PGE 383 – Machine Learning
Uncertainty Management

Lecture outline . . .

- Uncertainty Sources
- Uncertainty Philosophy
- The Bootstrap
- Monte Carlo Simulation
- Scenarios
- Sampling Uncertainty

Instructor: Michael Pyrcz, the University of Texas at Austin

Motivation

We cannot avoid uncertainty when modeling the subsurface.

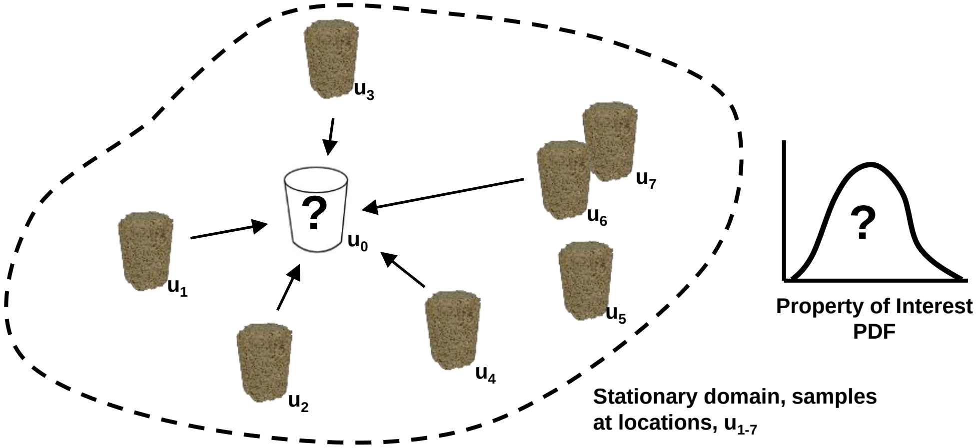
PGE 383 – Machine Learning
Uncertainty Management

Lecture outline . . .

- **Uncertainty Sources**

Instructor: Michael Pyrcz, the University of Texas at Austin

What is Uncertainty?



Uncertainty is not an intrinsic property of the subsurface.

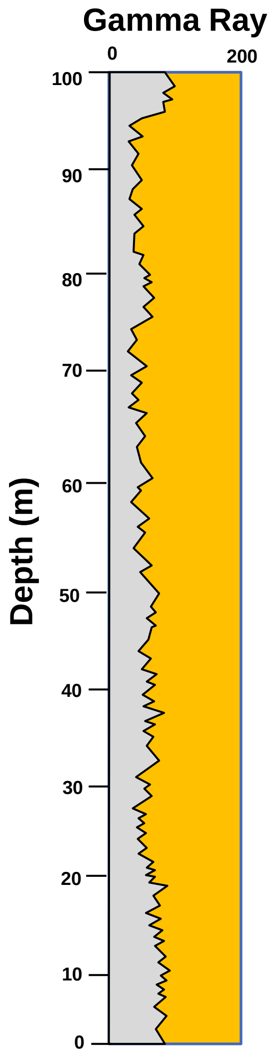
- At every location ($\mathbf{u}_{\alpha}$) within the volume of interest the true properties could be measured if we had access (facies, porosity etc.).
- **Uncertainty is a function of our ignorance**, our inability to observed and measure the subsurface with the coverage and scale required to support our scientific questions and decision making.

sparsity of sample data + heterogeneity = uncertainty

- If the subsurface was homogeneous, with a few measurements uncertainty would be reduced and estimates resolved to a sufficient degree of exactitude.

Uncertainty Sources

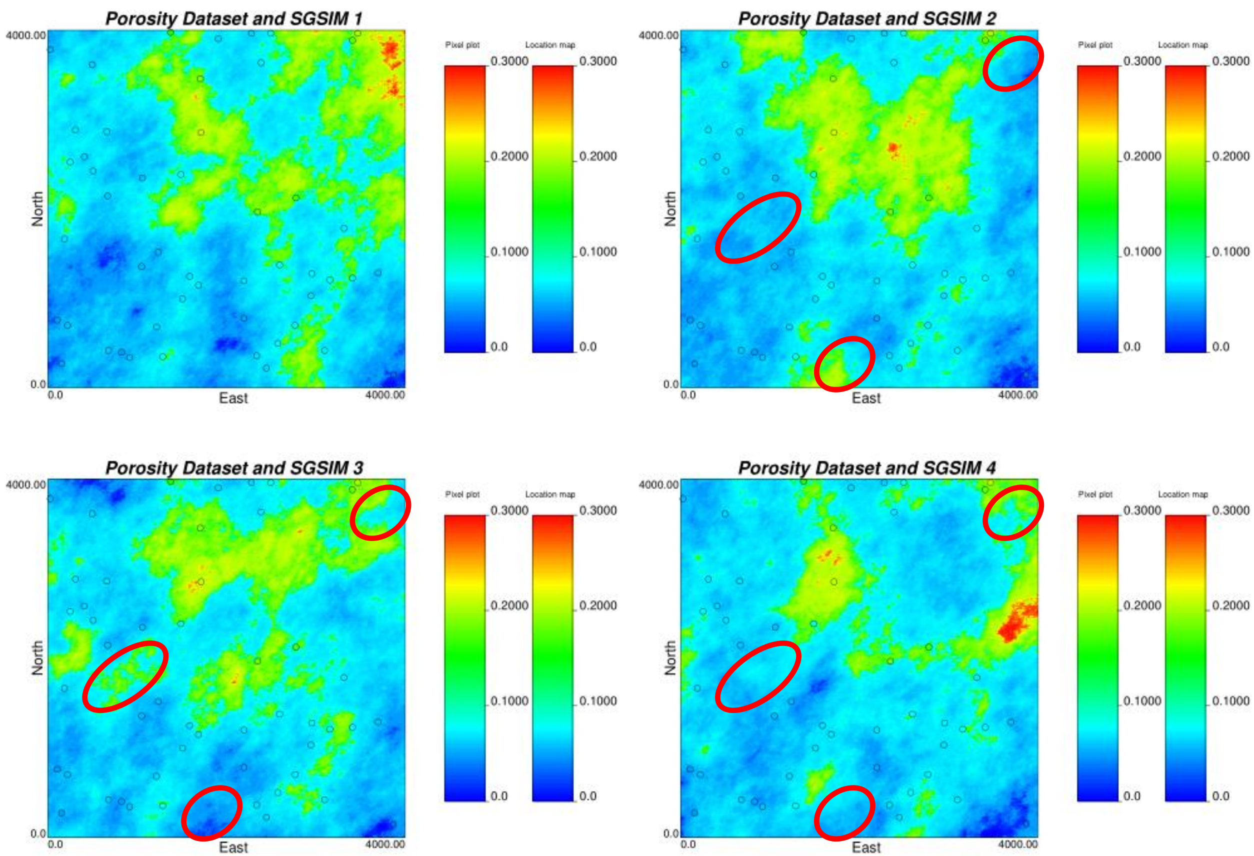
- **Measurement / Interpretation Error**
  - Formation evaluation – tool tolerance, calibration error, approximations / assumptions
  - Interpreter experience and prior model / assumptions
  - How to integrate it?
    - » Indicator method code as soft inputs
    - » Multiple data realizations in design of experiments



Schematic of a well log.

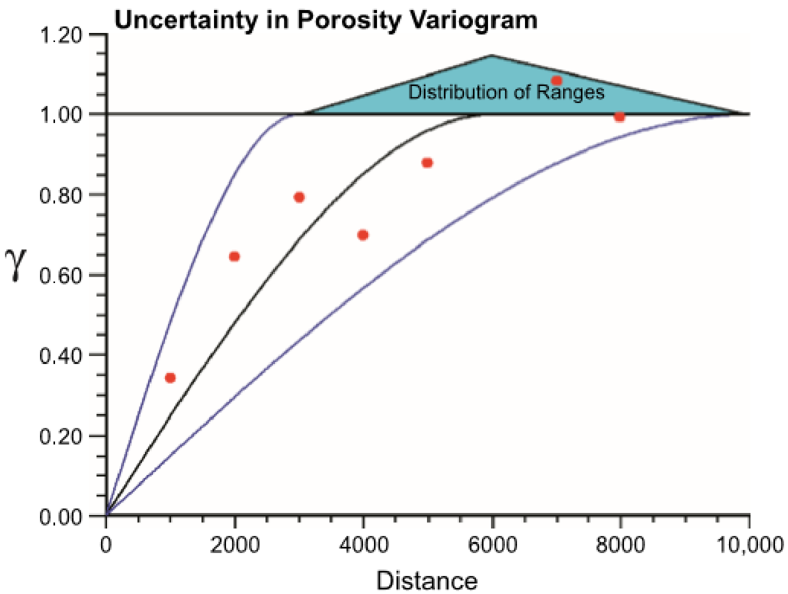
Uncertainty Sources

- **Spatial Uncertainty**
  - Uncertainty due to spatial offset from sampled locations
  - Integrate through multiple local geostatistical realizations and scenarios

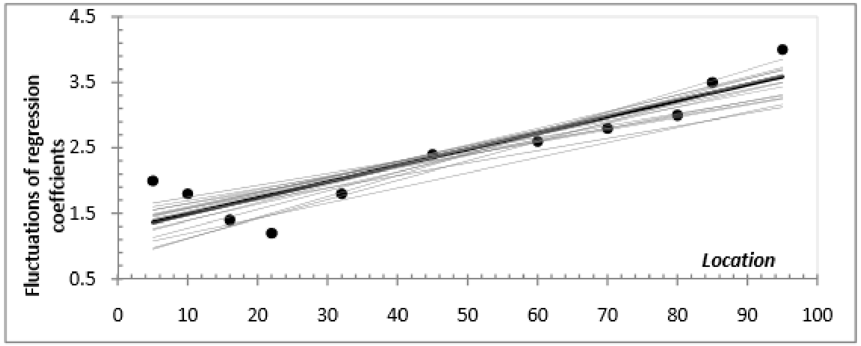


Uncertainty Sources

- **Model Parameter Uncertainty**
  - Uncertainty in the representativity of the sample data
  - Uncertainty in the trained model parameters
  - Use of analog studies, with more information.
  - Formulate distribution scenarios and bootstrap realizations of the data



Distribution of Variogram Ranges (Pyrcz et al., 2006)



Trend Uncertainty (Villalba and Deutsch., 2010)

PGE 383 – Machine Learning
Uncertainty Management

Lecture outline . . .

- **Uncertainty Philosophy**

Instructor: Michael Pyrcz, the University of Texas at Austin

Uncertainty Comments

Calculating Uncertainty in a Modeling Parameter: Use Bayesian methods, spatial bootstrap etc. You must account for the volume of interest, sample data quantity and locations, and spatial continuity.

If You Know It, Put It In. Use expert geologic knowledge and data to model trends. Any variability captured in a trend model is known and is removed from the unknown, uncertain component of the model. Overfit trend will result in unrealistic certainty.

Types of Uncertainty: (1) data measurement, calibration uncertainty, (2) decisions and parameters uncertainty, and (3) spatial uncertainty in estimating away from data. Your job is to hunt for and include all significant sources of uncertainty.

Be an uncertainty detective! Discover and evaluate various sources.

Uncertainty Comments

What about Uncertainty in the Uncertainty? Don’t go there! Use defendable choices in your uncertainty model, be conservative about what you known, document and move on.

Uncertainty Depends on Scale. It is much harder to predict a property of tea spoon vs. a house-sized volume at a location (u_α) in the subsurface. Ensure that scale and heterogeneity are integrated.

You Cannot Hide From It. Ignoring uncertainty assumes certainty and is often a very extreme and dangerous assumption.

Decision Making with Uncertainty. Apply all the models to the transfer function to calculate uncertainty in subsurface outcome to support decision making in the presence of uncertainty.

Ignoring uncertainty is assuming certainty.

Uncertainty Comments

What about Uncertainty in the Uncertainty? Don’t go there! Use defendable choices in your uncertainty model, be conservative about what you known, document and move on.

Uncertainty Depends on Scale. It is much harder to predict a property of tea spoon vs. a house-sized volume at a location (u_α) in the subsurface. Ensure that scale and heterogeneity are integrated.

You Cannot Hide From It. Ignoring uncertainty assumes certainty and is often a very extreme and dangerous assumption.

Decision Making with Uncertainty. Apply all the models to the transfer function to calculate uncertainty in subsurface outcome to support decision making in the presence of uncertainty.

Ignoring uncertainty is assuming certainty.

PGE 383 – Machine Learning
Uncertainty Management

Lecture outline . . .

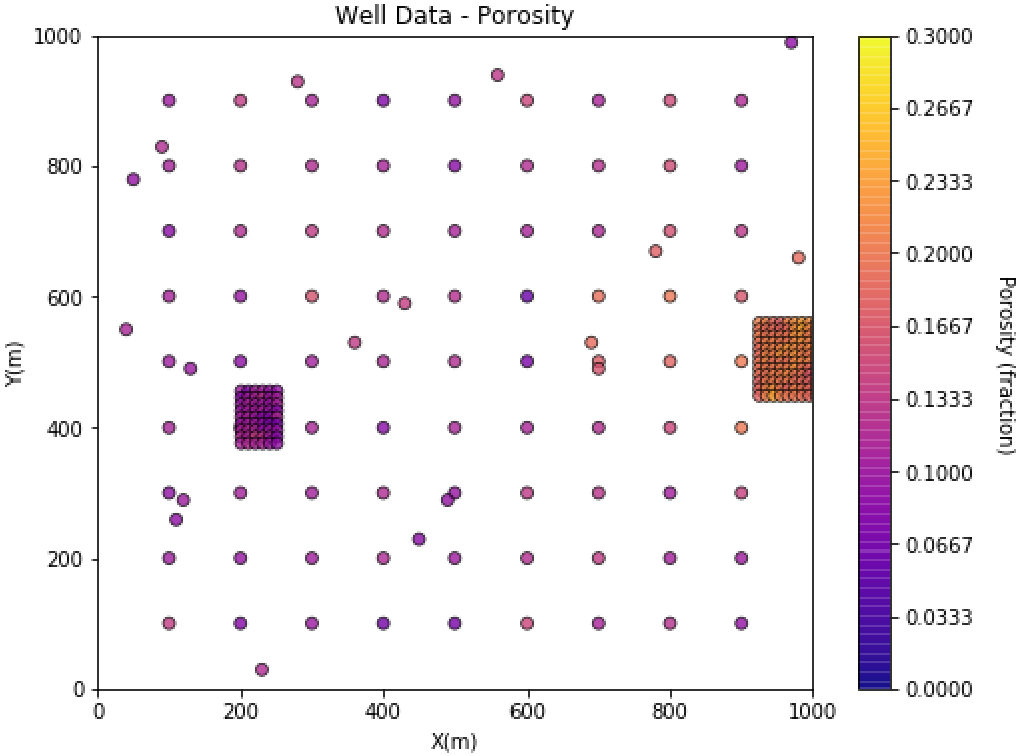
- **The Bootstrap**

Instructor: Michael Pyrcz, the University of Texas at Austin

Bootstrap Motivation

Uncertainty in the Sample Statistics and Models

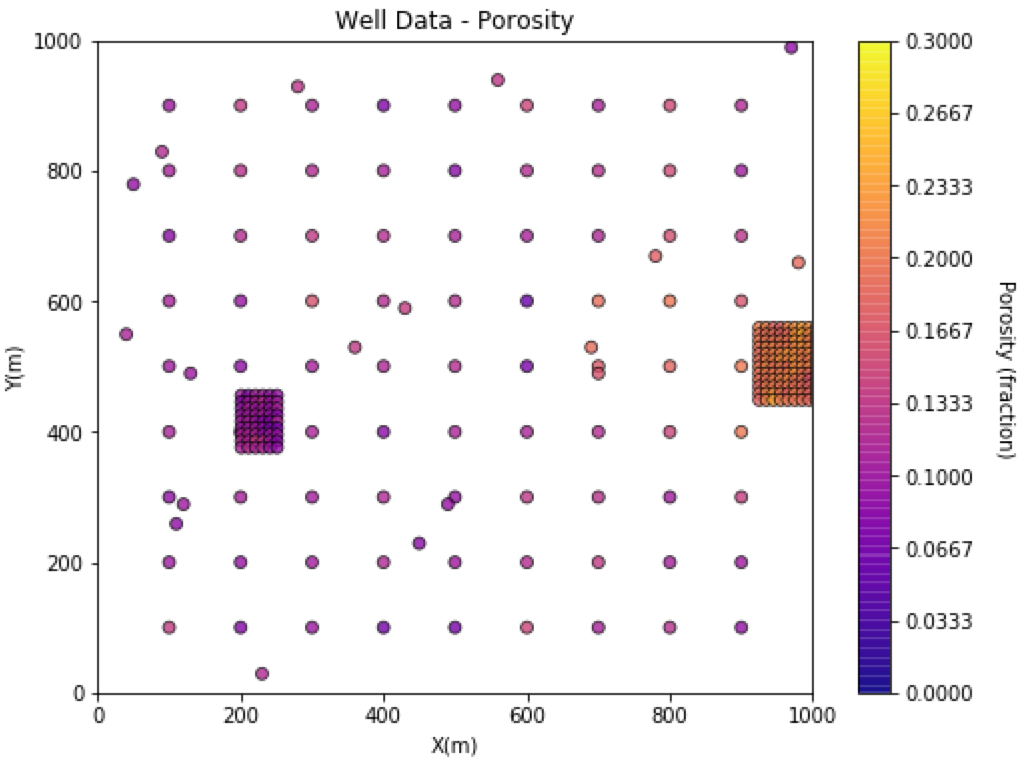
- One source of uncertainty is the paucity of data.
- Do these 200 or so wells provide a precise (and accurate estimate) of the mean? standard deviation? skew? P13?



Bootstrap Motivation

Would it be Useful to Know the Uncertainty in these Statistics Due to Limited Sampling?

- What is the impact of uncertainty in the mean porosity e.g. 20%+/-2%?
- How would our model change with a different dataset?



Bootstrap Definition

Bootstrap

- method to assess the uncertainty in a sample statistic or model (known as model bagging) by repeated random sampling of the data with replacement

Assumptions

- sufficient, representative sampling

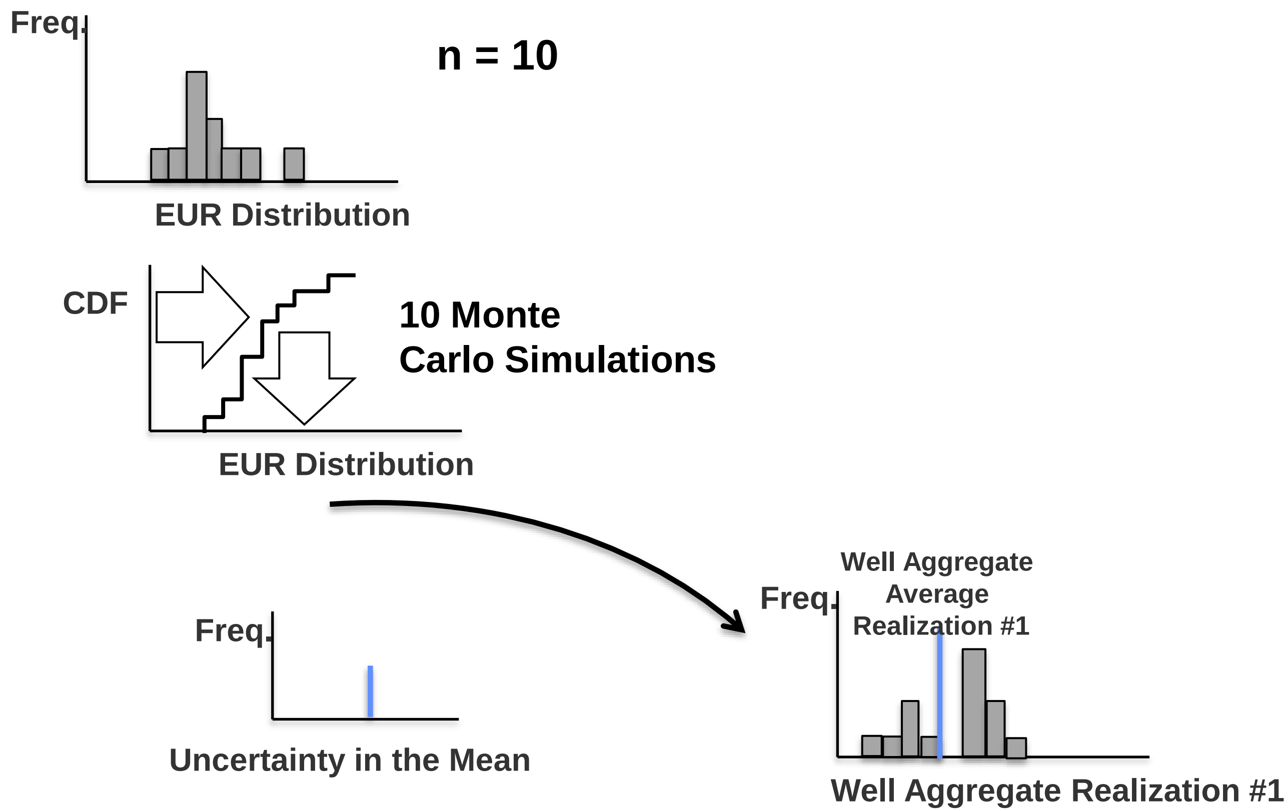
Limitations

- assumes the samples are representative
- assumes stationarity
- only accounts for uncertainty due to too few samples, e.g. no uncertainty due to changes away from data
- does not account for area of interest
- assumes the samples are independent
- does not account for other local information sources

No spatial Context

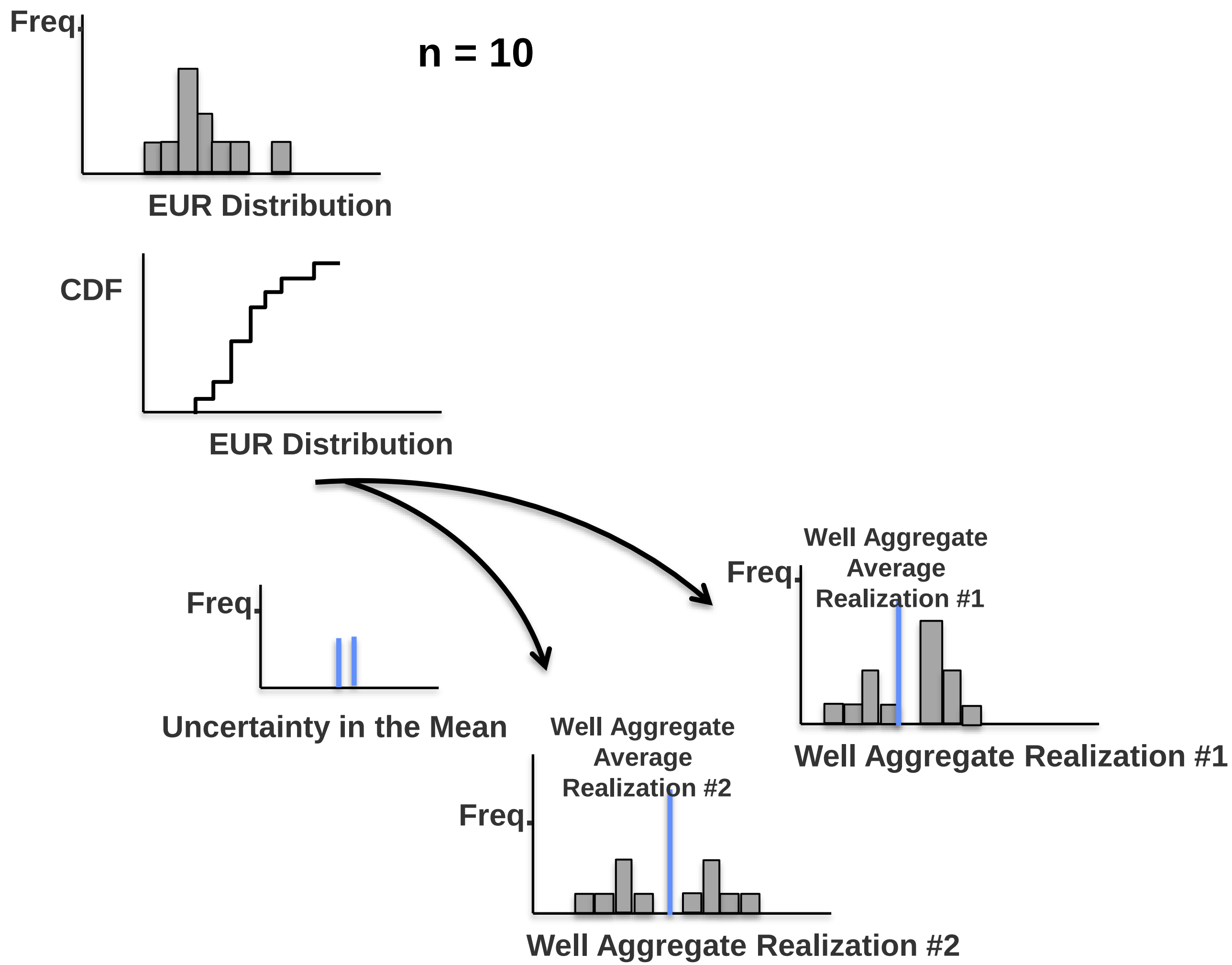
Univariate Statistics
Bootstrap

Bootstrap for Uncertainty in the Mean



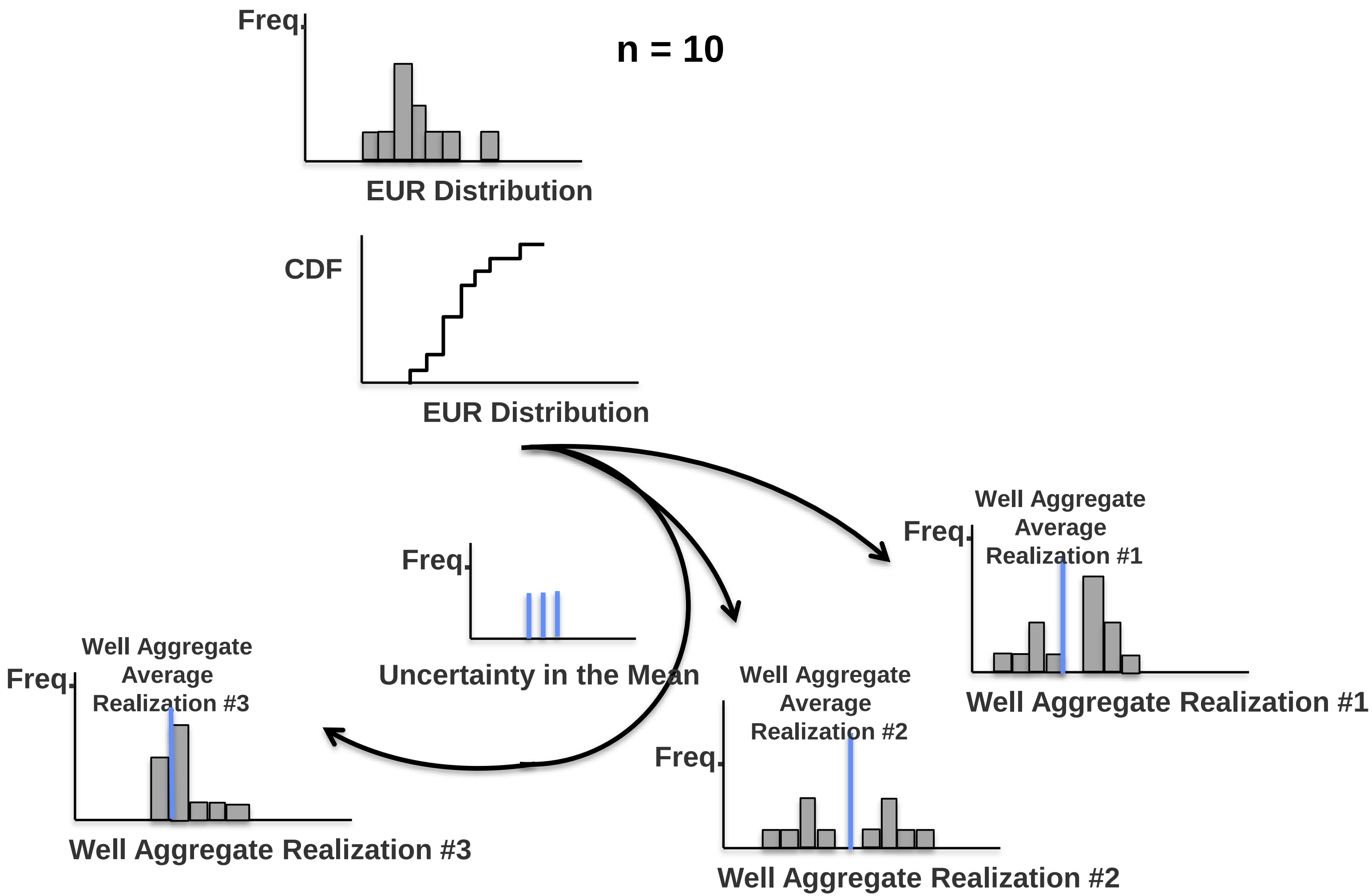
Univariate Statistics
Bootstrap

Bootstrap for Uncertainty in the Mean



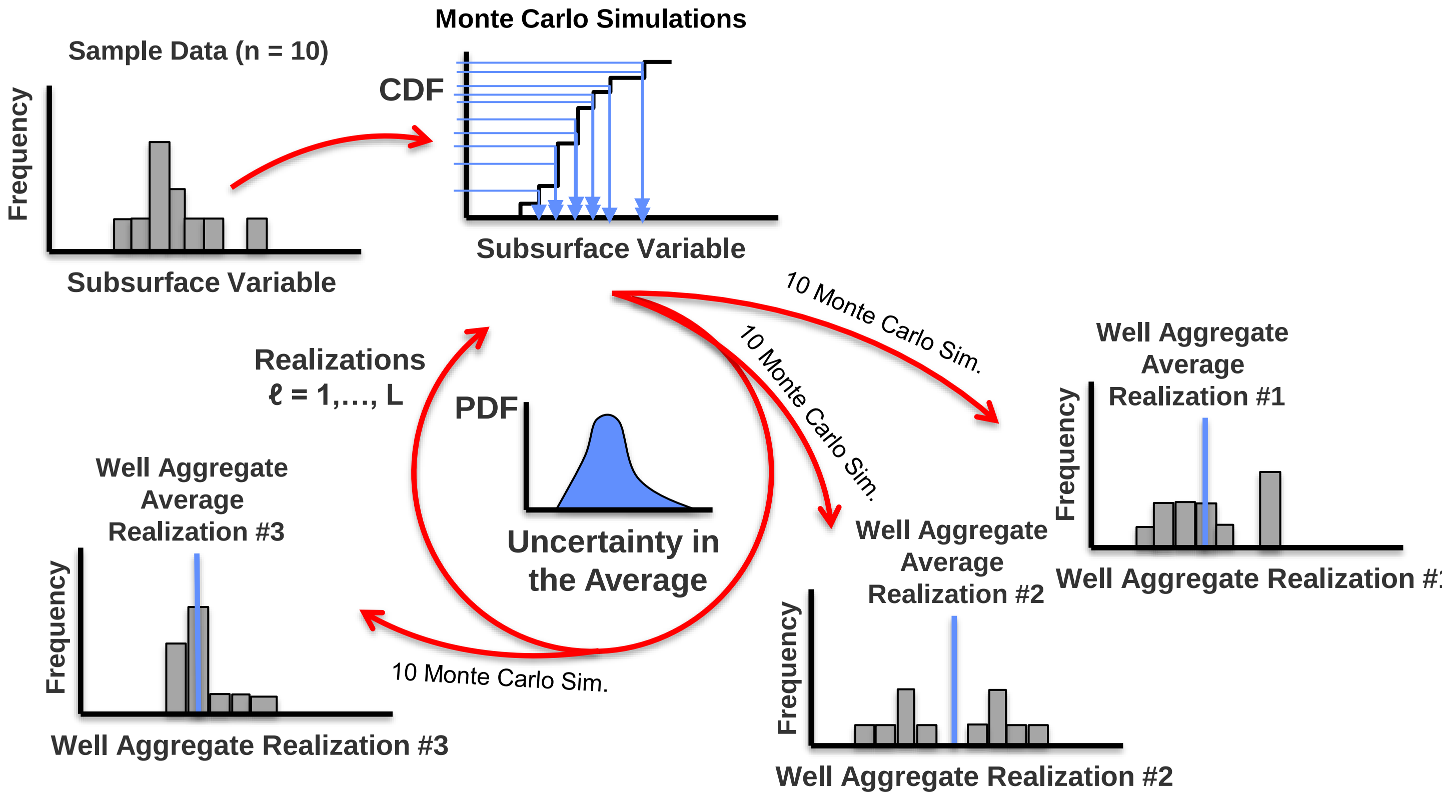
Univariate Statistics
Bootstrap

Bootstrap for Uncertainty in the Mean



Univariate Statistics
Bootstrap

Bootstrap for Uncertainty in the Mean

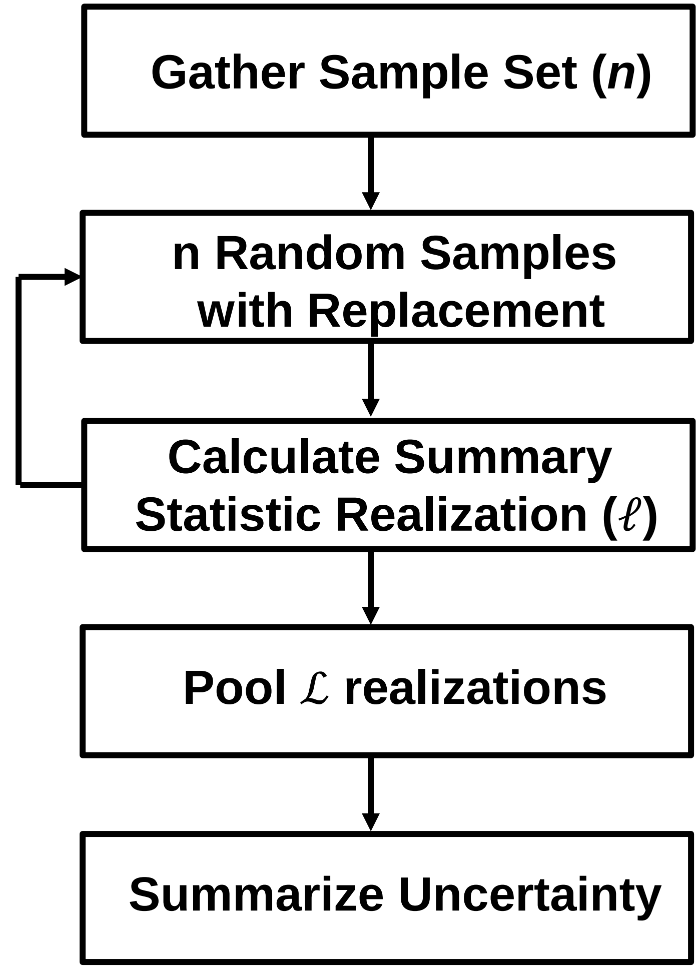


Univariate Statistics
Bootstrap

- Bootstrap Approach (Efron, 1982)
- Statistical resampling procedure to calculate uncertainty in a calculated statistic from the data itself.
- For uncertainty in the mean solution is standard error:

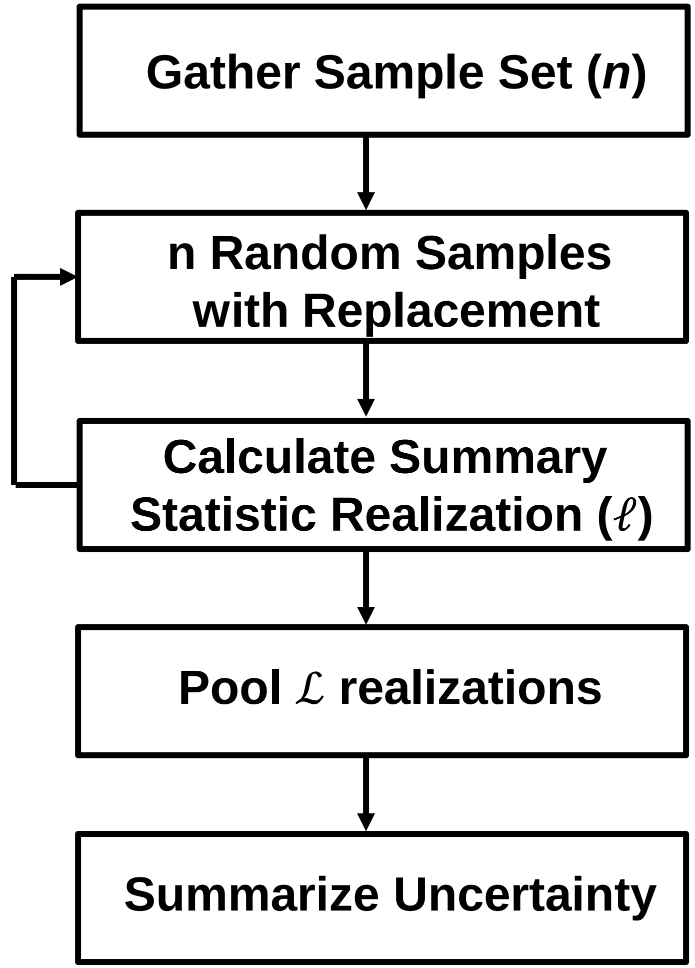
$$\sigma_{\bar{x}}^2 = \frac{\sigma_s^2}{n}$$

- Extremely powerful. Could get uncertainty in any statistic! e.g. P13, skew etc.
- Would not be possible without bootstrap.
- Advanced forms account for spatial information and strategy (game theory).



Univariate Statistics
Bootstrap

- You now know about one of the most powerful tools ever!
- Caveats:
  - assumes the sample set is representative
  - unbiased and covers the full range
  - assumes all samples are independent if not consider Journel's spatial bootstrap (1993).
- You can do bootstrap in Excel.



Bootstrap Demonstration in Python

Demonstration workflow for boostrap for uncertainty in statistics and models.

File:
SubsurfaceDataAnalytics_bootstrap.ipynb
at https://git.io/fhgUW.

Subsurface Data Analytics

Bootstrap for Subsurface Data Analytics in Python

Michael Pyrcz, Associate Professor, University of Texas at Austin

Twitter | GitHub | Website | GoogleScholar | Book | YouTube | LinkedIn

Exercise: Bootstrap for Subsurface Data Analytics in Python

Here's a simple workflow, demonstration of bootstrap for subsurface modeling workflows. This should help you get started with building subsurface models that integrate uncertainty in the sample statistics.

Bootstrap

Uncertainty in the sample statistics

of data.

ise (and accurate estimate) of the mean? standard deviation? skew? P13?

these statistics due to limited sampling?

nean porosity e.g. 20%+/-2%?

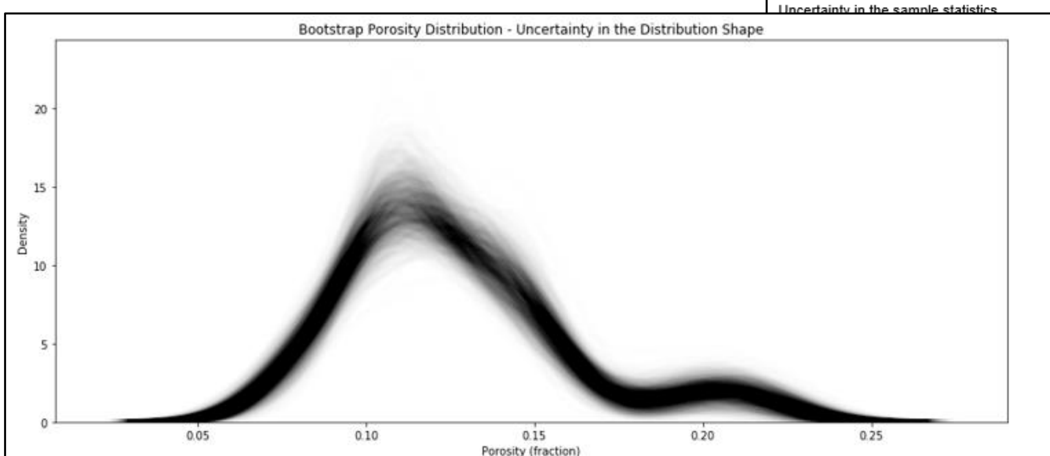
ainty in a sample statistic by repeated random sampling with replacement.

tical, independent samples

e

few samples, e.g. no uncertainty due to changes away from data

f interest



PGE 383 – Machine Learning
Uncertainty Management

Lecture outline . . .

- **Monte Carlo Simulation**

Instructor: Michael Pyrcz, the University of Texas at Austin

Monte Carlo Simulation

Random sampling from a distribution

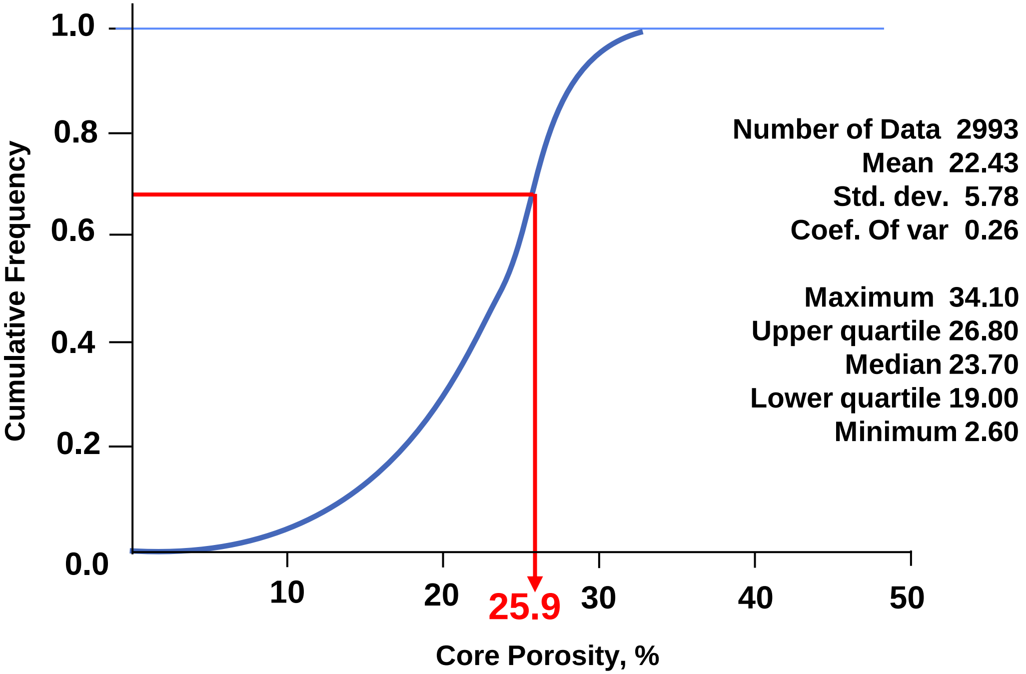
- Procedure:

1. Model the distribution
2. Draw random value from a uniform [0,1] distribution (p-value).
3. Apply the inverse of the CDF to calculate the associated realization.

$$X^l = F_x^{-1}(p^l)$$

4. Repeat to calculate enough realizations for analysis.

The method is very powerful. You can simulate distributions that could not be calculated analytically.

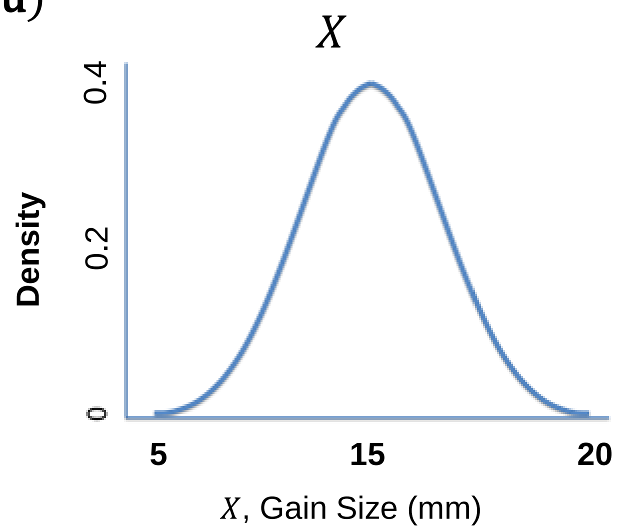


CDF of 2,993 porosity samples.

Random Variable (RV) Definition

Recall: Random Variable

- we do not know the value at a location / time, **it can take on a range of possible values**, fully described with a PDF.
- represented as an **upper case variable, e.g. X, while possible outcomes or data measures are represented with lower case, e.g. x.**
- in spatial context common to use a location vector, $\mathbf{u}$, to describe a location, e.g. $x(\mathbf{u})$, $X(\mathbf{u})$

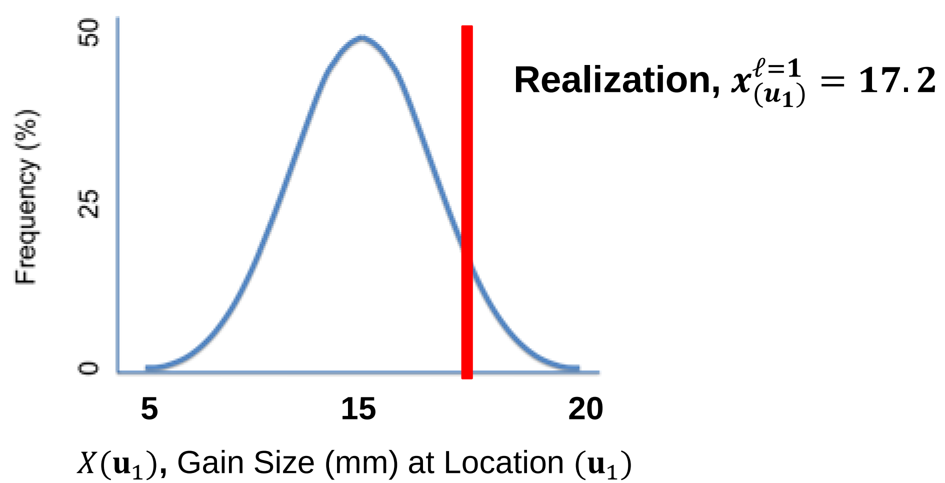


Realization Definition

Realization

- an outcome from a random variable (or joint set of outcomes from a random function – we will cover this later)
- represented with lower case, e.g. x.
- in spatial context common to use a location vector, $\mathbf{u}$, to describe a location, e.g. $x(\mathbf{u})$, $X(\mathbf{u})$
- resulting from simulation, e.g. Monte Carlo simulation, sequential Gaussian simulation ← a method to sample (jointly) from the RV (RF)
- each realization is considered equiprobable

Random Variable, X



Transfer Function Definition

- Any operation(s) applied to a set of realizations of random variables (functions) to access the behavior over a spatial / temporal interval of interest.

  - Examples:
    - Pore volume – average porosity x volume of reservoir
    - Oil-in-place – pore volume x oil saturation
    - Recovery factor – fluid flow rates, fractions by flow simulation

Monte Carlo Simulation Motivaiton

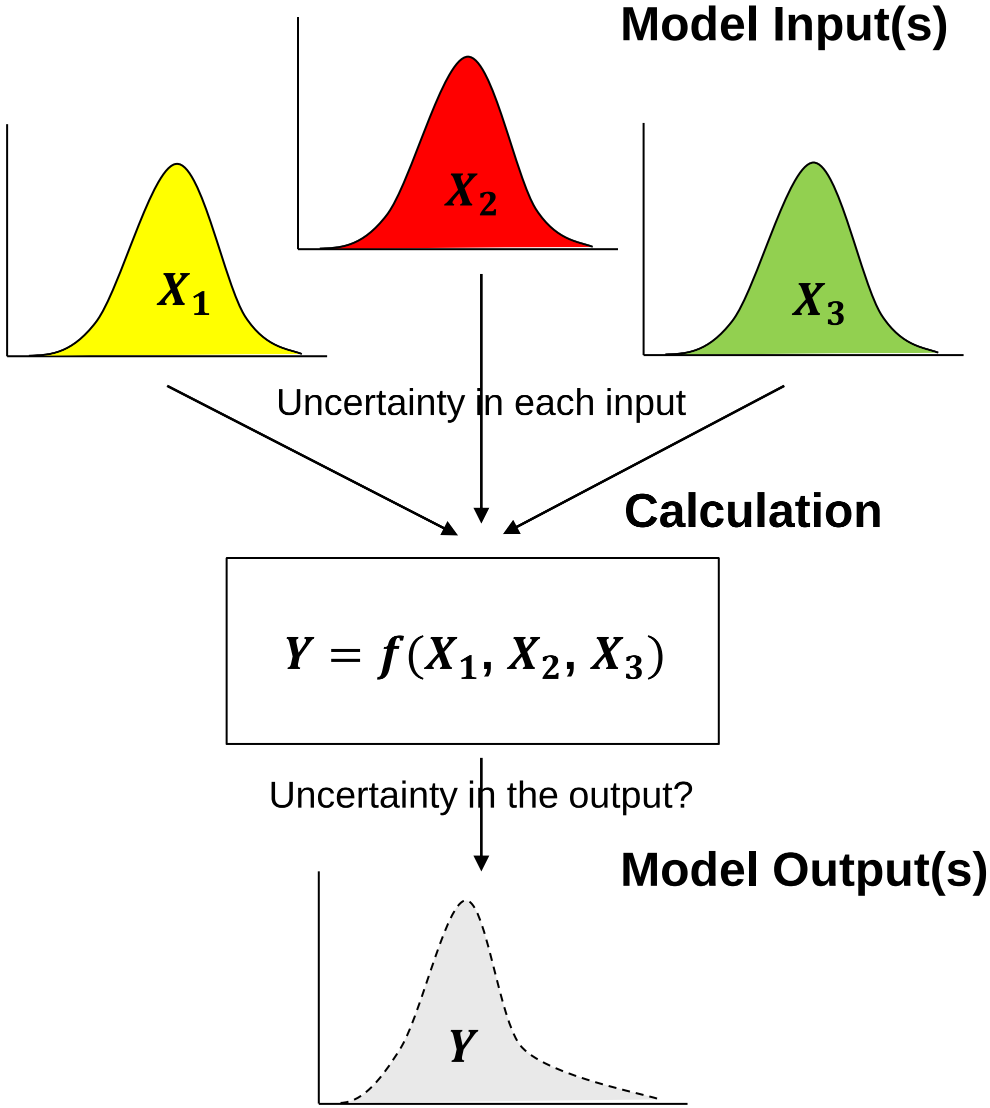
Very Simple and Flexible

- Method for propagating the distribution of uncertainty through a calculation.

- Volumetrics
- Flow Forecasts
- Economics

- We can only do this analytically for simple cases.

We need Monte Carlo Simulation to build practical uncertainty models

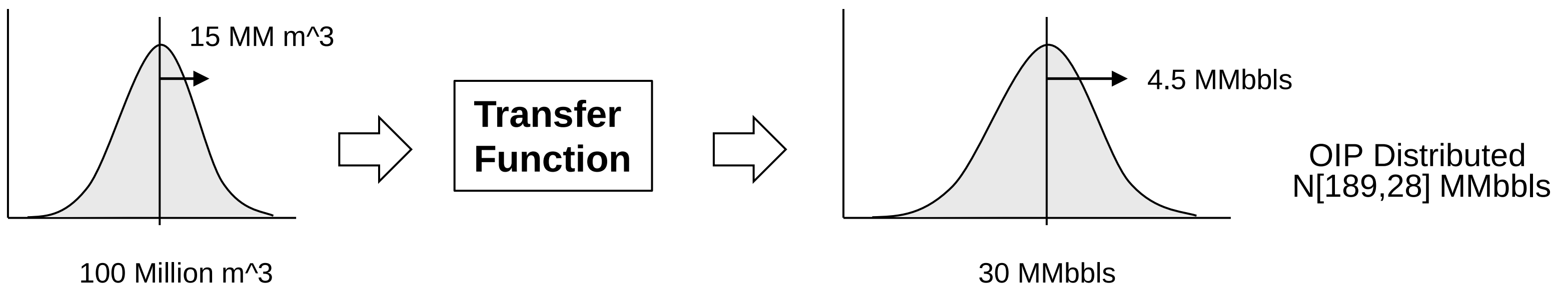


Monte Carlo Simulation Motivation

- **Problem 1:**
  - The reservoir oil in place (OIP) uncertainty is Gaussian distributed with:
    - mean of 100 million barrels and standard deviation of 15 million barrels.

  The recovery factor (RF) is estimated at 0.3, 30% of the oil can technically be extracted.
  - **Calculate the uncertainty in the recoverable oil (RO):**

$$\boldsymbol{RO = RF \times OIP}$$



With statistical expectation we can calculate the μ_{cX} and σ_{cX}. Scaling by a constant won't change the distribution shape.

$$c \cdot X, c = 0.3$$
$$\mu_{cX} = E[c \cdot X] = c \cdot E[X]$$
$$\sigma_{cX} = St.Dev.[c \cdot X] = c \cdot SD[X]$$

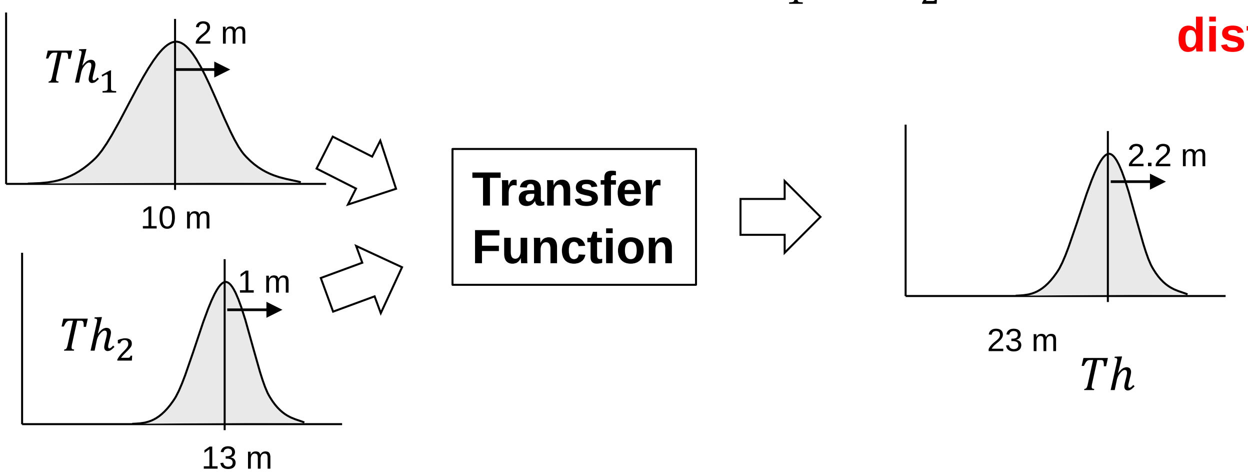
We can calculate the resulting uncertainty distribution analytically!

Monte Carlo Simulation Motivation

- **Problem 2:**
  - The reservoir is comprised of 2 units with thickness uncertainty for each:
    - » Unit 1 N[10,2] m (recall, Gaussian with mean = 10m and standard deviation = 2)
    - » Unit 2 N[13,1] m
  - **Calculate the uncertainty in the total thickness (Th)**

We can calculate the resulting uncertainty distribution analytically!

$$Th = Th_1 + Th_2$$



Total Distributed N[23,2.2] m

With expectation we can calculate $\mu_{X_1+X_2}$ and $\sigma_{X_1+X_2}$ if X_1 and X_2 are independent. Adding two Gaussian distributions results in a Gaussian distribution.

$$E[X_1 + X_2] = E[X_1] + E[X_2]$$

$$Var[X_1 + X_2] = Var[X_1] + Var[X_2] \quad \textbf{if } X_1 \perp\!\!\!\perp X_2$$

(independent random variables)

Monte Carlo Simulation Motivation

- **Problem 3:**
  - The reservoir has the following features with uncertainty:
    - Average Porosity ($\bar{\phi}$) ~ U[8,15] %
    - Thickness (Th) ~ LogN[1,1] m
    - Area (A) = 100,000 m^2 and Oil Saturation (s_o) = 0.8
  - **Calculate the uncertainty in the oil in place (OIP):**

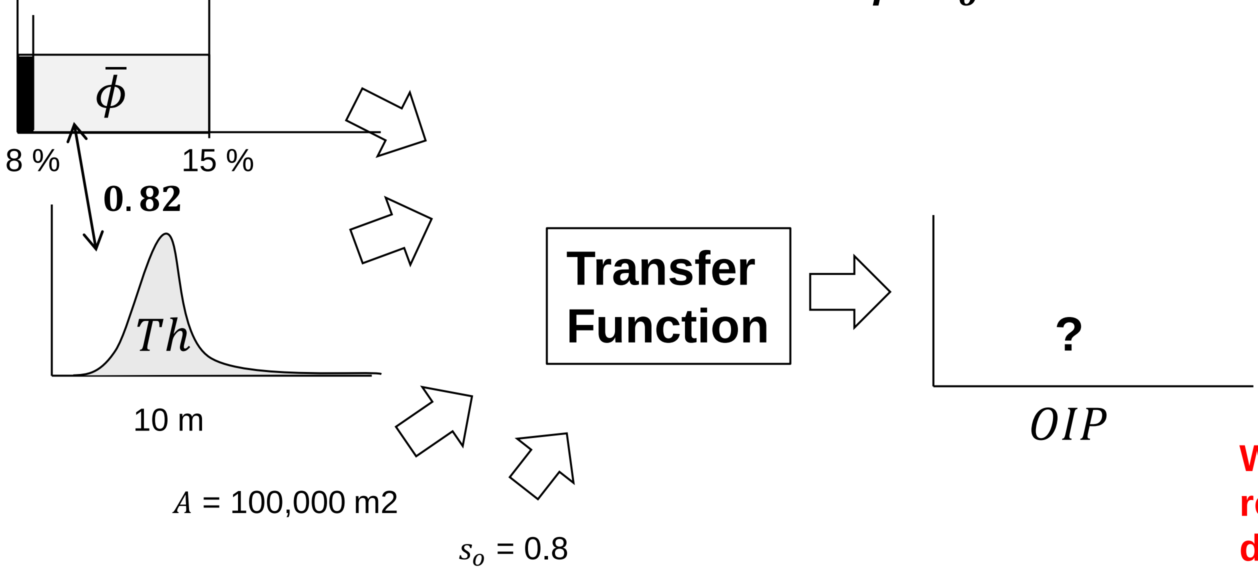
Correlation between features

$$\rho_{\bar{\phi},thick} = 0.82$$

Threshold on a feature

$$if\, \bar{\phi} < 9\% \text{ then } \bar{\phi} = 0$$

$$\boldsymbol{OIP = Th \cdot A \cdot \bar{\phi} \cdot s_o}$$



We CANNOT calculate the resulting uncertainty distribution analytically!

Monte Carlo Simulation Motivation

- **Problem 4:**
  - Solitaire Game – probability of winning:
    - » Large combinatorial of possible outcomes with each hand and cards on the table
    - » Card ordering matters
    - » Correlation / constraints imposed on hands from cards played
    - » Strategy / Discrete choices

Calculate the uncertainty distribution, probability of winning:

The Shuffled Deck ⇨ **Transfer Function** ⇨ Probability of Winning

The Game of Solataire

This is not trivial, many practical problems have large combinatorials, complicated correlations, ordering, correlation, constraints, discrete choices.

We CANNOT calculate this analytically. Why not just get the computer to play enough solitaire and observe / count the outcomes?

This is the idea of Monte Carlo Simulation.

Monte Carlo Simulation Uncertainty Modeling Workflow

- **The Common Steps for Monte Carlo Simulation:**

1. **Specify the model**/equation, f, and all required independent variables / predictor features, X, and dependent variable(s) / response feature(s), Y, $Y = f(X_1, \ldots, X_m)$

2. Specify the **uncertainty distribution for each predictor feature**, $X_1, \ldots, X_m$

3. Perform **Monte Carlo simulation from each distribution** to draw a realization from all $1, \ldots, m$ predictor feature distributions to get one realization of all predictor features, $x_1^\ell, \ldots, x_m^\ell$

4. Apply the realization of predictor features to the model to get a **realization of the output**, dependent variable, response feature, $Y^\ell = f(x_1^\ell, \ldots, x_m^\ell)$

5. **Repeat for $\ell = 1, \ldots, L$ realizations** to sufficiently sample the distribution of the response feature, this is your uncertainty model.

Monte Carlo Simulation Demonstration in Python

Things to Try:

1. **50% increase in the average porosity**
2. **Set saturation max as 0.9**
3. **Let the small L size to 10!**

Monte Carlo Simulation in Python demo in file: GeostatsPy_Monte_Carlo_simulation.ipynb

GeostatsPy: Monte Carlo Simulation for Subsurface Data Analytics in Python

Michael Pyrcz, Associate Professor, University of Texas at Austin

Twitter | GitHub | Website | GoogleScholar | Book | YouTube | LinkedIn

PGE 383 Exercise: Monte Carlo Simulation for Subsurface Data Analytics in Python

Here's a simple workflow, demonstration of Monte Carlo simulation for subsurface uncertainty modeling workflows. This should help you get started with building subsurface models that integrate uncertainty sources.

Monte Carlo Simulation

Definition: random sampling from a distribution

Procedure:

1. Model the representative distribution (CDF)
2. Draw a random value from a uniform [0,1] distribution (p-value)
3. Apply the inverse of the CDF to calculate the associated realization

In practice, Monte Carlo simulation refers to the workflow with multiple realizations drawn to buld an uncertainty model.

$$X^\ell = F_x(p^\ell), \ \forall \ \ell = 1, \dots, L$$

where X^ℓ is the realization of the variable X drawn from its CDF, F_x, with cumulative probability, p-value, p^ℓ.

It would be trivial to apply Monte Carlo simulation to a single variable, after many realizations one would get back the original distribution. The general approach is to:

1. Model all distributions for the input, variables of interest $F_{x_1}, \dots, F_{x_m}$.
2. For each realization draw $p_1^\ell, \dots, p_m^\ell$, p-values
3. Apply the inverse of each distribution to calculate a realization of each variable, $X_j^\ell = F_{x_j^\ell}^{-1}(p_j^\ell), \ \forall \ j = 1, \dots, m$ variables.
4. Apply each set of variables for a ℓ realization to the transfer function to calculate the ouptput realization, $Y^\ell = F(X_1^\ell, \dots, X_m^\ell)$.

Monte Carlo Simulation (MCS) is extremely powerful

- Possible to easily simulate uncertainty models for complicated systems
- Simulations are conducted by drawing values at random from specified uncertainty distributions for each variable
- A single realization of each variable, $X_1^\ell, X_2^\ell, \dots, X_m^\ell$ is applied to the transfer function to calculate the realization of the variable of interest (output, decision criteria):

$$Y^\ell = F(X_1^\ell, \dots, X_m^\ell), \ \forall \ \ell = 1, \dots, L$$

- The MCS method builds empirical uncertainty models by random sampling

Let's take a simple example, OIP is oil-in-place calculated as the product of reservoir volume, V, average porosity, $\overline{\phi}$, and oil saturaton, $\overline{s_o}$:

PGE 383 – Machine Learning

Uncertainty Management

Lecture outline . . .

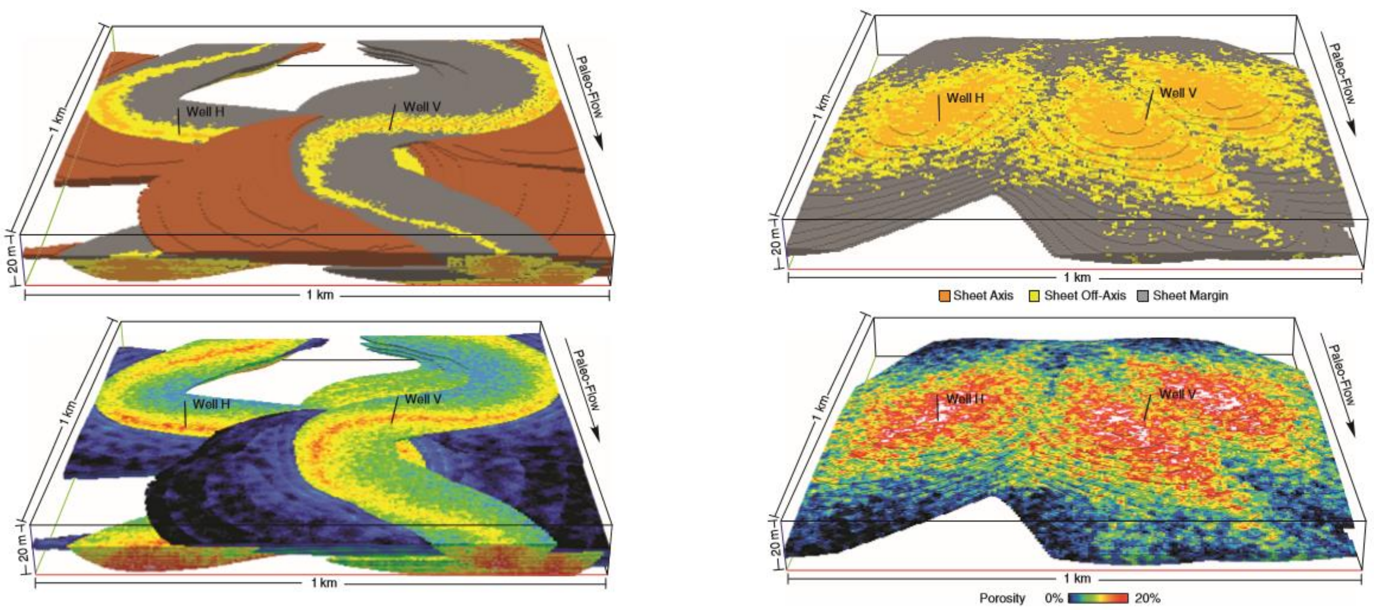
- **Scenarios**

Instructor: Michael Pyrcz, the University of Texas at Austin

Scenarios

Scenarios to Represent Discrete Possible Cases

- At times it is NOT possible to represent uncertainty as a continuous distribution.
- **Discrete scenarios are required**
  - Porosity compaction trend - yes or no.
  - Channelized or lobe uncertainty



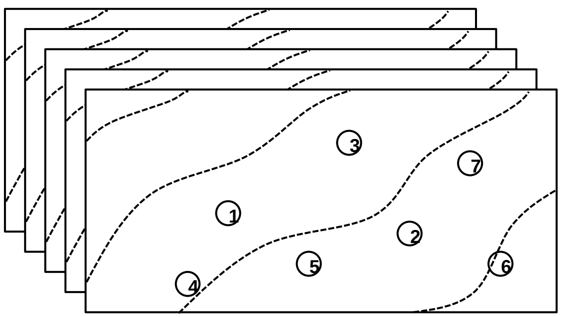
Uncertainty scenarios, left channels and right lobes.

How is Uncertainty Calculated?

Sample uncertainty through modeling scenarios and realizations:

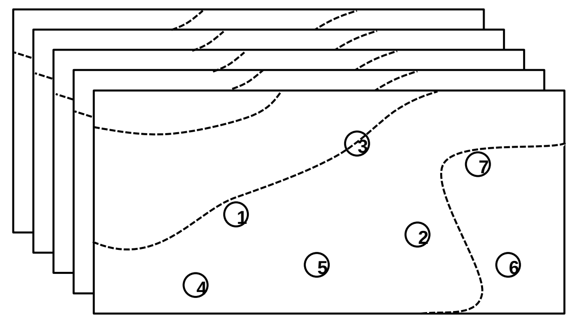
- **We can ask any question of the model by considering all scenarios and realizations jointly.**

Scenario 1



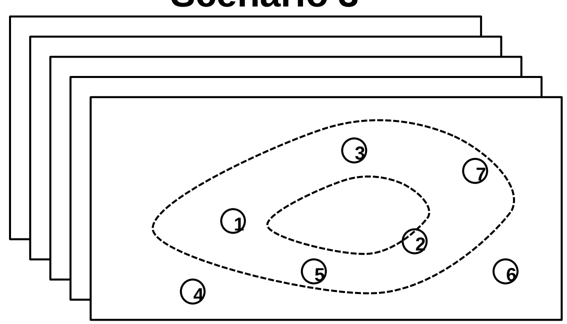
1,…,L realizations (all inputs the same)

Scenario 2



1,…,L realizations (all inputs the same)

Scenario 3



1,…,L realizations (all inputs the same)

PGE 383 – Machine Learning
Uncertainty Management

Lecture outline . . .

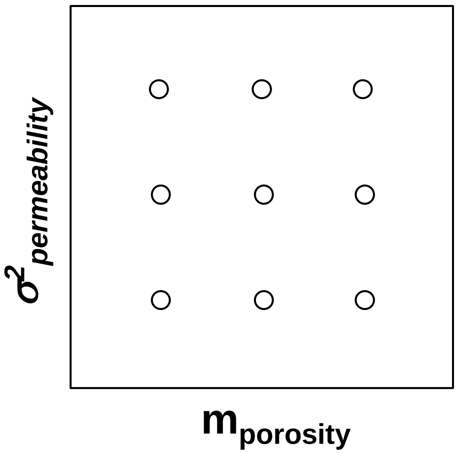
- **Sampling Uncertainty**

Instructor: Michael Pyrcz, the University of Texas at Austin

Sampling Uncertainty

Uncertainty Space is Vast!

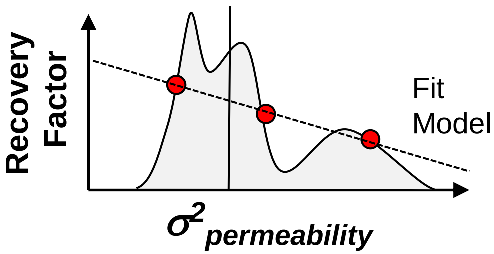
- Consider typical design



3 level design, 2 variables
level^{var} = 9 scenarios

- Consider there are typically are around 10 or more uncertain variables.
  - 3^{10} = 59,049 scenarios x 10 realizations of each scenario = 590,490 models
  - Variable screening is important!
  - 3 level may still poor sampling

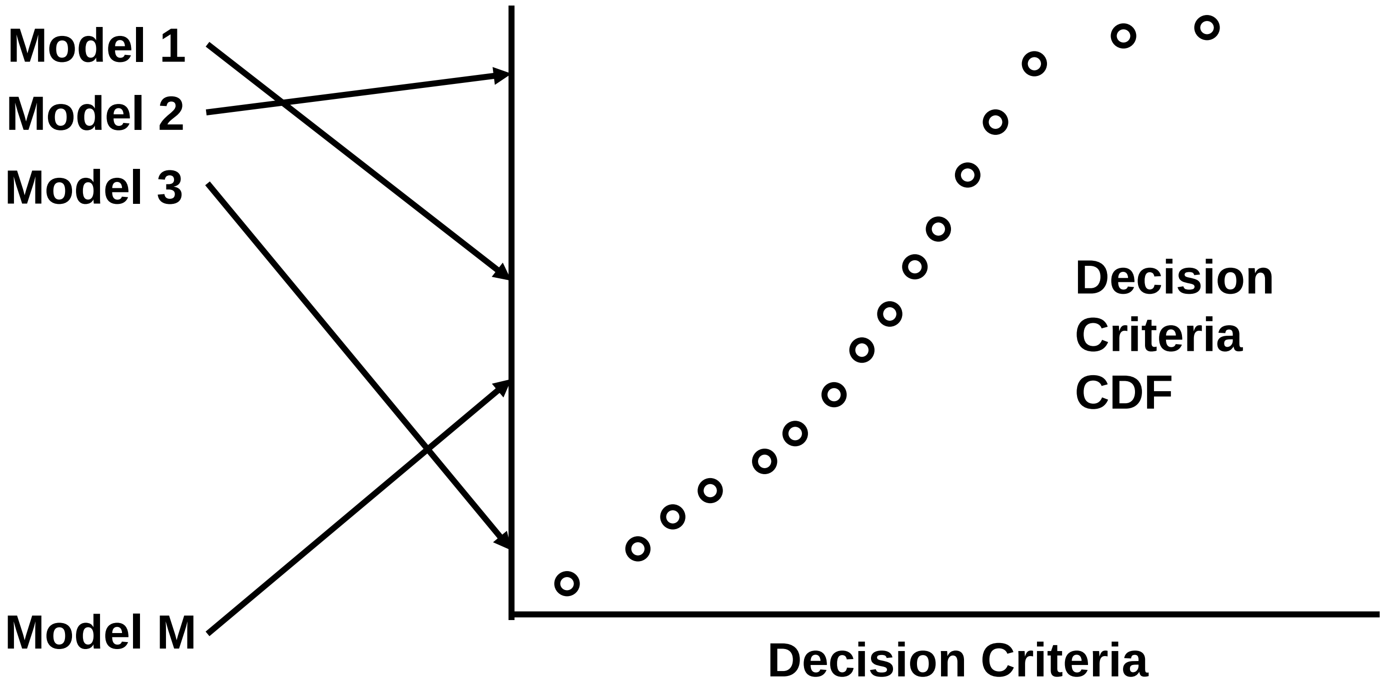
Univariate Response



Sampling Uncertainty

Sampling the Uncertainty Space

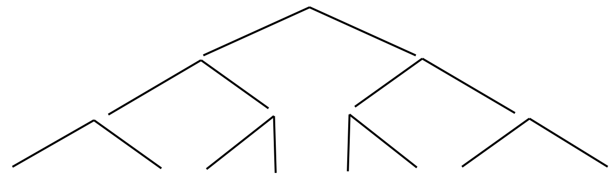
- Generally models are treated as equiprobably
- Equal-likely to be drawn realizations of the subsurface
- We may weight specific scenarios are more likely
- We are attempting to sample model the decision criteria CDF



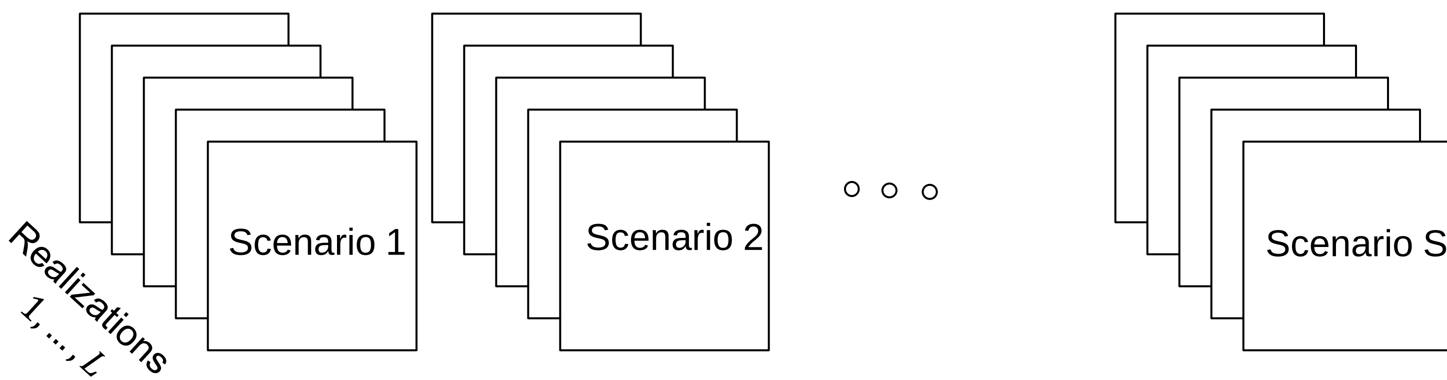
Sampling Uncertainty

Challenges:

1. Sampling a Vast Uncertainty Space



n model choices / parameters
k levels



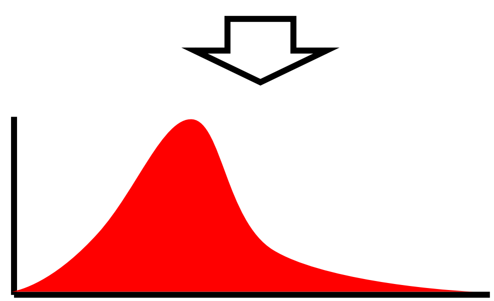
$S = k^n$

Scenarios $1, \ldots, S$

The Ensemble of Input(s) and Model(s)

- k^n scenarios
- L realizations

$$Models, M = k^n \cdot L$$



Model Predictions

2. Model Predictions

PGE 383 – Machine Learning
Uncertainty Management

Lecture outline . . .

- **Uncertainty Sources**
- **Uncertainty Philosophy**
- **The Bootstrap**
- **Monte Carlo Simulation**
- **Scenarios**
- **Sampling Uncertainty**

Instructor: Michael Pyrcz, the University of Texas at Austin