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

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

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

\[
\{ \hat{p}(c), \; \tilde{p}(c) \} = \arg \min_{u, v: u \leq v} \left\{ \mathbb{E}_x [\delta_y(x, v, u)] : \right.
\]

\[
y(x^{(i)}, v, u) \leq y^{(i)} \leq \bar{y}(x^{(i)}, v, u),
\]

\[
c(u, v) \leq 0, \; i = 1, \ldots N \}
\]

Additional set of constraints
Interval Predictor Models

N=1K, n_p=10
Interval Predictor Models

N=1K

n_p=20
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
  \[ \theta_z = [ \underline{z}, \bar{z}, \mu, m_2, m_3, m_4 ] \]

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

\[ g_1 = \bar{z} - \bar{z}, \]
\[ g_2 = \bar{z} - \mu, \]
\[ g_3 = \mu - \bar{z}, \]
\[ g_4 = -m_2, \]
\[ g_5 = m_2 - u \]
\[ g_6 = m_2^2 - m_2(\mu - \bar{z})^2 - m_3(\mu - \bar{z}), \]
\[ g_7 = m_3(\bar{z} - \mu) - m_2(\bar{z} - \mu)^2 + m_2^2, \]
\[ g_8 = 4m_2^3 + m_3^2 - m_2^2(\bar{z} - \bar{z})^2, \]
\[ g_9 = 6\sqrt{3}m_3 - (\bar{z} - \bar{z})^3, \]
\[ g_{10} = -6\sqrt{3}m_3 - (\bar{z} - \bar{z})^3, \]
\[ g_{11} = -m_4, \]
\[ g_{12} = 12m_4 - (\bar{z} - \bar{z})^4, \]
\[ g_{13} = (m_4 - um_2 - um_3)(v - m_2) + (m_3 - um_2)^2, \]
\[ g_{14} = m_3^2 + m_2^3 - m_4m_2, \]

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 $\theta_z$
- 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

\[ y = p^\top \phi(x) \]
Random Predictor Models

- Taking the expected value of the model equation we have

\[ \mu_y(x) = \mathbb{E}_p [p]^T \varphi(x), \]
\[ \mathbb{E}_y [y^2] = \varphi^T(x) \mathbb{E}_p \left[ pp^T \right] \varphi(x), \]
\[ \mathbb{E}_y [y^3] = \varphi^T(x) \mathbb{E}_p \left[ pp^T \varphi(x) p^T \right] \varphi(x), \]
\[ \mathbb{E}_y [y^4] = \varphi^T(x) \mathbb{E}_p \left[ pp^T \varphi(x) \varphi(x)^T pp^T \right] \varphi(x). \]

which can be combined to obtain the moment functions

\[ \mu_y(x) = h_\mu (\mu, x), \]
\[ m_{2,y}(x) = h_{m2} (\mu, m_2, x), \]
\[ m_{3,y}(x) = h_{m3} (\mu, m_2, m_3, x), \]
\[ m_{4,y}(x) = h_{m4} (\mu, m_2, m_3, m_4, x). \]

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:

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

- The predicted moments, given by

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

depend upon the design variables:

$$\theta_p = [p, \bar{p}, \mu, m_2, m_3, m_4]$$
Moment-Matching RPMs

- **Solution Approach:** a sequence of optimization programs for moments of increasing order.

1. Solve for the mean

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

2. Find a feasible support set using IPMs

3. Solve for the variance

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

   for \( c_2 = g_{b}\bigg|_{z=p, \bar{z} = \bar{p}, \mu = \hat{\mu}} \) and \( b = \{4, 5\} \)

4. Find a feasible support set using IPMs….
Moment-Matching RPMs

• **Outcome:**

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

\[
\hat{\mu}_y(x) = h_\mu(\hat{\mu}, x),
\]

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

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

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

• **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: \( y = p^\top \varphi(x) \)

\[
\varphi_{\text{force}}(x) = \begin{cases} 
\frac{x^2}{6EI} (3a - x) & \text{if } 0 \leq x \leq a, \\
\frac{a^2}{6EI} (3x - a) & \text{if } x \geq a 
\end{cases}
\]

\[
\varphi_{\text{moment}}(x) = \begin{cases} 
\frac{x^2}{2EI} & \text{if } 0 \leq x \leq a, \\
\frac{a}{2EI} (2x - a) & \text{if } x \geq a 
\end{cases}
\]

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

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

![Graph showing skewed response with IPM with and without probabilistic constraints.]
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

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

  where

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

  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 $I_\alpha$ as a function of $\theta$

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 $I_\alpha$ 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 $\rho$...
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 $\Delta_z = [z_{\text{min}}, z_{\text{max}}]$

- The central moments, defined as

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

are assumed to exist

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

$$\Delta_z \subset \Omega_z = [\underline{z}, \overline{z}] \quad \theta_z = [\underline{z}, \overline{z}, \mu, m_2, m_3, m_4]$$
\( \theta_z = [ z, \bar{z}, \mu, m_2, m_3, m_4 ] \)

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

$\theta$-Feasibility: equations

\[ g_1 = z - \bar{z}, \]
\[ g_2 = \bar{z} - \mu, \]
\[ g_3 = \mu - \bar{z}, \]
\[ g_4 = -m_2, \]
\[ g_5 = m_2 - v, \]
\[ g_6 = m_2^2 - m_2(\mu - z)^2 - m_3(\mu - \bar{z}), \]
\[ g_7 = m_3(\bar{z} - \mu) - m_2(\bar{z} - \mu)^2 + m_2^2, \]
\[ g_8 = 4m_2^3 + m_3^2 - m_2^2(\bar{z} - z)^2, \]
\[ g_9 = 6\sqrt{3}m_3 - (\bar{z} - z)^3, \]
\[ g_{10} = -6\sqrt{3}m_3 - (\bar{z} - z)^3, \]
\[ g_{11} = -m_4, \]
\[ g_{12} = 12m_4 - (\bar{z} - z)^4, \]
\[ g_{13} = (m_4 - vm_2 - um_3)(v - m_2) + (m_3 - um_2)^2, \]
\[ g_{14} = m_3^2 + m_2^3 - m_4m_2, \]
$\Theta$-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
\( \theta \)-Feasibility

- Feasible domain

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

- This set is non-convex
- Standard random variables cannot realize most of \( \Theta \)
- There might exist infinitely many random variables able to realize a feasible point
\( \Theta \)-Feasibility

- Feasible domain

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

- This set is non-convex
- Standard random variables cannot realize most of \( \Theta \)
- 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 \( \Theta \)?
Staircase random variables

- Staircase variables have a piecewise constant PDF over a uniform partition of $\Omega_z : n_b$ 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 $\ell$ is given by
Staircase random variables

\[
\hat{\ell} = \arg \min_{\ell \geq 0} \left\{ J : \sum_{i=1}^{n_b} \int_{z_i}^{z_{i+1}} z \ell_i \, dz = \mu, \right. \\
\left. \sum_{i=1}^{n_b} \int_{z_i}^{z_{i+1}} (z - \mu)^r \ell_i \, dz = m_r, \ r = 0, 2, 3, 4 \right\}
\]

• 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
    \[ J(\ell) = -E(\ell) \equiv \kappa \log(\ell)^T \ell \]
  – Minimal squared likelihood
  – Optimal target matching
    \[ J(\ell) = H(\ell, Q, f) \equiv \ell^T Q \ell + f^T \ell \]
• 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

\[ P[z > z_1 | \theta] = [\alpha, \beta] \]
Staircase variables: feasibility

• The staircase feasible space is defined as

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

• How much of \( \Theta \) can staircase variables represent?
\( n_b = 10 \)

\( m_3 \)
\[ n_b = 10 \]
The images depict graphs with variables $m_3$, $m_2$, and $n_b$.

- For $n_b = 10$:
  - The $m_3$ graph shows a yellow region with the labels $\Theta$ and $S$.
  - The $m_2$ graph displays a similar pattern.
  - The $m_3$ graph for $n_b = 10$ has a different configuration.

- For $n_b = 100$:
  - The $m_3$ graph still shows the yellow region with $\Theta$ and $S$.
  - The $m_2$ graph maintains the same pattern.
  - The $m_3$ graph for $n_b = 100$ has a unique shape.

- For $n_b = 200$:
  - The $m_3$ graph continues to display the yellow region with $\Theta$ and $S$.
  - The $m_2$ graph keeps the same configuration.
  - The $m_3$ graph for $n_b = 200$ presents a distinct shape.

The graphs consistently show the variables $m_3$ and $m_2$ with $n_b$ as a parameter, indicating a study of how these variables change with different values of $n_b$. The yellow regions with $\Theta$ and $S$ labels suggest a focus on specific parts of the graph.
Staircase variables: feasibility
Setting Target Functions
Setting Target Functions
Setting Target Functions
Setting Target Functions
True Moments vs. Approximation