

www.inl.gov



RAVEN Forward Sampling and statistical analysis

RAVEN Workshop



Objectives

- Learn the “Entities” of a generic statistical analysis
- Learn how these “Entities” are implemented in RAVEN
- Learn how RAVEN Steps and Entities are assembled in the input file for statistical analysis
- Learn how to visualize output
- Learn how to perform a generic statistical analysis
- Learn how to perform a correlation analysis
- Basically, you should be able to start playing with RAVEN

How to Think About the Task

- RAVEN prospective
 - Prepare the environment: [`<RunInfo>`](#)
 - Describe the statistical property: [`<Distributions>`](#)
 - Decide the exploration strategy: [`<Samplers>`](#)
 - Set up the data containers: [`<DataObjects>`](#)
 - Define the actions: [`<Models>`](#)
 - Support files: [`<Files>`](#)
 - Define the exporting method [`<OutStreams>`](#)
 - Combine the “actors” [`<Steps>`](#)

Prepare the environment

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```
<Simulation>
  <RunInfo>
    <WorkingDir></WorkingDir>
    <Sequence>GenerateData,Plot</Sequence>
    <batchSize>4</batchSize>
  </RunInfo>

  ..
  ..
  ..

</Simulation>
```

Describe the statistical property

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```
<Distributions>
  <Normal name="Normal">
    <mean>0</mean>
    <sigma>0.2</sigma>
    <lowerBound>-1</lowerBound>
    <upperBound>1</upperBound>
  </Normal>
  <Uniform name="Uniform">
    <lowerBound>0</lowerBound>
    <upperBound>1000</upperBound>
  </Uniform>
</Distributions>
```

Decide the exploration strategy

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<Samplers>
  <MonteCarlo name="myMC">
    <samplerInit>
      <limit>2000</limit>
    </samplerInit>
    <variable name="X1">
      <distribution>Normal</distribution>
    </variable>
    <variable name="X2">
      <distribution>Uniform</distribution>
    </variable>
  </MonteCarlo>
</Samplers>

```

Set up the data containers

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```
<DataObjects>
  <PointSet name="DummyIN">
    <Input>X1,X2</Input>
  </PointSet>
  <PointSet name="Out">
    <Input>X1,X2</Input>
    <Output>Y1,Y2</Output>
  </PointSet>
</DataObjects>
```

Define the actions

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```
<Models>
  <ExternalModel ModuleToLoad="ForwardSamplerExampleModel" name="PythonModule" subType="">
    <variables>X1,X2,Y1,Y2</variables>
  </ExternalModel>
</Models>
```

```
def initialize(self , runInfoDict , inputFiles):
  self.const1 = 3.5
  return

def run(self , Input):
  self.Y1 = self.X1 * self.X2 + self.const1
  self.Y2 = 0.7 * self.X1 + self.X2 * self.const1
```

Define the exporting method

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<OutStreams>
  <Print name="fileOut">
    <type>csv</type>
    <source>Out</source>
  </Print>
  <Plot name="myPlot">
    <plotSettings>
      <plot>
        <type>scatter</type>
        <x>Out|Input|X1</x>
        <y>Out|Input|X2</y>
        <z>Out|Output|Y1</z>
        <colorMap>Out|Output|Y2</colorMap>
      </plot>
      <xlabel>X1</xlabel>
      <ylabel>X2</ylabel>
      <zlabel>Y1</zlabel>
    </plotSettings>
    <actions><how>png, screen</how></actions>
  </Plot>
</OutStreams>

```

Combine the “actors”

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<Steps>
  <MultiRun name="GenerateData">
    <Sampler class="Samplers" type="MonteCarlo">myMC</Sampler>
    <Input class="DataObjects" type="PointSet">DummyIN</Input>
    <Model class="Models" type="ExternalModel" >PythonModule</Model>
    <Output class="DataObjects" type="PointSet">Out</Output>
  </MultiRun>
  <IOStep name="Plot" pauseAtEnd="True">
    <Input class="DataObjects" type="PointSet">Out</Input>
    <Output class="OutStreams" type="Print" >fileOut</Output>
    <Output class="OutStreams" type="Plot" >myPlot</Output>
  </IOStep>
</Steps>

```

Running??

Adding a Distribution

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```
<Distributions>
...
<Exponential name="Exp">
  <lambda>8.7E-4</lambda>
</Exponential>
<Triangular name="DistTri">
  <apex>1</apex>
  <min>-0.1</min>
  <max>3</max>
</Triangular>
</Distributions>
```

Changing the Used Distributions

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<Samplers>
  <MonteCarlo name="myMC">
    <samplerInit>
      <limit>1000</limit>
    </samplerInit>
    <variable name="X1">
      <distribution>Exp</distribution>
    </variable>
    <variable name="X2">
      <distribution>DistTri</distribution>
    </variable>
  </MonteCarlo>
</Samplers>

```

Running??

Adding a New Sampler

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<Samplers>
.....
<Stratified name="myLHS">
  <samplerInit><initialSeed>142323</initialSeed></samplerInit>
  <variable name="X1">
    <distribution>Normal</distribution>
    <grid construction="equal" steps="1000" type="CDF">0. 1.</grid>
  </variable>
  <variable name="X2">
    <distribution>Uniform</distribution>
    <grid construction="equal" steps="1000" type="value">0. 2000.</grid>
  </variable>
</Stratified>
</Samplers>

```

Changing the Used Sampler

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<Steps>
  <MultiRun name="GenerateData">
    <Sampler class="Samplers" type="Stratified ">myLHS</Sampler>
    <Input class="DataObjects" type="PointSet">DummyIN</Input>
    <Model class="Models" type="ExternalModel" >PythonModule</Model>
    <Output class="DataObjects" type="PointSet">Out</Output>
  </MultiRun>
  <IOStep name="Plot" pauseAtEnd="True">
    <Input class="DataObjects" type="PointSet">Out</Input>
    <Output class="OutStreams" type="Print" >fileOut</Output>
    <Output class="OutStreams" type="Plot" >myPlot</Output>
  </IOStep>
</Steps>

```

Running??

Adding a New Sampler

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<Samplers>
...
<EnsembleForward name="myEnse">
  <MonteCarlo name="theMC">
    <samplerInit>
      <limit>100</limit>
    </samplerInit>
    <variable name="X2">
      <distribution>Uniform</distribution>
    </variable>
  </MonteCarlo>
  <Grid name="theGrid">
    <variable name="X1">
      <distribution>Normal</distribution>
      <grid construction="equal" type="CDF" steps="10">0 1</grid>
    </variable>
  </Grid>
</EnsembleForward>
</Samplers>

```

Changing the Used Sampler

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<Steps>
  <MultiRun name="GenerateData">
    <Sampler class="Samplers" type="EnsembleForward">myEnse</Sampler>
    <Input class="DataObjects" type="PointSet">DummyIN</Input>
    <Model class="Models" type="ExternalModel" >PythonModule</Model>
    <Output class="DataObjects" type="PointSet">Out</Output>
  </MultiRun>
  <IOStep name="Plot" pauseAtEnd="True">
    <Input class="DataObjects" type="PointSet">Out</Input>
    <Output class="OutStreams" type="Print" >fileOut</Output>
    <Output class="OutStreams" type="Plot" >myPlot</Output>
  </IOStep>
</Steps>

```

Running??

Basic Statistic Recall

- Expected Value
- Minimum
- Maximum
- Median
- Variance
- Sigma
- Variation Coefficient
- Skewness
- Kurtosis
- Samples
- percentile_5
- percentile_95

Sensitivity Recall

- Sensitivity: derivate
- Covariance: measure the degree of correlation in the variable dispersion
- Pearson, aka correlation (sigma normalized covariance): linearity measure
- Normalized sensitivity: derivative normalized by the mean
- Variance Dependent Sensitivity:

Basic Statistic Analysis and Sensitivity Inputs

- Starting from test1.xml
 - Change event sequence: [**<RunInfo>**](#)
 - Adding a post processor action [**<Models>**](#)
 - Adding the export file [**<Files>**](#)
 - Remove the plotting step [**<Steps>**](#)
 - Add the post processor step [**<Steps>**](#)

Changing event sequence (test5.xml)

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```
<Simulation>
  <RunInfo>
    <WorkingDir></WorkingDir>
    <Sequence>GenerateData,PlotPP</Sequence>
    <batchSize>4</batchSize>
  </RunInfo>

  ..
  ..
  ..

</Simulation>
```

Adding a post processor action

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<Models>
...
<PostProcessor name="Stat" subType="BasicStatistics">
...
  <expectedValue prefix="mean" > X1,X2,Y1,Y2 </expectedValue>
  <variance prefix="var" > X1,X2,Y1,Y2 </ variance>
  <kurtosis prefix="kurt" > X1,X2,Y1,Y2 </ kurtosis>
  <pearson>
    <targets>X1,X2,Y1,Y2</targets>
    <features>X1,X2,Y1,Y2</features>
  </pearson>
...
</PostProcessor>
</Models>

```

Adding Export File

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```
<Files>
  <Input name="Stat" type="">StaFile.csv</Input>
</Files>
```

Changing the Used Sampler (test4.xml)

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				

```

<Steps>
  <MultiRun name="GenerateData">
    <Sampler class="Samplers" type="EnsembleForward">myEnse</Sampler>
    <Input class="DataObjects" type="PointSet">DummyIN</Input>
    <Model class="Models" type="Dummy" >PythonModule</Model>
    <Output class="DataObjects" type="PointSet">Out</Output>
  </MultiRun>
  <IOStep name="Plot" pauseAtEnd="True">
    <Input class="DataObjects" type="PointSet">Out</Input>
    <Output class="OutStreams" type="Print" >fileOut</Output>
    <Output class="OutStreams" type="Plot" >myPlot</Output>
  </IOStep>
  <!-- -->
  <PostProcess name="PP">
    <Input class="DataObjects" type="PointSet">Out</Input>
    <Model class="Models" type="PostProcessor">Stat</Model>
    <Output class="Files" type="">Stat</Output>
  </PostProcess>
</Steps>

```

Running??

Playing with the Model

Changing the Analysis

RunInfo	Distributions	Samplers	DataObjects	Models	Files
OutStreams	Steps				