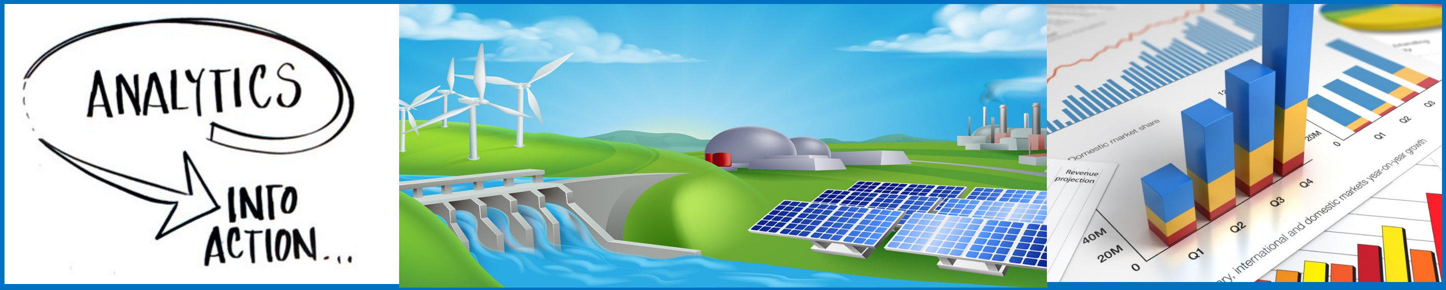


ENV797 - TIME SERIES ANALYSIS FOR ENERGY AND ENVIRONMENT APPLICATIONS

Module 12 - Scenario Generation

Prof. Luana Medeiros Marangon Lima, Ph.D.

Master of Environmental Management Program
Nicholas School of the Environment - Duke University

Learning Goals

- Understand the concept of scenario generation
- Connect optimization and decision making under uncertainty
  - Intro to stochastic optimization
- Learn how to represented uncertainty with a scenario tree
  - Give an idea of how the tree is generated
  - How to generate correlated scenarios
- Learn how to generate scenarios based on the time series models we learned in R

Scenario Generation

Motivation

"Wide range of real-world problems involve decision-making under uncertainty."

"If a statistical model can be used to describe this uncertainty, the decision problem can be modeled as a stochastic optimization problem."

Source: Nils L¨ohndorf , "An empirical analysis of scenario generation methods for stochastic optimization"

Stochasticity or Uncertainty

- Origin
  - Future information (e.g. prices or demand)
  - Lack of reliable data
  - Measurement errors
- In electric energy systems planning
  - Demand (yearly, seasonal or daily variation, load growth
  - Hydro, Wind and Solar (natural resources)
  - Availability of generation or network elements
  - Electricity or Fuel Prices

Source: Andres Ramos, Comillas, Madrid

Stochastic Optimization

- Optimizing or making decisions under **uncertainty**
- Why uncertainty?
  - Exact data is unavailable or expensive
  - Instead, data is specified by a probability distribution
- Obj.: Make the **best decisions** given the uncertainty
- Approach: Multi-stage Model

Given

Probability distribution over inputs

Stage I

Make **advance decisions** – plan ahead or **hedge against uncertainty**

Observe the actual input scenario

Stage II

Take **recourse**

Augment earlier solution paying a **recourse cost**

Decision Under Uncertainty

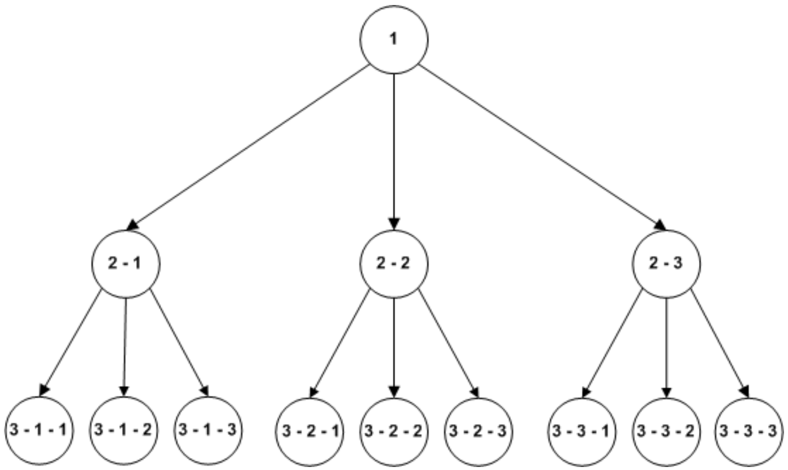
- Determinist optimization
  - **Best decision** when future is known

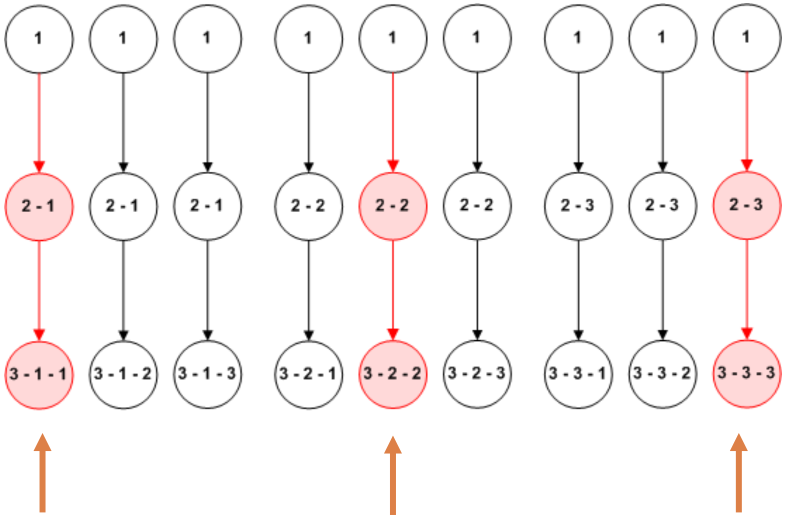
- Stochastic Optimization
  - **Better decision** when future is uncertain but with a known probability

- But how??
  - Scenario Analysis or scenario tree

Scenario Tree

- **Tree:** represents how the stochasticity is revealed over time, i.e., the different states of the random parameters
- **Nodes**: where decisions are taken
- **Scenarios**: path going from the root to the leaves
- Allow the solution of huge problem by solving **iteratively small size** problems





Scenario Tree Generation

- Correlation among random parameters should be considered
- Number of scenarios generated should be enough for observing parameter variability
- Common methods
  - **Monte Carlo sampling methods** ⇒ **Simulation**
  - Quasi-Monte Carlo methods
  - Optimal quantization of probability distributions
  - And others….

Simulations in R

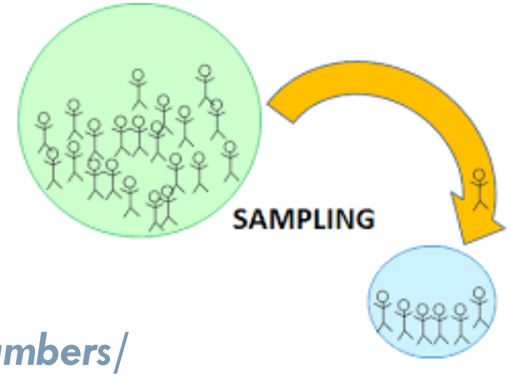
- Possible to simulate data with R using random number generators of different kinds of variables
- Sampling from

Multinomial distributions *sample(1:4,1000,rep=TRUE,prob=c(.2,.3,.2,.3))*

Uniform distribution *runif(n, min = 0, max = 25)*

Normal distribution *rnorm(n, mean = 0, sd = 1)*

Exponential distributions *rexp(n, rate = 1)*



Other examples available at:

http://uc-r.github.io/generating_random_numbers/

Sampling from Multivariate Normal Distribution

- When sampling the scenarios for multiple variables one need to take into account correlation
- Easiest way to deal with this is to draw independently $N[\mu, \sigma^2]$ and then pass the correlation through Cholesky decomposition
- Let R be correlation matrix $(n_{var}\ x\ n_{var})$ among the variables. The Cholesky decomposition of R is a lower triangular matrix such that

$$R = LL^T$$

- How to get L?

More info: https://en.wikipedia.org/wiki/Cholesky_decomposition

Sampling from Multivariate Normal Distribution (c'ed)

- Let X be a matrix $(n_{var}\ x\ n_{step})$ with independent identically distributed draws from a N[0, 1]
- Define Y such that

$$Y = LX$$

Recall L is the Cholesky decomposition of R

- Note that the resulting matrix Y will have order $n_{var}\ x\ n_{step}$
- Y corresponds to the correlated draws

Connecting Scenario and Models Learned in TSA

ARIMA Forecasting

- Recall the ARMA(1,1) model equation

$$Y_t = \boxed{\phi_1 Y_{t-1}} + a_t \boxed{- \theta_1 a_{t-1}} \qquad for\ t = 1,2,\ldots,n$$

where $a_t \sim N.I.D.(0, \sigma^2)$

- From the estimation step you have $\phi = (\phi_1\ \phi_2\ \ldots \phi_p)'$ and σ^2
- One can rewrite this equation as

$$Y_t \sim N.I.D.(\phi_1 Y_{t-1} - \theta_1 a_{t-1}, \sigma^2)$$

- Same principle is extended for the more general class of ARIMA Models

State Space BSM

- Model equations

$$y_t = \mu_t + \gamma_t + \varepsilon_t \qquad \varepsilon_t \sim \mathcal{NID}(0, \sigma_\varepsilon^2)$$

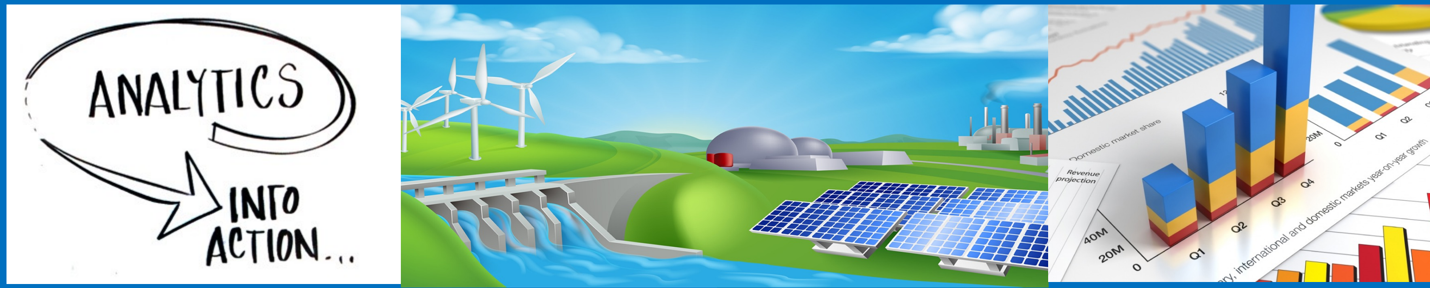
$$\mu_{t+1} = \beta_t + \mu_t + \eta_t \qquad \eta_t \sim \mathcal{NID}(0, \sigma_\eta^2)$$

$$\beta_{t+1} = \beta_t + \xi_t \qquad \xi_t \sim \mathcal{NID}(0, \sigma_\xi^2)$$

$$\gamma_{t+1} = -\sum_{j=1}^{s-1} \gamma_{t+1-j} + \omega_t \qquad \omega_t \sim \mathcal{NID}(0, \sigma_\omega^2)$$

- The observation equation can be rewritten

$$y_t \sim \mathcal{NID}(\mu_t + \gamma_t, \sigma_\varepsilon^2)$$



THANK YOU !

luana.marangon.lima@duke.edu