Minimal-Dispersion and Maximum-Likelihood Predictors with a Linear Staircase Structure

Luis G. Crespo, Sean P. Kenny, and Daniel P. Giesy

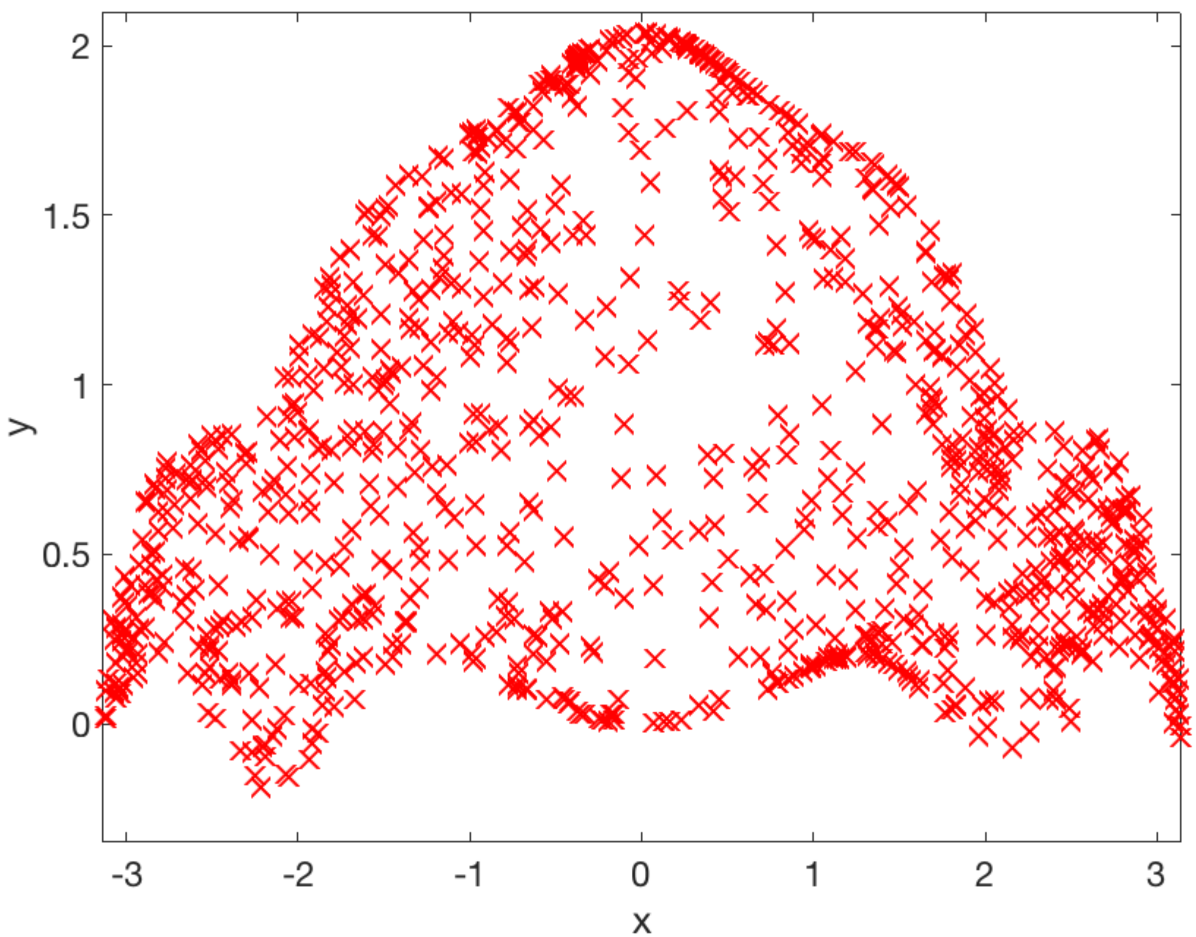
Dynamic Systems and Controls Branch
NASA Langley Research Center, Hampton, VA, 23666

Outline

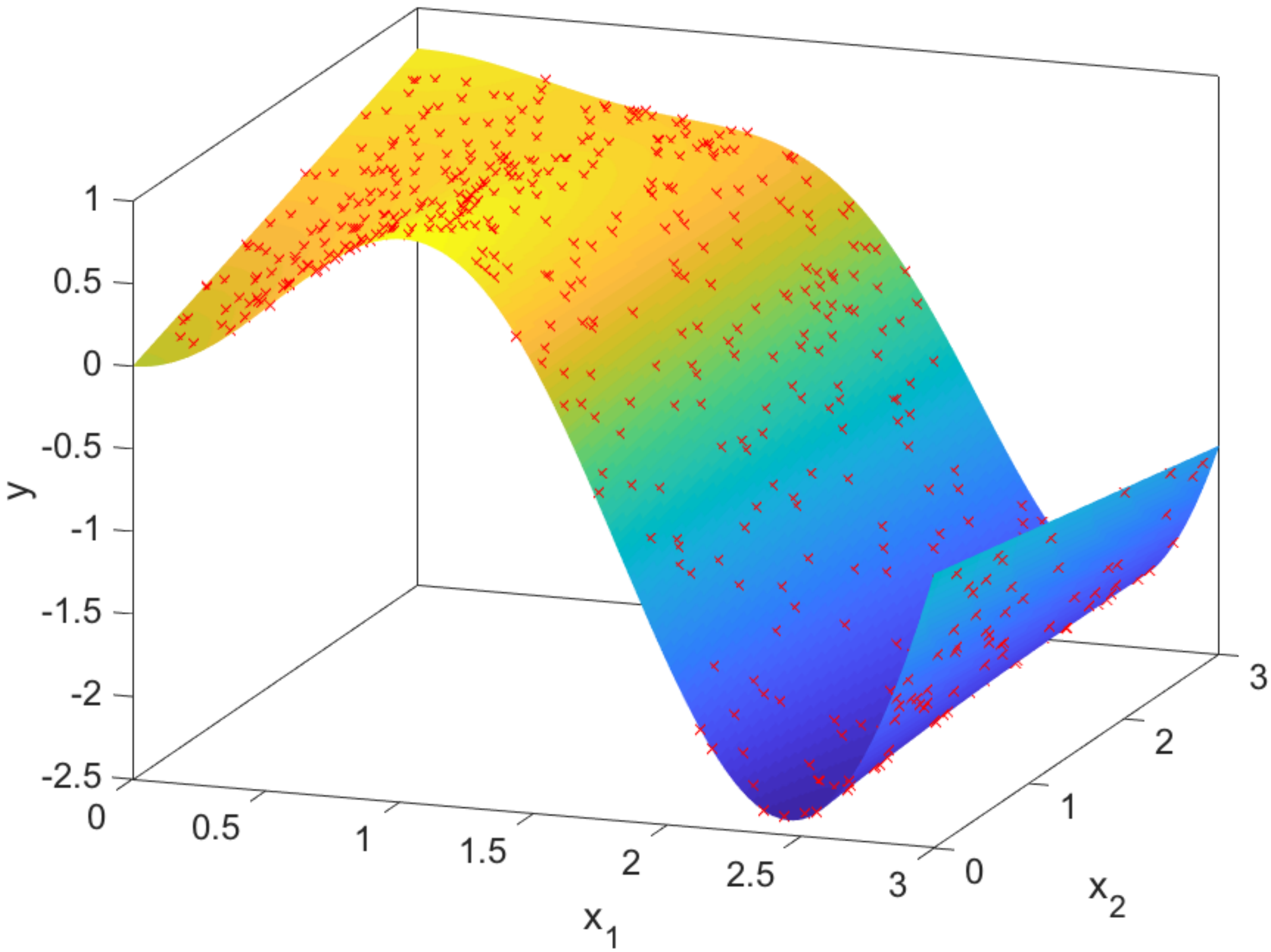
- Problem statement
- Background
- Random predictor models
- Conclusions

Problem Statement

- Goal: create a computational model of a Data Generating Mechanism (DGM) given N input-output pairs $D=\{x^{(i)}, y^{(i)}\}$

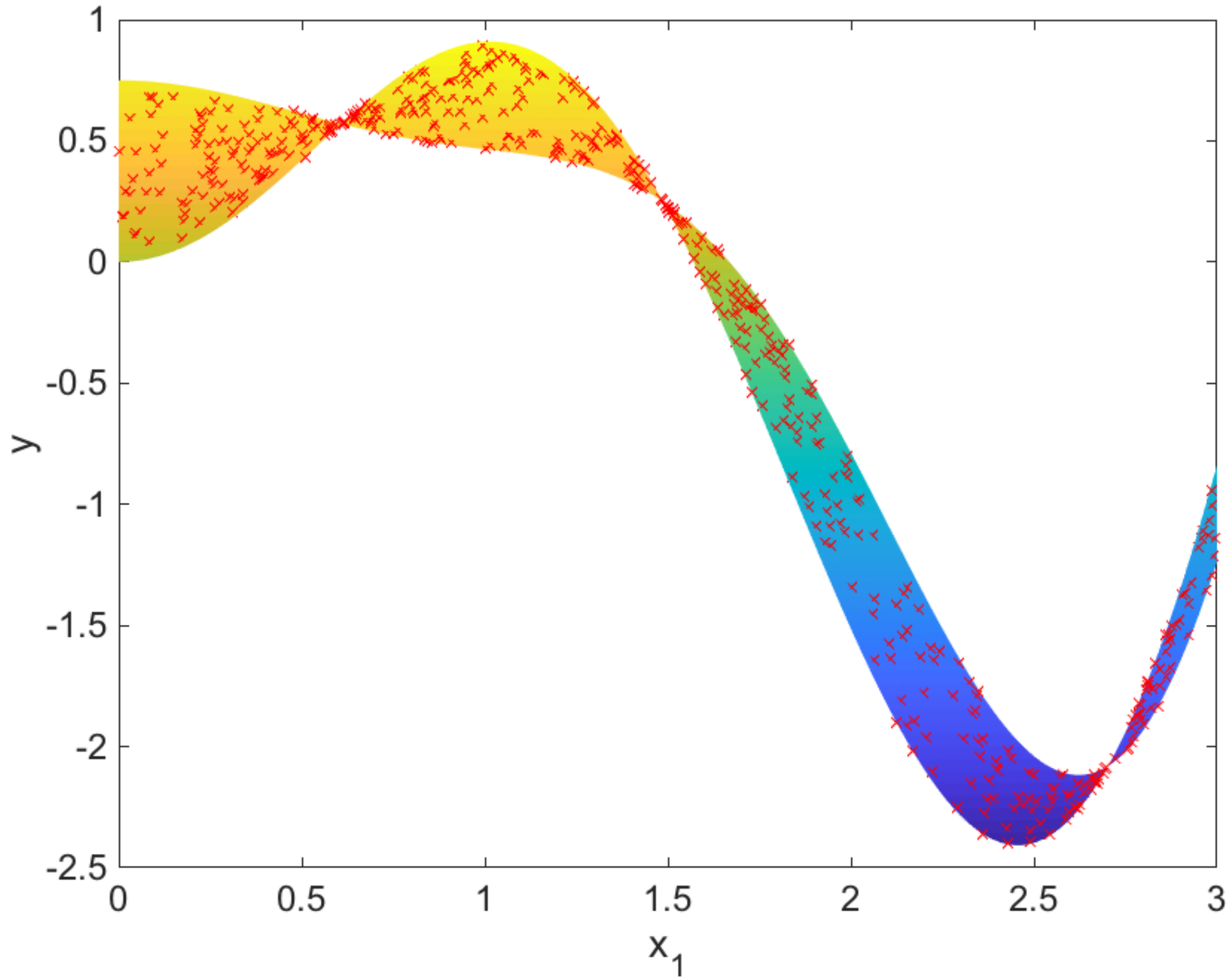


Problem Statement: On the DGM



DGM is a deterministic function of 2 inputs without noise

Problem Statement: On the DGM



Model form uncertainty vs. deterministic function + colored noise

Problem Statement

- Parametric models vs. non-parametric models
- This paper focuses on the parametric model

$$y = p^\top \varphi(x)$$

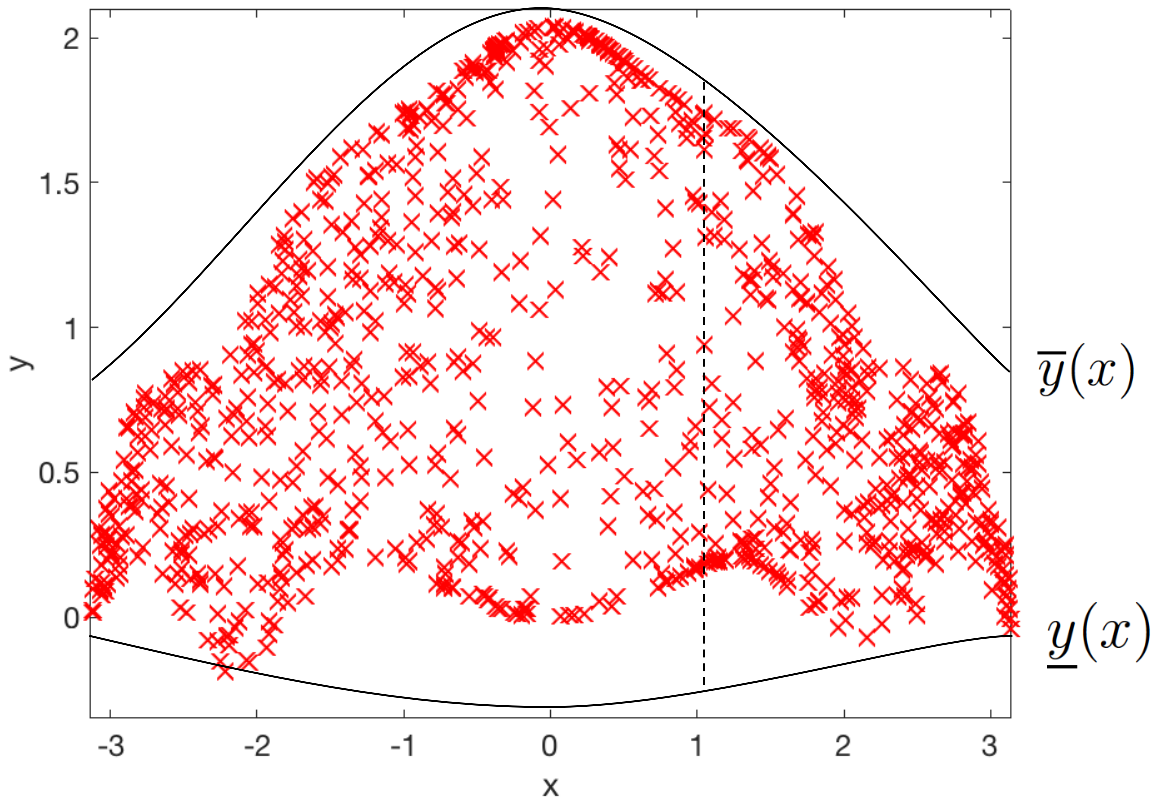
- This form is implied by the superposition property of linear system theory
- The calibration problem of interest is not standard since the calibrated variable is unobservable

Outline

- Problem statement
- Computational models
- Staircase variables
- Random predictor models
- Conclusions

Computational Models

- Interval Predictor Models (IPM)



Interval Predictor Models

- The output is an interval valued function of the input
- IPM considered here are given by

$$y = p^\top \varphi(x), \quad P = \{p : \underline{p} \leq p \leq \overline{p}\}$$

- This leads to

$$I_y(x, P) = \left[\, \underline{y}(x, \overline{p}, \underline{p}),\; \overline{y}(x, \overline{p}, \underline{p}) \,\right],$$

where the IPM boundaries are known analytically

- Interval and functional representation
- The spread of the IPM is

$$\delta_y(x, \overline{p}, \underline{p}) = (\overline{p} - \underline{p})^\top |\varphi(x)|.$$

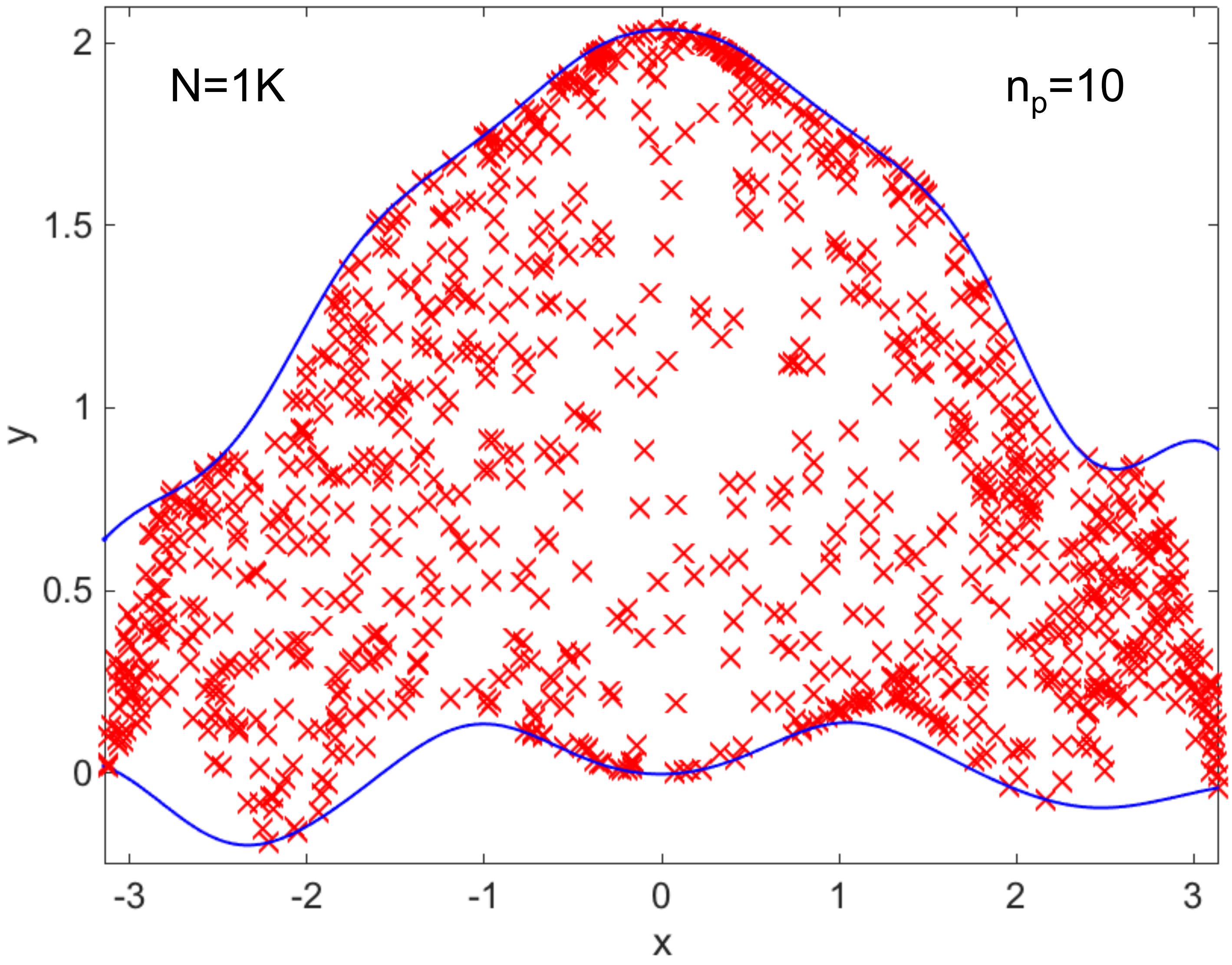
Interval Predictor Models

- IPMs are calculated by solving the convex program

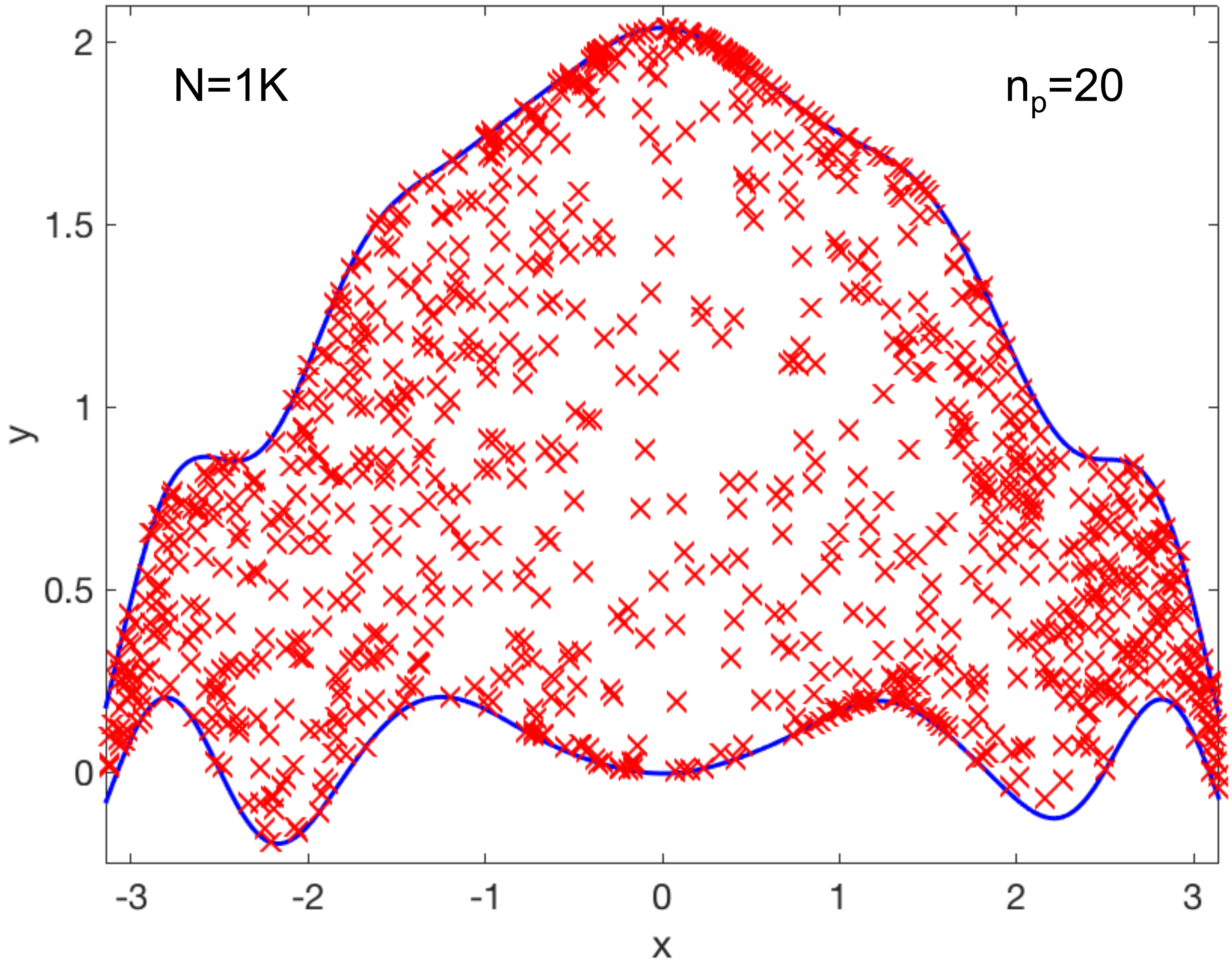
$$
\begin{aligned}
\left\{\underline{\hat{p}}(c),\ \hat{\bar{p}}(c)\right\} = \arg\min_{u,v:\,u\leq v} \Big\{ & \mathbb{E}_x[\delta_y(x,v,u)] : \\
& \underline{y}\left(x^{(i)},v,u\right) \leq y^{(i)} \leq \bar{y}\left(x^{(i)},v,u\right), \\
& c(u,v) \leq 0,\ i=1,\ldots N \Big\}
\end{aligned}
$$

Additional set of constraints

Interval Predictor Models



Interval Predictor Models



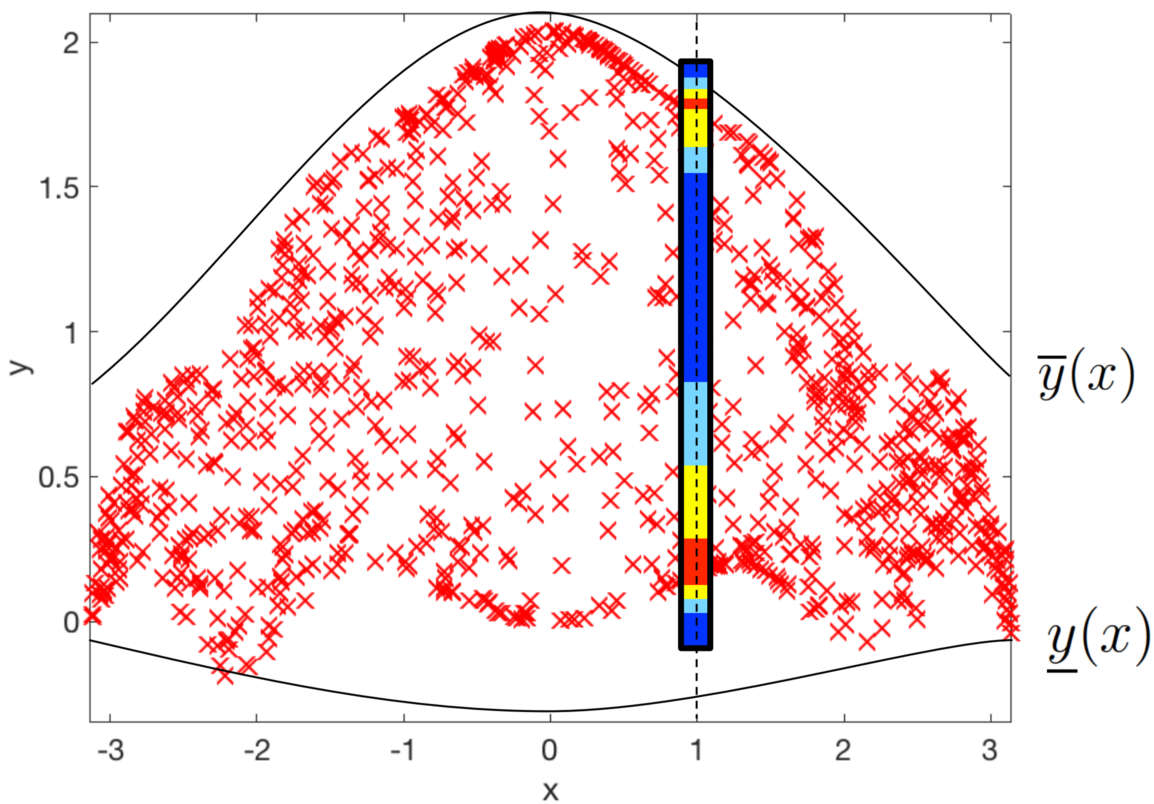
Interval Predictor Models: Reliability

- Reliability of the Predictor: scenario theory enables bounding the probability of a future observation falling outside the IPM: distribution-free, non-asymptotic

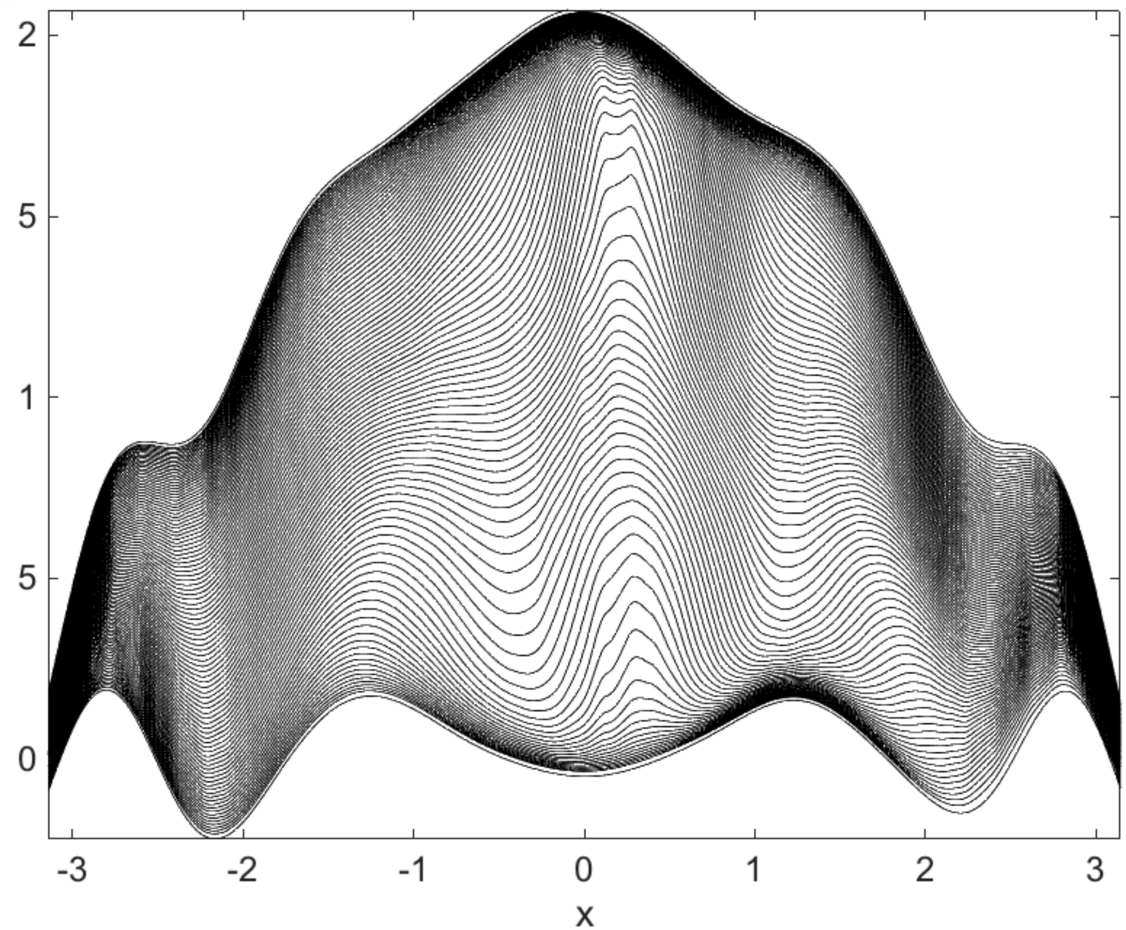
- This is a probabilistic certificate of correctness prescribing the interplay between the amount of information available, the complexity of the model, a confidence parameter, and the reliability of the model

Computational Models

- Random Predictor Models (RPM)



Random Predictor Models



Maximal-entropy Staircase RPM

The modality and skewness vary strongly with the input

Outline

- Problem statement
- Computational models
- Staircase variables
- Random predictor models
- Conclusions

Background

- Hyper-parameters

$$\theta_z = [\, \underline{z}, \overline{z}, \mu, m_2, m_3, m_4 \,]$$

- Desired variables must match these constraints
- Only some θ_z are feasible
- Polynomial feasibility constraints: $g(\theta_z) \leq 0$

Background: θ-Feasibility Equations

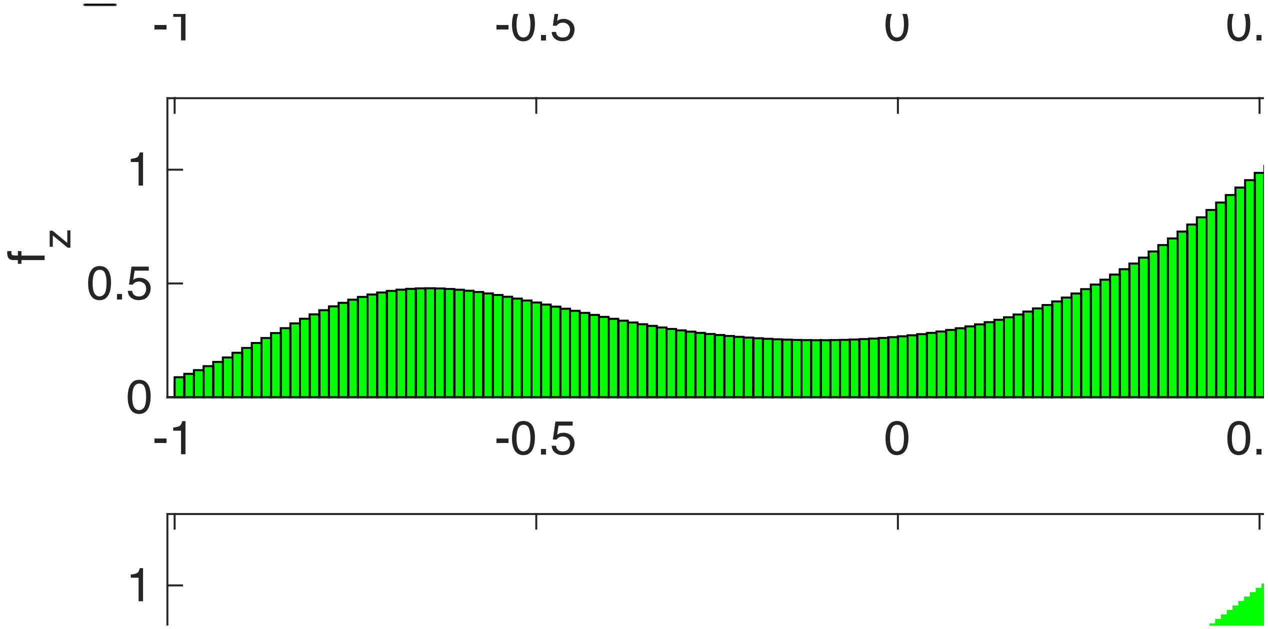
$$
\begin{aligned}
g_1 &= \underline{z} - \overline{z},\\
g_2 &= \underline{z} - \mu,\\
g_3 &= \mu - \overline{z},\\
g_4 &= -m_2,\\
g_5 &= m_2 - v\\
g_6 &= m_2^2 - m_2(\mu - \underline{z})^2 - m_3(\mu - \underline{z}),\\
g_7 &= m_3(\overline{z} - \mu) - m_2(\overline{z} - \mu)^2 + m_2^2,\\
g_8 &= 4m_2^3 + m_3^2 - m_2^2(\overline{z} - \underline{z})^2,\\
g_9 &= 6\sqrt{3}m_3 - (\overline{z} - \underline{z})^3,\\
g_{10} &= -6\sqrt{3}m_3 - (\overline{z} - \underline{z})^3,\\
g_{11} &= -m_4,\\
g_{12} &= 12m_4 - (\overline{z} - \underline{z})^4,\\
g_{13} &= (m_4 - vm_2 - um_3)(v - m_2) + (m_3 - um_2)^2,\\
g_{14} &= m_3^2 + m_2^3 - m_4 m_2,
\end{aligned}
$$

Distribution free

Staircase Variables

- A staircase random variables has a piecewise constant density function over a uniform partition of the domain that match the constraints imposed by θ_z
- Staircases are found by solving the convex program

$$\hat{\ell} = \arg\min_{\ell \geq 0} \{ J(\theta, n_b) : A(\theta, n_b)\ell = b(\theta), \theta \in \Theta \}$$



Staircase Variables: Key Attributes

- Able to represent a wide range of density shapes by using different optimality criteria
  - Max entropy
  - Max likelihood
  - Max degree of unimodality, etc
- Able to represent most of the feasible space
- Low-computational cost: from convex optimization

Outline

- Problem statement
- Computational models
- Staircase variables
- Random predictor models
  - Moment-matching
  - Minimal dispersion
- Conclusions

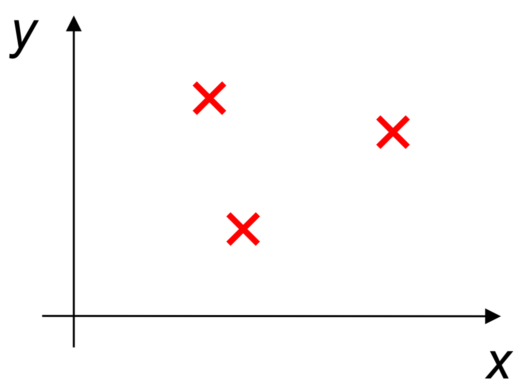
Random Predictor Models

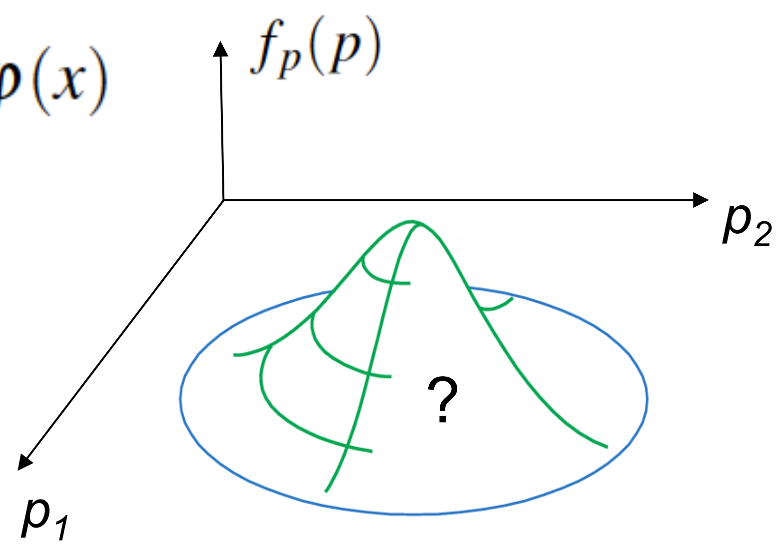
- The output is a random process
- RPM considered here are given by

$$R_y\,(x, f_p) = \{y = p^\top \varphi(x),\; p \sim f_p(p),\; p \in P\}$$

- Goal: given the data sequence D={$x^{(i)}$, $y^{(i)}$} we want to characterize the distribution of *p*

$$y = p^\top \varphi(x)$$

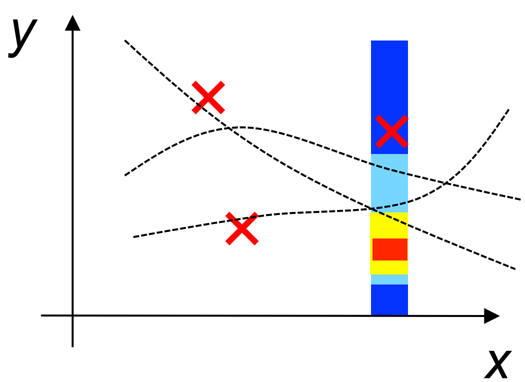


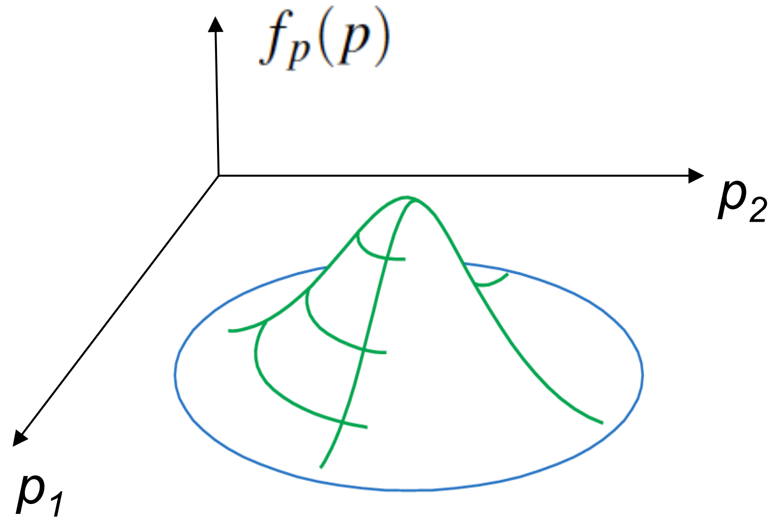


Random Predictor Models

- Bayesian/Maximum likelihood approach
  - Pros: any model, any distribution
  - Cons: expensive, tight to assumed distribution

$$y = p^\top \varphi(x)$$





Random Predictor Models

- Taking the expected value of the model equation we have

$$\begin{aligned}
\mu_{y(x)} &= \mathbb{E}_p[p]^\top \varphi(x), \\
\mathbb{E}_y\left[y^2\right] &= \varphi^\top(x)\mathbb{E}_p\left[pp^\top\right]\varphi(x), \\
\mathbb{E}_y\left[y^3\right] &= \varphi^\top(x)\mathbb{E}_p\left[pp^\top\varphi(x)p^\top\right]\varphi(x), \\
\mathbb{E}_y\left[y^4\right] &= \varphi^\top(x)\mathbb{E}_p\left[pp^\top\varphi(x)\varphi(x)^\top pp^\top\right]\varphi(x).
\end{aligned}$$

which can be combined to obtain the moment functions

$$\begin{aligned}
\mu_{y(x)} &= h_\mu(\mu, x), \\
m_{2,y(x)} &= h_{m_2}(\mu, m_2, x), \\
m_{3,y(x)} &= h_{m_3}(\mu, m_2, m_3, x), \\
m_{4,y(x)} &= h_{m_4}(\mu, m_2, m_3, m_4, x).
\end{aligned}$$

Parameter independency is assumed hereafter

Moment-Matching RPMs

- Idea: Find the moments of *p* leading to a prediction that minimizes the offset between the predicted moments and the empirical moments
- A sliding-window approach is used to estimate the empirical moments

$$\tilde{m}_{y(x)} = \left[\tilde{\mu}_{y(x)}, \tilde{m}_{2,y(x)}, \tilde{m}_{3,y(x)}, \tilde{m}_{4,y(x)}\right]$$

- The predicted moments, given by

$$m_{y(x)} = \left[\mu_{y(x)}, m_{2,y(x)}, m_{3,y(x)}, m_{4,y(x)}\right]$$

depend upon the design variables:

$$\theta_p = \left[\underline{p}, \overline{p}, \mu, m_2, m_3, m_4\right]$$

Moment-Matching RPMs

- <u>Solution Approach:</u> a sequence of optimization programs for moments of increasing order.
  1. Solve for the mean

$$\hat{\mu} = \arg\min_{\mu} \left\{ \sum_{i=1}^{N} \left(\tilde{\mu}_{y(x^{(i)})} - h_{\mu}\left(\mu, x^{(i)}\right) \right)^2 \right\}$$

  2. Find a feasible support set using IPMs
  3. Solve for the variance

$$\hat{m}_2 = \arg\min_{m_2} \left\{ \sum_{i=1}^{N} \left(\tilde{m}_{2,y(x^{(i)})} - h_{m_2}\left(\hat{\mu}, m_{2,}, x^{(i)}\right) \right)^2 : c_2(m_2) \leq 0 \right\}$$

for $c_2 = g_{\mathbf{b}}|_{\underline{z}=p,\overline{z}=\overline{p},\mu=\hat{\mu}}$ and $\mathrm{b} = \{4,5\}$

  4. Find a feasible support set using IPMs….

Moment-Matching RPMs

- Outcome:

$$\hat{\theta}_p = \left[\, \hat{\underline{p}}, \hat{\overline{p}}, \hat{\mu}, \hat{m}_2, \hat{m}_3, \hat{m}_4 \,\right]$$

$$\hat{\mu}_{y(x)} = h_\mu(\hat{\mu}, x),$$

$$\hat{m}_{2,y(x)} = h_{m_2}(\hat{\mu}, \hat{m}_2, x),$$

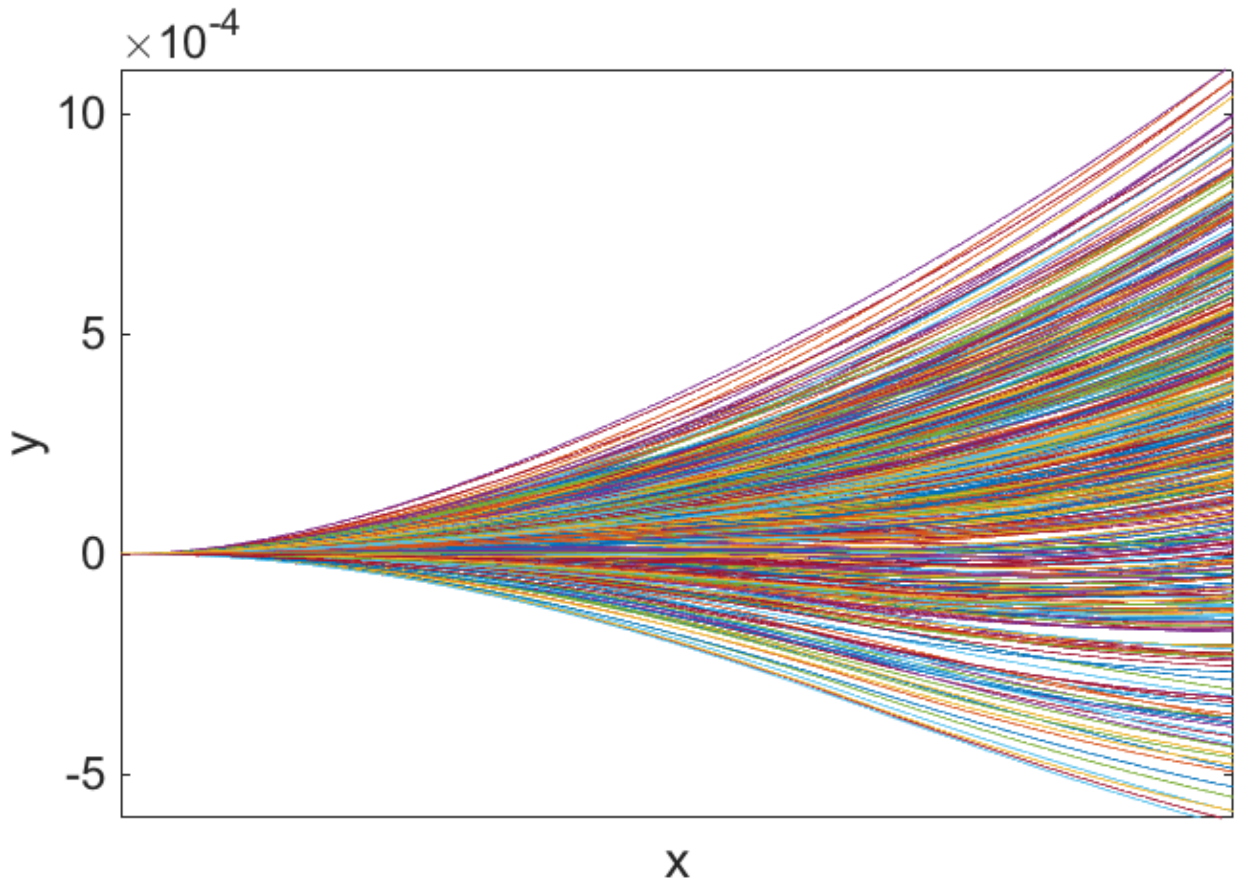
$$\hat{m}_{3,y(x)} = h_{m_3}(\hat{\mu}, \hat{m}_2, \hat{m}_3, x),$$

$$\hat{m}_{4,y(x)} = h_{m_4}(\hat{\mu}, \hat{m}_2, \hat{m}_3, \hat{m}_4, x).$$

- Advantage: approach is distribution-free: no need to assume a distribution for *p* upfront
- Setting a particular uncertainty model: use staircase variables to realize the optimal moments

Moment-Matching Example

- Goal: to characterize the unknown loading of a cantilever beam from displacement measurements
- A datum in the sequence is a set of measurements



Moment-Matching Example

- Basis chosen from Euler-beam theory: $y = p^\top \varphi(x)$

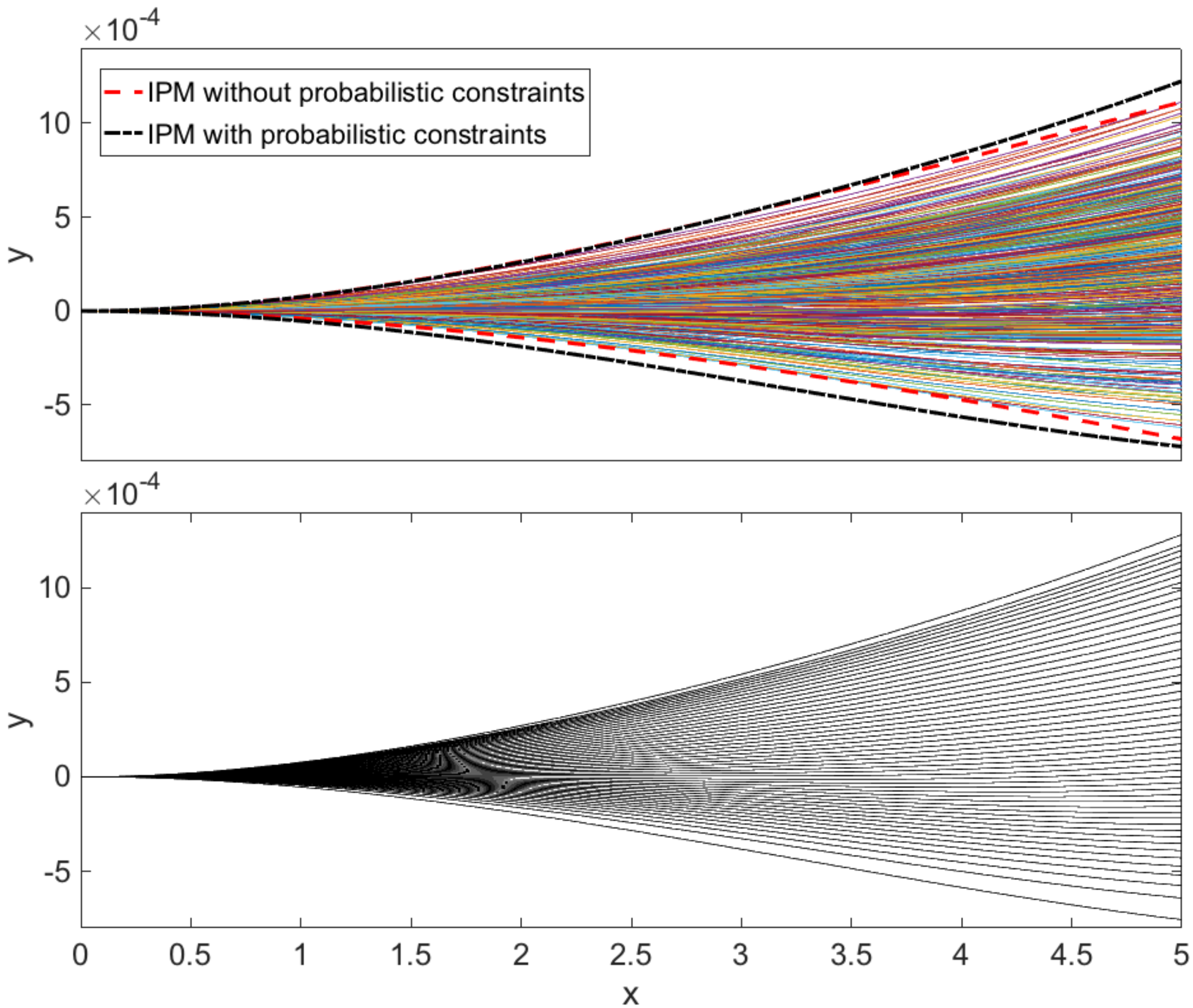
$$\varphi_{\text{force}}(x) = \begin{cases} \frac{x^2}{6EI}(3a - x) & \text{if } 0 \le x \le a, \\ \frac{a^2}{6EI}(3x - a) & \text{if } x \ge a \end{cases}$$

$$\varphi_{\text{moment}}(x) = \begin{cases} \frac{x^2}{2EI} & \text{if } 0 \le x \le a, \\ \frac{a}{2EI}(2x - a) & \text{if } x \ge a \end{cases}$$

$$\varphi(x)_{\text{uniform}} = \frac{x^2}{24EI}(x^2 + 6L^2 - 4Lx),$$

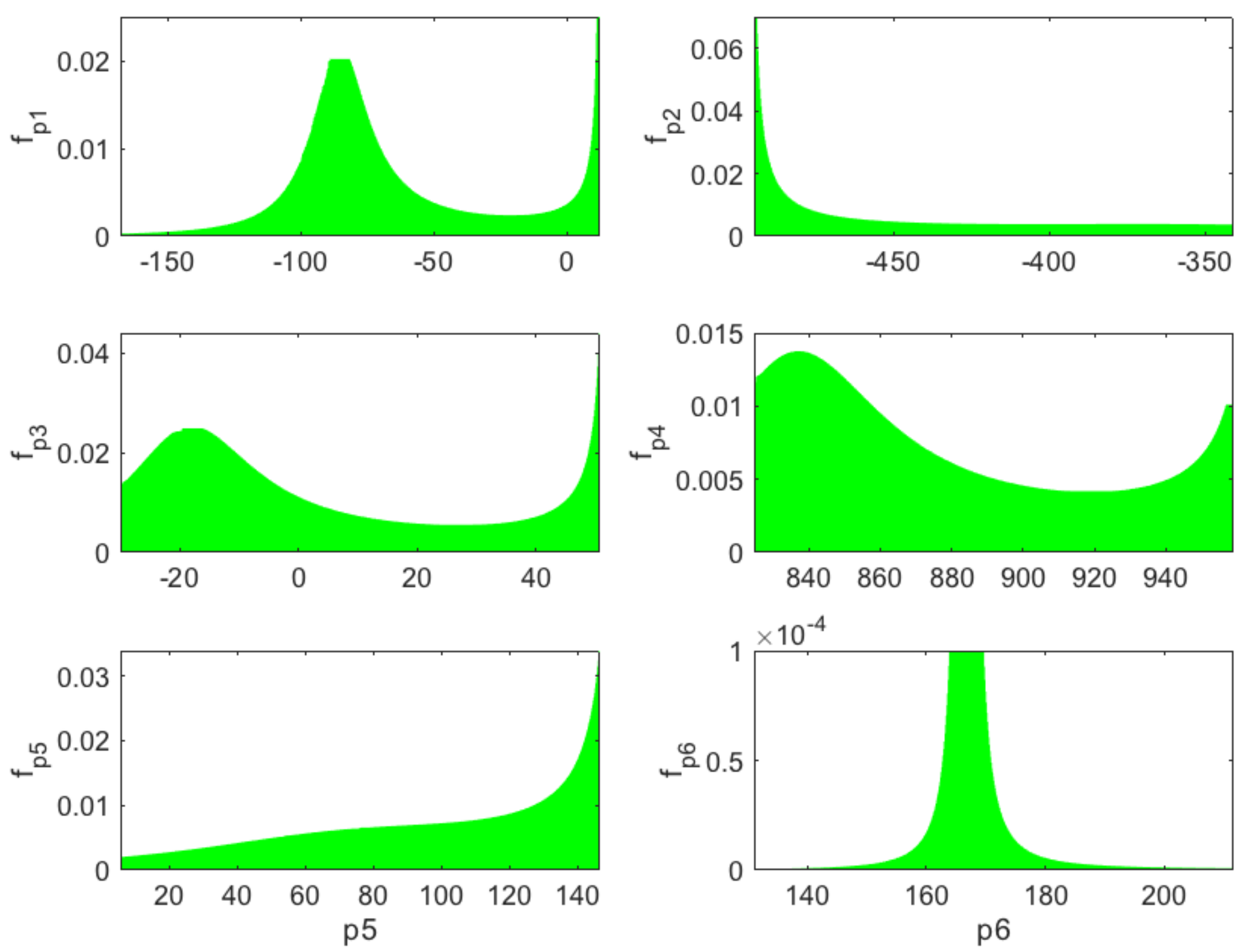
$$\varphi(x)_{\text{triangular increasing}} = \frac{x^3}{120EIL}(20L^3 - 10L^2x + x^3),$$

Moment-Matching Example



Skewed response

Moment-Matching Example



Minimal-Dispersion RPM

- Idea: find the moments of *p* leading to a prediction that concentrates the response as close as possible to the data while enclosing it into a high-probability region (trade-off)

- Solution approach: solve the optimization program

$$\min_{\theta_{p_1},\dots\theta_{p_{n_p}}} \left\{ \frac{\|c\|}{N} : g(\theta_{p_i}) \leq 0,\ y^{(j)} \in I_\alpha\left(x^{(j)}\right),\ i = 1,\dots n_p,\ j = 1,\dots N \right\}$$

where
$$c_j = \left(y^{(j)} - \mu_{y(x^{(j)})}\right)^2 + m_{2,\,y\left(x^{(j)}\right)}$$

and the high-probability region is

$$I_\alpha(x) = [y_\alpha(x), y_{1-\alpha}(x)]$$

Minimal-Dispersion RPM

- Same outcome and advantage as the previous approach
- When to use: unimodal DGM
- Challenge: characterizing I_{α} as a function of $\boldsymbol{\theta}$

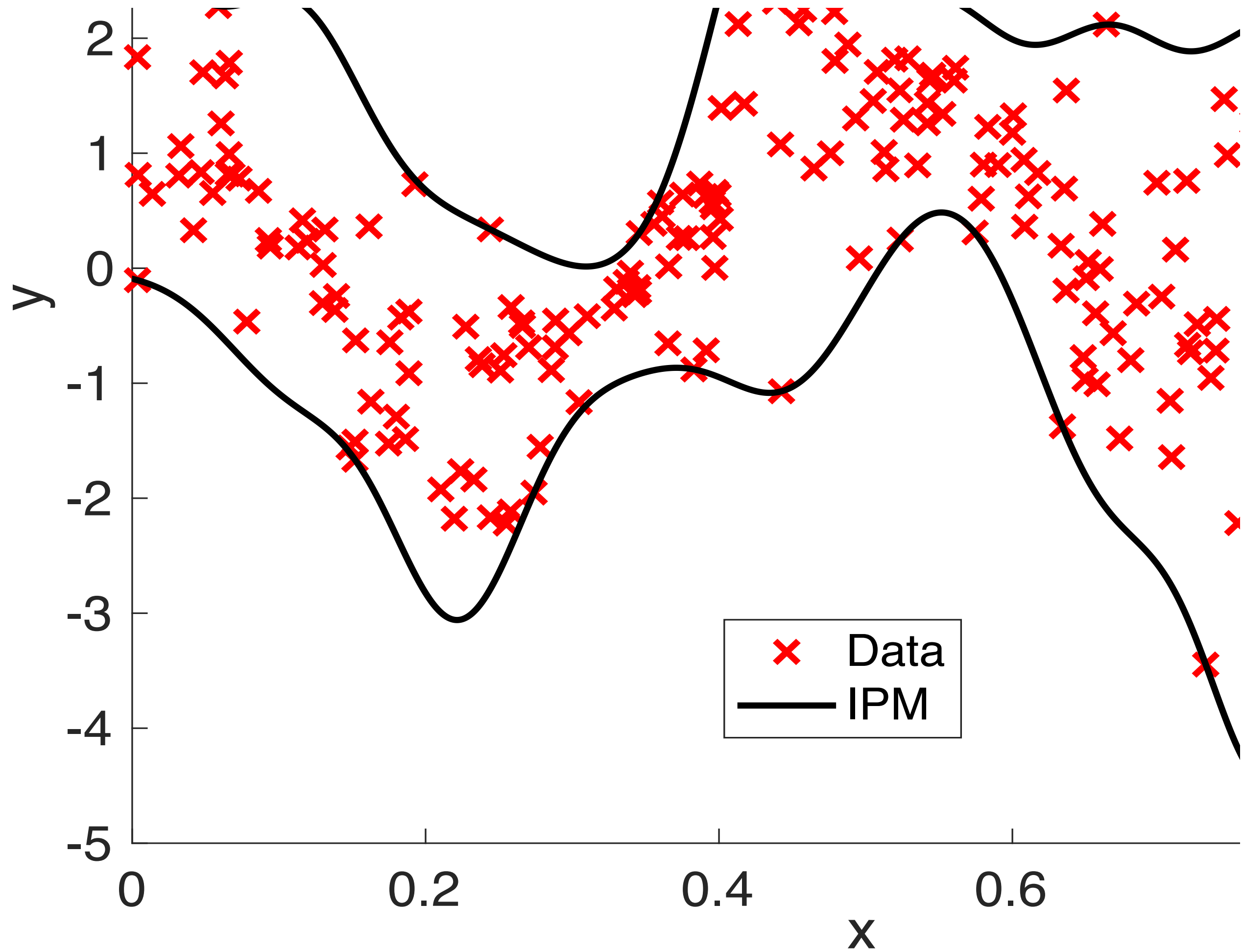
In the paper we use:

$$y_{\alpha}(x) = \mu_{y(x)} - n_1 \sqrt{m_{2,\, y(x)}} - n_2 \sqrt[3]{m_{3,\, y(x)}}$$
$$y_{1-\alpha}(x) = \mu_{y(x)} + n_1 \sqrt{m_{2,\, y(x)}} - n_2 \sqrt[3]{m_{3,\, y(x)}}$$

but a better I_{α} can be derived using regression/staircases

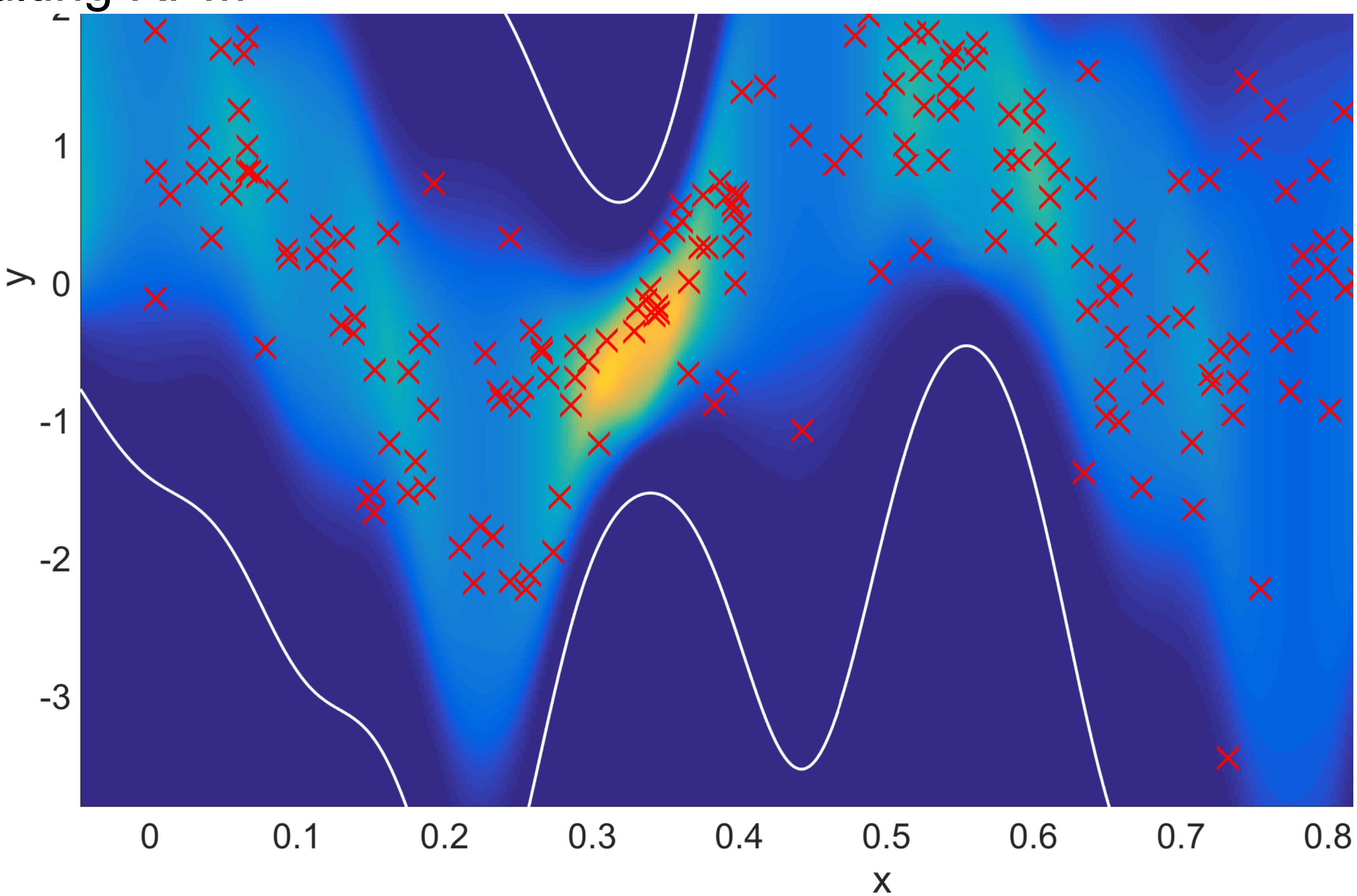
Minimal-Dispersion: Example

- Consider the data-cloud, and an arbitrary basis



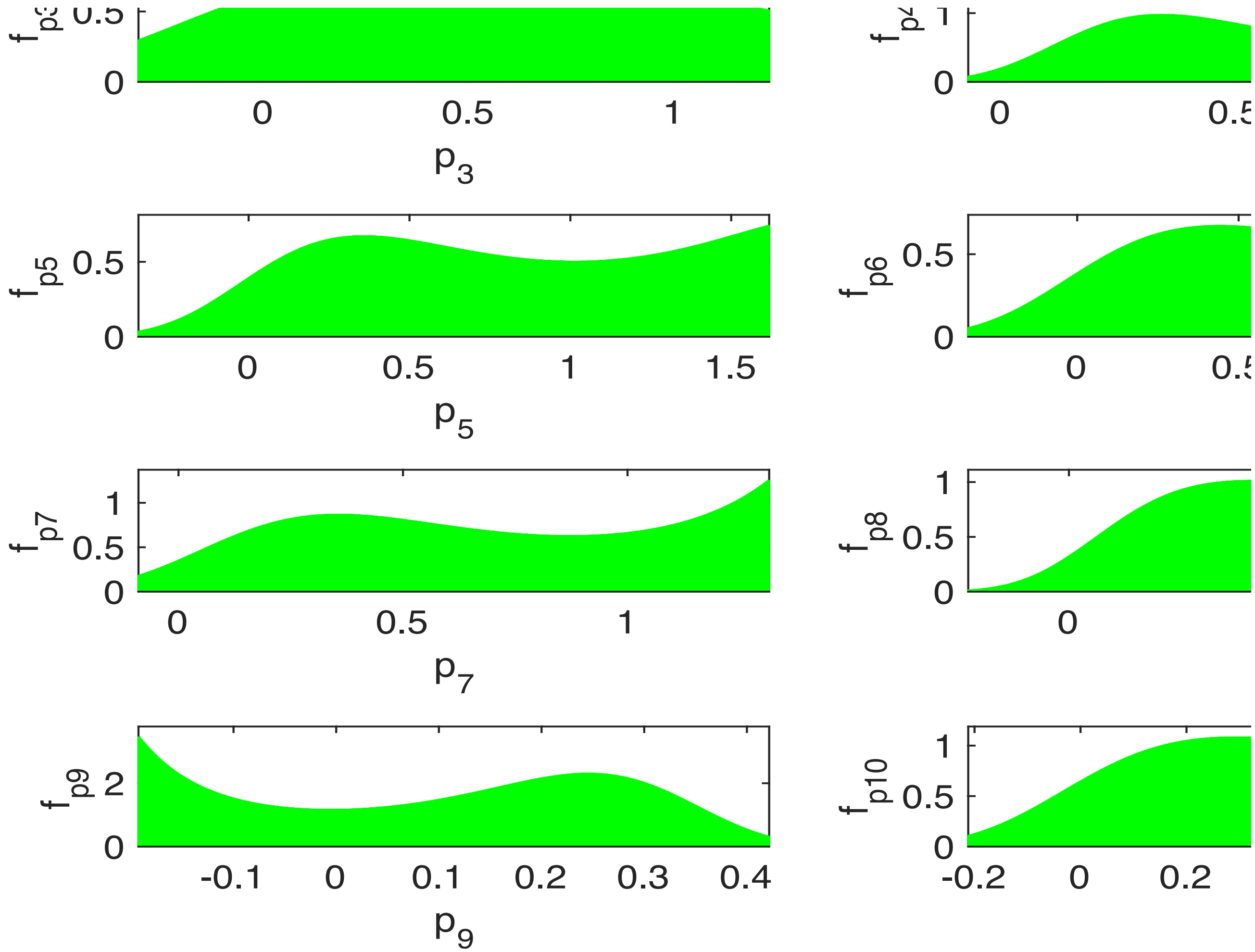
Minimal-Dispersion: Example

- Resulting RPM



Minimal-Dispersion: Example

- Distribution of the staircase parameters



Conclusions

- A framework for calibrating affine probabilistic models was developed
- Technique is moment-based and distribution-free
- Computational demands are considerably lower than maximum/likelihood based approaches
- Eliminates the need for assuming a distribution of the uncertainty upfront
- Analytical propagation of moments is possible when dependency is a known polynomial (we only did linear)

Conclusions

- Parameter dependencies can be accounted for (not done here, cumbersome)
- All sources of uncertainty and error are lumped into the resulting characterization of *p*...

Random Predictor Models with a Linear Staircase Structure

Luis G. Crespo, Sean P. Kenny, and Daniel P. Giesy

Dynamic Systems and Controls Branch
NASA Langley Research Center, Hampton, VA, 23666

Staircases

- Consider a random variable z with probability density function (PDF) $f_z : \Delta_z \subset \mathbb{R} \rightarrow \mathbb{R}^+$ and support set $\Delta_z = [z_{\min}, z_{\max}]$
- The central moments, defined as

$$m_r = \int_{\Delta_z} (z-\mu)^r f_z dz, \; r = 0, 1, 2, \ldots$$

are assumed to exist

- Goal: to calculate a random variable with a bounded support given values for the first four moments

$$\Delta_z \subseteq \Omega_z = [\underline{z}, \overline{z}] \qquad \theta_z = [\, \underline{z}, \overline{z}, \mu, m_2, m_3, m_4 \,]$$

θ-Feasibility

$$\theta_z = [\, \underline{z}, \overline{z}, \mu, m_2, m_3, m_4 \,]$$

- Does there exist a random variable that meets the constraints imposed by θ_z?
- Distribution-free vs. distribution fixed
- Such a random variable(s) exist if the set of polynomial constraints $g(\theta_z) \leq 0$ is satisfied

[1] - Sharma 2015, *ArXiv 1503.03786*; Kumar 2012

θ-Feasibility: equations

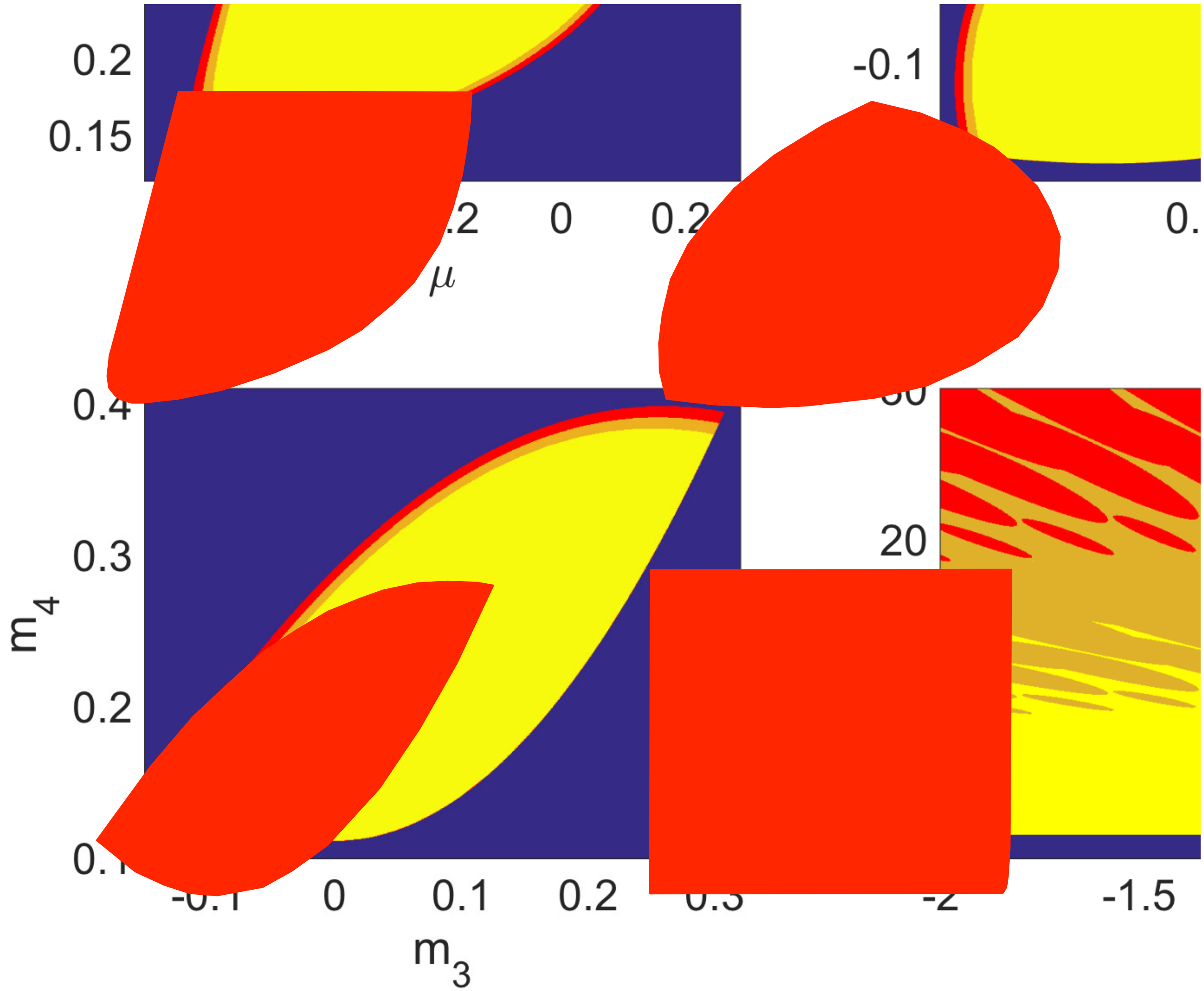
$$
\begin{aligned}
g_1 &= \underline{z} - \overline{z}, \\
g_2 &= \underline{z} - \mu, \\
g_3 &= \mu - \overline{z}, \\
g_4 &= -m_2, \\
g_5 &= m_2 - v \\
g_6 &= m_2^2 - m_2(\mu - \underline{z})^2 - m_3(\mu - \underline{z}), \\
g_7 &= m_3(\overline{z} - \mu) - m_2(\overline{z} - \mu)^2 + m_2^2, \\
g_8 &= 4m_2^3 + m_3^2 - m_2^2(\overline{z} - \underline{z})^2, \\
g_9 &= 6\sqrt{3}m_3 - (\overline{z} - \underline{z})^3, \\
g_{10} &= -6\sqrt{3}m_3 - (\overline{z} - \underline{z})^3, \\
g_{11} &= -m_4, \\
g_{12} &= 12m_4 - (\overline{z} - \underline{z})^4, \\
g_{13} &= (m_4 - vm_2 - um_3)(v - m_2) + (m_3 - um_2)^2, \\
g_{14} &= m_3^2 + m_2^3 - m_4 m_2,
\end{aligned}
$$

θ-Feasibility

- Feasible domain

$$\Theta = \{\theta : g(\theta) \leq 0\}$$

θ-Feasibility: intersections



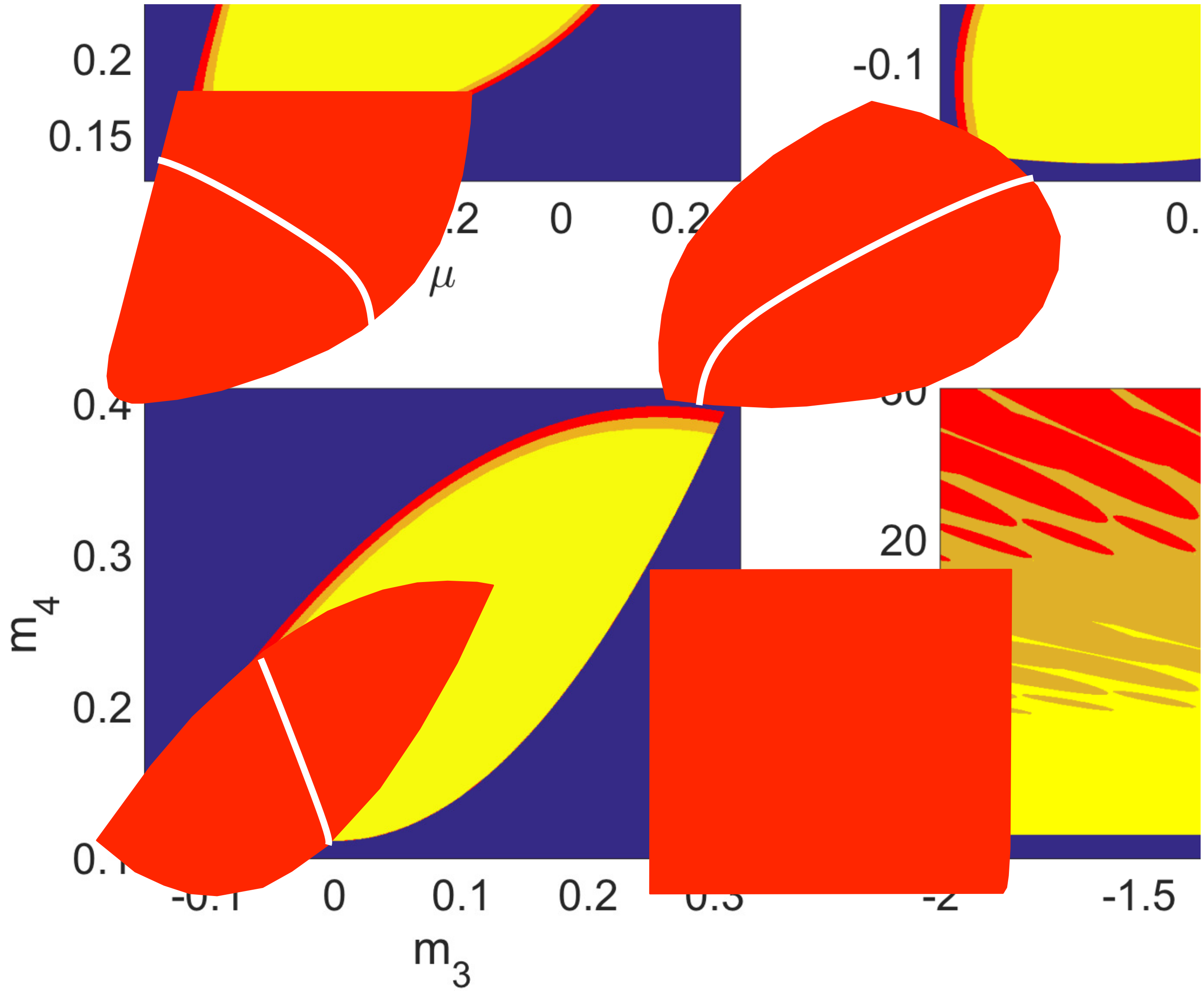
θ-Feasibility

- Feasible domain

$$\Theta = \{\theta : g(\theta) \leq 0\}$$

- This set is non-convex
- Standard random variables cannot realize most of Θ

θ-Feasibility: intersections



θ-Feasibility

- Feasible domain

$$\Theta = \{\theta : g(\theta) \leq 0\}$$

- This set is non-convex
- Standard random variables cannot realize most of Θ
- There might exist infinitely many random variables able to realize a feasible point

θ-Feasibility

- Feasible domain

$$\Theta = \{\theta : g(\theta) \leq 0\}$$

- This set is non-convex
- Standard random variables cannot realize most of Θ
- There might exist infinitely many random variables able to realize a feasible point
- How to construct a family of random variables that can realize most of Θ?

Staircase random variables

- Staircase variables have a piecewise constant PDF over a uniform partition of Ω_z : n_b bins
- The PDF of a staircase variable is given by

$$f_z(z,h) = \begin{cases} \ell_i & \forall z \in (z_i, z_{i+1}],\, i = 1, \ldots n_b \\ 0 & \text{otherwise}, \end{cases}$$

where ℓ is given by

Staircase random variables

$$\hat{\ell} = \arg\min_{\ell \geq 0} \left\{ J : \sum_{i=1}^{n_b} \int_{z_i}^{z_{i+1}} z \ell_i dz = \mu, \right.$$

$$\left. \sum_{i=1}^{n_b} \int_{z_i}^{z_{i+1}} \left(z - \mu\right)^r \ell_i dz = m_r,\; r = 0, 2, 3, 4 \right\}$$

- Cost to be defined later
- Hyper-parameter: $h = [\theta_z, n_b]$
- The above equation can be written as

$$\hat{\ell} = \arg\min_{\ell \geq 0} \{J(\theta, n_b) : A(\theta, n_b)\ell = b(\theta), \theta \in \Theta\}$$

Staircase random variables

- If the cost function is convex, calculating a staircase variable entails solving a convex optimization program: efficiently done for hundreds of thousands of constraints/design variables
- This optimization problem might be infeasible: distribution-fixed

Staircase variables: cost function

- Does not affect staircase-feasibility
- Three classes considered
  - Maximal entropy

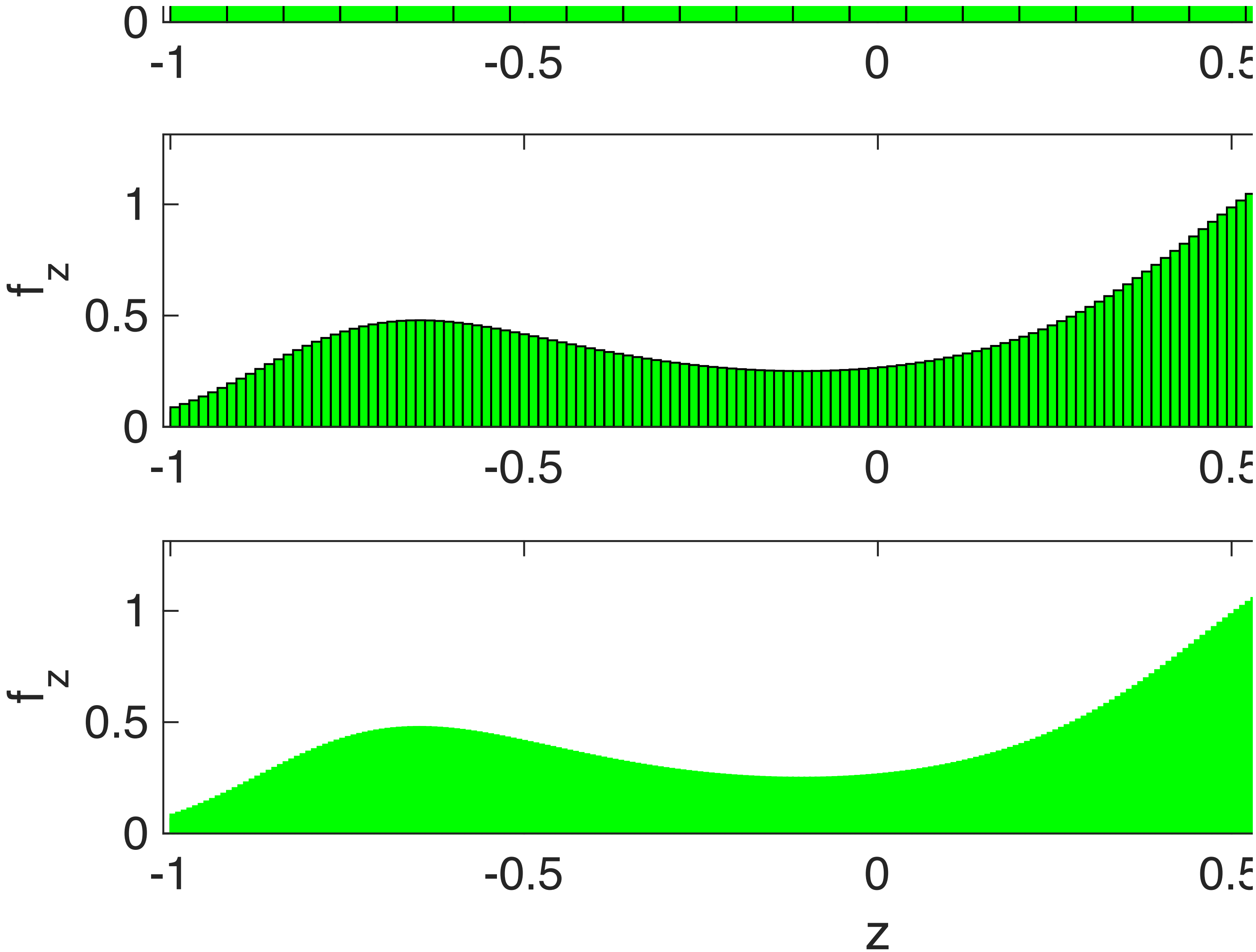
$$J(\ell) = -E(\ell) \triangleq \kappa \log(\ell)^\top \ell$$

  - Minimal squared likelihood
  - Optimal target matching

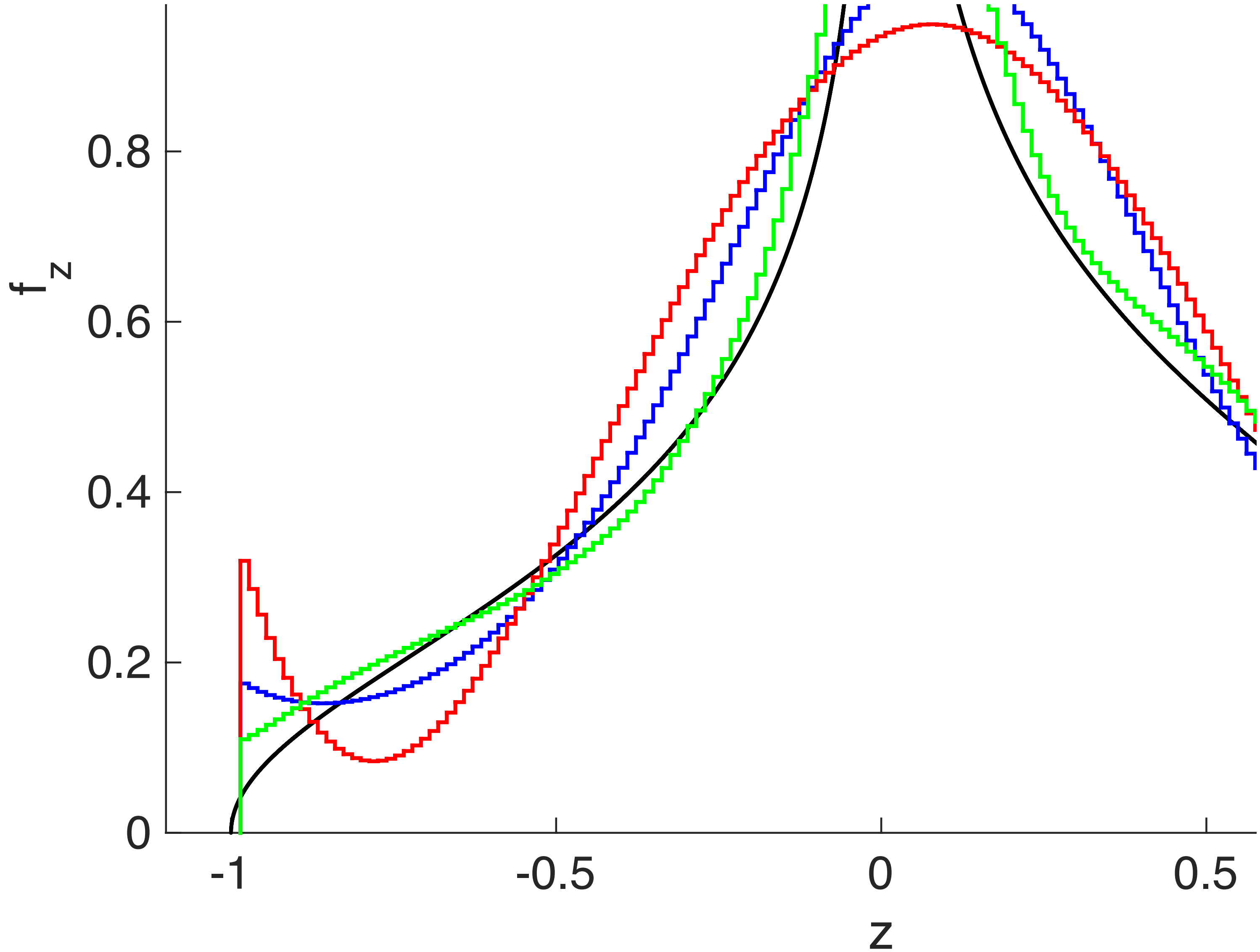
$$J(\ell) = H(\ell, Q, f) \triangleq \ell^\top Q \ell + f^\top \ell$$

- Other costs: max/min likelihood, min support, etc.
- Let's explore their structure and dependencies

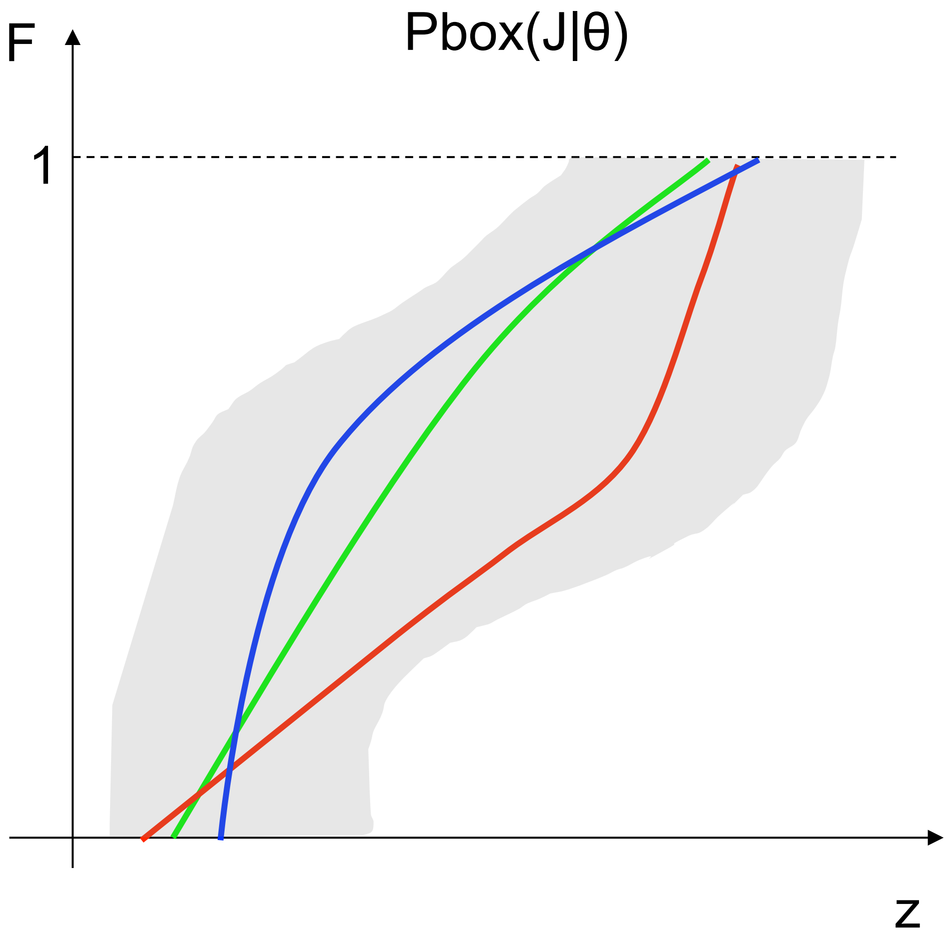
Staircase random variables: n_b



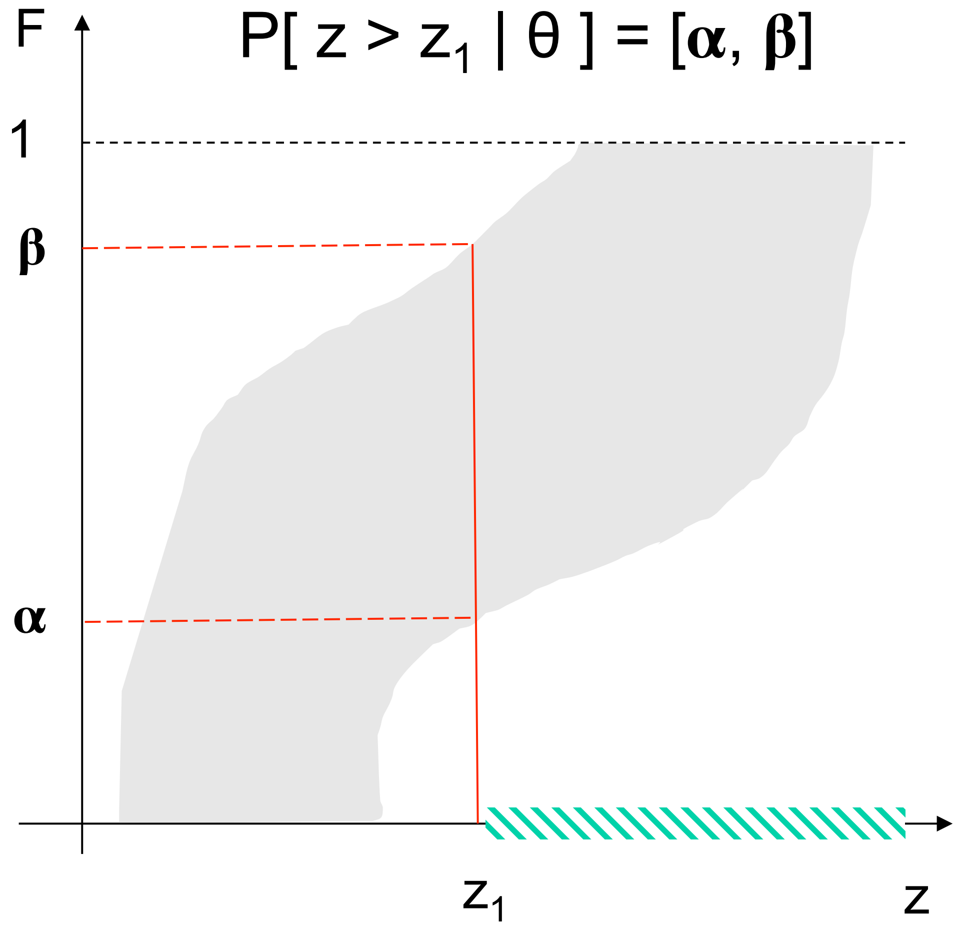
Staircase random variables: cost J



Staircase variables: worst-case variable



Staircase variables: worst-case PDF

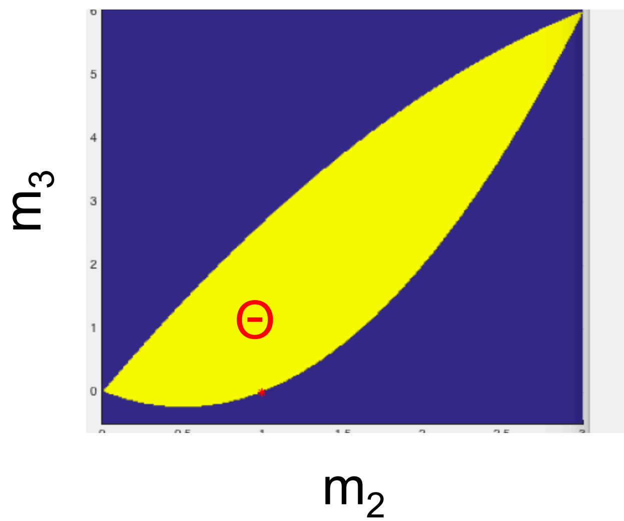


Staircase variables: feasibility

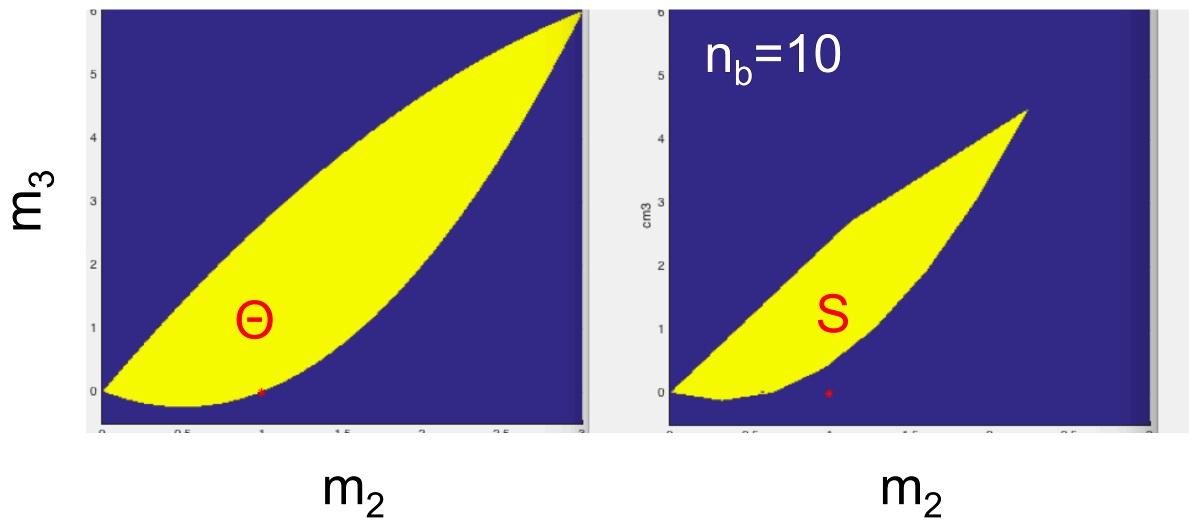
- The staircase feasible space is defined as

$$\mathcal{S}(n_b) = \{\theta : A(\theta, n_b)\ell = b(\theta),\ \ell \geq 0,\ \ \theta \in \Theta\}$$

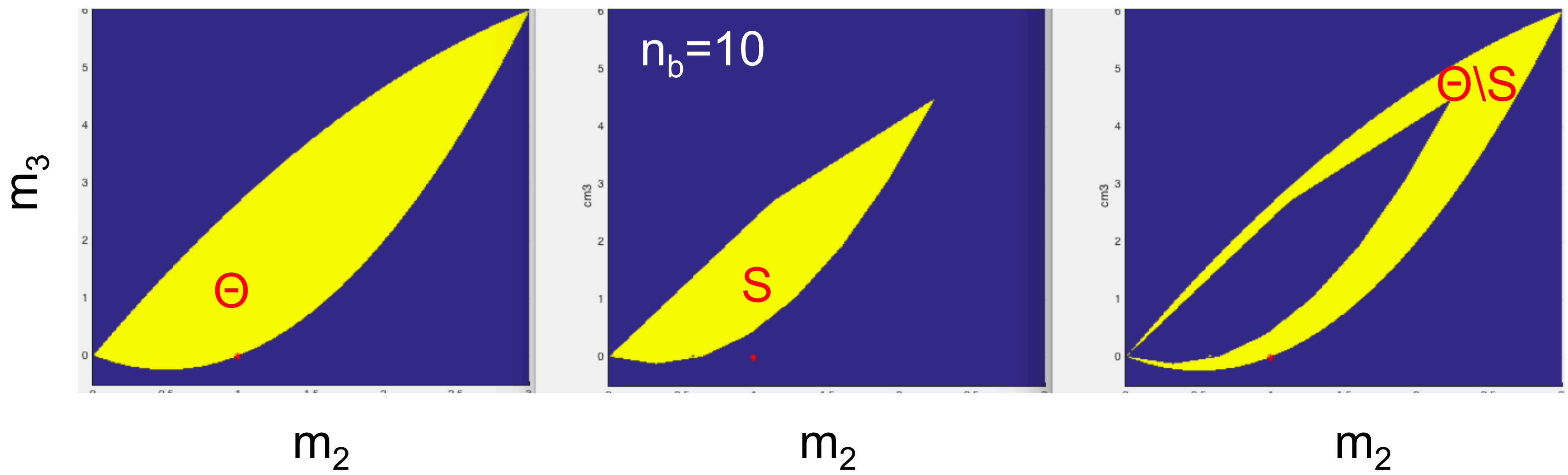
- How much of Θ can staircase variables represent?



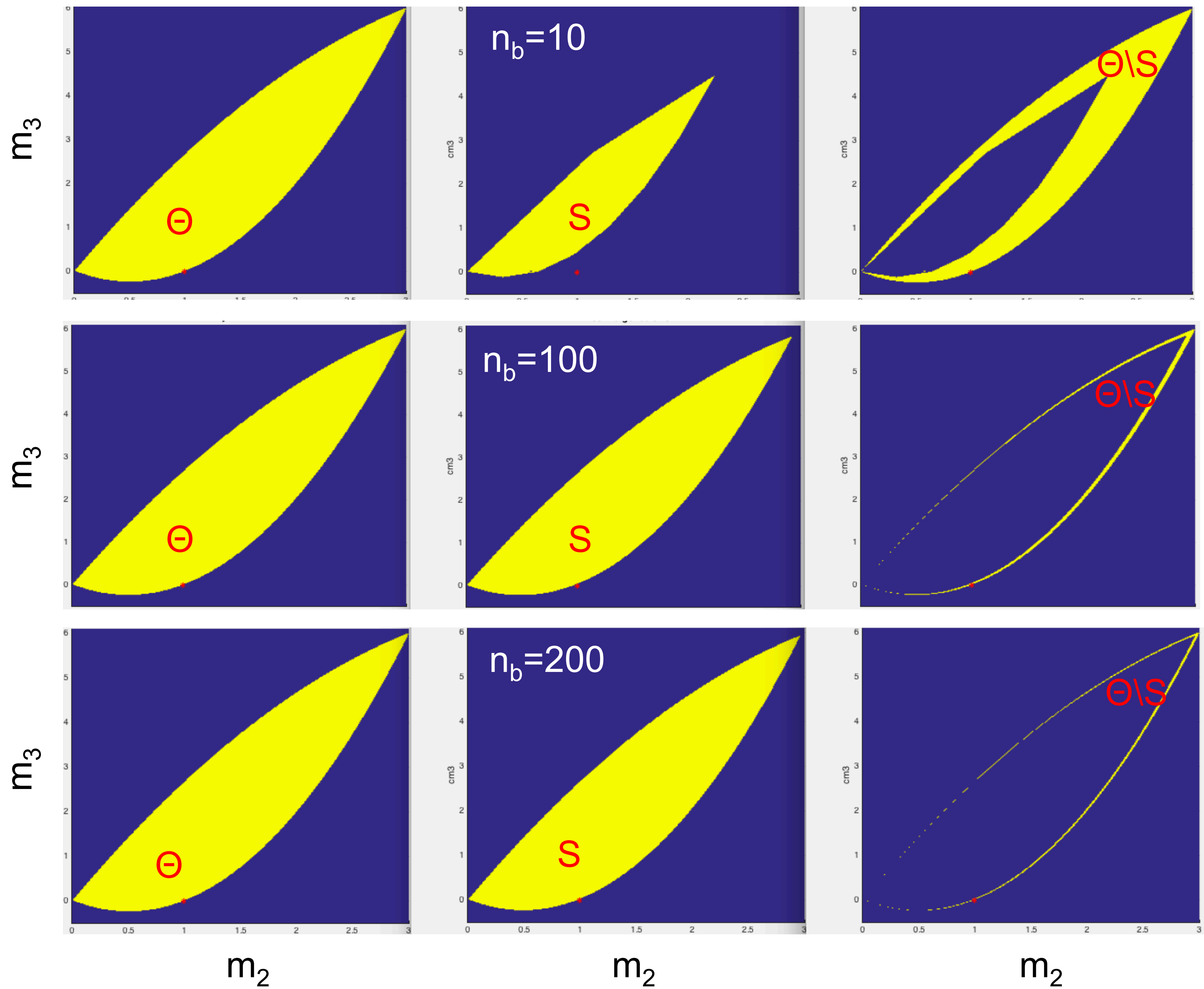
m3
m2
Θ



m3
Θ
m2
nb=10
cm3
S
m2

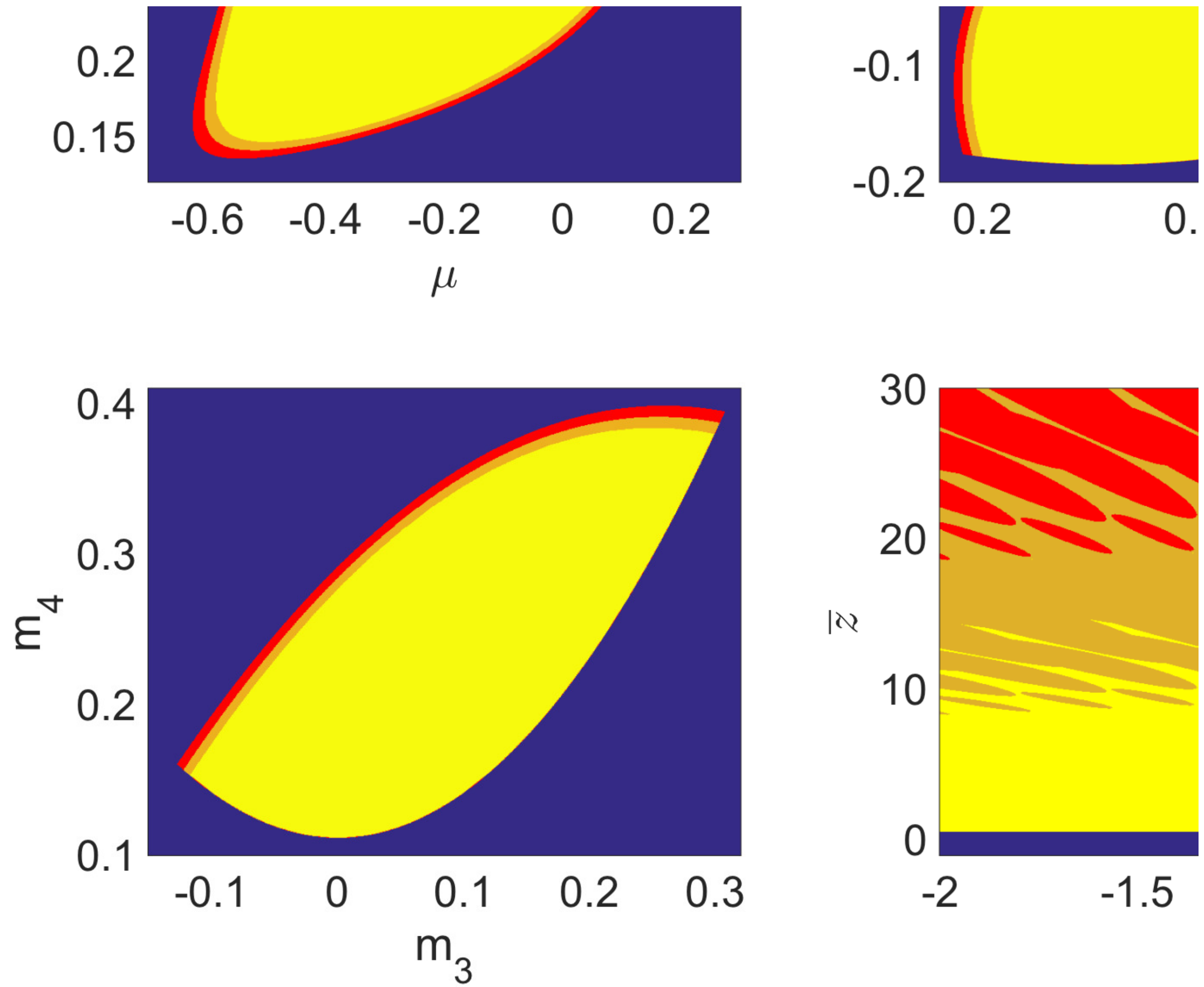


m_3
Θ
n_b=10
S
Θ\S
m_2
m_2
m_2

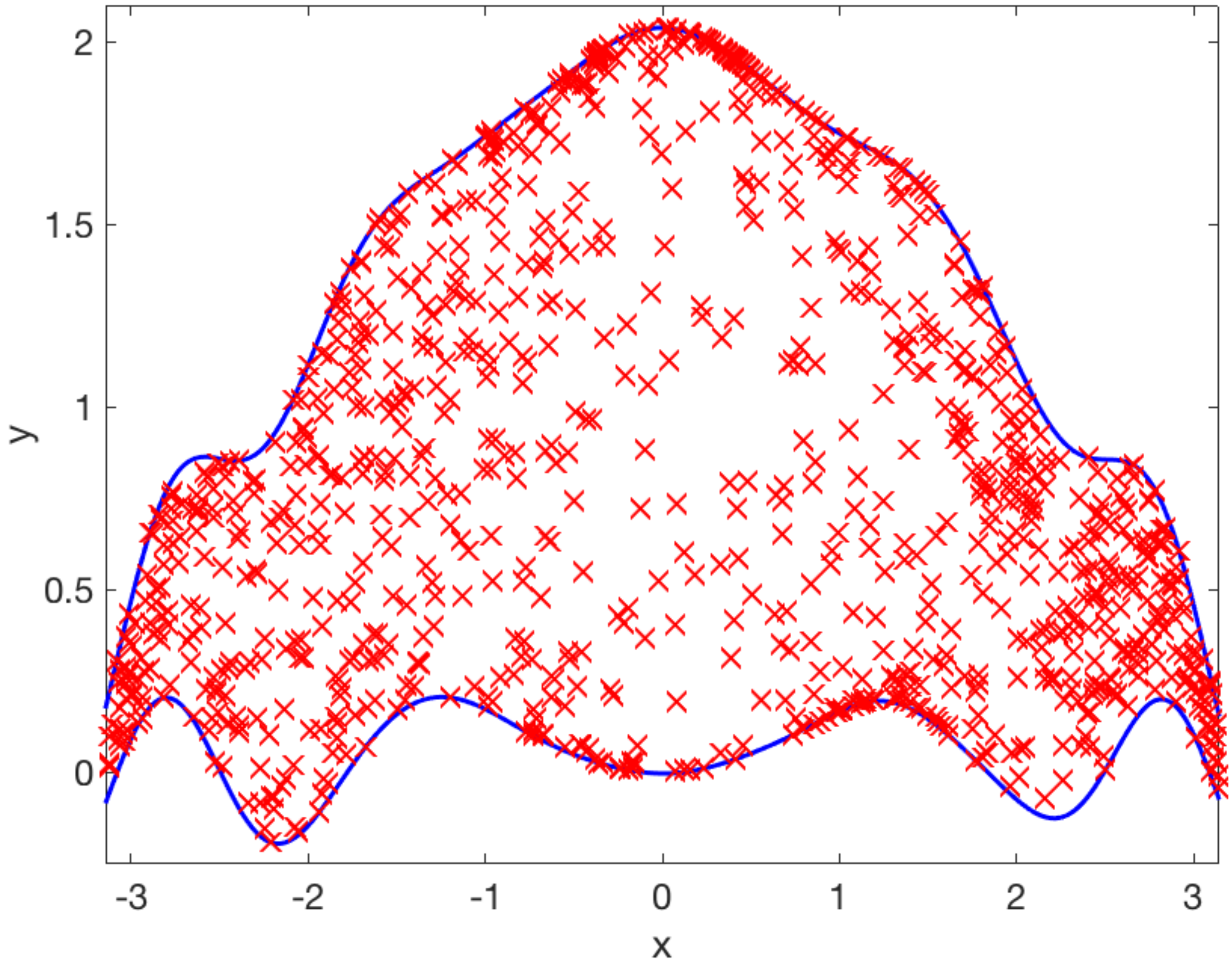


n_b=10
n_b=100
n_b=200
Θ
S
Θ\S
m_3
m_2
cm3

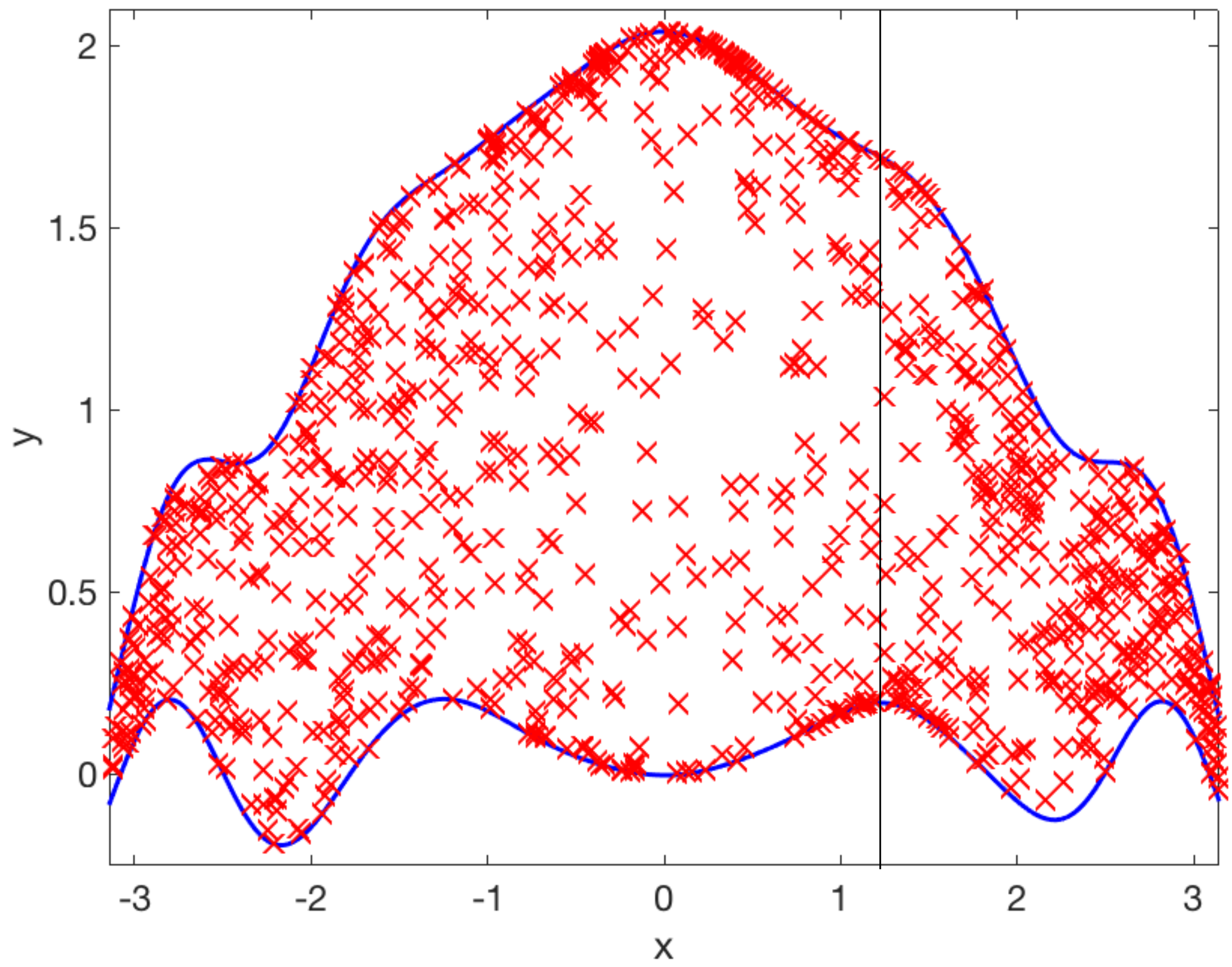
Staircase variables: feasibility



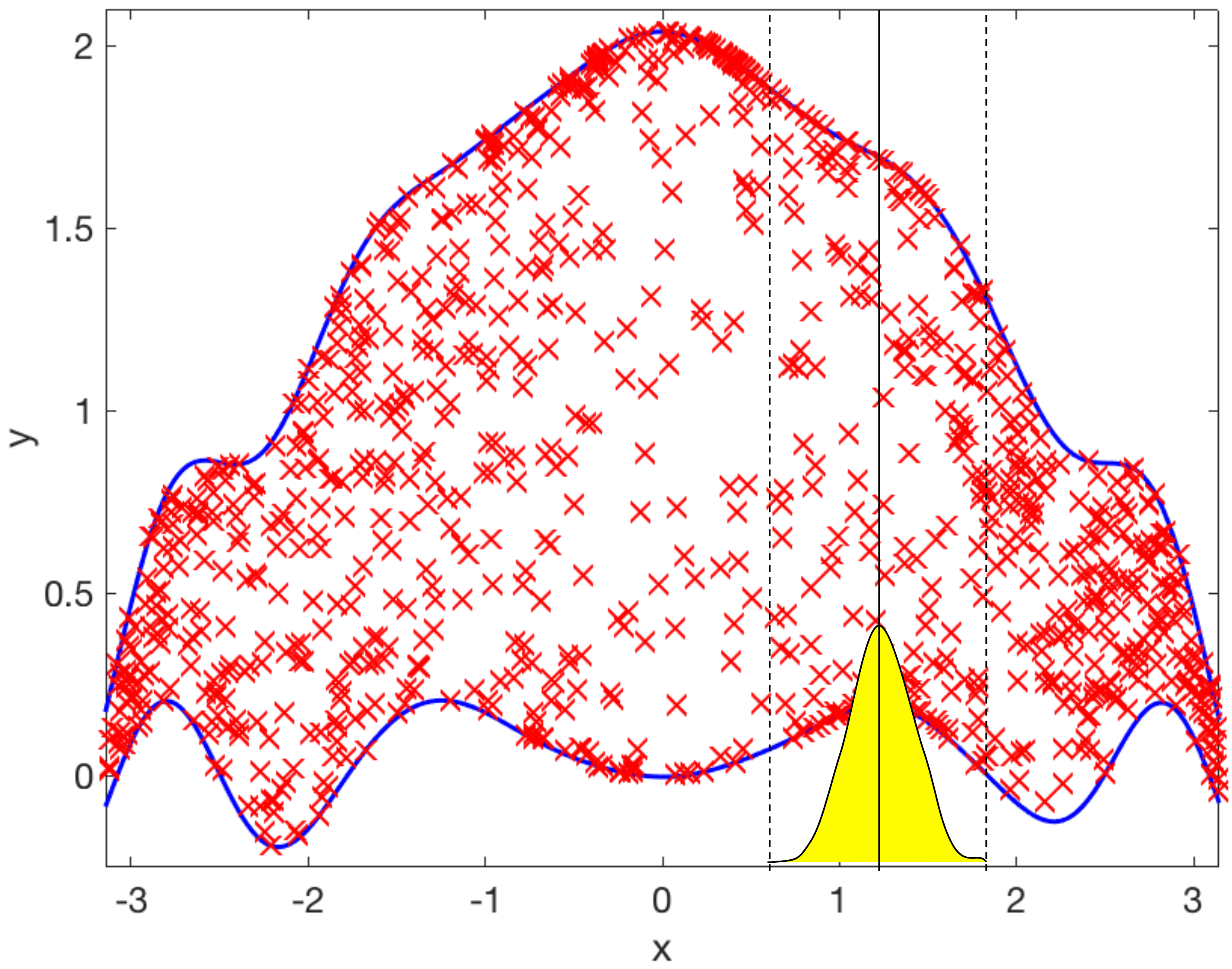
Setting Target Functions



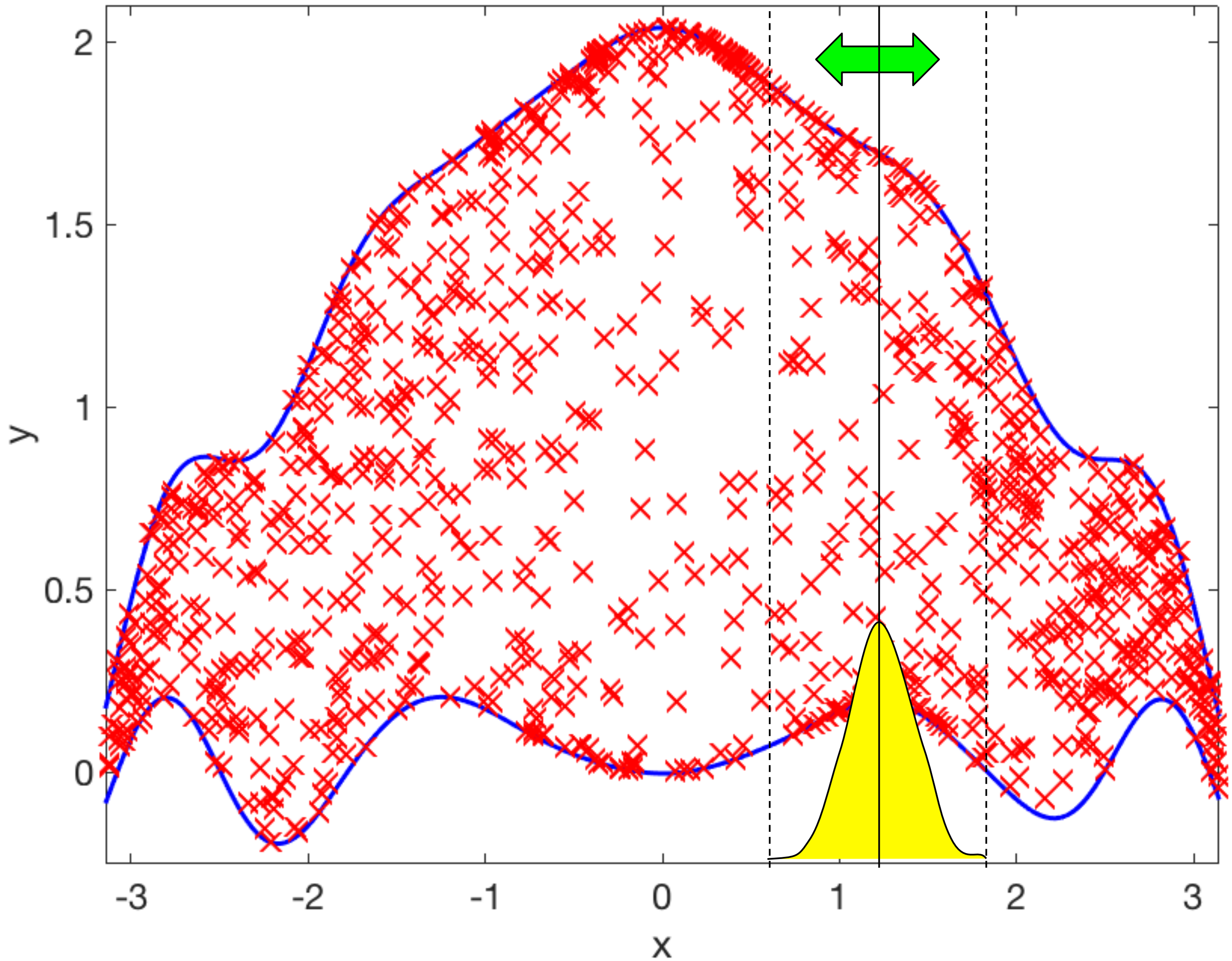
Setting Target Functions



Setting Target Functions



Setting Target Functions



True Moments vs. Approximation