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

September 4th – 5th, 2013

Motivation (I)

- **Uncertainty in Optimization**
  - In reality, the decision-making process of an enterprise involves multiple sources of uncertainty
    - Product demand
    - Production yield
    - Product selling price
    - Unplanned plant shutdown
  - Stochastic Programming with Recourse is a powerful modeling framework to explicitly account for uncertainty and provide corrective actions
  - For example: Two-Stage Stochastic Program (TSSP)

Motivation (II)

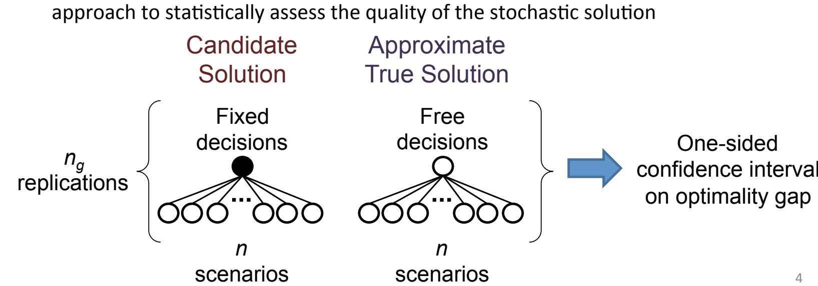
$$\begin{aligned} \max_{x} \quad & c^T x + \mathbb{E}_{\tilde{\xi}}\left[\max_{y(\tilde{\xi})} c(\tilde{\xi})^T y(\tilde{\xi})\right] \\ \text{s.t.} \quad & Ax = b \\ & T(\tilde{\xi})x + W y(\tilde{\xi}) = h(\tilde{\xi}) \\ & x \geq 0,\; y(\tilde{\xi}) \geq 0 \end{aligned}$$

Birge & Louveaux (2011)

- Some shortcomings of the above formulation:
  - In practice, the true distributions of $\tilde{\xi}$ are not known
  - Except for some trivial cases, it cannot be solved using continuous distributions
- Hence, we need to work with discretized distributions
- Due to limited computing power, approximate the distributions of the stochastic parameters with a limited number of outcomes => **scenario tree**

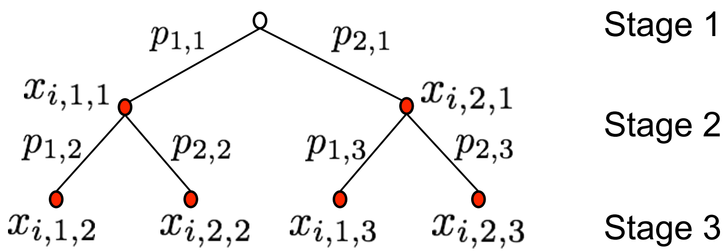
Quality of Scenario Tree

- The goal of the generated scenario tree is to *approximately* represent the real problem's uncertainties via discrete outcomes
- Therefore, the quality of the scenario tree is a case of GIGO (Garbage-In Garbage-Out), *i.e.* the solution of a SP based on a scenario tree that does not effectively represent the data in the real problem may be of little practical value
- Kaut (2003) claims there is not a universally good scenario tree generation method for all possible models
- Bayraksan & Morton (2006) propose a simulation-based Monte Carlo sampling approach to statistically assess the quality of the stochastic solution



Scenario Generation via Optimization

- Decision variables in an optimization formulation
  - Probabilities of outcomes
  - Values of outcomes



Moment Matching Problem (MMP)

- Determine *x* and *p* vectors such that the moments calculated from the tree match those estimated from the data (Høyland & Wallace, 2001)

$$\min_{x,\,p} \quad z^{L^2}_{\text{MMP}} = \sum_{i\in I}\sum_{k\in K} \mathrm{w}_{i,k}(m_{i,k} - \mathrm{M}_{i,k})^2 + \sum_{\substack{(i,\,i')\in I\\ i<i'}} \mathrm{w}_{i,i'}(c_{i,i'} - \mathrm{C}_{i,i'})^2 \qquad (2a)$$

Min weighted error between tree and data

s.t.

$$\sum_{j=1}^{\mathrm{N}} p_j = 1 \qquad (2b)$$

Probabilities add up to 1

$$m_{i,1} = \sum_{j=1}^{\mathrm{N}} x_{i,j}p_j \qquad \forall\, i \in I \qquad (2c)$$

$$m_{i,k} = \sum_{j=1}^{\mathrm{N}} (x_{i,j} - m_{i,1})^k p_j \qquad \forall\, i \in I,\ k > 1 \qquad (2d)$$

Moments calculated from the tree

$$c_{i,i'} = \sum_{j=1}^{\mathrm{N}} (x_{i,j} - m_{i,1})(x_{i',j} - m_{i',1})p_j \qquad \forall\,(i,\,i') \in I,\ i < i' \qquad (2e)$$

Covariances calculated from the tree

$$x_{i,j} \in [\mathrm{x}^{\mathrm{LB}}_{i,j},\ \mathrm{x}^{\mathrm{UB}}_{i,j}] \qquad \forall\, i \in I,\ j = 1, \ldots, \mathrm{N} \qquad (2f)$$

$$p_j \in [0,\ 1] \qquad \forall\, j = 1, \ldots, \mathrm{N} \qquad (2g)$$

Bounds on variables

Distribution Matching Problem (DMP) (I)

- Under-specification is common in the MMP
- Also, higher-order target moments may not be available or obtained with reasonable accuracy
- DMP: match moments and (Empirical) Cumulative Distribution Function information

$$\min_{x,\,p} \quad z^{L^2}_{\text{DMP}} = z^{L^2}_{\text{MMP}} + \sum_{i\in I}\sum_{j=1}^{\mathrm{N}} \omega_{i,j}\delta^2_{i,j} \qquad (8a)$$

Min weighted error between tree and data

s.t.

Constraints (2b) – (2g)

L² MMP constraints

$$\widehat{ECDF}(x_{i,j}) - \sum_{j'=1}^{j} p_{j'} = \delta_{i,j} \qquad \forall\, i \in I,\ j = 1, \ldots, \mathrm{N} \qquad (8b)$$

ECDF data

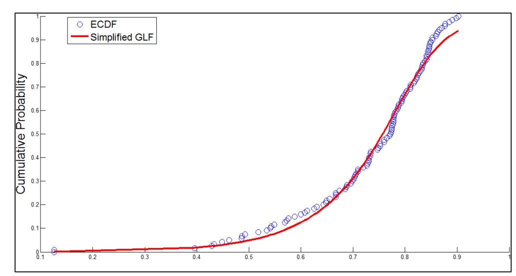
$$x_{i,j} \le x_{i,j+1} \qquad \forall\, i \in I,\ j = 1, \ldots, \mathrm{N} - 1 \qquad (8c)$$

Ordered nodes

The authors acknowledge Alex Kalos from Dow for the fruitful discussions with respect to considering (E)CDF information.

Distribution Matching Problem (DMP) (II)

- Approximation to CDF may be used if closed-form expression is not available or ECDF with smooth curve

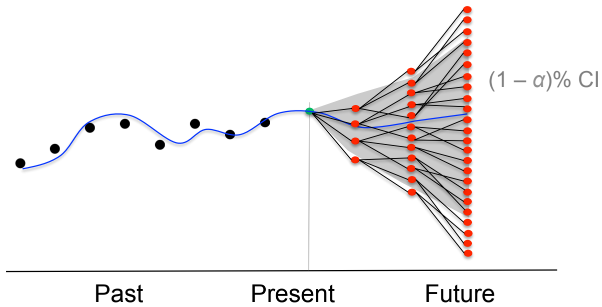


Generalized Logistic Function (GLF)

$$GLF(x) = \beta_0 + \frac{\beta_1 - \beta_0}{\left(1 + \beta_2 e^{-\beta_3 x}\right)^{1/\beta_4}}$$

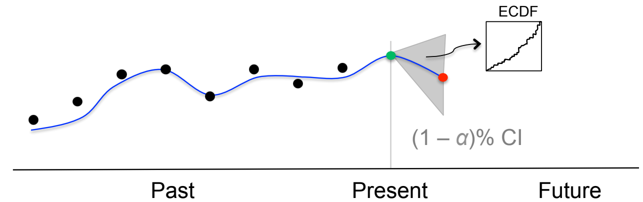
Multi-Stage Scenario Tree Generation (I)

- **Stochastic Processes**
  - Random data that are autocorrelated in time
  - DMP is aided by time series forecasting

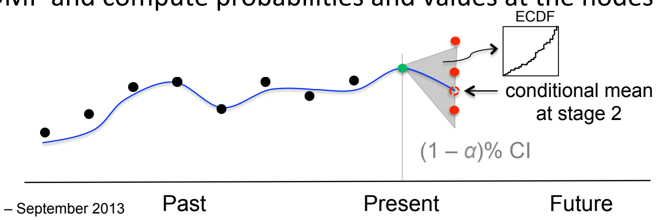


Multi-Stage Scenario Tree Generation (II)

- **Step 1:** Forecast conditional moments for next stage

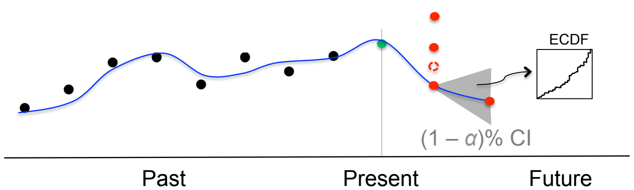


- **Step 2:** For a given number of nodes (outcomes), solve L^2 or L^1 DMP and compute probabilities and values at the nodes



Multi-Stage Scenario Tree Generation (III)

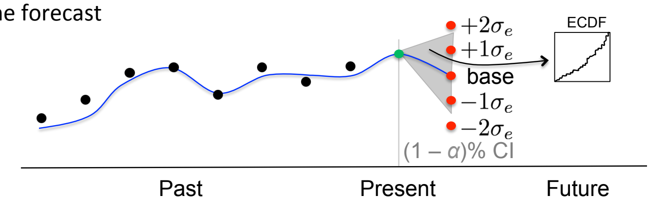
- **Step 3:** For each node generated in previous Step, forecast one-step ahead. For example, for the bottom node:



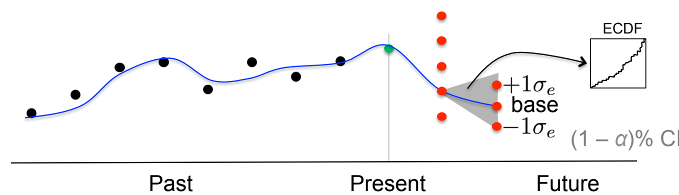
- **Step 4:** Repeat Steps 1 – 3 recursively for all nodes until the final stage is achieved
- Other published works:
  - Used expressions for updating estimation of moments (mean reversion, volatility clumping, Geometric Brownian Motion (GBM), other updating rules), in some cases fitted the node data to models

Multi-Stage Scenario Tree Generation (IV)

- **Step 1:** Forecast the base or most likely or expected value and create other nodes by adding and subtracting multiples of the standard error of the forecast

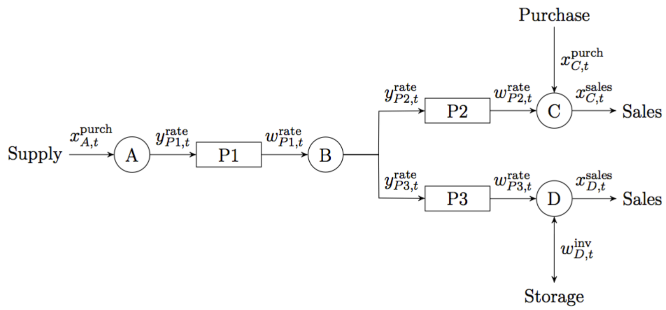


- **Step 2:** Repeat Step 1 for every node generated



Motivating Example: Network

- 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)

Motivating Example: Model

- Deterministic multi-period production planning LP model

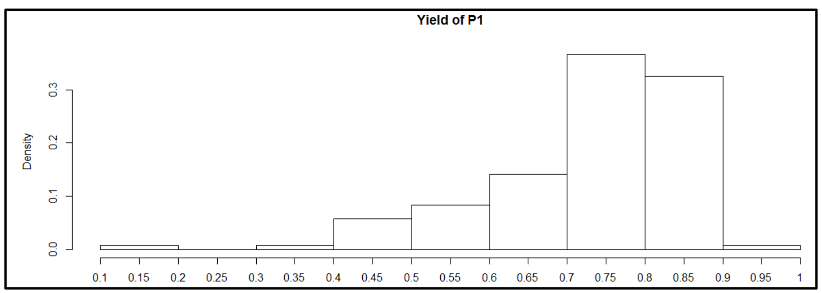
Production yield

$$
\begin{aligned}
\max\ & w^{\text{profit}} & \\
\text{s.t.}\ & w^{\text{rate}}_{f,t} = \theta_f y^{\text{rate}}_{f,t} & \forall f \in F,\ t \in T \\
& x^{\text{sales}}_{C,t} = w^{\text{rate}}_{P2,t} + x^{\text{purch}}_{C,t} & \forall t \in T \\
& w^{\text{inv}}_{D,t} = w^{\text{inv}}_{D,t-1} + w^{\text{rate}}_{P3,t} - x^{\text{sales}}_{D,t} & \forall t \in T \\
& w^{\text{rate}}_{P1,t} = y^{\text{rate}}_{P2,t} + y^{\text{rate}}_{P3,t} & \forall t \in T \\
& x^{\text{purch}}_{A,t} = y^{\text{rate}}_{P1,t} & \forall t \in T \\
& x^{\text{sales}}_{m,t} + slack^{\text{sales}}_{m,t} = \xi_{m,t} & \forall m \in FP,\ t \in T \\
& w^{\text{rate}}_{f,t} \le w^{\text{rate,max}}_{f,t} + slack^{\text{max,cap}}_{f,t} & \forall f \in F,\ t \in T \\
& w^{\text{rate}}_{f,t} \ge w^{\text{rate,min}}_{f,t} - slack^{\text{min,cap}}_{f,t} & \forall f \in F,\ t \in T \\
& w^{\text{inv}}_{D,t} \le w^{\text{inv,max}}_{D,t} & \forall t \in T \\
& x^{\text{purch}}_{A,t} \le x^{\text{purch,max}}_{A,t} & \forall t \in T
\end{aligned}
$$

where profit includes sales, operating costs, purchase costs, inventory costs, and penalties for unmet demands and capacity violation

Test Case: Uncertain Yield

- Historical data for production yield of facility *P1*



Generated in R
(R Core Team, 2013)

- Skewed to the right
- Tail effects (extreme values) are not negligible
- Approaches: Heuristic and L^2 DMP (2 moments + ECDF)
- TSSP, where first stage is t = 1 and second stage is t = 2, ..., 4

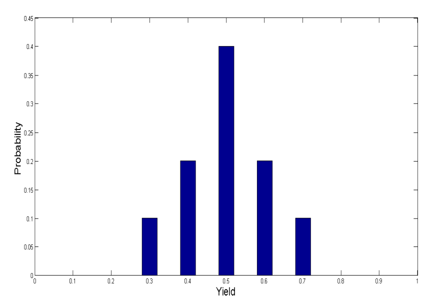
Probability Profiles

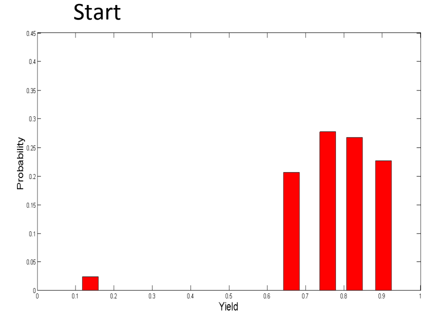
Heuristic Approach

- From possible minimum and maximum yield, took ± 20% and ± 40% deviations from the mean, and assigned probabilities arbitrarily

L^2 DMP Approach

- Solved L^2 DMP to determine the p and x vectors
- Implemented in AIMMS 3.13, solver: IPOPT 3.10.1 with Multi-Start





Results and Discussion

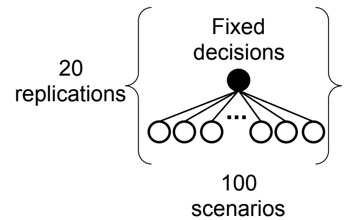
- Heuristic scenario tree does not represent well the distribution of the uncertain parameter
- Expected profits by using the trees from each approach

Approach	Expected Profit [$]
Heuristic	65.43 MM
L^2 DMP	73.23 MM

- Production plan obtained by using the Heuristic approach predicts higher total inventory levels of product D and higher purchase levels of product C
- Solution obtained from Heuristic approach is misleading

Solution Quality

- Stochastic simulation via Monte Carlo sampling to approximate the "true" solution
- Subject first-stage decisions obtained from MMA and L^2 DMP approaches to generated i.i.d. scenarios
- Compute optimality gap and its one-sided confidence interval
- Compute the squared deviations of 1^{st}-stage decisions from each approach from average 1^{st}-stage decisions in simulation



Approach	1-S CI [$ MM]	SDev
Heuristic	0.21 + [0, 0.07]	19.9
L^2 DMP	0.04 + [0, 0.07]	0.51

Better quality solution

Conclusions

- Distribution matching provides a systematic, data-driven way of generating scenarios
- Quality and practicality of solution is a function of the scenario tree used
- Facilitates better decision making by modeling uncertainty

References

- Birge, J. R., & Louveaux, F. (2011). *Introduction to Stochastic Programming*. Springer, second edition.
- Bayraksan, G., & Morton, D. P. (2006). *Assessing Solution Quality in Stochastic Programs*. **Mathematical Programming**, 108(2-3), 495–514.
- 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.
- R Core Team. (2013) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.
- Richards, F. J. (1959) *A Flexible Growth Function for Empirical Use*. **Journal of Experimental Botany**. 10(2):290–301.