Module 11: Monte Carlo Risk Assessment

11.2 Quantifying Variability and Uncertainty

Statistical Methods for Expressing Variability and Uncertainty

- Uncertainty can be expressed in a variety of ways:
  - Order of Magnitude
    - "This value is correct within an order of magnitude"
  - Percent Error
    - "We're probably off no more than plus or minus 50%"
  - Plausible Range
    - "It could be as low as … or as high as …"

Statistical Methods for Expressing Variability and Uncertainty

- Uncertainty and/or variability can be expressed in a variety of ways:
  - Plausible Range and Best Estimate
  - Probability Distribution:
    - Empirical Distribution Function (edf)
      - Based on data
    - Parametric Distribution Function (pdf)
      - Need to specify distribution form and parameters
    - Confidence Interval on a Parameter Estimate

Sources of Information for Quantifying Variability and Uncertainty

- Data
  - General
  - Site Specific
- Information/Knowledge
  - Experience
  - Scientific literature
  - Expert judgement

How to Quantify Variability and Uncertainty Based on Data

- Issue - Are the data reliable?
  - Were the appropriate quality assurance/quality control procedures used?
    - sample collection, transport, and storage
    - analytical laboratory handling and measurement
    - data input and manipulation

Module 11.2

How to Quantify Variability and Uncertainty Based on Data

- Issue - Are the data representative of the population being studied and do they cover an appropriate temporal/spatial scale?
  - Are they a random sample?
  - Do they reflect the full range of variability in
    - Demographic characteristics of the population
    - Spatial coverage and scale
    - Temporal coverage and scale

Module 11.2

How to Quantify Variability and Uncertainty Based on Data

- Issue - Are the data applicable to the current problem?
  - Data gathered on one population may not apply to a different population at another time or place

How to Quantify Variability and Uncertainty Based on Data

- Decide between using an empirical distribution and a parametric distribution
- Empirical
  - Summarize the data, usually as an empirical cumulative distribution function
  - Make adjustments to the distribution if appropriate

How to Quantify Variability and Uncertainty Based on Data

- Making adjustments to empirical data distributions:
  - If the number of data points is small, the full range of variability in the population has probably not been captured
  - In this case, tails would need to be added
    - based on other similar data sets
    - based on expert judgement

How to Quantify Variability and Uncertainty Based on Data

- Parametric
  - Select the form of the distribution
    - Choose a distribution based on where the data came from and what we know about that type of data
    - Use a distribution based on historical precedent
    - Select a distribution based on goodness-of-fit
  - Estimate the parameters of the distributions
  - Examples
    - Uniform(min, max), Normal (mean, standard.dev.), Lognormal (mean, standard deviation), ...

How to Quantify Variability and Uncertainty Based on Data

- Parametric
  - If you know little or nothing apart from the data
    - Use classical statistical techniques (maximum likelihood, method of moments, probability plotting)
  - If you have prior information
    - Bayesian techniques allow subjective judgements based on scientific knowledge, past experience, data, and intuition as the basis for estimation

How to Quantify Variability and Uncertainty Based on Data

- Important considerations
  - The range of the distribution should cover the theoretical range of the variable
    - Distribution may need to be wider than the data
    - Truncate the distribution if necessary
  - Some distributions have upper and lower bounds (uniform and triangular). Others have tails.
  - For variables that cover a wide range, transformations such as the log may be helpful.

How to Quantify Variability and Uncertainty Based on Literature

- Values from the literature usually define a range or a range and most likely value
- If other information about the underlying mechanism is absent, the most useful distributions are the uniform and triangular
- If the theoretical distribution is clear, use literature values to estimate distribution parameters

How to Quantify Variability and Uncertainty Based on Expert Judgement

- Often, it is difficult or impossible to obtain data or literature values that are completely reliable and representative of the population
  - no data exists
  - can't afford to collect the data
  - data that does exists is old or inadequate
- In these cases, expert opinion can be used to replace or augment data and literature values
- Usually results in empirical, uniform, or triangular distributions

How to Quantify Variability and Uncertainty Based on Expert Judgement

- There are some problems with using expert judgement, most notably over-confidence
- Commonly occurring biases:
  - Availability - probability estimates are controlled by the ease of remembering past occurrences or the ease of imagining future occurrences. If event is uncommon, the estimate may be affected by the timing of the last occurrence, media coverage, personal impact, and so on.
  - Example - estimates of the probability of a release from a transportation accident may increase if an accident just occurred

How to Quantify Variability and Uncertainty Based on Expert Judgement

- Commonly occurring biases:
  - Representativeness - Probability is estimated based on how alike a specific situation is to some class or group
  - Example - In a lottery, the probability of five consecutive numbers appearing (21, 22, 23, 24, 25) is judged less likely than five widely spaced numbers (8, 14, 26, 28, 43) even though the true probability is the same.

How to Quantify Variability and Uncertainty Based on Expert Judgement

- Commonly occurring biases:
  - Anchoring and Adjustment - Estimators begin at a value (most likely) and adjust up or down from there but adjustments are usually too small to encompass all uncertainty.
  - Example - If initial estimate is 30%, min and max may be estimated up and down 10% (20% to 40%). True uncertainty may be wider than this range.

How to Quantify Variability and Uncertainty Based on Expert Judgement

- Other problems
  - Personal agenda - expert has something to gain from a particular answer (research funding, forwarding a scientific hypothesis, enhanced reputation, etc)
  - Unwillingness to appear uncertain
  - Difficulty in imagining other scenarios or extreme events
  - Estimates of reality may be affected by goals

How to Quantify Variability and Uncertainty Based on Expert Judgement

- There are different techniques involved in doing expert elicitations
- These techniques have elements in common
- Conducting the elicitation
  - Prepare yourself and the experts in advance using written material and possibly meetings
  - Everyone needs to be aware of, and guard against, bias and over-confidence

How to Quantify Variability and Uncertainty Based on Expert Judgement

- Conducting the elicitation
  - Make sure there is a common understanding of the problem, all terminology, and the quantities to be elicited
  - Obtain estimates of distribution percentiles, or min/max/mean, or other values. Elicit extreme values first and then others in random order.
  - Encourage “stretching” the extremes to avoid over-confidence
  - When complete, allow the expert to review the probability estimates and adjust them

How to Quantify Variability and Uncertainty Based on Expert Judgement

- Recommendations/Pitfalls
  - It is sometimes helpful to allow the experts to discuss their views with one another and adjust probabilities
  - If fundamental disagreements exist, don't pool or average. Use each elicitation separately, assign probabilities to each, or strive to understand the fundamental disagreement between the experts.

How to Quantify Uncertainty in a Parameter Estimate

- Uncertainty in a parameter estimate is usually expressed as a confidence interval or an upper confidence bound
- These can be constructed using
  - classical parametric statistical techniques and equations
  - nonparametric techniques
  - resampling techniques such as the bootstrap

Classical vs Bayesian Statistics

- The classical, or frequentist, perspective is that the probability of an event is defined as the frequency with which it occurs in a long sequence of similar trials
- The Bayesian, or subjectivist, perspective is that the probability of an event occurring is the degree of belief a person has in the occurrence

Advantages of the Bayesian Approach

- No long sequence or representative population of similar trials may exist
- Allows probability distributions used as inputs to a PRA to be defined using expert judgement
- Allows the input distributions to be updated/refined if new information becomes available

Advantages of the Bayesian Approach

- Allows data to be combined with an input distribution based on expert judgement
- Allows site specific data to be combined with a generic distribution or a distribution based on another location
- Allows an output distribution from a PRA to be updated with new information

Conclusions

- Consider data reliability, representativeness, and applicability
- Characterize the distributions based on data, literature values, and/or expert judgement
- Distributions based on data can be empirical or parametric
- Bayesian analysis allows expert judgement to be used to create input distributions for a PRA. It also allows data to be combined with a distribution to give an updated information state.