SAND 2009-0260P

DAKOTA Training

Sensitivity Analysis

http://www.cs.sandia.gov/dakota

Learning goals:

- **Define sensitivity analysis and know when and why to apply it**
- **Use a DAKOTA parameter study to perform local sensitivity analysis**
- **Use DAKOTA to perform LHS with correlation analysis; understand potential pitfalls**
- ***Understand use of orthogonal arrays***

*Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sensitivity Analysis and UQ Terminology

- **Sensitivity Analysis (SA):**
  - How do code outputs vary due to changes in code inputs?
  - Local sensitivity: code output gradient (derivative) data for a specific set (or sets) of code input parameter values
  - Global sensitivity: the general trends of the code outputs over the full range of code input parameter values (linear, quadratic, etc.)

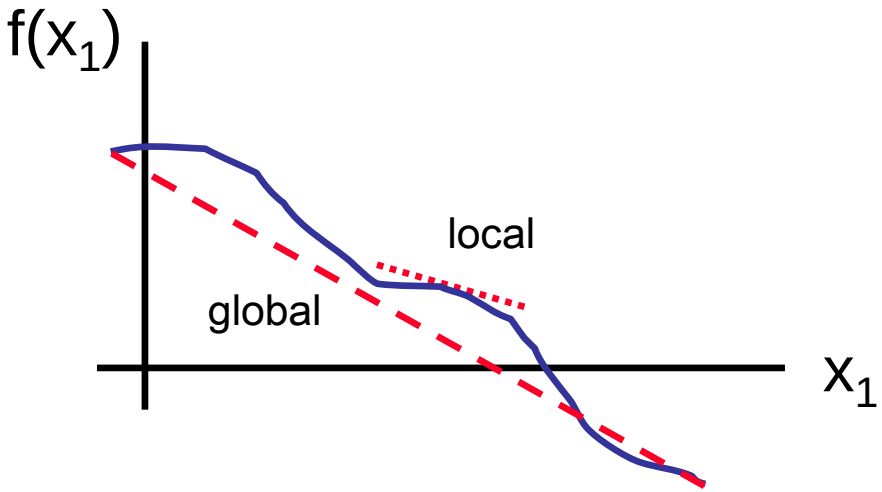
- **Uncertainty Quantification (UQ):**
  - What are the statistics or probability distributions on code outputs, given probability distributions on code inputs? (aleatoric UQ)
    - Estimate Probability[f > f*], e.g., probability of system failure
  - What are the possible/plausible code outputs? (epistemic UQ)

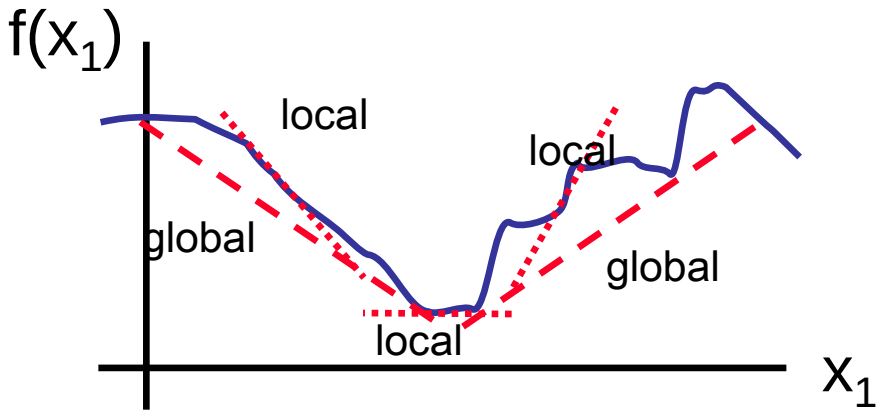
- **Validation:**
  - How "close" are my code output predictions to experimental data, including UQ for both?

- **Quantification of margins and uncertainties (QMU):**
  - How "close" are my code output predictions (including UQ) to the system's required performance level?

Sandia National Laboratories

Examples of Sensitivity Analysis





- **Sensitivity analysis examines variations in $f(x_1)$ due to perturbations in x_1.**
  - **Local sensitivities are typically partial derivatives.**
    - **Given a specific x1, what is the slope at that point?**
  - **Global sensitivities are typically found via regression.**
    - **What is the trend of the function over all values of x1?**

Sensitivity Analysis Methods

- **Sensitivity analysis methods:**
  - **Simple 1-parameter and multi-parameter studies***
  - **Importance factors***
  - **Random sampling and correlation analysis***
  - **Random sampling and analysis of variance**
  - **Standardized Regression Coefficients**
  - **Variance based decomposition***
  - **Orthogonal arrays**
  - **Many others....**

Workhorse methods

**** Capability in DAKOTA***

- **Relevant software tools:**
  - **DAKOTA**
  - **Minitab statistics package (SNL site license)**
  - **JMP statistics package (30 licenses for ASC users – contact D. Lifke)**
  - **Mathematica**
  - **Matlab with Statistics Toolbox**
  - **Others (Origin, etc.)**

Sandia National Laboratories

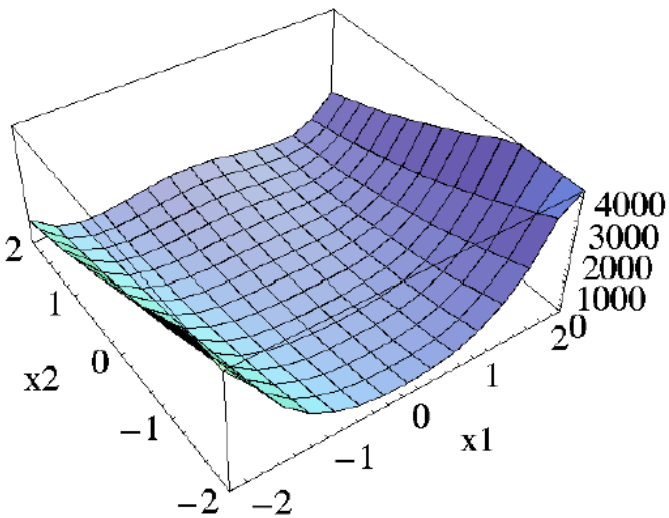
Sensitivity Analysis Example

- Explore sensitivity analysis concepts by applying DAKOTA to the Rosenbrock example:
  - Sensitivity analysis with gradients
  - Sensitivity analysis with sampling methods and correlation analysis

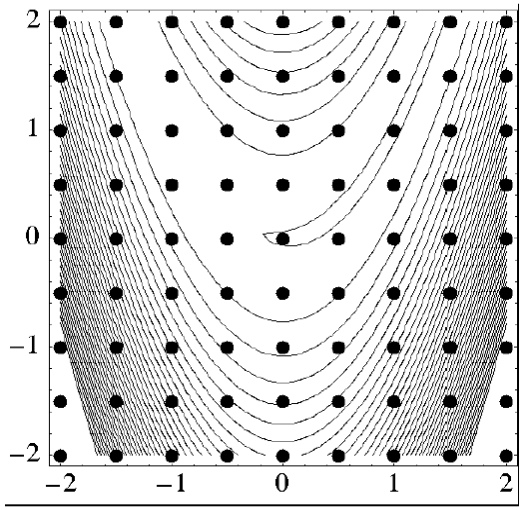
- Goal of sensitivity analysis: identify most important variables and their interactions
  - Provide a focus for resources for additional model and/or code development
  - Provide a basis for using surrogate models

Sensitivity Analysis with Derivatives

- Gradients (e.g., ∂R/∂x1 and ∂R/∂x2) typically estimated numerically via finite differences
- For inputs on radically different scales, use scaled derivatives
- Provides local sensitivity information
- Be wary of extrapolating trends
- No interaction data from this approach, but still useful
- Recommend against averaging derivatives across space: for Rosenbrock average ∂R/∂x1 = -2, ∂R/∂x2 = -533



<u>Goal:</u> Use a grid parameter study to understand how response f(x1,x2) varies with respect to the inputs x1 and x2



Rosenbrock Example: SA with Derivatives

```
strategy,
      single_method
        graphics
        tabular_graphics_data

method,
      multidim_parameter_study
        partitions = 2  2

model,
      single

variables,
      continuous_design = 2
        cdv_lower_bounds    -2.0     -2.0
        cdv_upper_bounds     2.0      2.0
        cdv_descriptors      'x1'     'x2'

interface,
      direct
        analysis_driver = 'rosenbrock‘

responses,
      num_objective_functions = 1
      numerical_gradients
      method_source dakota
      fd_gradient_step_size=1.e-6
      no_hessians
```

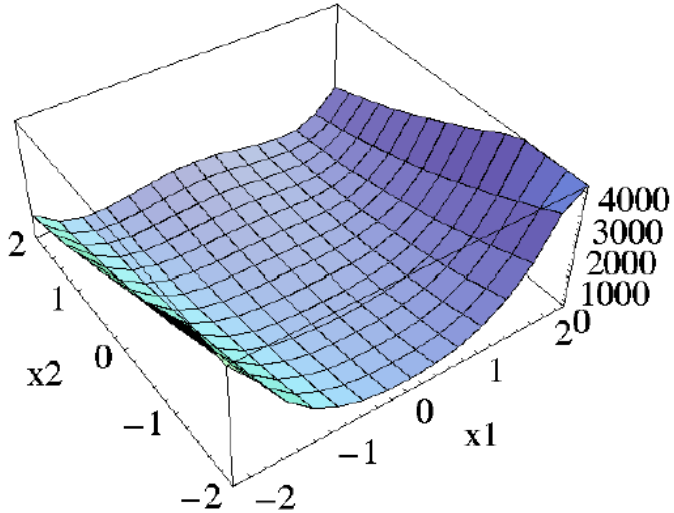
%eval_id	x1	x2	obj_fn
1	-2	-2	3609
2	0	-2	401
3	2	-2	3601
4	-2	-2	1609
5	0	-2	1
6	2	-2	1601
7	-2	-2	409
8	0	-2	401
9	2	-2	401

Parameter Study Output

```
Parameters for function evaluation 3:
                    -2.0000000000e+00 x1
                    -1.9999980000e+00 x2

Active response data for function evaluation 3:
Active set vector = { 1 }
                     3.6089976000e+03 response_fn_1

>>>>> Total response returned to iterator:

Active set vector = { 3 } Deriv vars vector = { 1 2 }
                     3.6090000000e+03 response_fn_1
 [ -4.8059943983e+03 -1.1999998003e+03 ] response_fn_1 gradient
```

Parameter study input: modified examples/tutorial/dakota_rosenbrock_2d.in

Results: Local SA for Rosenbrock Function

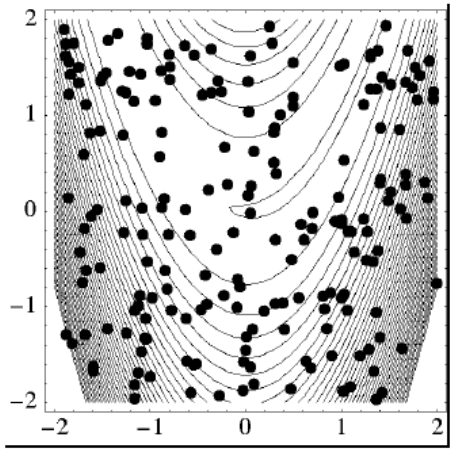


Goal:
Understand how response f(x1,x2) varies with respect to the inputs x1 and x2

x1	x1	x2	Rosenbrock	dR/dx1	dR/dx2
1	-2	-2	3609	-4.81E+03	-1.20E+03
2	0	-2	401	-2.00E+00	-4.00E+02
3	2	-2	3601	4.80E+03	-1.20E+03
4	-2	0	1609	-3.21E+03	-8.00E+02
5	0	0	1	-2.00E+00	9.99E-07
6	2	0	1601	3.20E+03	-8.00E+02
7	-2	2	409	-1.61E+03	-4.00E+02
8	0	2	401	-2.00E+00	4.00E+02
9	2	2	401	1.60E+03	-4.00E+02

Sensitivity Analysis with Sampling

- Assume inputs fall within lower and upper bounds
- Generate uniform random samples over these intervals
- Compute response value at each sample point
- Look at correlation results
  - Simple and partial correlations
  - Raw and rank correlations
- *Caution: correlation only measures the strength and direction of a linear relationship between variables*
- Correlation coefficient near
  - 0 indicates no relationship
  - 1 indicates strong positive relationship (as x increases, y increases)
  - -1 indicates strong negative relationship (as x increases, y decreases)



Rosenbrock Function: SA with Sampling

```
strategy,
    single_method
      graphics
      tabular_graphics_data

method,
      nond_sampling
      samples = 200 seed = 17
      sample_type random
      response_levels = 100.0

model,
    single

variables,
      uniform_uncertain = 2
       uuv_lower_bounds -2.0  -2.0
       uuv_upper_bounds  2.0   2.0
       uuv_descriptor        'x1'  'x2'

interface,
    direct
      analysis_driver = 'rosenbrock‘

responses,
      num_response_functions = 1
    no_gradients
    no_hessians
```

Sampling Input

```
Statistics based on 200 samples:

Moments for each response function:
response_fn_1:  Mean = 4.43855e+02  Std. Dev. = 5.88920e+02  Coeff. of
Variation = 1.32683e+00

95% confidence intervals for each response function:
response_fn_1:  Mean = ( 3.61736e+02, 5.25973e+02 ), Std Dev = ( 5.36308e+02,
6.53068e+02 )

Probabilities for each response function:
Cumulative Distribution Function (CDF) for response_fn_1:
     Response Level      Probability Level  Reliability Index  General Rel Index
     --------------      -----------------  -----------------  -----------------
   1.0000000000e+02   3.7000000000e-01

Simple Correlation Matrix among all inputs and outputs:
                    x1          x2 response_fn_1
          x1  1.00000e+00
          x2 -4.33667e-03  1.00000e+00
response_fn_1  6.45646e-02 -4.81363e-01  1.00000e+00

Partial Correlation Matrix between input and output:
             response_fn_1
          x1  7.12791e-02
          x2 -4.82094e-01
```

Sampling Output

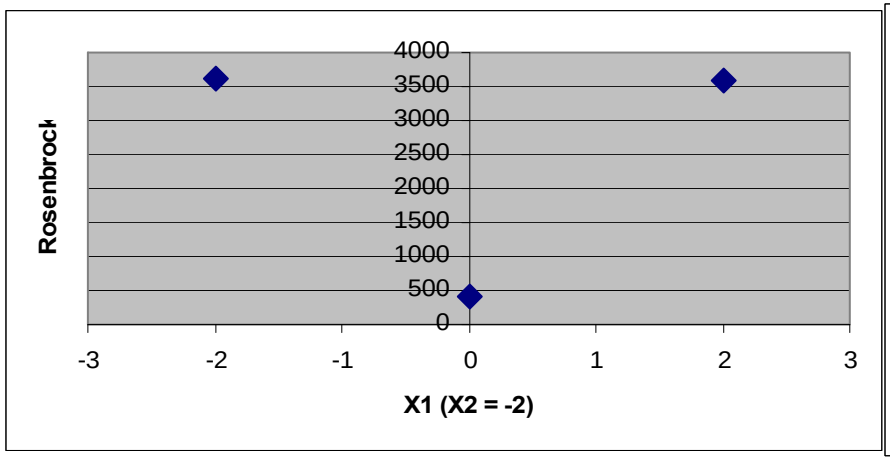
Global SA Results for Rosenbrock Function

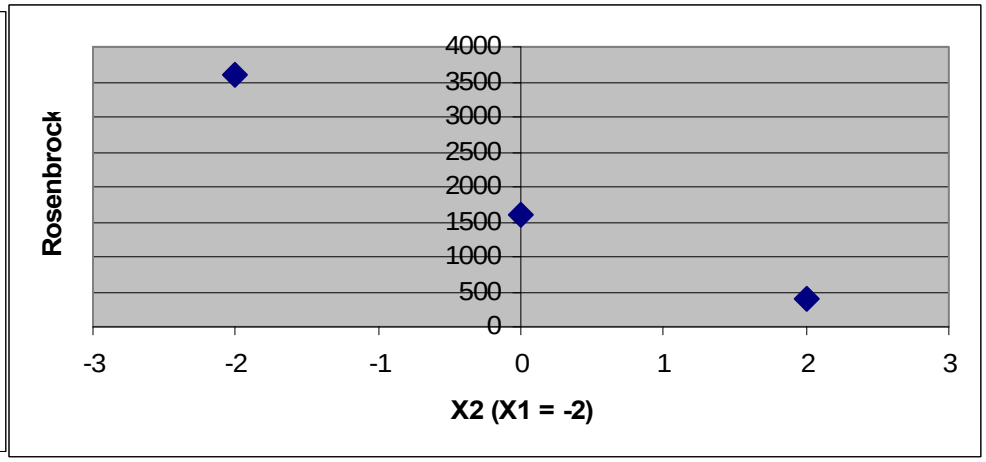
Simple Correlations

	x1	x2	response_fn_1
x1	1.00E+00		
x2	-8.15E-03	1.00E+00	
response_	-4.91E-02	-5.06E-01	1.00E+00

Partial Correlations

	response_fn_1
x1	-6.17E-02
x2	-5.07E-01





Global SA Results for Rosenbrock Function

- **Correlation between R and x1 is low because it is strongly negative as x1 goes from -2 to 0, then strongly positive as x1 goes from 0 to 2.**
- **If we re-define the input bounds between 0 and 2 instead of -2 to 2, we get a very different picture:**

	x1	x2	response_fn_1
x1	1.00E+00		
x2	-8.15E-03	1.00E+00	
response_fn_1	5.60E-01	-1.30E-01	1.00E+00

	response_fn_1
x1	5.64E-01
x2	-1.52E-01

Correlations are a great SA tool and very useful to screen out unimportant variables, but use with care!

Design of Experiments

- Design of Experiments (DOE) is sometimes used to help understand variable importance.

- Design and Analysis of Computer Experiments (DACE) refers to DOE for computer models.
- Big difference between physical and computer experiments: many of our codes are deterministic (e.g., same input settings will produce same outputs under replication), whereas physical experiments are usually not.

- DACE can be used to help understand range of outputs and important variables. It is generally NOT an uncertainty propagation method.

- Prototypical method: orthogonal arrays

Orthogonal Arrays

- For each level of one factor, all levels of the other factors occur an equal number of times: "cancel out" effect.
- Orthogonality: statistical independence between the columns of the experimental design matrix
- Standard analysis involves comparison of main effects: Is the mean of factor 1 at level 1 different than the mean of factor 1 at level 2?
- Large databases of OAs have been compiled by various industry and statistical organizations.
- Example:

Exp. No	Var. 1	Var. 2	Var. 3	Var. 4	Var. 5	Var. 6	Var. 7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	1	2
6	2	1	2	2	1	2	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

Example SA with Orthogonal Array

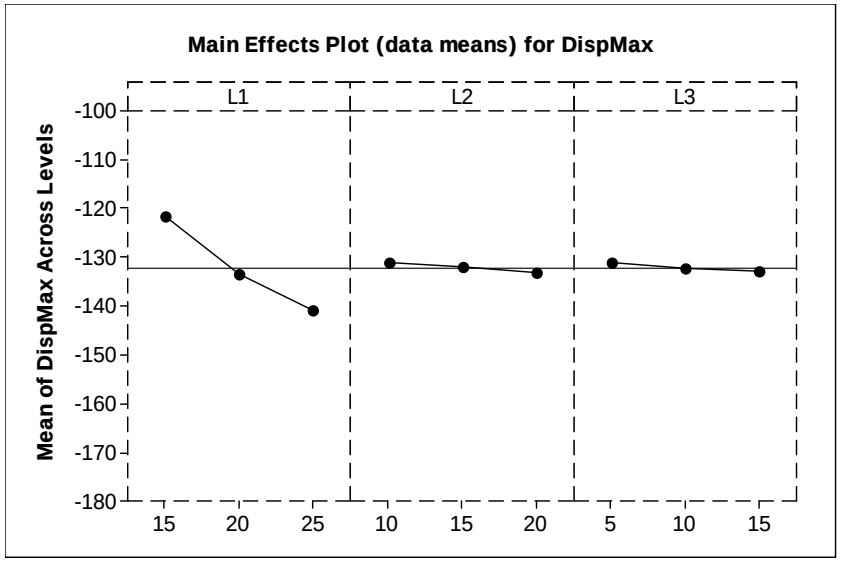
```
strategy,
      single_method
        graphics
        tabular_graphics_data

method,
      dace oas
        samples = 49 symbols = 7
        main_effects

model,
      single

variables,
      continuous_design = 2
        cdv_lower_bounds    -2.0     -2.0
        cdv_upper_bounds     2.0      2.0
        cdv_descriptors      'x1'     'x2'

interface,
      direct
        analysis_driver = 'rosenbrock‘

responses,
      num_objective_functions = 1
      no_gradients
      no_hessians
```

OA Input

```
>>>>> Running Single Method Strategy.

ANOVA Table for Factor (Variable) 1
Source of                    Sum of        Mean Sum
Variation          DoF       Squares       of Squares      Fdata          p-value
Between Groups       6   8.41566e+06   1.40261e+06   4.80464e+00   1.27778e-03
Within Groups       42   1.22610e+07   2.91929e+05
Total               48   2.06767e+07

ANOVA Table for Factor (Variable) 2
Source of                    Sum of        Mean Sum
Variation          DoF       Squares       of Squares      Fdata          p-value
Between Groups       6   6.20553e+06   1.03425e+06   3.00175e+00   2.66272e-02
Within Groups       42   1.44711e+07   3.44551e+05
Total               48   2.06767e+07
<<<<< Single Method Strategy completed.
```

OA Output

Orthogonal Arrays



Source	DF	SS	MS	F	P
L1	2	8430	4215	4.49	0.013
Error	132	123922	939		
Total	134	132352			

Level	N	Mean	StdDev
15	45	-121.83	31.06
20	45	-133.56	31.02
25	45	-141.03	29.82

Individual 95% CIs for Mean Based on Pooled StdDev=30.64

```
  +---------+---------+---------+---------
                     (--------*--------)
         (--------*--------)
(--------*--------)
  +---------+---------+---------+---------
-150      -140      -130      -120
```