SAND2016-1198PE

# Dakota Software Training

## Uncertainty Quantification

http://dakota.sandia.gov

Sandia National Laboratories

*Exceptional service in the national interest*

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.

# Familiarize Yourself with Key Statistics Ideas: Moments of Random Variables

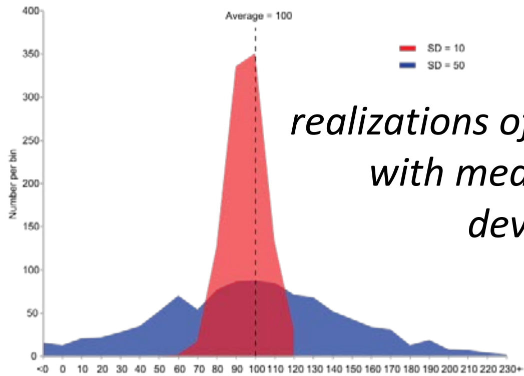
Understanding the following basic concepts will help with Dakota UQ

- Concept of a random variable *X*
- Mean (*m*, *μ*): expected or average value of *X*, *e.g., mean of sample of size N:*

$$\mu_T = \frac{1}{N}\sum_{i=1}^{N} T(u^i)$$

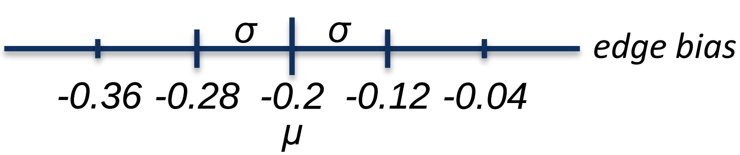
- Standard deviation (*s*, *σ*): measure of dispersion / variability of *X:*

$$\sigma_T = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[T(u^i) - \mu_T\right]^2}$$




*realizations of random variables with mean μ=100, standard deviation σ=10, σ=50*

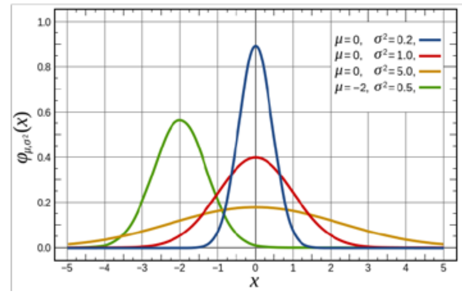
*In the earlier MEMS application, the manufactured edge has a mean bias of -0.2 μm, with standard deviation 0.08 μm:*



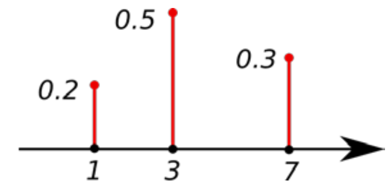
# Familiarize Yourself with Key Statistics Ideas: PDFs, CDFs, Intervals

Understanding the following basic concepts will help with Dakota UQ

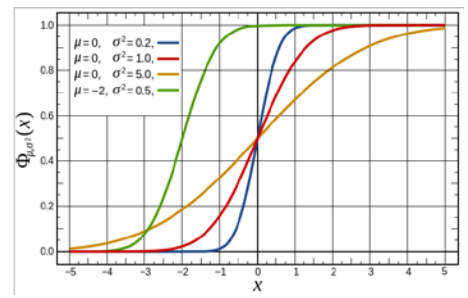
- Probability density / probability mass function: relative likelihood of a given value of *X*
- Cumulative distribution function: probability that *X* will take on a value less than or equal to *x:* **P**($X \leq x$)
- Interval-valued uncertainty: *X* can take on any value in the interval *[a,b]*, but no probability or likelihood of one value vs. another




*probability density functions*




*probability mass function*



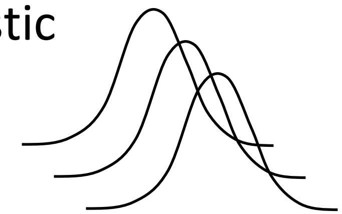

*cumulative distribution functions*

*For the earlier thermal application, a PDF or CDF can answer questions about the probability of exceeding a critical temperature.*

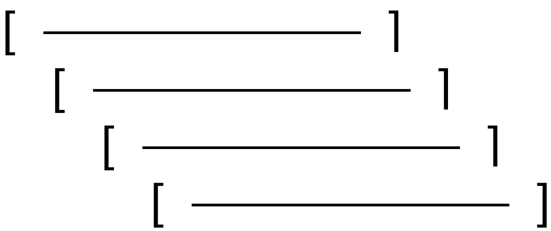
# Categories of Uncertainty

This distinction can help in selecting Dakota variable types and method

- Aleatory *(think probability density function, frequency; sufficient data)*
  - Inherent variability (e.g., in a population), type-A, stochastic
  - Irreducible: further knowledge won't help
  - Ideally simulation would incorporate this variability



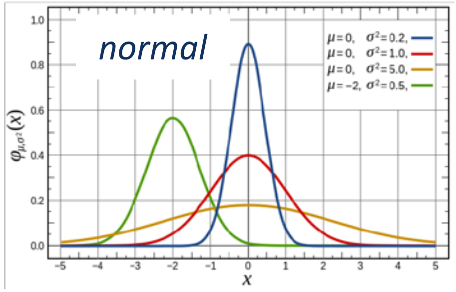
- Epistemic *(e.g., bounded intervals, distribution with uncertain parameters)*
  - Subjective, type-B, state of knowledge uncertainty
  - Reducible: more data or information, would make uncertainty estimation more precise
  - Fixed value in simulation, e.g., elastic modulus, but not well known for this material

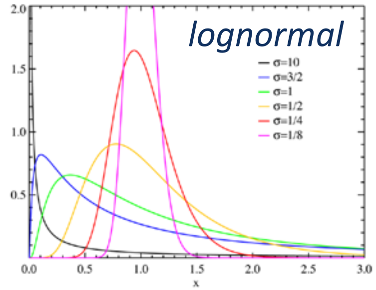


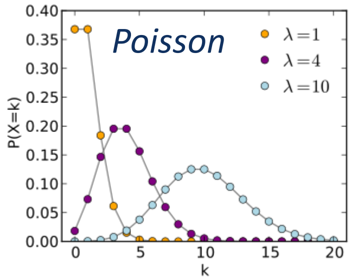
*See separate course on motivation for aleatory vs. epistemic uncertainty*

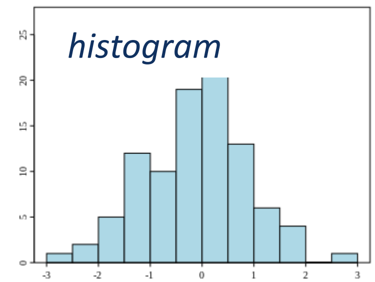
# Characterizing Uncertainties to Dakota

- Must characterize each variable's uncertainty and (optionally) any correlation between pairs of variables. Need not be normal (or uniform)!
- May require processing data with math/stats tool to fit distributions, performing literature searches, or querying experts










Dakota uncertain variable types:

- Aleatory continuous: normal, lognormal, uniform, loguniform, triangular, exponential, beta, gamma, Gumbel, Frechet, Weibull, histogram
- Aleatory discrete: Poisson, binomial, negative binomial, hypergeometric, histogram point (integer, real, string)
- Epistemic: continuous interval, discrete interval, discrete set

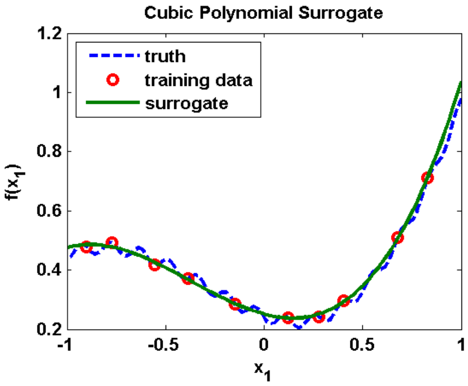
# Specifying Dakota Uncertain Variables

- UQ problems are specified to Dakota using uncertain variables (keywords *_uncertain)
- Typically generic response functions are used
- Thermal UQ example: here is a possible Dakota input file fragment for the uncertain variable types shown on the previous slide
- See the Reference Manual variables section for all variable types and their parameters

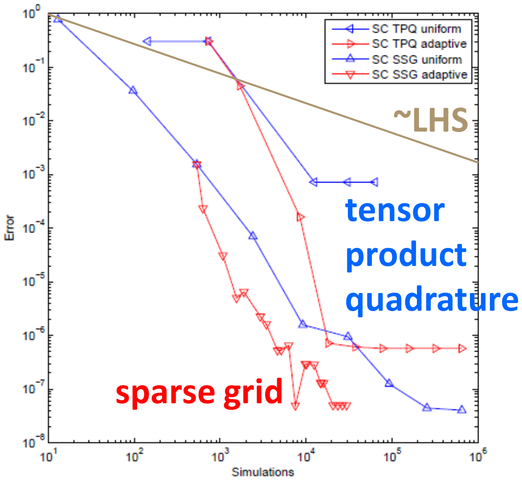
```
variables
  normal_uncertain 1
    descriptors     'density'
    means            8.1
    std_deviations   1.7
  lognormal_uncertain 1
    descriptors    'specific_heat'
    means           2.7
    error_factors   1.1
  poisson_uncertain
    descriptors   'fire_strength'
    lambdas       1.5
  histogram_bin_uncertain 1
    descriptors   'foam_thickness'
    num_pairs     4
      abscissas   2.5 3.0 3.5 4.0
      counts      15  11  20  0

responses
  response_functions 2
  descriptors 'pressure' 'temperature'
  ...
```

# Stochastic Expansions: What Are They?

- General-purpose UQ methods that build UQ-tailored polynomial approximations of the output responses
- Perform particularly well for smooth model responses
- Resulting convergence of statistics can be considerably faster than sampling methods




- Need to specify the Dakota method:
  - Polynomial Chaos (polynomial_chaos): specify the type of orthogonal polynomials and coefficient estimation scheme, e.g., sparse grid or linear regression.
  - Stochastic Collocation (stoch_collocation): specify the type of polynomial basis and the points at which the response will be interpolated; supports piecewise local basis

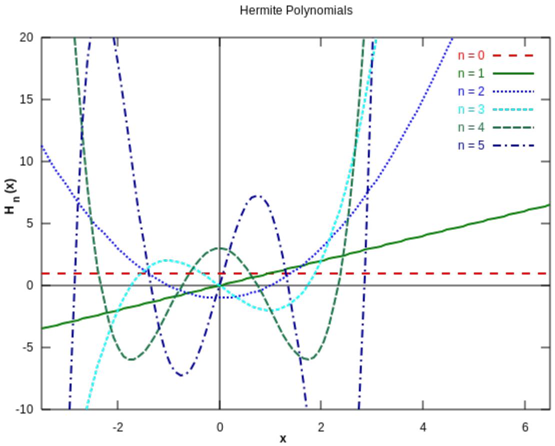


# Polynomial Chaos: How Does It Work?

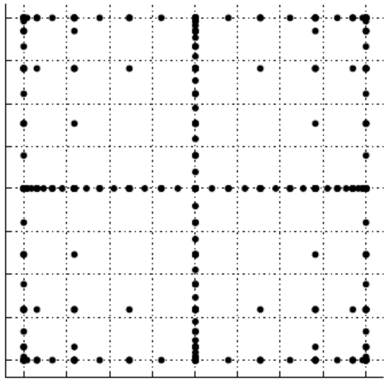
- Uses an orthogonal polynomial basis $\varphi_i(u)$, e.g., Wiener-Askey basis, with Hermite polynomials orthogonal w.r.t. normal density, Legendre polynomials orthogonal w.r.t. uniform density
- Evaluates the model in a strategic way (sampling, quadrature, sparse grids, cubature)…
- …to efficiently approximate the coefficients of an orthogonal polynomial approximation of the response

$$f(u) \approx p(u) = \sum_i c_i \varphi_i(u)$$

- And analytically calculates statistics from the approximation instead of approximating the statistics with MC samples




*Hermite Polynomials*



*Sparse Grid*

# Dakota UQ Methods Summary

| character | method class | problem character | variants |
|---|---|---|---|
| aleatory | probabilistic sampling | nonsmooth, multimodal, modest cost, # variables | Monte Carlo, LHS, importance |
| | local reliability | smooth, unimodal, more variables, failure modes | mean value and MPP, FORM/SORM, |
| | global reliability | nonsmooth, multimodal, low dimensional | EGRA |
| | stochastic expansions | nonsmooth, multimodal, low dimension | polynomial chaos, stochastic collocation |
| epistemic | interval estimation | simple intervals | global/local optim, sampling |
| | evidence theory | belief structures | global/local evidence |
| both | nested UQ | mixed aleatory / epistemic | nested |

*Also see Usage Guidelines in User's Manual*

# Using Dakota-generated Data

- Users commonly work with the Dakota tabular data file (dakota_tabular.dat by default)
- Import tabular data into Excel, Minitab, Matlab, R, SPlus, JMP, Python to
  - Generate histogram or other probability plots
  - Generate scatterplots to assess variability or see outliers / extreme behavior
  - Fit distributions to generated model outputs
  - Post-process samples to generate other statistics, e.g., probability of failure, ANOVA, variance-based decomposition, Sobol indices, safety factors
- Use Dakota results to refine characterization of variables and repeat study

- Decision making considerations
  - Consider what form your customers needs the information in to have impact
  - Consider engaging a Dakota team member in conversation with analyst and decision maker

Method-oriented

# BACKUP SLIDES

# Generalized Polynomial Chaos Expansions (PCE)

*Approximate response with Galerkin projection using multivariate orthogonal polynomial basis functions defined over standard random variables*

$$R = \sum_{j=0}^{P} \alpha_j \Psi_j(\xi)$$

*R(ξ) ≈ f(u)*

$$\alpha_j = \frac{\langle R, \Psi_j \rangle}{\langle \Psi_j^2 \rangle} = \frac{1}{\langle \Psi_j^2 \rangle} \int_\Omega R \, \Psi_j \, \varrho(\xi) \, d\xi$$

- Intrusive or non-intrusive
- Wiener-Askey Generalized PCE: optimal basis selection leads to exponential convergence of statistics

| Distribution | Density function | Polynomial | Weight function | Support range |
|---|---|---|---|---|
| Normal | $\frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$ | Hermite $He_n(x)$ | $e^{\frac{-x^2}{2}}$ | $[-\infty, \infty]$ |
| Uniform | $\frac{1}{2}$ | Legendre $P_n(x)$ | $1$ | $[-1, 1]$ |
| Beta | $\frac{(1-x)^\alpha (1+x)^\beta}{2^{\alpha+\beta+1} B(\alpha+1, \beta+1)}$ | Jacobi $P_n^{(\alpha,\beta)}(x)$ | $(1-x)^\alpha (1+x)^\beta$ | $[-1, 1]$ |
| Exponential | $e^{-x}$ | Laguerre $L_n(x)$ | $e^{-x}$ | $[0, \infty]$ |
| Gamma | $\frac{x^\alpha e^{-x}}{\Gamma(\alpha+1)}$ | Generalized Laguerre $L_n^{(\alpha)}(x)$ | $x^\alpha e^{-x}$ | $[0, \infty]$ |

- Can also numerically generate basis orthogonal to empirical data (PDF/histogram)

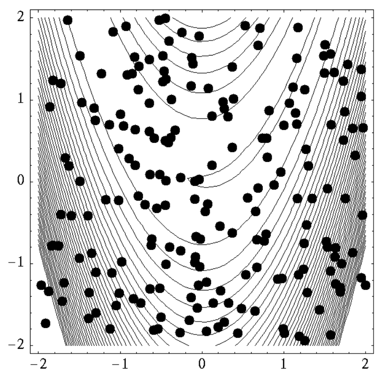
# Sample Designs to Form Polynomial Chaos or Stochastic Collocation Expansions

## Random sampling: PCE

***Expectation (sampling):***
- Sample w/i distribution of $x$
- Compute expected value of product of $R$ and each $Y_j$

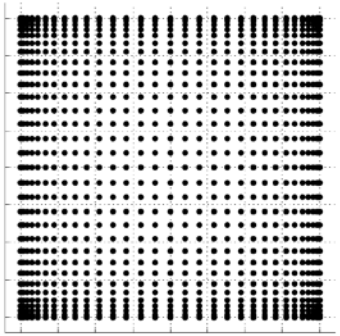
***Linear regression ("point collocation"):***

$$\Psi \alpha = R$$

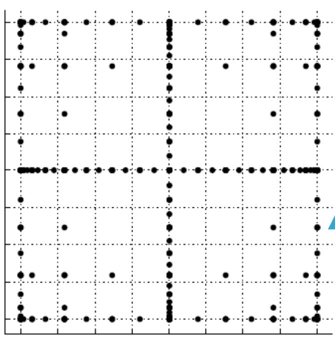


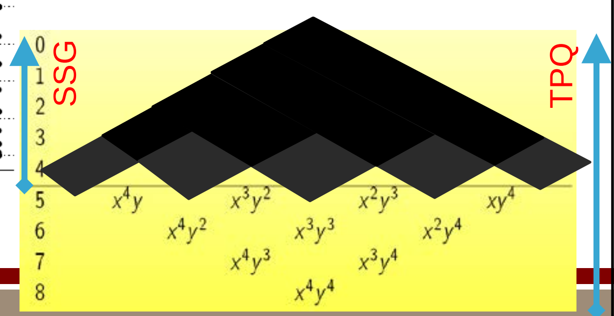

## Tensor-product quadrature: PCE/SC

**Tensor product of 1-D integration rules, e.g., Gaussian quadrature**



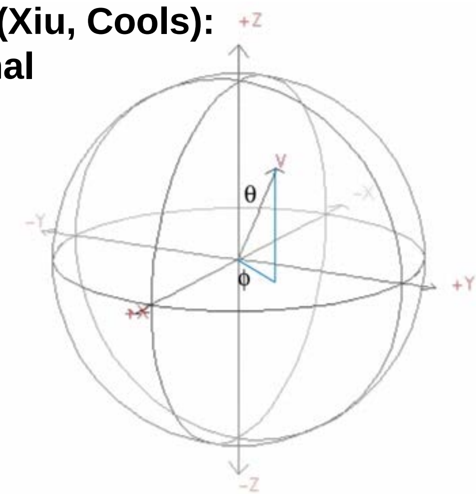
## Smolyak Sparse Grid: PCE/SC






## Cubature: PCE

**Stroud and extensions (Xiu, Cools): optimal multidimensional integration rules**



# Adaptive PCE/SC: Emphasize Key Dimensions

- Judicious choice of new simulation runs
- Uniform p-refinement
  - Stabilize 2-norm of covariance
- Adaptive p-refinement
  - Estimate main effects/VBD to guide
- h-adaptive: identify important regions and address discontinuities
- h/p-adaptive: p for performance; h for robustness