Probabilistic scenarios as input for prescriptive analytics

CLSAC 2017
Annapolis

David L. Woodruff
Graduate School of Management
UC Davis
Davis CA USA 95616
DLWoodruff@UCDavis.edu

Partial list of senior co-authors

- Andrea Staid, Sandia National Laboratories
- Jean-Paul Watson, Sandia National Laboratories
- Roger Wets, UC Davis

This work was funded, in part, by DOE through ARPA-e and Office of Science and by the Bonneville Power Administration.

Data

earth.jpg

Our interest is in quantifiable results, so we view the world as a generator of data. At time t, the world generates a vector of "observations" $\mathcal{O}(t)$.

The Objective

The goal is to come up with a way to do well in the future, perhaps by taking into account the nature of the world and observations from the past. We quantify "doing well" using a function

$$\tilde{f}_t(x, \{\mathcal{O}(\tau), \tau \geq t\})$$

where x is a vector of decision values made just before time t.

To Use a Computer to Optimize

- One common approximation is to look at parts of $\mathcal{O}(\tau)$ and we will call these vectors $\xi(\tau)$. This is typically done for $\tau \geq t_{now}$, where t_{now} is the time "now."
- One also typically approximates $\tilde{f}_t(x, \{\mathcal{O}(\tau),\ \tau \geq t\})$ using some function $f_t(x, \{\xi(\tau),\ \tau \geq t\})$ that is easier to work with or at least possible to write down.

The Future Matters

- Only the decisions for "now" can be implemented, but typically the decisions impact the world in the future (i.e., $\{\mathcal{O}(\tau), \tau \geq t_{now}\}$ depend on x) so it is usually a good idea to take that into account.

- At time $t = t_{now}$ one might use a computer to find

$$\operatorname*{argmin}_{x} f_t(x, \{\xi(\tau),\ \tau \geq t\})$$

  and maybe explicitly require $x \in \Omega(\xi)$.

But the Future is Uncertain

and the uncertainty is not so simple

dice.jpeg

An Obvious Example

ShanghaiStockExchange.jpeg

"SSE bases its development on the principles of rule of law, regulation, self-discipline, and compliance in order to create a transparent, open, secure and efficient marketplace."

Example: Forest Harvest Planning

- x is a binary vector indicating cutting blocks in the forest
- and may include road building, etc.
- Time stages may be months or years.
- ξ may include costs, timber values, fire characteristics, etc.

2004-Wildlandfire2nd85-28.jpg

Example: Day-ahead Unit Commitment

- First stage x is a binary vector indicating on-off state for each electricity generator for each hour
- Second stage is generation levels.
- ξ may include costs, demand, wind power, solar power, etc.

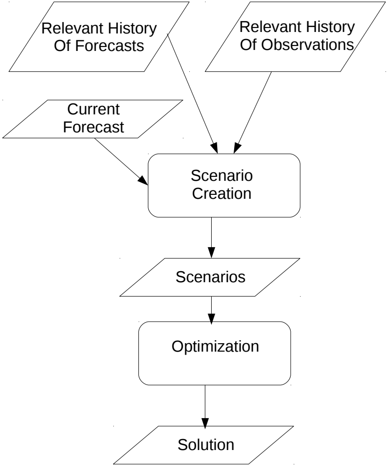
wind_v_coal_power.jpg

A Few More Examples

- Sensor location
- Network Interdiction and defense
- Emergency equipment location

Software Architecture

We will talk first about Optimization, then Scenario Creation



Abstract Notation

- We use x^t to represent the part of the decision vector that corresponds to stage t.
- $\vec{x}^{\,t}$ for $1 \leq t \leq T$ represents the decisions for all stages up to, and including, stage t.
- For the first stage we write $f_1(x^1)$ and for subsequent stage $f_t(x^t; \vec{x}^{\,t-1}, \vec{\xi}^{\,t})$.

$$\min_x f_1\left(x^1\right) + \mathbb{E} \sum_{t=2}^{T} f_t\left(x^t; \vec{x}^{\,t-1}, \vec{\xi}^{\,t}\right) \tag{1}$$

Discrete Scenarios

For Practical Reasons

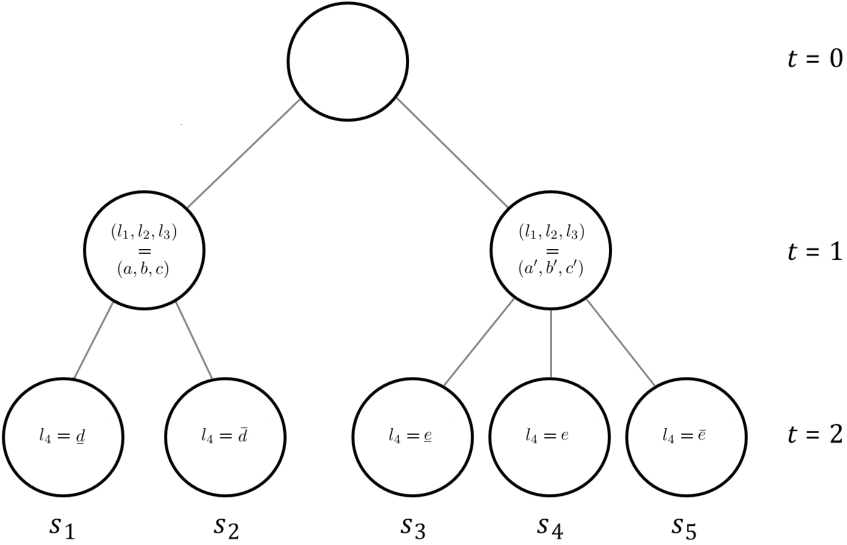
- Define $\xi \equiv \{\xi(t)\}_{t=1}^{T}$ on a discrete probability space.
- Use Ξ to represent the full set of scenarios.
- Each scenario, ξ, has probability π_ξ.
- Write simply ξ to represent the entire scenario.

Scenario Trees

ginkgo-plant-form.jpg

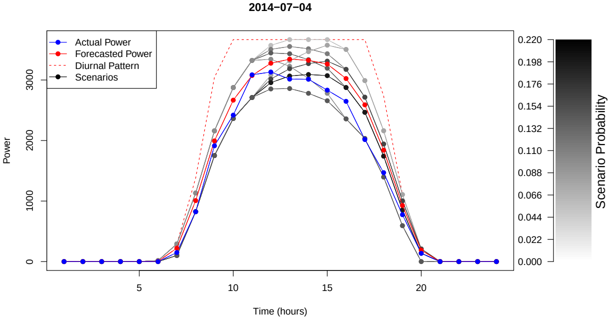
- We organize ξ into a tree with the property that scenarios with the same realization up to stage t share a node at that stage.
- So $\overrightarrow{\xi}^{t}$ refers also to a node in the scenario tree.
- Let $\mathcal{G}_t$ be the set of all scenario tree nodes for stage t
- Let $\mathcal{G}_t(\xi)$ be the node at time t for a particular scenario, ξ.
- For a particular node $\mathcal{D}$ let $\mathcal{D}^{-1}$ be the set of scenarios that define the node.

A Scenario Tree



Example of Scenarios (two stage)

Probabilistic solar power scenarios for the northern region of the California Independent System Operator (CAISO) made using data available before the date.



Potential Uses for Scenarios

- Analysis of a plan or policy (x); e.g. simulation
- Optimization

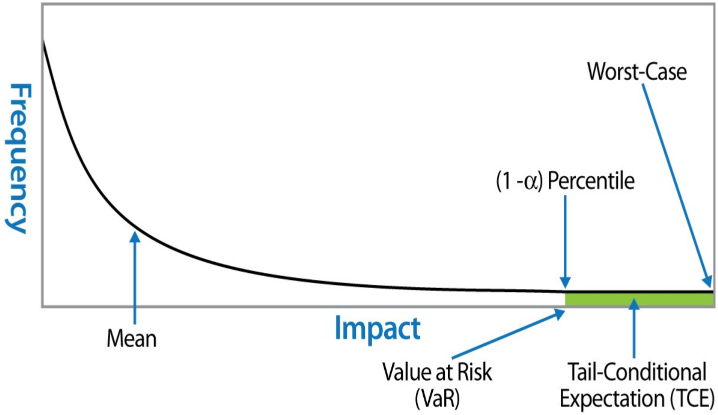
A Scenario Tree Formulation

Non-anticipativity must be enforced at each non-leaf node:

$$\min_{x,\hat{x}} \sum_{\xi \in \Xi} \pi_\xi \left[f_1(x^1(\xi)) + \sum_{t=2}^{T} f_t \left(x^t(\xi); \vec{x}^{\,t-1}, \vec{\xi}^{\,t} \right) \right] \tag{2}$$

$$\pi_\xi x^t(\xi) - \pi_\xi \hat{x}^t(\mathcal{D}) = 0, \quad t = 1, \ldots, T-1, \ \mathcal{D} \in \mathcal{G}_t, \ \xi \in \mathcal{D}^{-1} \tag{3}$$

CVaR (TCE)



Frequency
Worst-Case
(1 -α) Percentile
Mean
Impact
Value at Risk
(VaR)
Tail-Conditional
Expectation (TCE)

CVaR (TCE)

Some risk measures are easy to add directly

- The expected value of the $1 - \alpha$ tail.
- Bounds VaR and probably the way to get VaR if you need it.

CVaR can be implemented by augmenting this problem. Following the implementation given by Schultz and Tiedemann we add a real-valued, scalar variable, η, and for every scenario $s \in \mathcal{S}$ add a non-negative, real-valued, scalar variable ν_s.

Two Questions Remain for my talk:

0. How do we get scenarios, and

1. how do we know how good they are?

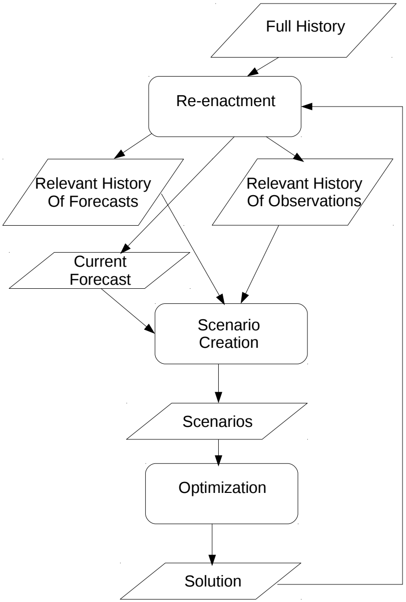
Some Ways to Get Scenarios

- Statistical models, perhaps obtained by data mining (e.g., in Finance)
  - Monte Carlo sampling
  - Moment matching
- Simulations (e.g., in Forest Harvesting with Fire Risk)
- Forecast Error Distributions (e.g., Unit Commitment)

Evaluating Scenarios

- Analyze the statistical properties
- Analyze the solutions obtained
  - Simulation
    - In-sample
    - Out-of-sample
    - Independent
  - Re-enactment

Software Architecture for Re-enactment



Sketch of Re-enactment

Details depend on the application

- $\tilde{f}(\cdot) \succ \texttt{Eval}(\cdot) \succeq f(\cdot)$
- $\texttt{Eval}(\hat{x}; \mathcal{O}(\tau), \tau = t_{now} + T_{lt}, T)$, ($T$ is end of data)
- T_{oper} = periods of use of the solution; T_0 is first time with data and T_{sc} is needed for scenarios.

1: **Initialization**:
2: **Scenario Creation**:
3: **Optimization:**
4: **Evaluation and Record Keeping:** Compute and store the results of $\texttt{Eval}(\hat{x}; \mathcal{O}(\tau), \tau = t_{now} + T_{lt}, T)$
5: **Iterate:**
6: **Termination:**

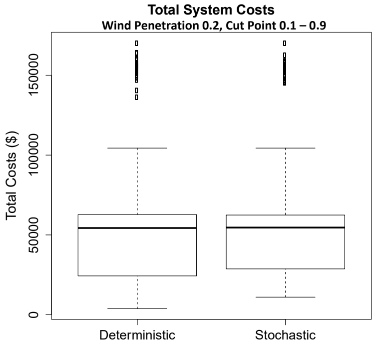
Stochastic Unit Commitment

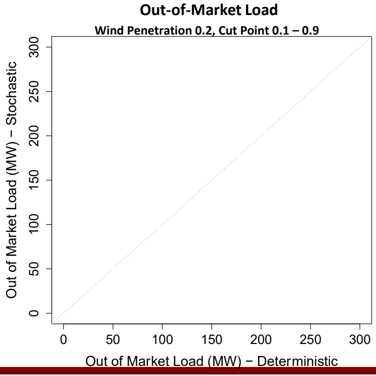
Sandia National Laboratories

- Power system unit commitment must incorporate variable generation resources (i.e., wind power)
- To account for the uncertainty of wind power, we can model this variable generation stochastically, using scenarios
- We perform day-ahead two-stage unit commitment, with scenarios created to represent the plausible range of wind power uncertainty throughout the day
    - Wind is ***not*** modeled as must-take, allowing for curtailment without penalty

Unit Commitment Performance

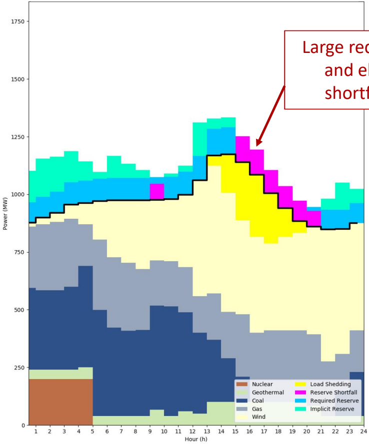
- Costs are comparable in deterministic and stochastic solutions
- However, we do not account for the cost of procuring additional generation in real-time to serve the out-of-market load (not met in day-ahead market)





Stochastic vs Deterministic

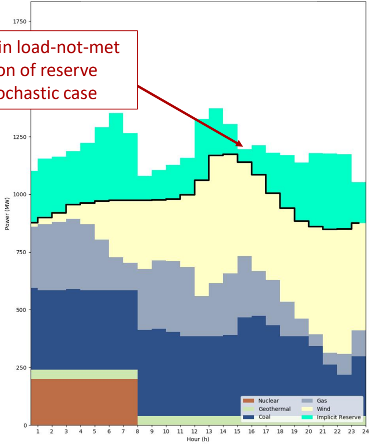
Deterministic: 2017-03-18
CP: 0 – 0.01 – 0.5 – 0.99 – 1



Variable costs: 227111.27
Fixed costs: 445983.41
Renewables penetration rate: 33.03%

Large reduction in load-not-met and elimination of reserve shortfall in stochastic case

Stochastic: 2017-03-18
CP: 0 – 0.01 – 0.5 – 0.99 – 1



Variable costs: 181086.81
Fixed costs: 571981.60
Renewables penetration rate: 32.88%

Conclusions

so far

- Optimized decision making operates in a uncertain world
- so methods are needed
  - for modeling stochastics
  - and optimization under uncertainty (www.pyomo.org), and
  - for evaluating the entire system.

- We have software for doing this for unit commitment;
- we are interested in extending to other domains.

rd100.png