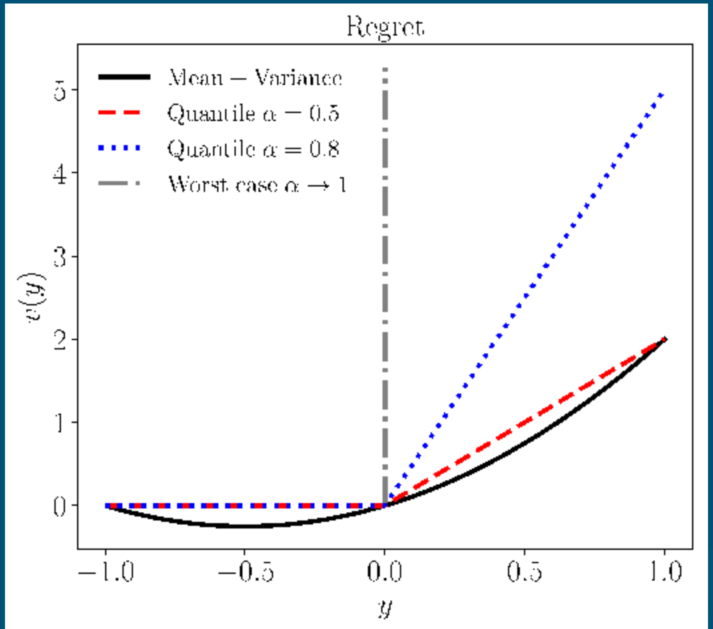
Loss functions (error functions) used to train surrogates should be tailored to the risk measure

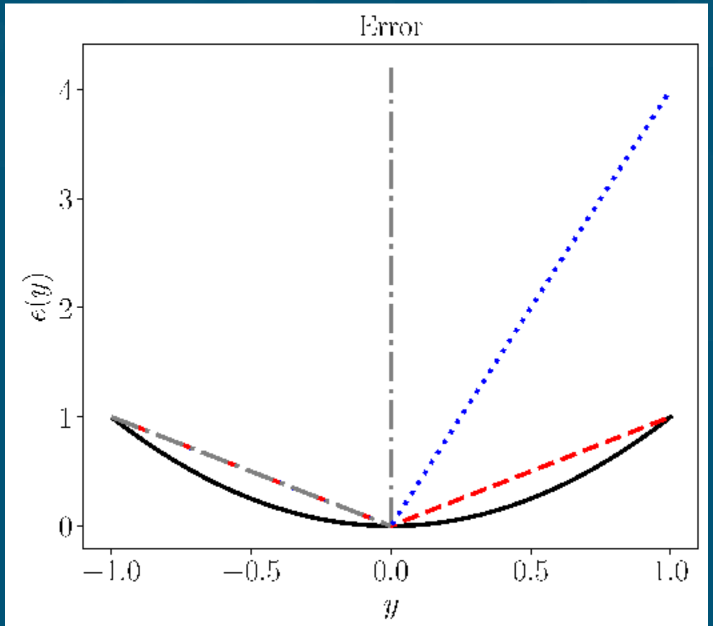
$$\mathcal{E}(Y) = \mathbb{E}[v(Y)] - \mathbb{E}[Y]$$

Mean+variance ($\lambda = 1$ LstSq loss)

$$v(y) = y + \lambda y^2$$

$$\mathrm{e}(y) = \lambda y^2$$





Tailoring Loss Functions To Estimate Risk Measures

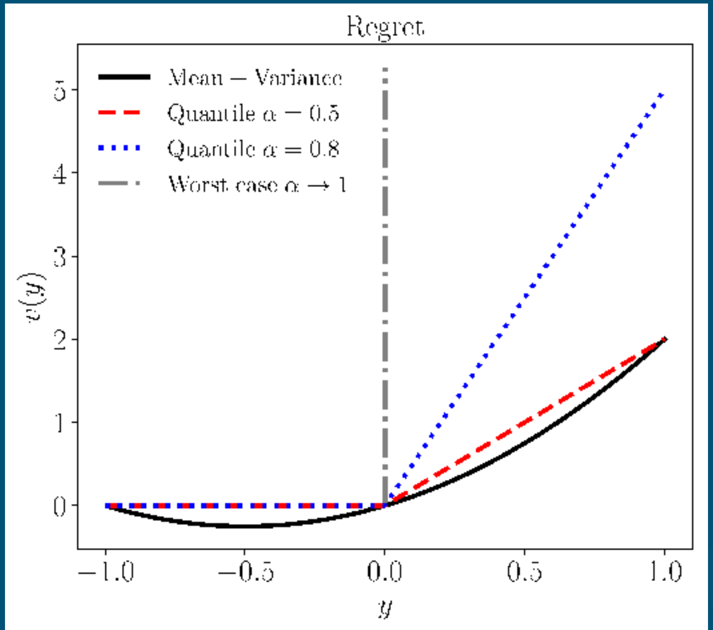
Loss functions (error functions) used to train surrogates should be tailored to the risk measure

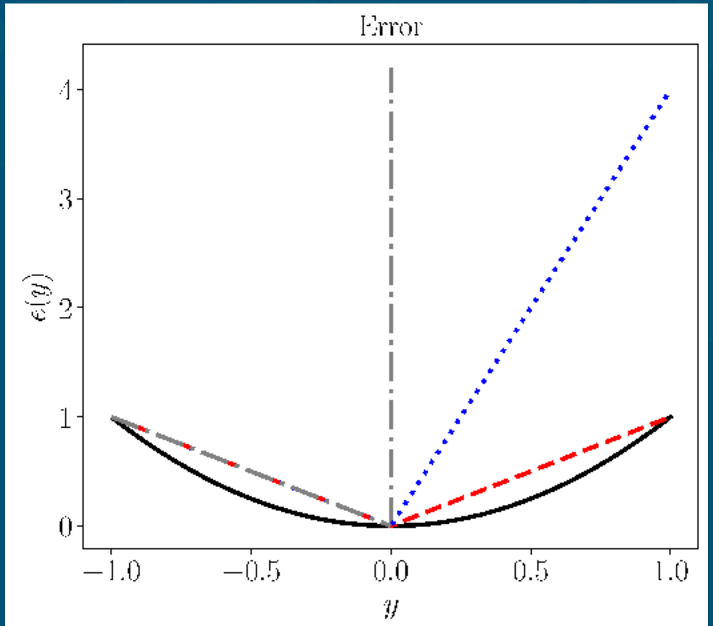
$$\mathcal{E}(Y) = \mathbb{E}[v(Y)] - \mathbb{E}[Y]$$

Quantile

$$v(y) = \frac{1}{1-p}\max[0, y]$$

$$\mathrm{e}(y) = \frac{p}{1-p}\max[0, y] + \max[0, -y]$$





Estimating A Single Risk Measure: Two Step Procedure [2]

Build surrogates of the form

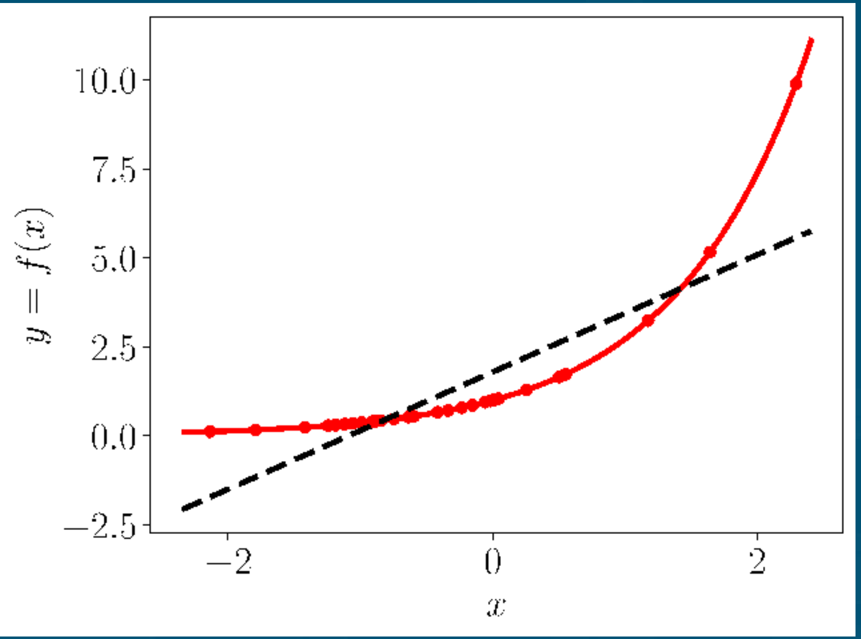
$$\hat{Y} = \theta_0 + g(X,\theta) \approx f(X) = Y$$

From *M* training data

$$X_M, Y_M$$

Step1: Solve the regression problem using the loss associated with the decision makers regret

$$\min_{\theta_0 \in \mathbb{R}, \theta \in \mathbb{R}^N} \mathcal{E}(Y_M - \theta_0 + g(X_M, \theta))$$



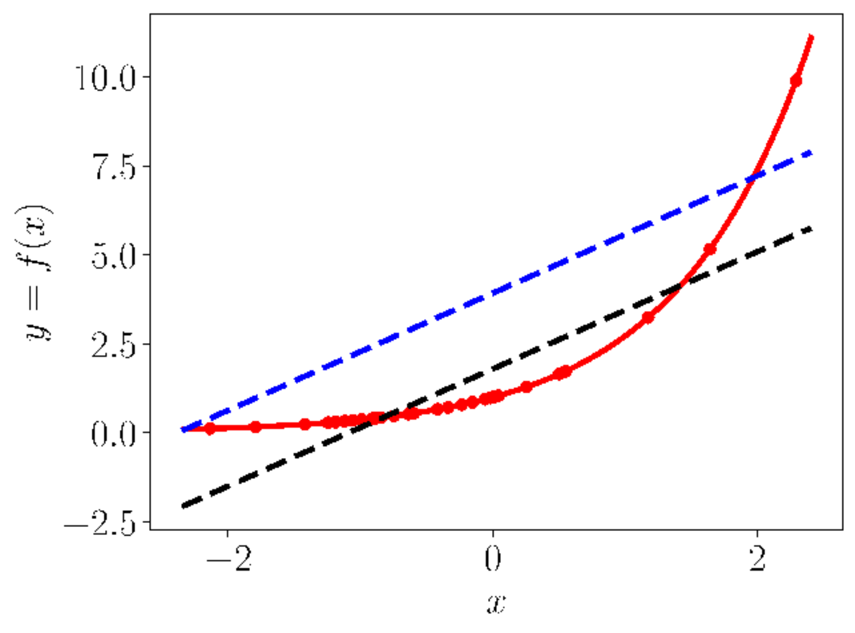
Estimating A Single Risk Measure: Two Step Procedure [2]

Build surrogates of the form

$$\hat{Y} = \theta_0 + g(X,\theta) \approx f(X) = Y$$

From *M* training data

$$X_M, Y_M$$



Step1: Solve the regression problem using the loss associated with the decision makers regret

$$\min_{\theta_0 \in \mathbb{R}, \theta \in \mathbb{R}^K} \mathcal{E}(Y_M - \theta_0 + g(X_M, \theta))$$

Step 2: Introduce the bias

$$\theta_0^\star = \mathcal{R}(Y_M - g(X,\theta))$$

Estimating A Single Risk Measure: Two Step Procedure [2]

Build surrogates of the form

$$\hat{Y} = \theta_0 + g(X, \theta) \approx f(X) = Y$$

From *M* training data

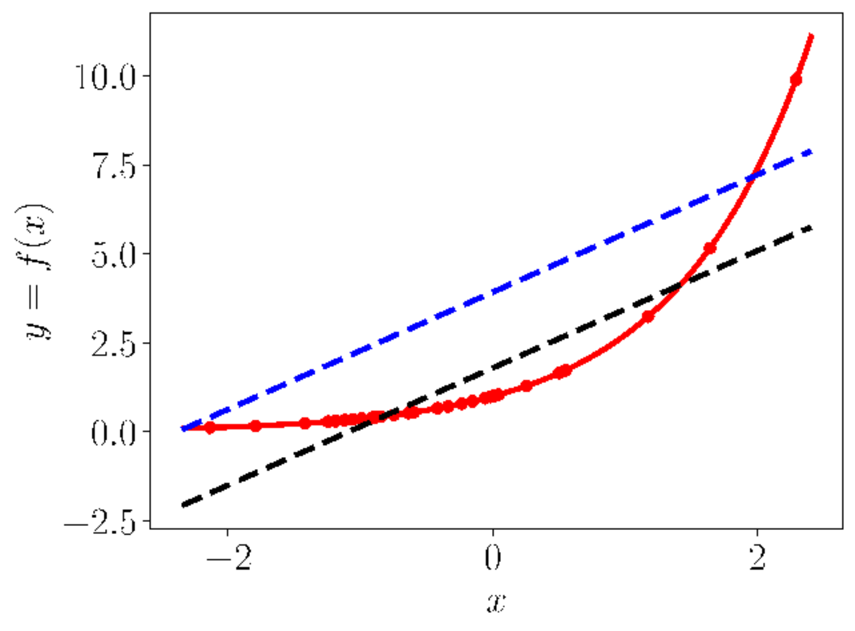
$$X_M, Y_M$$

Step1: Solve the regression problem using the loss associated with the decision makers regret

$$\min_{\theta_0 \in \mathbb{R}, \theta \in \mathbb{R}^K} \mathcal{E}(Y_M - \theta_0 + g(X_M, \theta))$$

Step 2: Introduce the bias

$$\theta_0^\star = \mathcal{R}(Y_M - g(X, \theta))$$



Surrogates are guaranteed to satisfy

$$\mathcal{R}\left(\theta_0^\star + g(X, \theta)\right) \geq \mathcal{R}(Y_M)$$

Convergence for orthonormal Systems

If bias is large enough any approach will produce conservative surrogates

We produces surrogates that both conservatively and accurately estimate risk

Consider orthonormal surrogates, e.g. PCEs

$$f(x) \approx \theta_0 + \sum_{k=1}^{K} \psi_k(x)\theta_k := \theta_0 + g_K(x,\theta)$$

If

$$\mathbb{E}[(Y - \bar{\theta}_0 - g_K(X,\bar{\theta}))^2] = \tau_K := \sum_{k=K+1}^{\infty} \bar{\theta}_k^2, \qquad \mathcal{E}(Z) \le C\mathbb{E}[Z^2]^{\alpha}$$

(and some other assumptions) then minimizing difference between true and surrogate risk measures yields

$$0 \le \theta_0^\star + \mathcal{R}(g_K(\theta^\star)) - \mathcal{R}(Y) \le 2C\tau_K^{\alpha}$$

COMPUTATIONAL ASPECTS

A number of regression problems arising from risk quadrangles can be solved efficiently for linear models. For example,

$$\mathcal{R}(Y) = \mathrm{AVaR}_p\,[Y]$$

We solve the following optimization problem via a linear program

$$\min_{\theta_0 \in \mathbb{R},\, \theta \in \Theta_k} \sum_{m=1}^{M} \pi^{(m)} \gamma_p \left[y^{(m)} - \theta_0 - g(x^{(m)}, \theta) \right]$$

$$\text{where} \quad \gamma_p[u] = p \max(0, u) + (1-p) \max(0, -u)$$

We the compute the shift

$$\theta_0^\star = \mathrm{AVaR}_p\,[Y - g(X, \theta^\star)]$$

$$\theta_0^\star = q_p(R) + \frac{1}{1-p} \sum_{m=k+1}^{M} \pi^{(m)} (r^{(m)} - q_p(R))$$

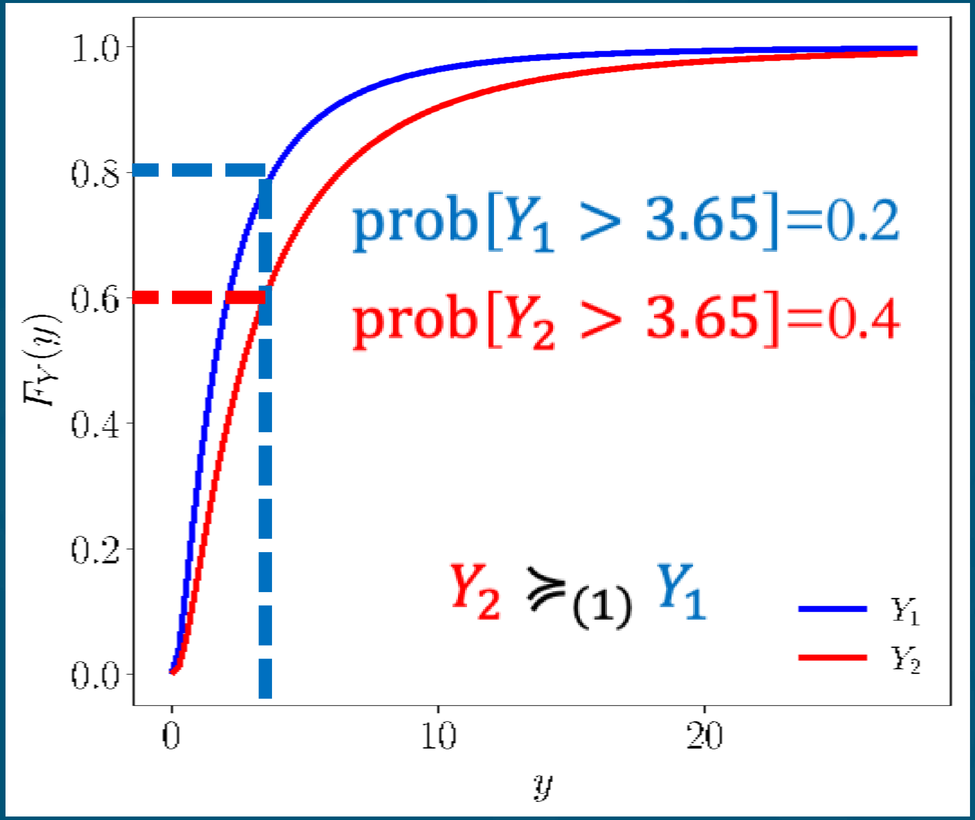
First-Order Stochastic Dominance (FSD)

- Risk preferences of two different stakeholders may differ
- Stochastic dominance can be used to construct surrogates that conservatively estimate a set of risk measures
- First-order stochastic dominance

$$Y' \succcurlyeq_{(1)} Y \iff 1 - F_{Y'(t)} \geq 1 - F_Y(t) \; \forall t \in \mathbb{R}$$

Guarantees that for all law invariant risk and monotonic risk measures

$$\mathcal{R}(Y') \geq \mathcal{R}(Y)$$



Increasing Convex order (ICX)

- FSD can be to risk averse
- ICX ordering is a weaker condition than FSD

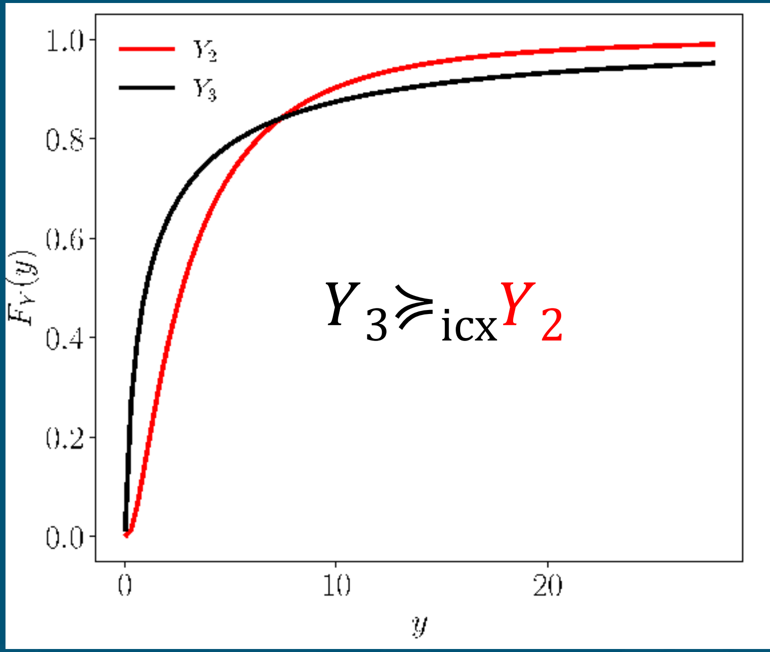
$$Y' \succcurlyeq_{\text{icx}} Y \iff \mathbb{E}[\max(0, Y' - t)] \geq \mathbb{E}[\max(0, Y - t)] \; \forall t \in \mathbb{R}$$

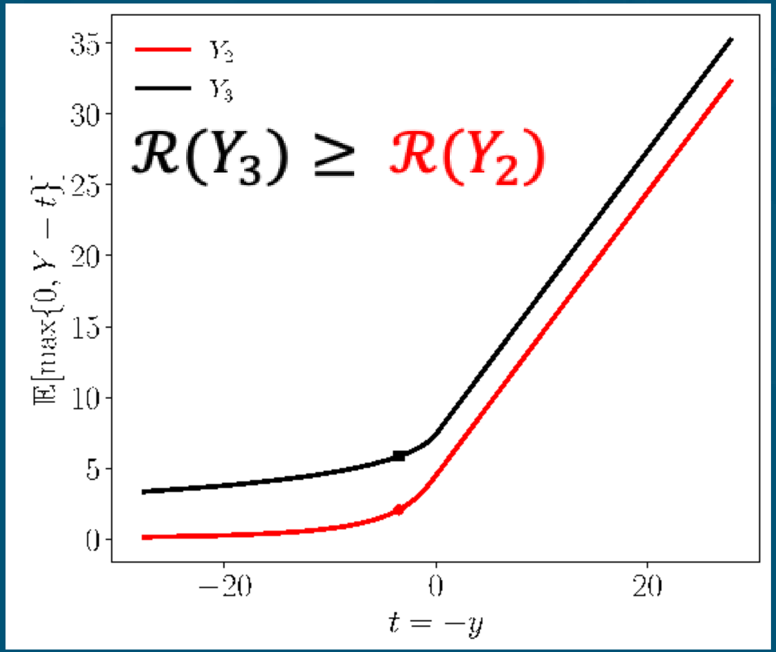
- ICX guarantees

$$\text{AVaR}_p[Y'] \geq \text{AVaR}_p[Y] \;\; \forall p \in (0,1)$$

- In fact for any convex, monotone, translation equivariant, law invariant risk measures

$$\mathcal{R}(Y') \geq \mathcal{R}(Y)$$





Computational Aspects

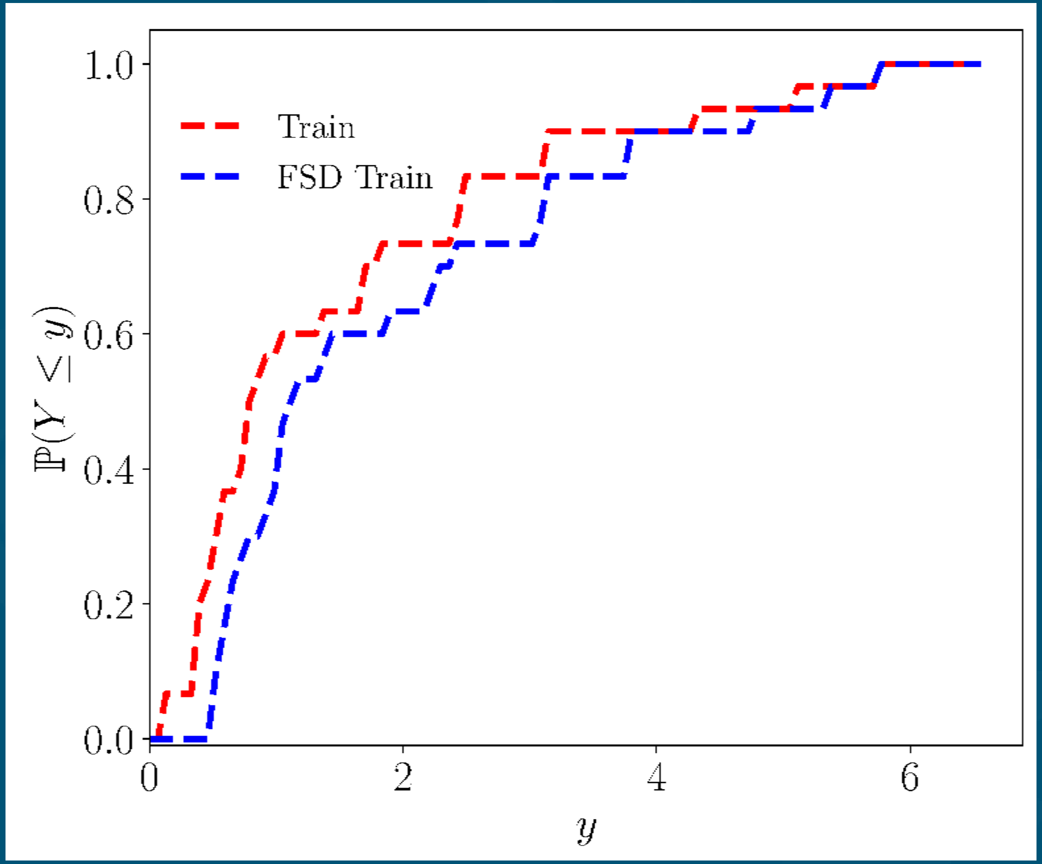
- We construct FSD surrogates using constrained least squares regression

$$\min_{\square_0 \in \mathbb{R},\ \square \in \square} \frac{1}{2} \sum_{m=1}^{M} \square^{(m)} \left(y^{(m)} - \square_0 - g(x^{(m)}, \square)\right)^2$$

subject to

$$\sum_{m=1}^{M} \square^{(m)} h_1\left(g^{(m)}(\square) - g^{(i)}(\square)\right) \leq \sum_{m=1}^{M} \square^{(m)} h_2\left(y^{(m)} - \square_0 - g^{(i)}(\square)\right)$$

$i = 1, \ldots, M.$



- Constraints are enforced at the training data
- Smooth the constraints to use gradient based optimization $\mathbb{1}_{(-1,0]}(t) \leq h_1(t)$
- We adopt similar smoothing approach for ICX

Computational Aspects

- We construct FSD surrogates using constrained least squares regression

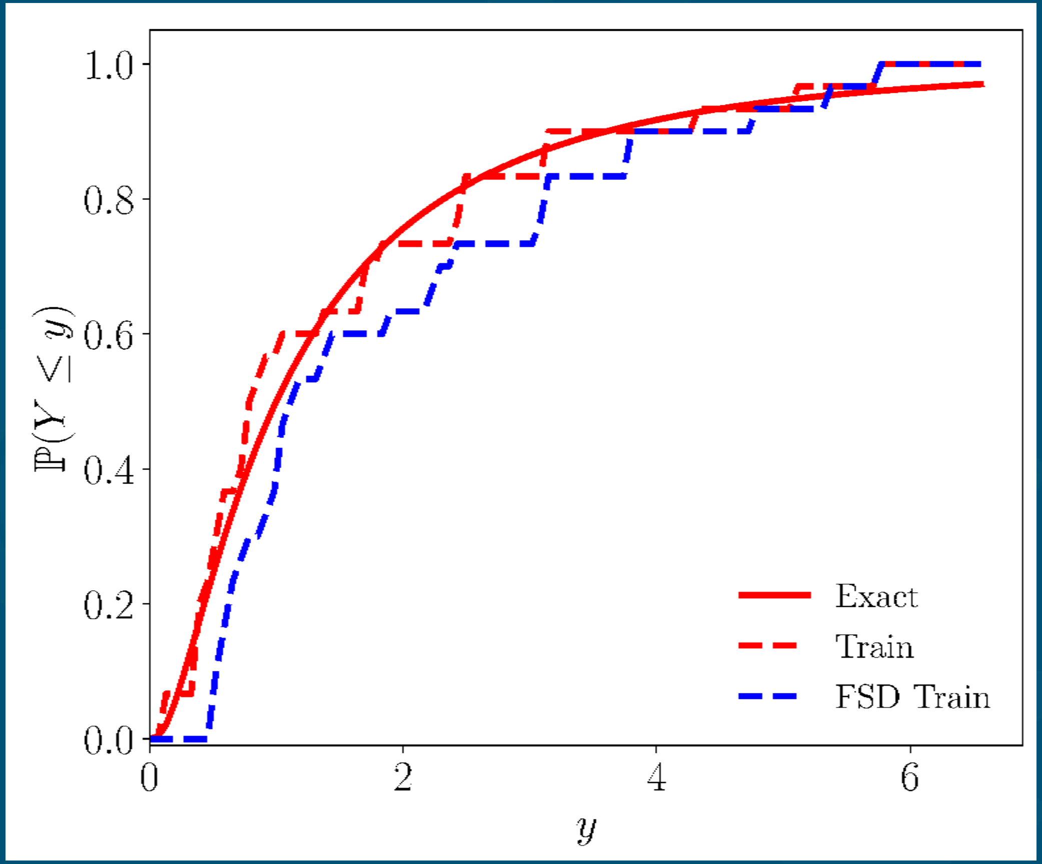
$$\min_{\square_0 \in \mathbb{R},\ \square \in \square} \frac{1}{2} \sum_{m=1}^{M} \square^{(m)} \left(y^{(m)} - \square_0 - g(x^{(m)}, \square)\right)^2$$

subject to

$$\sum_{m=1}^{M} \square^{(m)} h_1\left(g^{(m)}(\square) - g^{(i)}(\square)\right) \le \sum_{m=1}^{M} \square^{(m)} h_2\left(y^{(m)} - \square_0 - g^{(i)}(\square)\right)$$

$$i = 1, \dots, M.$$

- Constraints are enforced at the training data
- Smooth the constraints to use gradient based optimization $\mathbb{1}_{(-\infty,0]}(t) \le h_1(t)$
- We adopt similar smoothing approach for ICX



Convergence

- For any continuous function
- If we enforce ICX constraints on set S' (training data) and

$$\sup_{x \in f^{-1}(S')} |f(x) - (\bar{\theta}_0 + g_K(x, \bar{\theta}))| \leq \delta_K \qquad f^{-1}(S') := \{x \in \mathbb{R}^D \mid f(x) \in S'\}$$

- Then ICX and thus FSD surrogate $(\theta_0^\star, \theta^\star)$ satisfy

$$\mathbb{E}[(Y - \theta_0^\star - g_K(X, \theta^\star))^2] \leq \tau_K + \delta_K^2$$

- The convergence rate is no worse than mean-squared error plus the metric δ_K based upon the uniform approximation quality

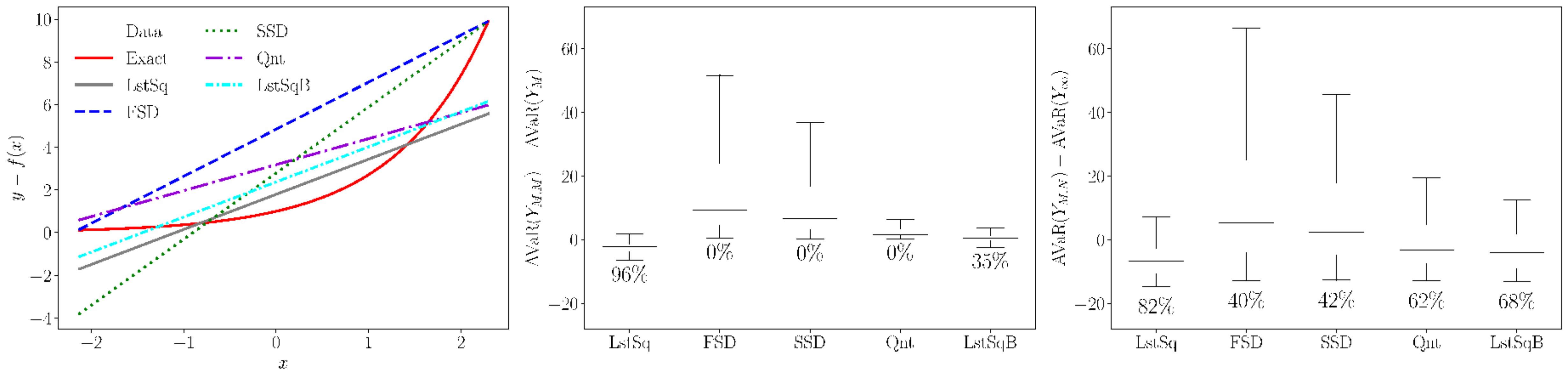
Numerical Examples: Estimating a Single Risk Measure

- Consider Gaussian random variables X

$$Y = f(x) = \exp(1^\top x)$$

- Y is lognormal (exact statistics known)

Linear PCE



D=1, M=30 D=3, M=100 D=3, M=100

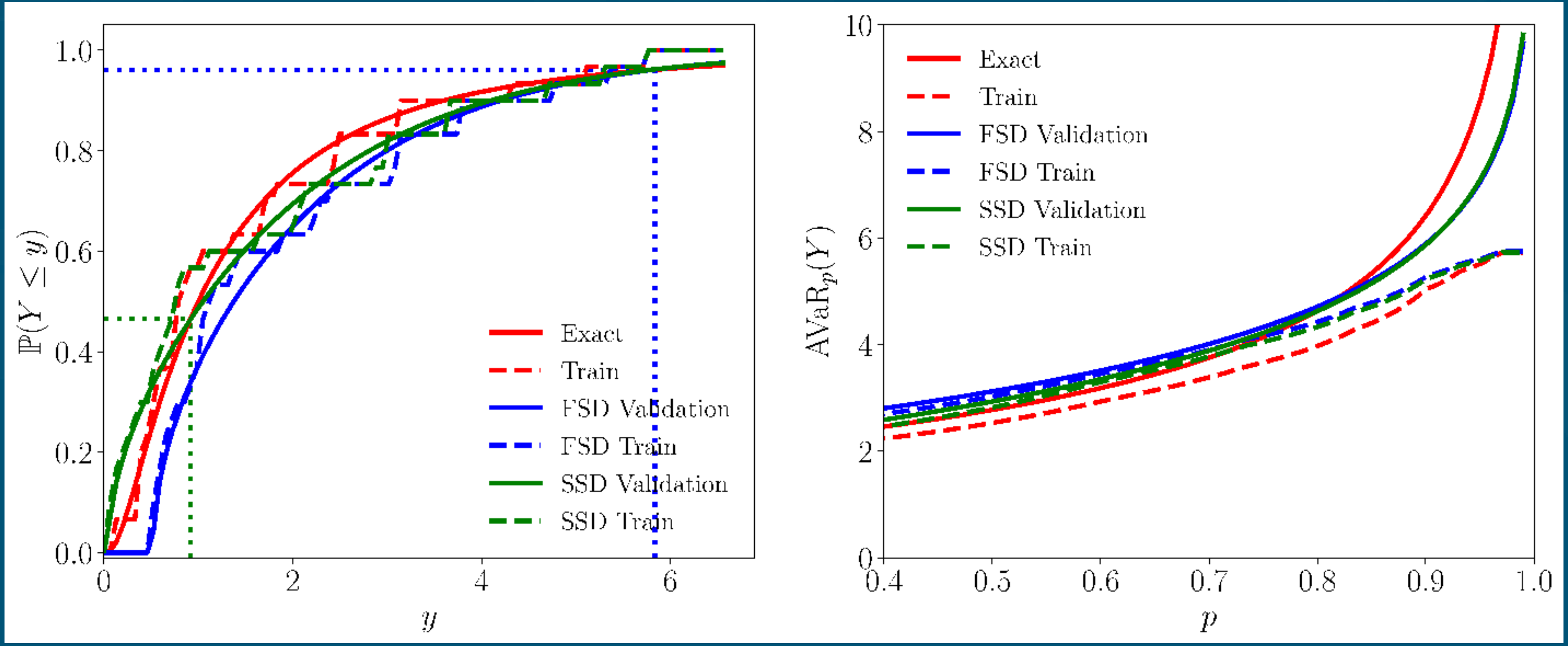
Numerical Examples: Estimating a Set of Risk Measures

- Consider Gaussian random variables X

$$Y = f(x) = \exp(1^{\top}x)$$

- Y is lognormal (exact statistics known)

Quadratic PCE
D=1, M=30



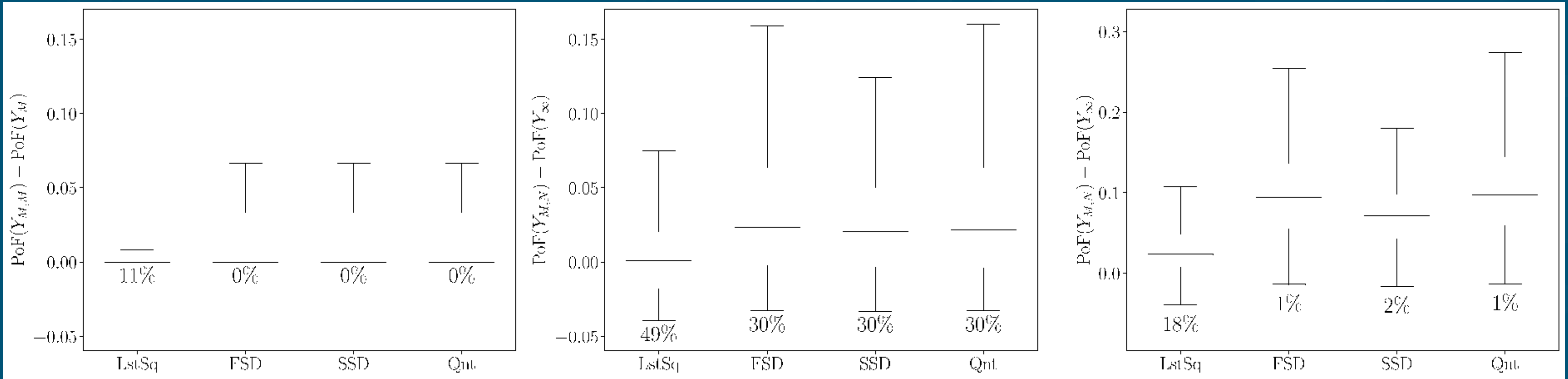
NUMERICAL EXAMPLES: ESTIMATING PROBABILITY OF FAILURE

- Consider Gaussian random variables *X*

$$Y = f(x) = \exp(1^\top x)$$

- *Y* is lognormal (exact statistics known)

Quadratic PCE



D=3, M=30

D=3, M=30

D=3, M=100

Numerical Examples: Convergence

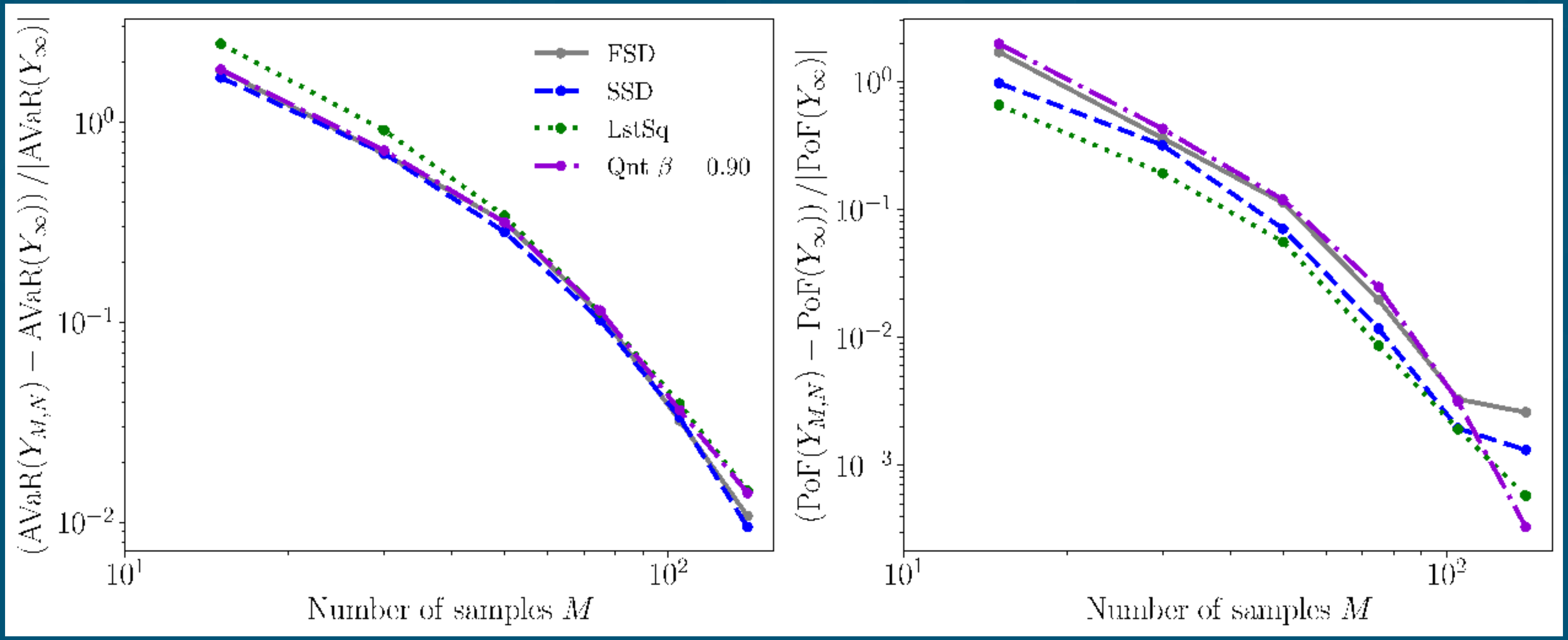
- Consider Gaussian random variables X

$$Y = f(x) = \exp(1^\top x)$$

- Y is lognormal (exact statistics known)

Quadratic PCE

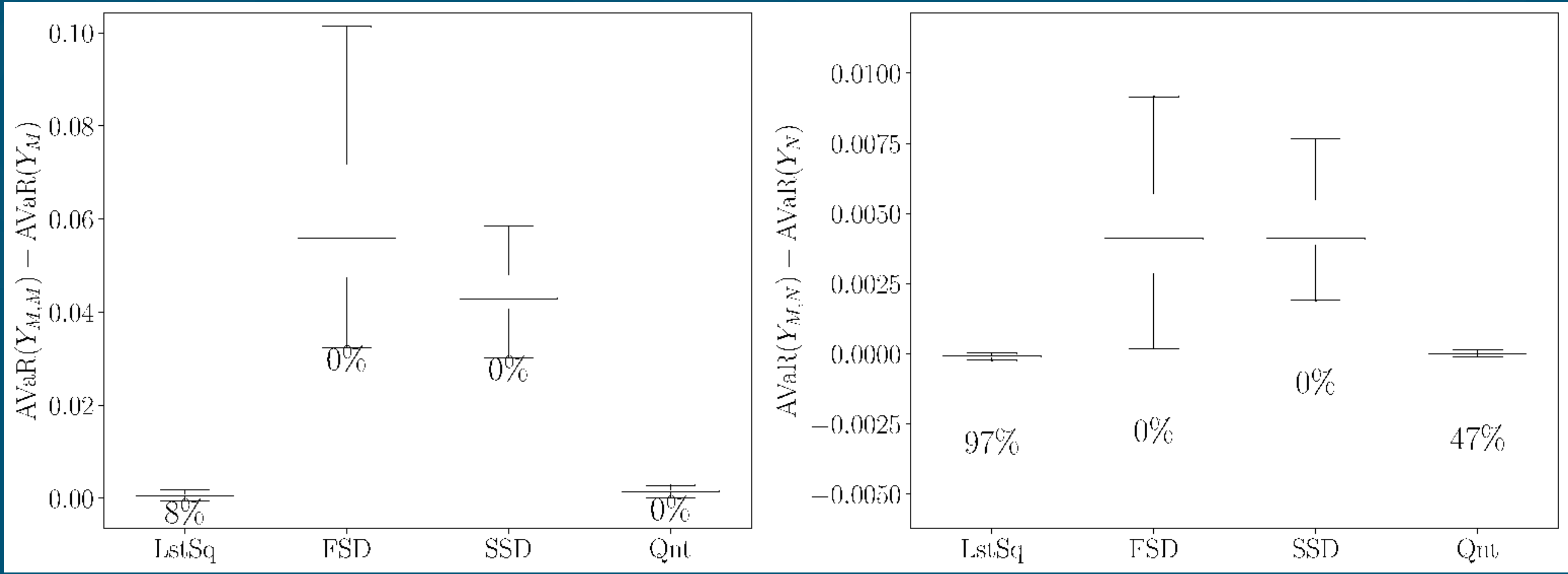
Over sample ratio: 5



Degree → Degree →

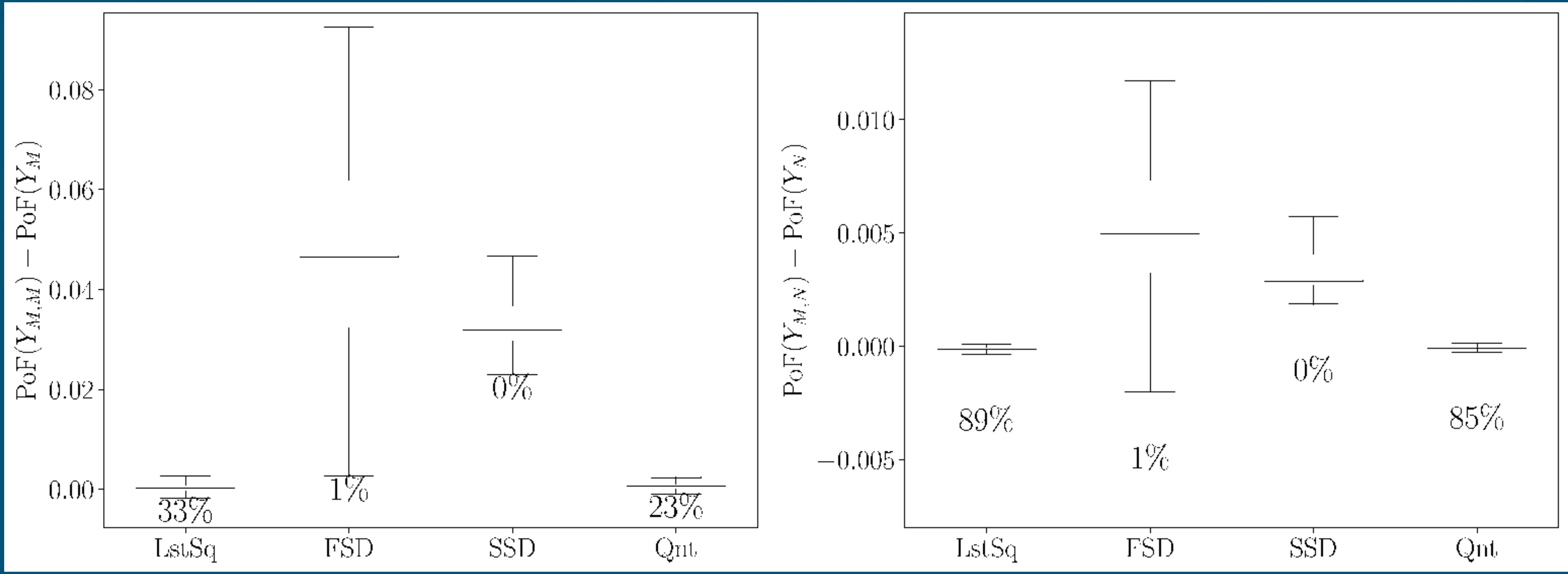
Numerical Examples: Truss - AVaR

- Use quadratic PCE in 10 dimensions with 66 terms
- Use 80 training samples
- 9,000 validation samples
- Compute AVaR with p=0.90



Numerical Examples: Truss - Probability of Failure

- Use quadratic PCE in 10 dimensions with 66 terms
- Use 80 training samples
- 9,000 validation samples
- Compute PoF with. $prob\left(Y > q_p(Y)\right) = 0.1$



Numerical Examples: Dimension Reduction

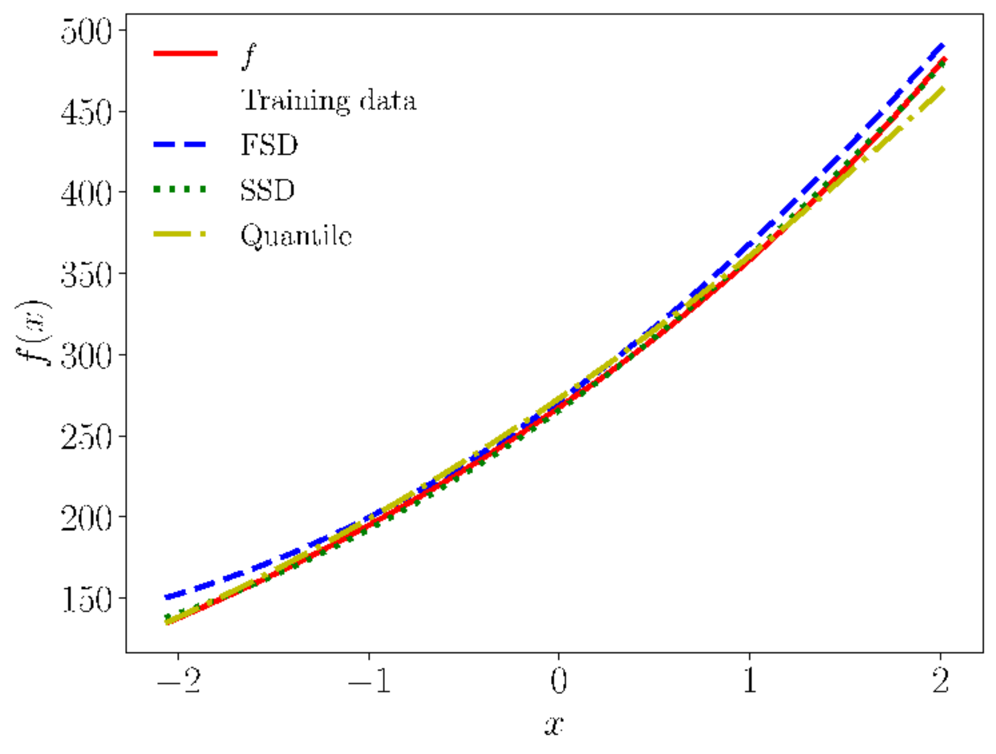
- Consider model of 10D wing weight

$$f(x) = 0.036\, S_w^{0.758} W_{fw}^{0.0035} \left(\frac{A}{\cos^2(\Lambda)}\right)^{0.6} q^{0.006} \lambda^{0.04} \left(\frac{100 t_c}{\cos(\Lambda)}\right)^{-0.3} (N_z W_{dg})^{0.49} + S_w W_p$$

- Compute 1D active subspace

$$W \Lambda W^\top = C = \frac{1}{M} \sum_{m=1}^{M} \nabla_x f(x^{(m)}) \nabla_x f(x^{(m)})^\top$$

- Signed relative error in AVaR (p=0.9) using 30 samples
  - FSD: 0.029
  - SSD: 0.008
  - Quantile: 0.010



Conclusions

- When estimating uncertainty the loss (error) function matters
- The risk measures should be elicited from stakeholder preferences
- The surrogate regression problem should be tailored to the risk measure (especially for limited data)
- The resulting surrogate is assured to conservatively (over-estimate) a chosen risk measure
- If single risk measure is agreed upon the risk quadrangle is most effective
- Stochastic dominance can be used to enforce conservativeness with respect to multiple risk measures

References

This talk is a summary of the paper

J.D. Jakeman, D. Kouri, and G. Huerta. Surrogate modeling for efficiently, accurately and conservatively estimating measures of risk (2021). Submitted

The procedures for constructing surrogates that estimate a single risk measure are based upon

[1] R. T. Rockafellar, S. Uryasev, M. Zabarankin, Risk tuning with generalized linear regression, Mathematics of Operations Research 33 (3) (2008) 712–729.

[2] R. T. Rockafellar, J. O. Royset, Measures of residual risk with connections to regression, risk tracking, surrogate models, and ambiguity, SIAM Journal on Optimization 25 (2) (2015) 1179– 1208.

[3] R. T. Rockafellar, S. Uryasev, The fundamental risk quadrangle in risk management, optimization and statistical estimation, Surveys in Operations Research and Management Science 18 (1–2) (2013) 33 – 53.

Risk Quadrangle

The risk quadrangle formulates risk from regret

The risk quadrangle tailors the error (loss) function used to build surrogates with regre

We can use these connections to ensure surrogates conservatively estimate risk

$$
\begin{array}{lrcccll}
 & \text{Risk} & \mathcal{R} & \longleftrightarrow & \mathcal{D} & \text{Deviation} & \\
\text{Optimization} & & \uparrow\downarrow & \mathcal{S} & \downarrow\uparrow & & \text{Estimation} \\
 & \text{Regret} & \mathcal{V} & \longleftrightarrow & \mathcal{E} & \text{Error} &
\end{array}
$$

$$\mathcal{R}(Y) = \mathbb{E}[Y] + \mathcal{D}(Y) \qquad \mathcal{D}(Y) = \mathcal{R}(Y) - \mathbb{E}[Y]$$

$$\mathcal{V}(Y) = \mathbb{E}[Y] + \mathcal{E}(Y) \qquad \mathcal{E}(Y) = \mathcal{V}(Y) - \mathbb{E}[Y]$$

$$\mathcal{R}(Y) = \inf_{t\in\mathbb{R}} \{t + \mathcal{V}(Y-t)\} \qquad \mathcal{D}(Y) = \inf_{t\in\mathbb{R}} \mathcal{E}(Y-t)$$

$$\mathcal{S}(Y) = \arg\min_{t\in\mathbb{R}} \{t + \mathcal{V}(Y-t)\} = \arg\min_{t\in\mathbb{R}} \mathcal{E}(Y-t)$$