Multi-Stage Scenario Tree Generation via Statistical Property Matching

Bruno A. Calfa, Ignacio E. Grossmann
Department of Chemical Engineering
Carnegie Mellon University
Pittsburgh, PA 15213

Anshul Agarwal, John M. Wassick, Scott J. Bury
The Dow Chemical Company
Midland, MI

March 12th, 2014

Motivation

- **Uncertainty in Optimization**
  - In reality, the decision-making process of an enterprise involves multiple sources of uncertainty
    - Product demand, selling price
    - Production yield
    - Unplanned plant shutdown
  - Stochastic Programming with Recourse is a powerful modeling framework to explicitly account for uncertainty and provide corrective actions
  - Practical issues
    - Unknown "true" distributions (scenarios usually assumed to be known)
    - Need meaningful scenarios and probabilities
    - Use of historical data and forecasts (data-driven approach)
    - Quality of scenario tree ➔ quality of the decisions

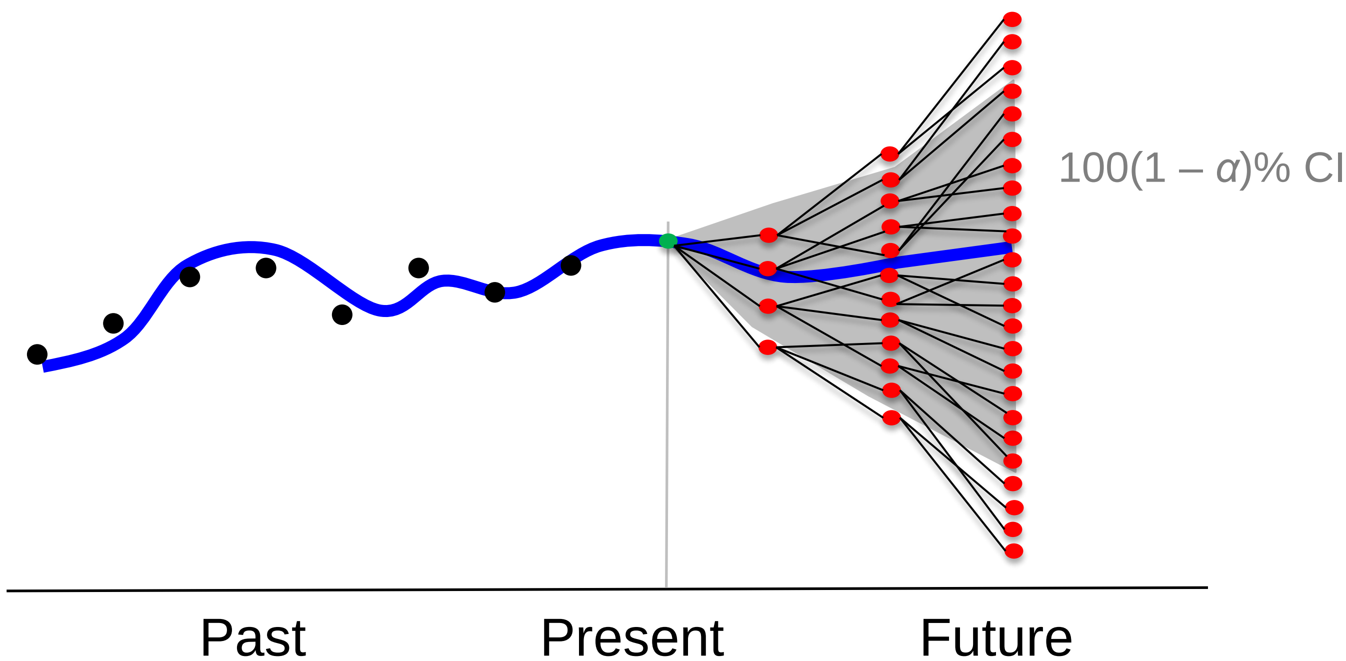
Problem Statement

- **Given**
  - Topology of the network of chemical plants
  - Deterministic multi-period production planning model with all its parameters (costs, capacities etc.)
  - Historical data and forecasts of product demand (uncertain parameter)

- **Objectives**
  - Generate multi-stage scenario tree from historical and forecast data
  - Formulate the multi-stage stochastic production planning model
  - Assess the quality of the stochastic solution with Monte Carlo simulation method

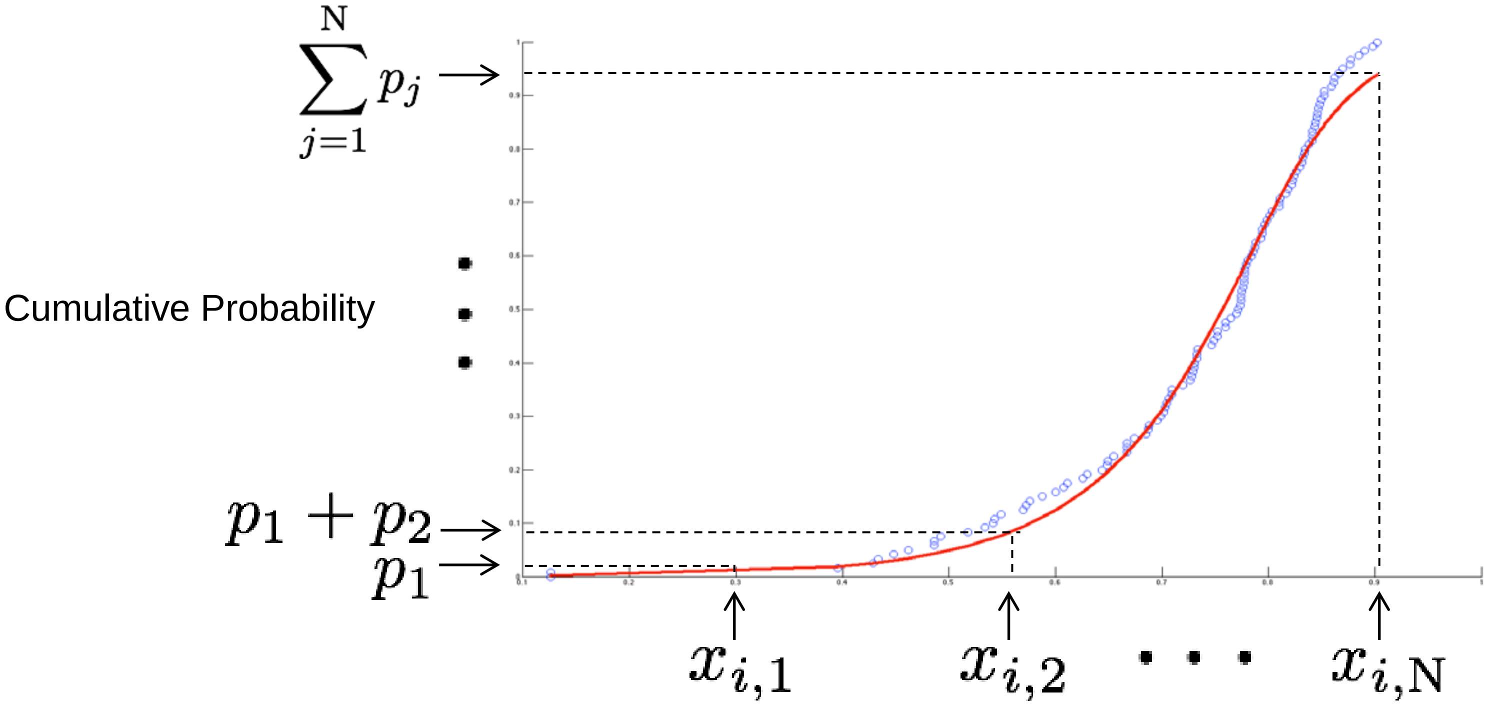
Multi-Stage Scenario Tree Generation

- **Stochastic Processes**
  - Random data that are autocorrelated in time
  - Scenario generation is aided by *time series forecasting*
  - Solve Distribution Matching Problem at each non-leaf node



New Contribution

- Only matching moments ➔ underdetermined problem
- Consequences:
  - Multiple combinations of demand and probability values
  - Probabilities may not capture the shape of the underlying distribution
- Additional information: marginal (Empirical) Cumulative Distribution



Distribution Matching Problem (DMP)

$$\min_{x,\, p} \quad z_{\text{DMP}}^{L^2} = \sum_{i \in I} \sum_{k \in K} \mathrm{w}_{i,k} (m_{i,k} - \mathrm{M}_{i,k})^2 + \sum_{\substack{(i,\, i') \in I \\ i < i'}} \mathrm{w}_{i,i'} (c_{i,i'} - \mathrm{C}_{i,i'})^2 + \sum_{i \in I} \sum_{j=1}^{\mathrm{N}} \omega_{i,j} \delta_{i,j}^2$$

Min weighted error between tree and data

s.t.

$$\sum_{j=1}^{\mathrm{N}} p_j = 1$$

Probabilities add up to 1

$$m_{i,1} = \sum_{j=1}^{\mathrm{N}} x_{i,j} p_j \qquad \forall\, i \in I$$

$$m_{i,k} = \sum_{j=1}^{\mathrm{N}} (x_{i,j} - m_{i,1})^k p_j \qquad \forall\, i \in I,\ k > 1$$

Moments calculated from the tree

$$c_{i,i'} = \sum_{j=1}^{\mathrm{N}} (x_{i,j} - m_{i,1})(x_{i',j} - m_{i',1}) p_j \qquad \forall\, (i,\, i') \in I,\ i < i'$$

Covariances calculated from the tree

$$x_{i,j} \in [\mathrm{x}_{i,j}^{\mathrm{LB}},\ \mathrm{x}_{i,j}^{\mathrm{UB}}] \qquad \forall\, i \in I,\ j = 1, \ldots, \mathrm{N}$$

$$p_j \in [0,\ 1] \qquad \forall\, j = 1, \ldots, \mathrm{N}$$

Bounds on variables

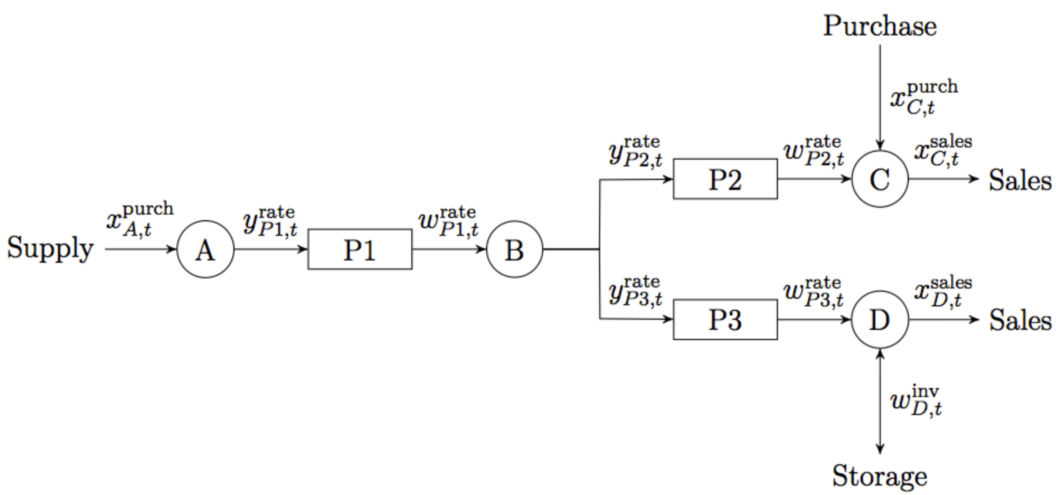
$$\widehat{ECDF}(x_{i,j}) - \sum_{j'=1}^{j} p_{j'} = \delta_{i,j} \qquad \forall\, i \in I,\ j = 1, \ldots, \mathrm{N}$$

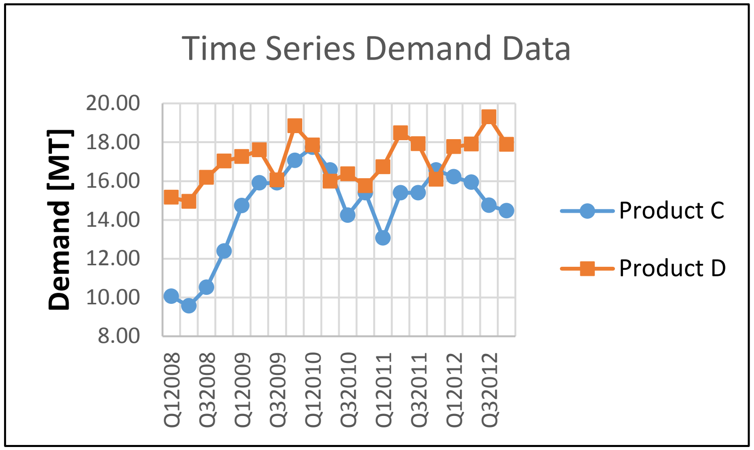
$$x_{i,j} \leq x_{i,j+1} \qquad \forall\, i \in I,\ j = 1, \ldots, \mathrm{N} - 1$$

ECDF information

Motivating Example

- Network of chemical plants





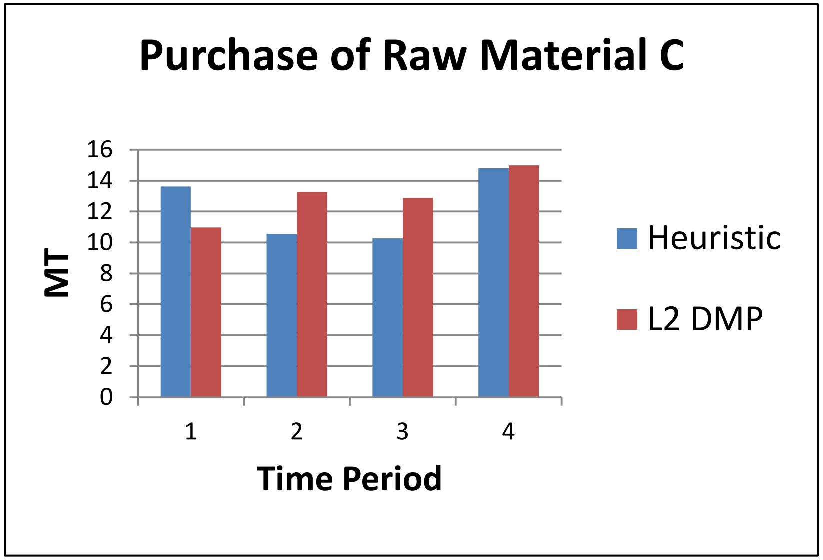
- 1 raw material (A), 1 intermediate product (B), two finished products (C and D), 1 site
- Only D can be stored and C can be purchased from elsewhere (may simulate inter-site transfers)

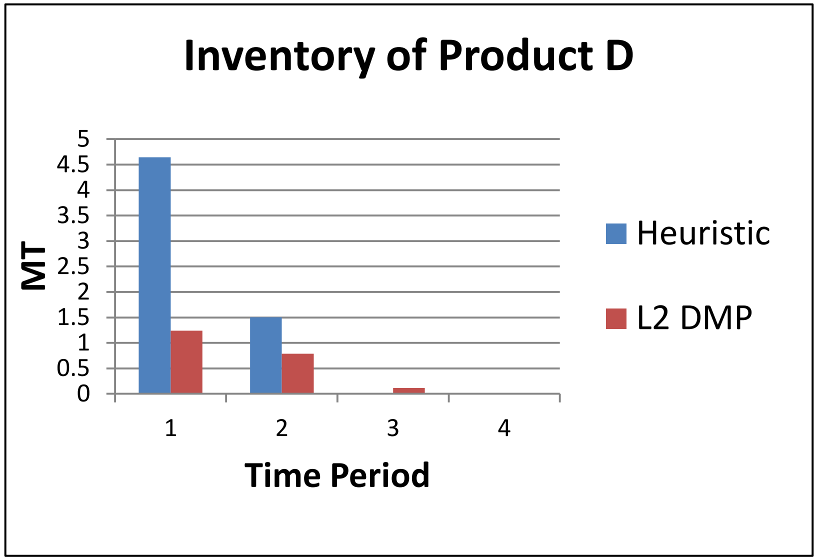
Results and Discussion

- Expected profits [$]

Heuristic	L^2 DMP
79.95	82.39

- Production plan variables





- The choice of the scenario tree impacts the decisions
- DMP accounts for correlations between the product demands

Literature and Practical Implications

- Current research on **scenario generation**
  - Mostly addressed by the OR&MS community
  - Predominantly financial applications (portfolio optimization)
  - Key idea: data-driven approach, *i.e.* no or very few assumptions about distributions (nonparametric rather than parametric models)
- This work
  - Additional information to be matched: (E)CDF
  - Improved numerics (degeneracy) when solving DMP
- Implementation in an industrial setting
  - Take advantage of forecasting and data mining software suites (e.g., SAS)
    1. Collect plant and market data (sources: Global Insight, Chemical Markets Associates Inc. (CMAI), government sources etc.)
    2. Obtain *good* forecasts (crucial part, see Rey *et al.*, 2012)
    3. Solve DMP (using information from step 2.) and generate tree
    4. Solve stochastic optimization model
  - Takeaway: poor quality data (scenario trees) yield poor quality decisions

References

- Bayraksan, G., & Morton, D. P. (2006). *Assessing Solution Quality in Stochastic Programs*. **Mathematical Programming**, 108(2-3), 495–514.
- Chiralaksanakul, A., & Morton, D. P. (2004). *Assessing Policy Quality in Multi-stage Stochastic Programming*. **Stochastic Programming E-Print Series**, 2004(12), 1–36. Retrieved from http://edoc.hu-berlin.de/docviews/abstract.php?id=26770.
- Høyland, K., & Wallace, S. W. (2001). *Generating Scenario Trees for Multistage Decision Problems*. **Management Science**, *47*(2), 295–307.
- Ji, X., Zhu, S., Wang, S., & Zhang, S. (2005). *A stochastic linear goal programming approach to multistage portfolio management based on scenario generation via linear programming*. **IIE Transactions**, 37(10), 957–969.
- Kaut, M. (2003). *Scenario Tree Generation for Stochastic Programming: Cases from Finance*. Ph.D. Thesis. Norwegian University of Science and Technology.
- Rey, T., Kordon, A., & Wells, C. (2012). *Applied Data Mining for Forecasting Using SAS®*. SAS Institute, Inc.