SAND 2011-5917 C

Reliability Methods: Determining small failure probabilities

Laura P. Swiler and Michael S. Eldred
Sandia National Laboratories
P.O. Box 5800, Albuquerque, NM
87185-1318

lpswile@sandia.gov
505-844-8093

SAMSI UQ Program: Methodology Opening Workshop
September 7-10, 2011
Raleigh, NC

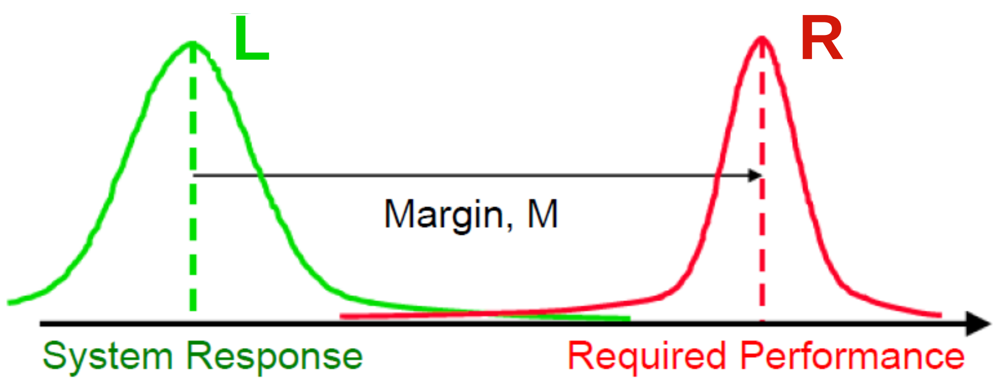
* Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Lecture Outline

- Safety Factor
- Mean Value Method
- FORM
- Advanced Mean Value
- SORM
- Limit State Approximations
- Probability Calculations
- Global Reliability Methods

Safety Factors

- Much of the early work on engineering reliability comes from the civil engineering field, concerned with reliability of structures
- In this lecture, the notation of L = load, R = resistance, we want L < R
- Nominal safety factor: SF = R_{nom}/L_{nom}, where $R_{nominal}$ is usually a conservative value (e.g. 2-3 standard deviations below the mean) and $L_{nominal}$ is also a conservative value (2-3 standard deviations above the mean)
- Problem: the nominal safety factor may not convey the true margin of safety in a design



Safety Factors

- Variety of approaches to improve a design
  - Increase the distance between the relative positions of the two curves: this reduces the probability of the overlapping area, and the probability of failure decreases
  - Reduce the dispersion of the two curves
  - Improve the shapes of the two curves

Probability of Failure

$$p_f = P(failure) = P(R < L)$$

$$p_f = \int_0^{\infty} \left[\int_0^{l} f_R(r)dr \right] f_L(l)dl$$

$$p_f = \int_0^{\infty} F_R(l) f_L(l)dl$$

In practice, this integration is hard to perform and doesn't always have an explicit form, except in some special cases

Probability of Failure

- **Special Case: R~N(μ_R, σ_R) , L~N(μ_L, σ_L)**
- **Define Z = R – L**

$$p_f = P(failure) = P(Z < 0)$$

$$p_f = \Phi\left[\frac{0-(\mu_R - \mu_L)}{\sqrt{\sigma_R^2 + \sigma_L^2}}\right]$$

$$p_f = 1 - \Phi\left[\frac{(\mu_R - \mu_L)}{\sqrt{\sigma_R^2 + \sigma_L^2}}\right]$$

- **There are also modifications which treat multiple loads, or lognormal distributions (Haldar and Mahadevan)**

Reliability Analysis

- Assume that the probability of failure is based on a specific performance criterion which is a function of random variables, denoted X_i.
- The performance function is described by Z:

  $Z = g(X_1, X_2, X_3, \ldots, X_n)$
- The failure surface or limit state is defined as Z = 0. It is a boundary between safe and unsafe regions in a parameter space.
- Now we have a more general form of $P_{failure}$

$$p_f = P(failure) = P(Z < 0)$$

$$p_f = \int \ldots \int_{g() < 0} f_X(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$$

Sandia National Laboratories

Reliability Analysis

- Note that the failure integral has the joint probability density function, f, for the random variables, and the integration is performed over the failure region

$$p_f = \int \ldots \int_{g() < 0} f_X(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n$$

- If the variables are independent, we can replace this with the product of the individual density functions
- In general, this is a multi-dimensional integral and is difficult to evaluate.
- People use approximations. If the limit state is a linear function of the inputs (or is approximated by one), first-order reliability methods (FORM) are used.
- If the nonlinear limit state is approximated by a second-order representation, second-order reliability methods (SORM) are used.

Mean Value Method (FOSM)

- Often called the First-Order Second-Moment (FOSM) method or the Mean Value FOSM method
- The FOSM method is based on a first-order Taylor series expansion of the performance function
- It is evaluated at the mean values of the random variables, and only uses means and covariances of the random variables
- The mean value method only requires one evaluation of the response function at the mean values of the inputs, plus *n* derivative values if one assumes the variables are independent → n+1 evaluations in the simplest approach (CHEAP!)

$$\mu_g = g(\mu_{\mathbf{x}})$$

$$\sigma_g^2 = \sum_{i=1}^{n} \sum_{i=1}^{n} Cov(i, j) \frac{dg}{dx_i}(\mu_{\mathbf{x}}) \frac{dg}{dx_j}(\mu_{\mathbf{x}})$$

$$\sigma_g^2 = \sum_{i=1}^{n} \left(\frac{dg}{dx_i}(\mu_{\mathbf{x}}) \right)^2 Var(x_i)$$

Mean Value Method (FOSM)

- Introduce the idea of a safety index β (think of this as how far in "normal space" that your design is away from failure)

$$\beta = \frac{\mu_g}{\sigma_g}$$

$$p_f = \Phi[-\beta] = 1 - \Phi[\beta]$$

- FOSM does not use distribution information when it is available
- When g(x) is nonlinear, significant error may be introduced by neglecting higher order terms in the expansion
- The safety index fails to be constant under different problem formulations
- It can be very efficient. When g(x) is linear and the input variables are normal, the mean value method gives exact results!

Mean Value Method (FOSM)

Some extensions/notation

$$\bar{z} \Rightarrow p, \beta$$

$$\bar{p}, \bar{\beta} \Rightarrow z$$

$$\beta_{cdf} = \frac{\mu_g - \bar{z}}{\sigma_g},$$

$$\beta_{ccdf} = \frac{\bar{z} - \mu_g}{\sigma_g}$$

$$z = \mu_g - \sigma_g \bar{\beta}_{cdf},$$

$$z = \mu_g + \sigma_g \bar{\beta}_{ccdf}$$

$$p_f = \Phi[-\beta] = 1 - \Phi[\beta]$$

p = probability of failure

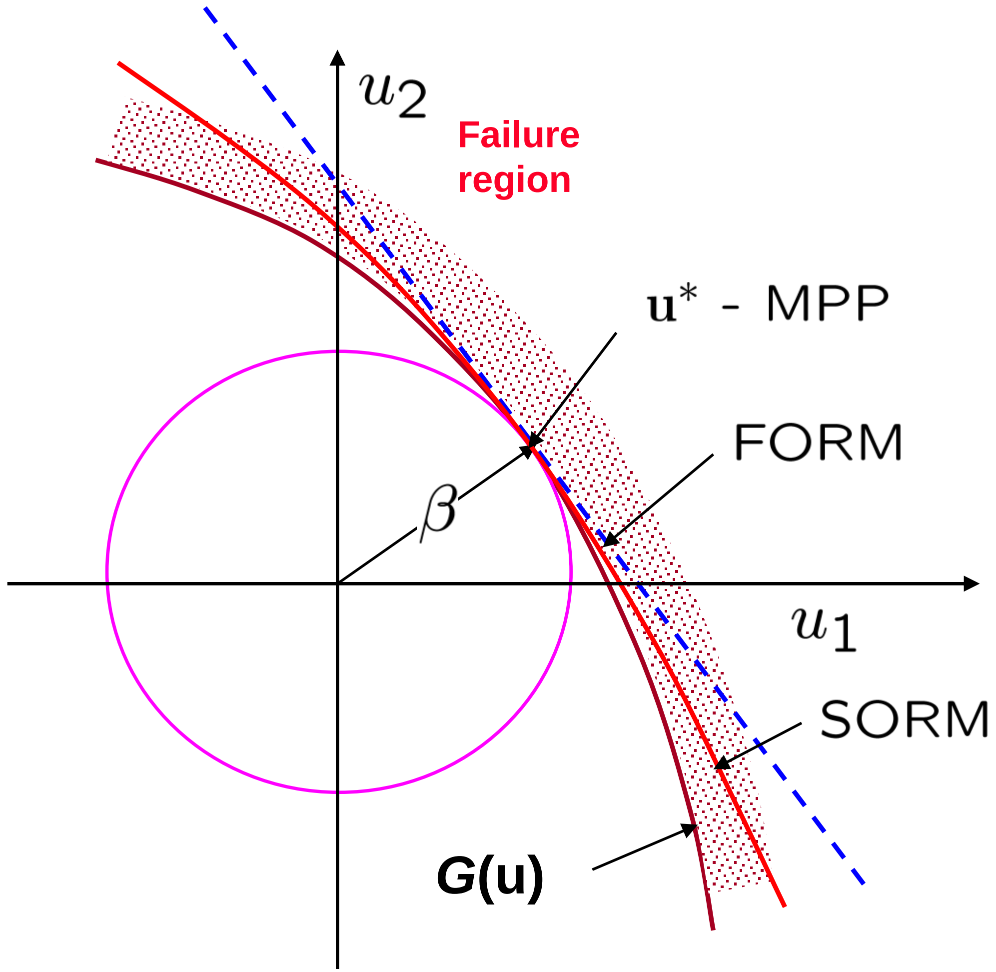
β**= reliability index**

z = response level

Most Probable Point Methods

- Transform the uncertainty propagation problem into an optimization one: first transform all of the non-normal random variables into independent, unit normal variables. Then, find the point on the limit state surface with minimum distance to the origin.
- The point is called the Most Probable Point (MPP). The minimum distance, β, is called the safety index or reliability index.
- X is often called the original space, U is the transformed space.

MPP Search Methods



Sandia National Laboratories

Uncertainty Transformations

- Want to go from correlated non-normals to uncorrelated standard normals (*u*)
- Several methods
  - Rosenblatt
  - Rackwitz-Fiesler
  - Chen-Lind
  - Wu-Wirshing
  - Nataf
- Rosenblatt: First transform a set of arbitrarily, correlated random variables $X_1 \ldots X_n$ to uniform distributions, then transform to independent normals.
- Nataf: First transform to correlated normals (*z*), then to independent normals *u.* L is the Cholesky factor of the correlation matrix

$$U_1 = F_{X_1}(X_1)$$
$$U_2 = F_{X_2|X_1}(X_2 \mid x_1)$$
$$\ldots$$
$$U_n = F_{X_n|X_1,X_2,\ldots}(X_n \mid x_1, x_2, \ldots x_{n-1})$$

$$u_1 = \Phi^{-1}(U_1)$$
$$u_2 = \Phi^{-1}(U_2)$$
$$\ldots$$
$$u_n = \Phi^{-1}(U_n)$$

$$\Phi(z_i) = F(x_i)$$
$$\mathbf{z} = \mathbf{L}\mathbf{u}$$

MPP Search Methods

Reliability Index Approach (RIA)

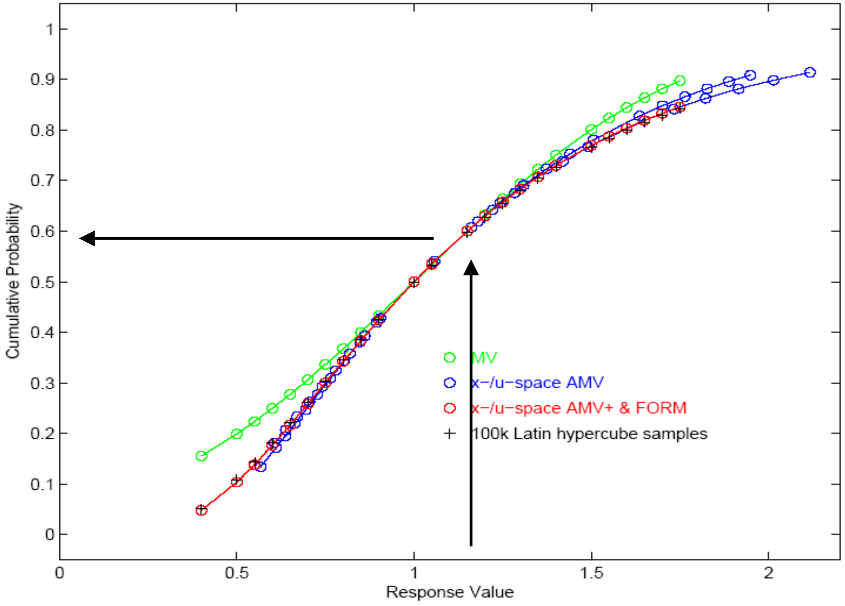
$$\text{minimize} \quad \mathbf{u}^T\mathbf{u}$$
$$\text{subject to} \quad G(\mathbf{u}) = \bar{z}$$

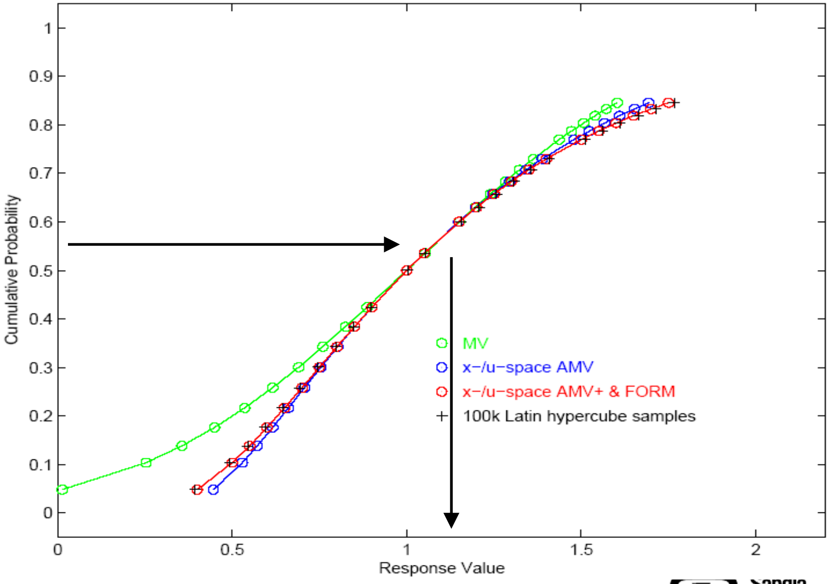
- **Find min dist to *G* level curve**
- **Used for fwd map *z* → *p*/*β***

Performance Measure Approach (PMA)

$$\text{minimize} \quad \pm G(\mathbf{u})$$
$$\text{subject to} \quad \mathbf{u}^T\mathbf{u} = \bar{\beta}^2$$

- **Find min *G* at *β* radius**
- **Used for inv map *p*/*β* → *z***





Sandia National Laboratories

Reliability Algorithm Variations: First-Order Methods

Limit state linearizations

AMV: $g(\mathbf{x}) = g(\mu_{\mathbf{x}}) + \nabla_x g(\mu_{\mathbf{x}})^T (\mathbf{x} - \mu_{\mathbf{x}})$

u-space AMV: $G(\mathbf{u}) = G(\mu_{\mathbf{u}}) + \nabla_u G(\mu_{\mathbf{u}})^T (\mathbf{u} - \mu_{\mathbf{u}})$

AMV+: $g(\mathbf{x}) = g(\mathbf{x}^*) + \nabla_x g(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*)$

u-space AMV+: $G(\mathbf{u}) = G(\mathbf{u}^*) + \nabla_u G(\mathbf{u}^*)^T (\mathbf{u} - \mathbf{u}^*)$

FORM: no linearization

Integrations

1st-order:
$$\begin{cases} p(g \le z) & = & \Phi(-\beta_{cdf}) \\ p(g > z) & = & \Phi(-\beta_{ccdf}) \end{cases}$$

MPP search algorithm

[HL-RF], Sequential Quadratic Prog. (SQP), Nonlinear Interior Point (NIP)

Warm starting

When: **AMV+ iteration increment, *z/p/β* level increment, or design variable change**

What: **linearization point & assoc. responses (AMV+) and MPP search initial guess**

Sandia National Laboratories

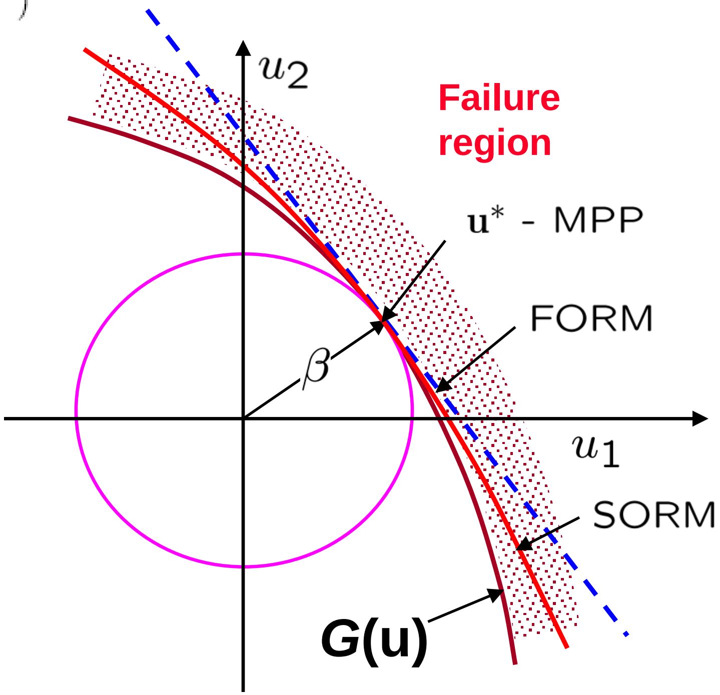
Reliability Algorithm Variations: Second-Order Methods

2nd-order local limit state approximations

- e.g., x-space AMV^2+:

$$g(\mathbf{x}) \cong g(\mathbf{x}^*) + \nabla_x g(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \nabla_x^2 g(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*)$$

- Hessians may be full/FD/Quasi
- Quasi-Newton Hessians may be BFGS or SR1



2nd-order integrations

$$p = \Phi(-\beta) \prod_{i=1}^{n-1} \frac{1}{\sqrt{1 + \beta\kappa_i}}$$

curvature correction

Synergistic features:

- Hessian data needed for SORM integration can enable more rapid MPP convergence
- [QN] Hessian data accumulated during MPP search can enable more accurate probability estimates

Reliability Algorithm Variations: Second-Order Methods

Multipoint limit state approximations

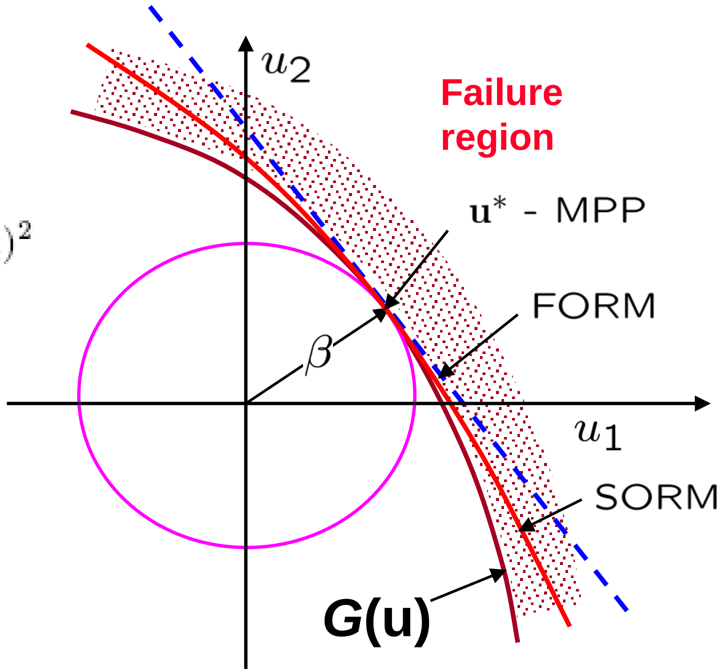
- **e.g., TPEA, TANA:**

$$g(\mathbf{x}) \cong g(\mathbf{x}_2) + \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{x}_2)\frac{x_{i,2}^{1-p_i}}{p_i}(x_i^{p_i} - x_{i,2}^{p_i}) + \frac{1}{2}\epsilon(\mathbf{x})\sum_{i=1}^{n}(x_i^{p_i} - x_{i,2}^{p_i})^2$$

$$p_i = 1 + \ln\left[\frac{\frac{\partial g}{\partial x_i}(\mathbf{x}_1)}{\frac{\partial g}{\partial x_i}(\mathbf{x}_2)}\right] \Big/ \ln\left[\frac{x_{i,1}}{x_{i,2}}\right]$$

$$\epsilon(\mathbf{x}) = \frac{H}{\sum_{i=1}^{n}(x_i^{p_i} - x_{i,1}^{p_i})^2 + \sum_{i=1}^{n}(x_i^{p_i} - x_{i,2}^{p_i})^2}$$

$$H = 2\left[g(\mathbf{x}_1) - g(\mathbf{x}_2) - \sum_{i=1}^{n}\frac{\partial g}{\partial x_i}(\mathbf{x}_2)\frac{x_{i,2}^{1-p_i}}{p_i}(x_{i,1}^{p_i} - x_{i,2}^{p_i})\right]$$

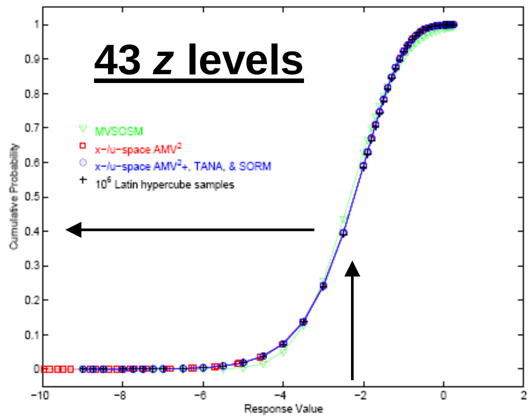


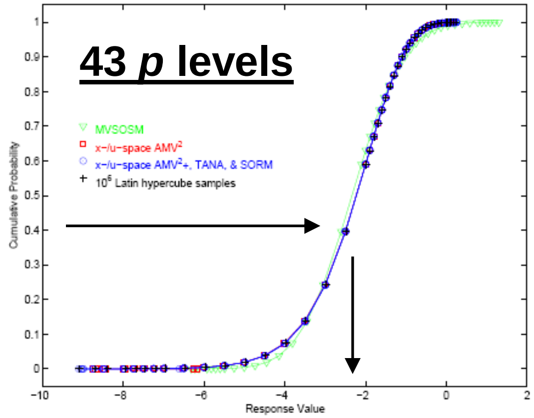
Importance Sampling

- **Use of importance sampling to calculate prob of failure:**
  - **After MPP is identified, sample around MPP to estimate P_f more accurately**

Reliability Algorithm Variations: Sample Results

Analytic benchmark test problems: lognormal ratio, short column, cantilever



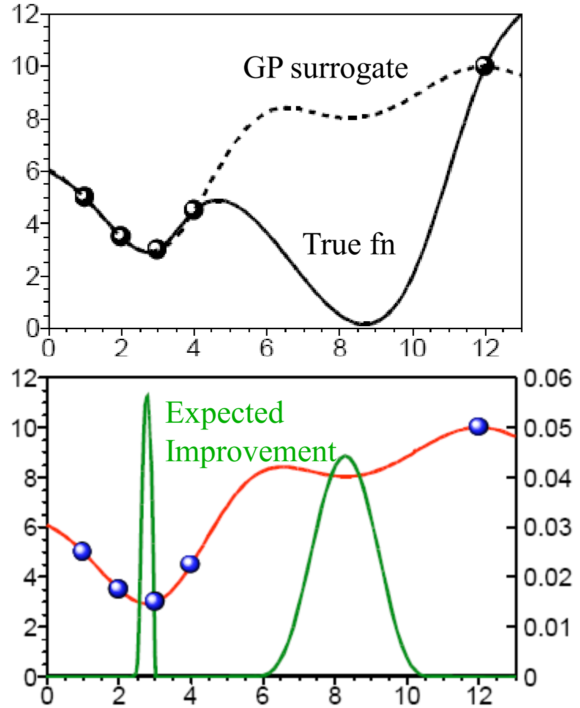


RIA Approach	SQP Function Evaluations	NIP Function Evaluations	CDF p Error Norm	Target z Offset Norm
MVFOSM	1	1	0.1548	0.0
MVSOSM	1	1	0.1127	0.0
x-space AMV	45	45	0.009275	18.28
u-space AMV	45	45	0.006408	18.81
x-space AMV^2	45	45	0.002063	2.482
u-space AMV^2	45	45	0.001410	2.031
x-space AMV+	192	192	0.0	0.0
u-space AMV+	207	207	0.0	0.0
x-space AMV^2+	125	131	0.0	0.0
u-space AMV^2+	122	130	0.0	0.0
x-space TANA	245	246	0.0	0.0
u-space TANA	296*	278*	6.982e-5	0.08014
FORM	626	176	0.0	0.0
SORM	669	219	0.0	0.0

PMA Approach	SQP Function Evaluations	NIP Function Evaluations	CDF z Error Norm	Target p Offset Norm
MVFOSM	1	1	7.454	0.0
MVSOSM	1	1	6.823	0.0
x-space AMV	45	45	0.9420	0.0
u-space AMV	45	45	0.5828	0.0
x-space AMV^2	45	45	2.730	0.0
u-space AMV^2	45	45	2.828	0.0
x-space AMV+	171	179	0.0	0.0
u-space AMV+	205	205	0.0	0.0
x-space AMV^2+	135	142	0.0	0.0
u-space AMV^2+	132	139	0.0	0.0
x-space TANA	293*	272	0.04259	1.598e-4
u-space TANA	325*	311*	2.208	5.600e-4
FORM	720	192	0.0	0.0
SORM	535	191*	2.410	6.522e-4

Efficient Global Reliability Analysis (EGRA)

- **Address known failure modes of local reliability methods:**
  - Nonsmooth: fail to converge to an MPP
  - Multimodal: only locate one of several MPPs
  - Highly nonlinear: low order limit state approxs. fail to accurately estimate probability at MPP
- **Based on EGO (surrogate-based global opt.), which exploits special features of GPs**
  - Mean and variance predictions: formulate expected improvement (EGO) or expected feasibility (EGRA)
  - Balance explore and exploit in computing an optimum (EGO) or locating the limit state (EGRA)

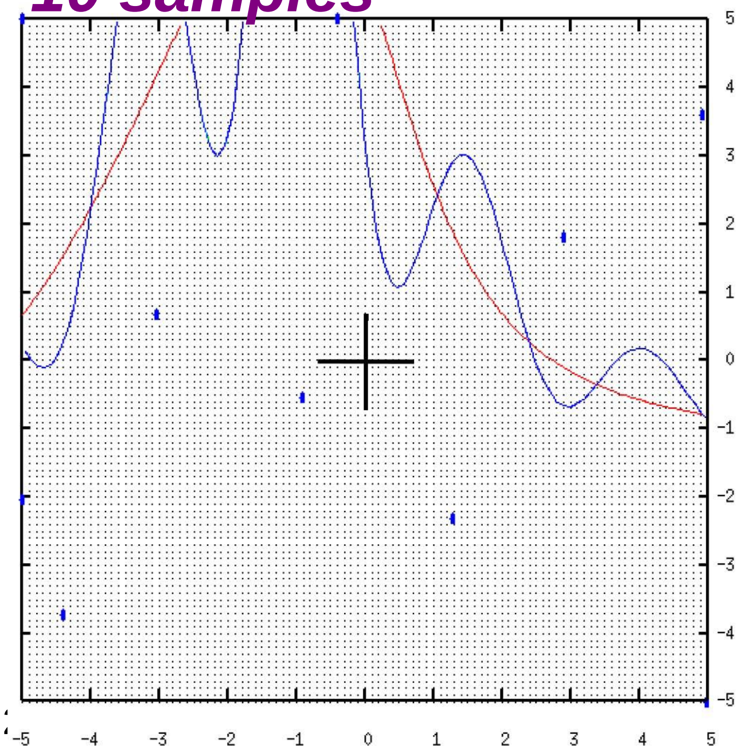


From Jones, Schonlau, Welch, 1998

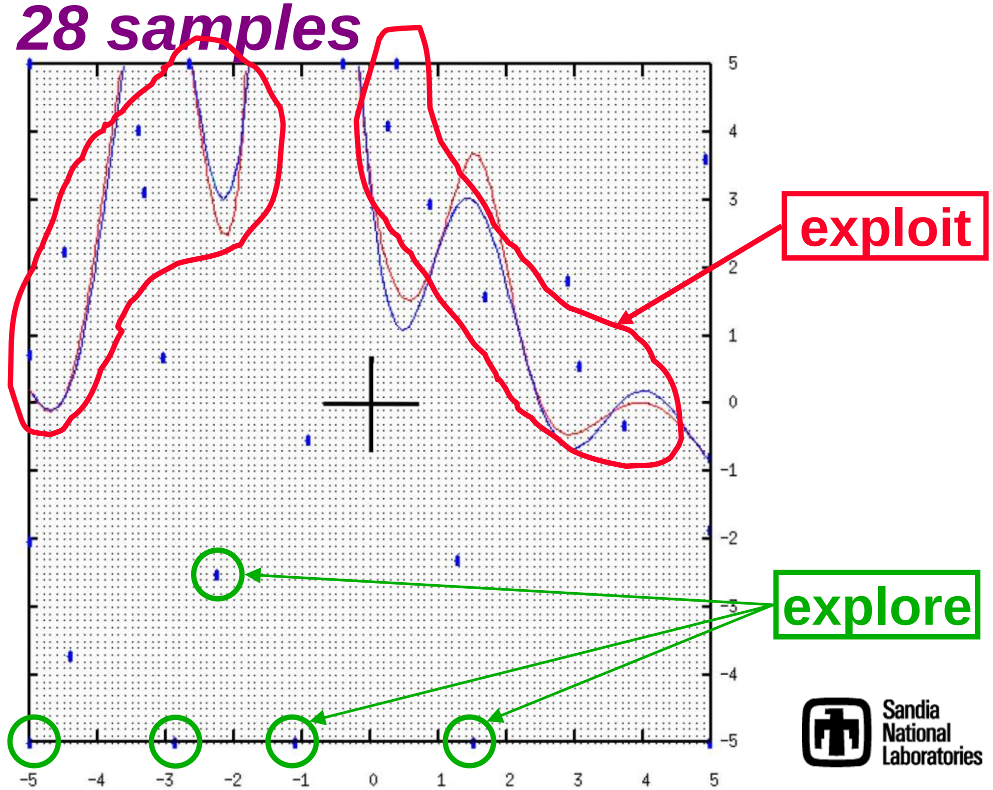
Efficient Global Reliability Analysis (EGRA)

- **Address known failure modes of local reliability methods:**
  - Nonsmooth: fail to converge to an MPP
  - Multimodal: only locate one of several MPPs
  - Highly nonlinear: low order limit state approxs. fail to accurately estimate probability at MPP
- **Based on EGO (surrogate-based global opt.), which exploits special features of GPs**
  - Mean and variance predictions: formulate expected improvement (EGO) or expected feasibility (EGRA)
  - Balance explore and exploit in computing an optimum (EGO) or locating the limit state (EGRA)

10 samples



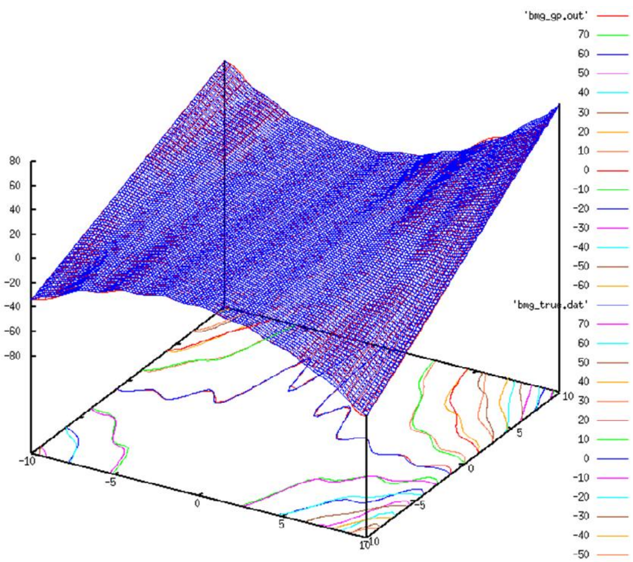
28 samples

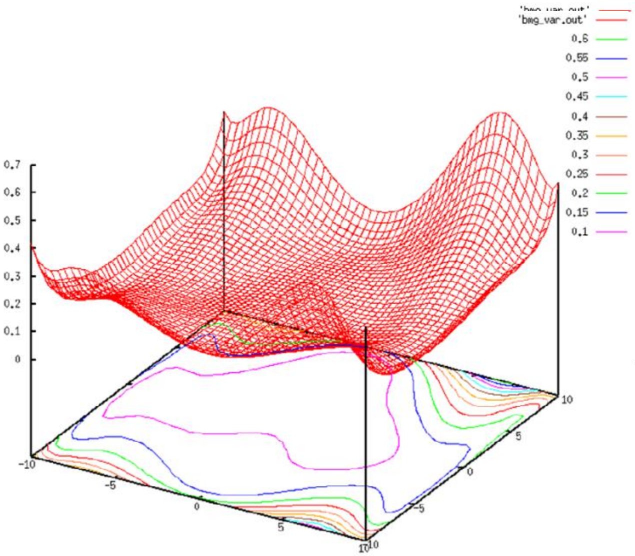


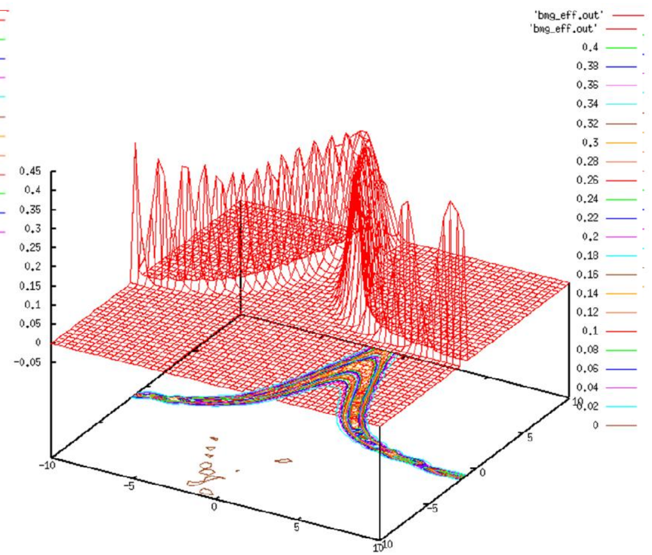
Sandia National Laboratories

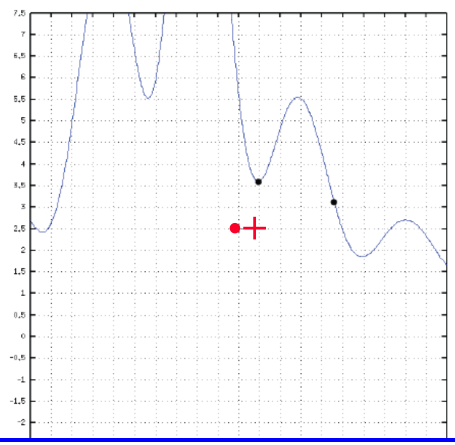
Efficient Global Reliability Analysis (EGRA)

Mean | **Variance** | **Expected Feasibility**







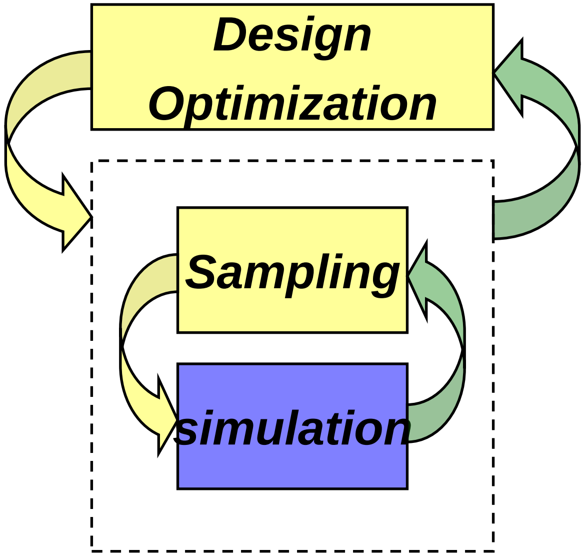


Reliability Method	Function Evaluations	First-Order p_f (% Error)	Second-Order p_f (% Error)	Sampling p_f (% Error, Avg. Error)
No Approximation	70	0.11797 (277.0%)	0.02516 (-19.6%)	—
x-space AMV^2+	26	0.11797 (277.0%)	0.02516 (-19.6%)	—
u-space AMV^2+	26	0.11777 (277.0%)	0.02516 (-19.6%)	—
u-space TANA	131	0.11797 (277.0%)	0.02516 (-19.6%)	—
LHS solution	10k	—	—	0.03117 (0.385%, 2.847%)
LHS solution	100k	—	—	0.03126 (0.085%, 1.397%)
LHS solution	1M	—	—	0.03129 (truth , 0.339%)
x-space EGRA	35.1	—	—	0.03134 (0.155%, 0.433%)
u-space EGRA	35.2	—	—	0.03133 (0.136%, 0.296%)

•Accuracy similar to exhaustive sampling at cost similar to local reliability assessment

Optimization under Uncertainty

- Design for reliability is a classic OUU problem, often called RBDO (reliability-based design optimization)
- Nice properties in that the reliability formulation itself generates quantities such as derivatives of performance function with respect to uncertain variables
- Variety of approaches (next page)
- Simplest case: think of a “nested” algorithm, with an optimization outer loop and sampling inner loop



Sandia National Laboratories

RBDO Algorithms

Bi-level RBDO

- **Constrain RIA $z \rightarrow p/\beta$ result**
- **Constrain PMA $p/\beta \rightarrow z$ result**

RIA RBDO

$$\begin{cases} \text{minimize} & f \\ \text{subject to} & \beta \geq \bar{\beta} \\ \text{or} & p \leq \bar{p} \end{cases}$$

PMA RBDO

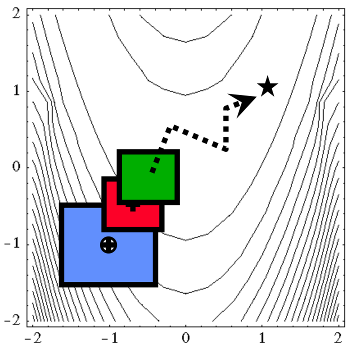
$$\begin{cases} \text{minimize} & f \\ \text{subject to} & z \geq \bar{z} \end{cases}$$

Sequential/Surrogate-based RBDO:

- **Break nesting: iterate between opt & UQ until target is met. Trust-region surrogate-based approach is non-heuristic.**

$$\begin{array}{rl} \text{minimize} & f(\mathbf{d}_0) + \nabla_d f(\mathbf{d}_0)^T(\mathbf{d} - \mathbf{d}_0) \\ \text{subject to} & \beta(\mathbf{d}_0) + \nabla_d \beta(\mathbf{d}_0)^T(\mathbf{d} - \mathbf{d}_0) \geq \bar{\beta} \\ & \| \mathbf{d} - \mathbf{d}_0 \|_\infty \leq \Delta^k \end{array}$$

1st-order (also 2nd-order w/ QN)



Unilevel RBDO:

- **All at once: apply KKT conditions of MPP search as equality constraints**
  - **Opt. increases in scale (*d*,*u*)**
  - **Requires 2nd-order info for derivatives of 1st-order KKT**

$$\min_{\mathbf{d}_{aug}=(\mathbf{d},\mathbf{u}_1,..,\mathbf{u}_{N_{hard}})} : f(\mathbf{d},\mathbf{p},\mathbf{y}(\mathbf{d},\mathbf{p}))$$

$$\text{s. t.} \quad : G_i^R(\mathbf{u}_i,\eta) = 0$$

$$\beta_{allowed} - \beta_i \geq 0$$

$$\|\mathbf{u}_i\| \, \|\nabla_{\mathbf{u}} G_i^R(\mathbf{u}_i,\eta)\| + \mathbf{u}_i^T \nabla_{\mathbf{u}} G_i^R(\mathbf{u}_i,\eta) = 0$$

$$\beta_i = \|\mathbf{u}_i\|$$

$$\mathbf{d}^l \leq \mathbf{d} \leq \mathbf{d}^u$$

KKT of MPP

Sandia National Laboratories

References

- Haldar, A. and S. Mahadevan. *Probability, Reliability, and Statistical Methods in Engineering Design* (Chapters 7-8). Wiley, 2000.
- Eldred, M.S. and Bichon, B.J., "Second-Order Reliability Formulations in DAKOTA/UQ," paper AIAA-2006-1828 in Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference (8th AIAA Non-Deterministic Approaches Conference), Newport, Rhode Island, May 1 - 4, 2006.
- Eldred, M.S., Agarwal, H., Perez, V.M., Wojtkiewicz, S.F., Jr., and Renaud, J.E. "Investigation of Reliability Method Formulations in DAKOTA/UQ," *Structure & Infrastructure Engineering: Maintenance, Management, Life-Cycle Design & Performance*, Vol. 3, No. 3, Sept. 2007, pp. 199-213.
- Bichon, B.J., Eldred, M.S., Swiler, L.P., Mahadevan, S., and McFarland, J.M., "Efficient Global Reliability Analysis for Nonlinear Implicit Performance Functions," *AIAA Journal*, Vol. 46, No. 10, October 2008, pp. 2459-2468.
- Li, J., Li, J. and D. Xiu. "An Efficient Surrogate-based Method for Computing Rare Failure Probability" *Journal of Computational Physics*, 2011.
- RIAC (Reliability Information Analysis Center): DoD site with useful information, guides on failure rates, accepted practices, etc.: http://www.theriac.org/

New Topic:
Importance Sampling for
Black-Box Simulators

Motivation

- USE CASE: We have DAKOTA users who take an initial set of Latin Hypercube samples, then would like to perform some additional samples to help refine a failure probability estimate
- They want to do this with relatively small number of samples: 100-200 initial samples and 100-200 samples from an importance sampling density.
- We developed a customized importance sampler where the importance sampling densities are constructed based on kernel density estimators.

Is importance sampling efficient and accurate for situations where we can only afford small numbers of samples?

Does importance sampling require the use of surrogate methods to generate a sufficient number of samples to increase the accuracy of the failure probability estimate?

Background

- **Importance sampling is a method used to sample random variables from different densities than originally defined.**
- **These importance sampling densities are constructed to pick "important" values of input random variables to improve the estimation of a statistical response of interest, such as a mean or probability of failure.**

$$E(r(X)) = \int r(x) f_X(x) dx$$

$$\hat{\mathrm{E}}_{\mathrm{n}}(r(X)) = \frac{1}{n}\sum_{i=1}^{n} r(x_i)$$

$$\mathbf{E}(r(X)) = \int r(x)\frac{f(x)}{h(x)}h(x)dx = \hat{\mathbf{E}}_h\left[r(X)\frac{f(X)}{h(X)}\right] = \frac{1}{n}\sum_{i=1}^{n}\frac{r(x_i)f(x_i)}{h(x_i)}, x_i \sim h(X)$$

Selection of importance density

- The variance of the importance sampling estimator is minimized when $h(x) \propto |r(x)f(x)|$.
- For black-box simulations that have multiple uncertain inputs which may come from a wide variety of random input distribution types, **we cannot generally assume that the importance sampling density will be normal or have a parametric form.**

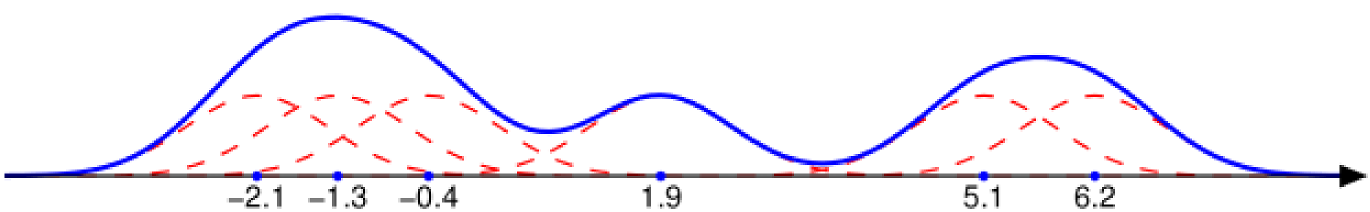
"The most difficult part in parametric importance sampling is choosing a suitable distribution family to start with. *There is no general recipe*, and *the issue remains largely a matter of art* in the literature. Most parametric distributions fail to include *g* (the optimal importance sampling density) as a member."

- Nonparametric methods: Zhang (1996) demonstrates increased convergence but higher computational cost of nonparametric methods.

➔ **Kernel density estimators**

Kernel density estimators

- **Kernel Density Estimation (KDE) is a technique used to estimate the density of a random variable *X* given *n* independent samples $X_1, \ldots, X_n$ of it.**



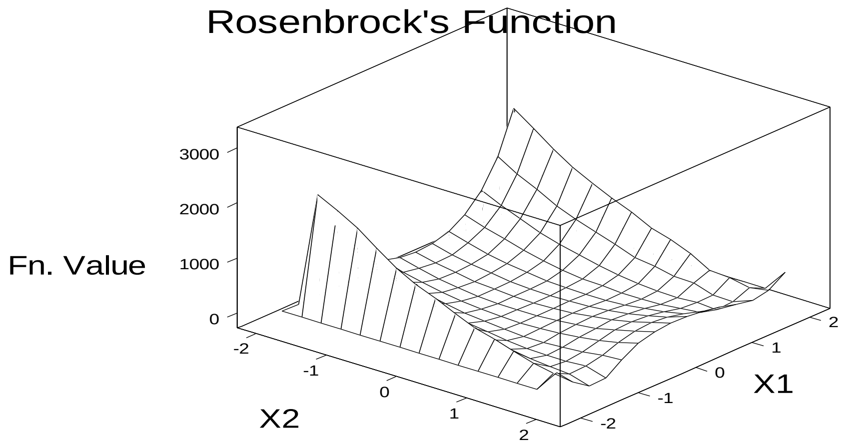
$$f_n^{\{KDE\}}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

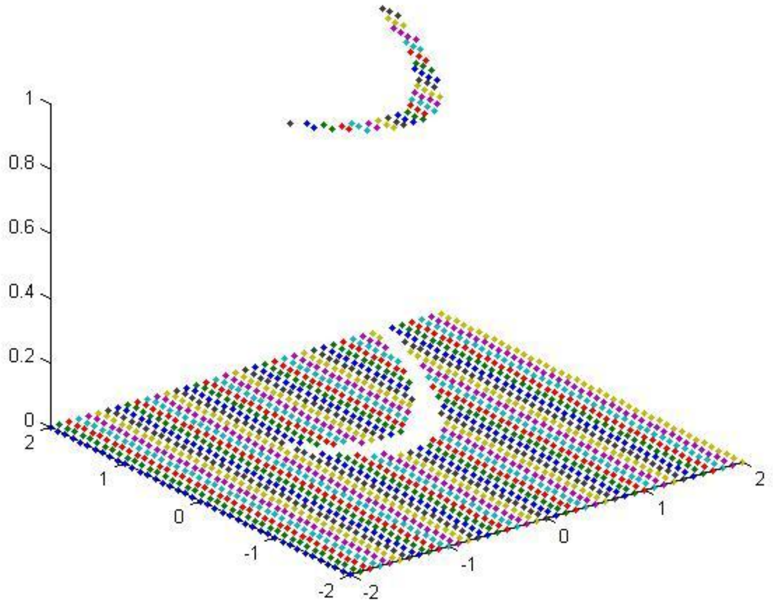
Used a Gaussian Kernel $$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Figure taken from http://en.wikipedia.org/wiki/Kernel_density_estimation

Sandia National Laboratories

Rosenbrock test function





Failure region defined as (Rosenbrock function < 3)

Probability of failure = 0.0383

Importance Sampling Test Results

- Rosenbrock function: X1, X2 both distributed as U[-2,2]
- Want to obtain estimate of failure probability (Rosenbrock < 3)
- For each combination of initial LHS/ IS points, we ran 100 replicates
- True value ≈ 0.0383

Number of initial LHS samples	Number of Importance samples	LHS Mean Failure probability	LHS Std dev. Failure Probability	IS Mean Failure probability	IS Std dev. Failure Probability	Mean percentage of IS that "fail"
100	200	0.039	0.01096	0.03798	0.01139	0.1283
50	100	0.0432	0.01429	0.03854	0.01665	0.1159
100	100	0.036	0.01241	0.03986	0.01567	0.133
200	200	0.03785	0.00955	0.03774	0.00823	0.1412
200	400	0.03702	0.00843	0.03868	0.00768	0.1432

- Note that failure probabilities not significantly different but we have increased percentage of points that fail by a factor of three
- Additional analysis showed that the accuracy of the failure estimates is not greatly improved IS due to the limited number of samples.
- The main benefit is that KDE IS estimators provide a quick way to generate more samples in the failure region.

Next steps

- We tested this approach on a variety of test problems, focusing on situations with small sample sizes to be representative of expensive computational simulations.
- We have looked at this approach for 5-D problems and a problem with a discontinuity in the response space: it worked fine in both cases
- Further investigation into scaling up to multiple dimensions
- Adding points "near" the response threshold can help
- Also investigated surrogate methods: need to have accurate surrogates
- This approach is reasonably robust and can produce failure probability estimates that are comparable to failure estimates produced by small numbers of LHS sample points.
- The main benefit we see by using this approach is that the kernel density estimators provide a quick way to generate more samples in the failure region. We found that importance sampling increased the number of samples in the failure region by a factor of 3 to 8 for our test cases.

Summary

- How can we identify "rare events" using computational simulations?
  - Importance Sampling
    - Challenges for black-box, high-dimensional problems
  - Reliability Methods
    - Designed to explicitly calculate probability of failure
    - Not widely used: limitations for very nonlinear problems, problems with multiple failure modes
  - Optimization methods in general
    - Find parameters that correspond to extreme situations
  - Use of surrogate models
    - very powerful and often necessary, but errors in tails, near boundaries
  - Model errors