Uniformly Ordered Binary Decision Algorithm for Benchmark Experiment Correlations in Whisper Validation

Brian Kiedrowski
Nuclear Engineering and Radiological Sciences Department
University of Michigan
Los Alamos National Laboratory Subcontract

Nuclear Criticality Safety Program Technical Program Review
February 22-25, 2021
Overview & Motivation

- Whisper is a statistical analysis package developed and maintained by LANL to support NCS validation.

- The original methodology in Whisper does not account for benchmark experiment correlations when estimating the bias plus bias uncertainty (calculational margin).
  - Correlations between benchmarks reduce the amount of information in the population and its statistical predictive power.

- The Uniformly Ordered Binary Decision Algorithm has been developed to address this shortcoming.
Outline

• Summary of Original Whisper Algorithm

• Uniformly Ordered Binary Decision Algorithm

• Results
  – HEU Solutions
  – LEU Lattices
Whisper Overview

- Whisper uses MCNP6.2 sensitivity profiles with ENDF/B-VII.1 and covariance data to compute $c_k$ to assess similarity between benchmark experiment and application models.

- The $c_k$ similarity values are used to assign statistical weighting factors $w_i$ for each benchmark.
  - Weighting factors can be thought of as a probability of including a particular benchmark experiment in a hypothetical validation exercise.
  - Assigned proportional to the similarity coefficients to ensure an adequate effective sample size.
Whisper Overview

• Whisper ensures there is an adequate population or sample size in the benchmark set
  – The effective sample size is the expected (mean) population of benchmarks in a hypothetical validation exercise
  – In the uncorrelated case, this is the sum of the weighting factors
  – Requires the effective sample size is greater than some threshold
    • nominally 25, but increases if the maximum $c_k$ is too low
Whisper Overview

• Whisper uses the weighting factors to compute an extreme-value distribution
  – Probability distribution for the maximum negative bias
  – Cumulative distribution is the product of the weighted cumulative distribution functions of the normally distributed biases for each benchmark
  – Calculational Margin is the 99% upper confidence limit of the extreme-value distribution
Handling Correlations

• This approach does not account for benchmark experiment correlations

• Questions about the method with correlations
  – How to assign weighting factors?
  – How to quantify an effective sample size?

• Done with the Uniformly Ordered Binary Decision Algorithm
Uniformly Ordered Binary Decision Algorithm

- Find correlated clusters in the population, and, for each cluster, determine if an hypothetical analyst would consider pairs of benchmark experiments as redundant
  - Randomly order the cluster with uniform probability
  - Start with the first benchmark and include with a probability equal to the weighting factor $w_i$
  - For the next benchmarks, include with a probability equal to their weight, but decide whether to treat them as redundant with more similar benchmarks previously in the list
  - If both are included and redundant:
    1. Assign the worst case bias (done automatically by including both with extreme-value distribution)
    2. Count these two benchmarks as a single benchmark toward the effective sample size
  - If not redundant, include as an independent benchmark
  - Repeat for all possible permutations
Uniformly Ordered Binary Decision Algorithm

• The adjust weighting factor is the probability that a benchmark experiment will be included as independent or non-redundant
  – The unadjusted weighting factor is the probability that it is included either way

• The effective sample size is now the expected (mean) number of non-redundant benchmarks in the population
  – This is the sum of the adjusted weighting factors

• Since the adjusted weight is less than the original, the effective sample size is smaller
  – Causes Whisper to expand its benchmark search
Example

- Consider a cluster of 3 correlated benchmarks {1,2,3} in a random order

- Here we have three levels to the decision algorithm with each level giving the adjusted weight
  - $\rho_j$ is the adjusted sample weight for the $j$th benchmark
  - $w_j$ is the original or unadjusted sample weight
  - $\rho_{ij} = \text{is the redundancy probability:}$

\[
\rho_{ij} = r_{ij} \min\left\{ 1, \frac{w_i}{w_j} \right\}, \quad r_{ij} \geq 0
\]

- $r_{ij} = \text{benchmark correlation coefficient}$
Example: Level 1

Begin

Include \{1\}?

No: \( \xi_1 \geq w_1 \)

Do not include \{1\}

Continue to Level 2

Yes: \( \xi_1 < w_1 \)

Include \{1\} +1 to sample size
Example: Level 2

Include {2}?

Yes: $\xi_2 < w_2$

No

Do not include {2}

Yes: $\xi_3 < 1 - \rho_{12}$

{2} independent of {1}?

Include {2} +1 to sample size

{2} redundant

Continue to Level 3
Example: Level 3

1. Include \{3\}? Yes: \( \xi_4 < w_3 \)
   - No
     - No
       - \{3\} independent of \{1\}? Yes: \( \xi_5 < 1 - \rho_{13} \)
         - No
           - \{3\} redundant
     - \{3\} independent of \{2\}? Yes: \( \xi_6 < 1 - \rho_{23} \)
       - Include \{3\} +1 to sample size
     - Do not include \{3\}
Adjusted Sample Weight

• We can deduce the probability for all permutations that a benchmark experiment is included as independent:

$$p_j = \frac{1}{n!} \sum_{\sigma} w_j \prod_{i=1}^{\sigma_j - 1} (1 - \rho_{ij})$$

  – n = cluster size
  – \(\sigma\) = all possible permutations of the cluster
  – \(\sigma_j\) = the position of j within the permutation

• The effective sample size is the sum of the \(p_j\)
Adjusted Sample Weight

• The summation in the expression for the adjusted sample weight contains $n!$ terms
  – Computationally tractable to compute directly for cluster sizes $n \leq 10$ on a modern computer
  – For larger clusters, use Monte Carlo sampling with random permutations and estimate
Effective Sample Size

• The computed effective sample size with correlations is compared against the sample size requirement
  – If effective sample size falls below requirement, decrease the acceptance criterion on $c_k$ expanding the sample population, compute new weights, and repeat until the requirement is satisfied

• Once sample size is met, use the unadjusted weighting factors to find the extreme value distribution and calculational margin
  – Ensures the calculational margin is assigned conservatively
Results

• Implemented Uniformly Ordered Binary Decision Algorithm in a research version of Whisper1.1
  – Deliverable for the subcontract

• Transport calculations performed with MCNP6.2 and ENDF/B-VII.1 nuclear data

• Whisper calculations use data from Whisper1.1 library (except where otherwise noted)
Results

• HEU Solutions
  – Use application models identical to Cases 1-10 for HEU-SOL-THERM-001
  – Correlations available in DICE
  – Cases 9 and 10 are excluded by $\chi^2$ rejection algorithm

• LEU Lattices
  – Use application models identical to Cases 1-3 of LEU-COMP-THERM-007 and Cases 1-10 of LEU-COMP-THERM-039
  – Correlations obtained from Scenario A from W.J. Marshall’s doctoral dissertation (used to show impact of a highly correlated cluster)
  – Case 3 of LCT-007 is excluded by $\chi^2$ rejection algorithm
To further analyze these results, details for the application are shown; while this number in itself is not used, it does assume to be independent because no correlation values are currently available. Therefore, a more accurate representation of the correlation will likely lead to a higher estimate of the bias plus bias uncertainty. There is a significant variation in the change in the bias plus bias uncertainty between the cases because the computation with the extreme-value method is sensitive to the worst case. Should significant variation in the change in the bias plus bias uncertainty (calculational margin) at a significant weight, then this will be the primary reason for the increase.

As expected the population size increases because the expansion of the population cause Whisper to include more benchmark experiments within the population as well as including more of them. By nature of the extreme value distribution, this increases the upper confidence limit or bias uncertainty. For the same reason, using the unadjusted weights increases the bias plus bias uncertainty. The reason for using the adjusted weights is to produce a conservative estimate. By using the adjusted weights, and the change in the bias plus bias uncertainty is a modulation of the population size using the effective sample size and the unadjusted weights.

### Table I. HEU-SOL-THERM-001 Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>HST-001-001</th>
<th>HST-001-002</th>
<th>HST-001-003</th>
<th>HST-001-004</th>
<th>HST-001-005</th>
<th>HST-001-006</th>
<th>HST-001-007</th>
<th>HST-001-008</th>
</tr>
</thead>
<tbody>
<tr>
<td>HST-001-001</td>
<td>1.00</td>
<td>0.47</td>
<td>0.46</td>
<td>0.44</td>
<td>0.42</td>
<td>0.42</td>
<td>0.46</td>
<td>0.57</td>
</tr>
<tr>
<td>HST-001-002</td>
<td>0.47</td>
<td>1.00</td>
<td>0.42</td>
<td>0.58</td>
<td>0.42</td>
<td>0.42</td>
<td>0.41</td>
<td>0.44</td>
</tr>
<tr>
<td>HST-001-003</td>
<td>0.46</td>
<td>0.42</td>
<td>1.00</td>
<td>0.46</td>
<td>0.43</td>
<td>0.43</td>
<td>0.46</td>
<td>0.46</td>
</tr>
<tr>
<td>HST-001-004</td>
<td>0.44</td>
<td>0.58</td>
<td>0.46</td>
<td>1.00</td>
<td>0.44</td>
<td>0.42</td>
<td>0.42</td>
<td>0.44</td>
</tr>
<tr>
<td>HST-001-005</td>
<td>0.42</td>
<td>0.42</td>
<td>0.43</td>
<td>0.44</td>
<td>1.00</td>
<td>0.54</td>
<td>0.48</td>
<td>0.47</td>
</tr>
<tr>
<td>HST-001-006</td>
<td>0.42</td>
<td>0.42</td>
<td>0.43</td>
<td>0.42</td>
<td>0.54</td>
<td>1.00</td>
<td>0.48</td>
<td>0.47</td>
</tr>
<tr>
<td>HST-001-007</td>
<td>0.46</td>
<td>0.41</td>
<td>0.46</td>
<td>0.42</td>
<td>0.48</td>
<td>0.48</td>
<td>1.00</td>
<td>0.51</td>
</tr>
<tr>
<td>HST-001-008</td>
<td>0.57</td>
<td>0.44</td>
<td>0.46</td>
<td>0.44</td>
<td>0.47</td>
<td>0.47</td>
<td>0.51</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The method was tested for HEU solution and LEU latin hypercube sampling with benchmark models for different benchmark experiments. The benchmark correlation matrix for cases 1-8 are given for the HEU-SOL-THERM-001 cases 1 through 10, which have also shown; while this number in itself is not used, it does assume to be independent because no correlation values are currently available. Therefore, a more accurate representation of the correlation will likely lead to a higher estimate of the bias plus bias uncertainty. There is a significant variation in the change in the bias plus bias uncertainty between the cases because the computation with the extreme-value method is sensitive to the worst case. Should significant variation in the change in the bias plus bias uncertainty (calculational margin) at a significant weight, then this will be the primary reason for the increase.

As expected the population size increases because the expansion of the population cause Whisper to include more benchmark experiments within the population as well as including more of them. By nature of the extreme value distribution, this increases the upper confidence limit or bias uncertainty. For the same reason, using the unadjusted weights increases the bias plus bias uncertainty. The reason for using the adjusted weights is to produce a conservative estimate. By using the adjusted weights, and the change in the bias plus bias uncertainty is a modulation of the population size using the effective sample size and the unadjusted weights.

### Table II. Results for HEU Solution Application Cases

<table>
<thead>
<tr>
<th>Application</th>
<th># (orig.)</th>
<th># (corr.)</th>
<th>% (orig.)</th>
<th>% (corr.)</th>
<th>w</th>
<th>m</th>
<th>c</th>
<th>k</th>
</tr>
</thead>
<tbody>
<tr>
<td>HST-001-001</td>
<td>41</td>
<td>66</td>
<td>100</td>
<td>100</td>
<td>0.0020</td>
<td>32.0</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>HST-001-002</td>
<td>53</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>0.0002</td>
<td>32.2</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>HST-001-003</td>
<td>49</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>0.0017</td>
<td>32.2</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>HST-001-004</td>
<td>53</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>0.0020</td>
<td>32.2</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>HST-001-005</td>
<td>43</td>
<td>63</td>
<td>100</td>
<td>100</td>
<td>0.0017</td>
<td>32.0</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>HST-001-006</td>
<td>53</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>0.0002</td>
<td>32.2</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>HST-001-007</td>
<td>49</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>0.0017</td>
<td>32.2</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>HST-001-008</td>
<td>53</td>
<td>67</td>
<td>100</td>
<td>100</td>
<td>0.0020</td>
<td>32.2</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>
The presence of correlations causes the number of benchmarks included to increase. The increase in the total sampling weight increases from 25 to around 30-35. The calculational margin shows increases up to a few 100 pcm. This is because more benchmarks must be included.

### HEU Solution Results

<table>
<thead>
<tr>
<th>Application</th>
<th># (orig.)</th>
<th># (corr.)</th>
<th>$\sum w_i$</th>
<th>$\Delta m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HST-001-001</td>
<td>55</td>
<td>68</td>
<td>34.5</td>
<td>0.0022</td>
</tr>
<tr>
<td>HST-001-002</td>
<td>52</td>
<td>62</td>
<td>31.0</td>
<td>0.0002</td>
</tr>
<tr>
<td>HST-001-003</td>
<td>49</td>
<td>69</td>
<td>34.7</td>
<td>0.0023</td>
</tr>
<tr>
<td>HST-001-004</td>
<td>53</td>
<td>66</td>
<td>31.2</td>
<td>0.0002</td>
</tr>
<tr>
<td>HST-001-005</td>
<td>41</td>
<td>72</td>
<td>34.1</td>
<td>0.0040</td>
</tr>
<tr>
<td>HST-001-006</td>
<td>43</td>
<td>73</td>
<td>34.0</td>
<td>0.0040</td>
</tr>
<tr>
<td>HST-001-007</td>
<td>49</td>
<td>69</td>
<td>34.6</td>
<td>0.0024</td>
</tr>
<tr>
<td>HST-001-008</td>
<td>51</td>
<td>69</td>
<td>34.6</td>
<td>0.0020</td>
</tr>
<tr>
<td>HST-001-009</td>
<td>53</td>
<td>66</td>
<td>31.4</td>
<td>0.0002</td>
</tr>
<tr>
<td>HST-001-010</td>
<td>41</td>
<td>72</td>
<td>34.2</td>
<td>0.0049</td>
</tr>
</tbody>
</table>
There are three clusters of correlated benchmarks in a sample population including 72 benchmarks

The second and third columns show the reduction in the effective sample size because of correlation for each cluster

As expected, the adjusted sample weight or inclusion probability decreases proportionate to the weight and size and degree of correlation within the cluster.
Case Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>LCT-007-001</th>
<th>LCT-007-002</th>
<th>LCT-007-003</th>
<th>LCT-039-001</th>
<th>LCT-039-002</th>
<th>LCT-039-003</th>
<th>LCT-039-004</th>
<th>LCT-039-005</th>
<th>LCT-039-006</th>
<th>LCT-039-007</th>
<th>LCT-039-008</th>
<th>LCT-039-009</th>
<th>LCT-039-010</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCT-007-001</td>
<td>1.00</td>
<td>0.93</td>
<td>0.39</td>
<td>0.98</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.98</td>
<td>0.97</td>
</tr>
<tr>
<td>LCT-007-002</td>
<td>0.93</td>
<td>1.00</td>
<td>0.56</td>
<td>0.92</td>
<td>0.92</td>
<td>0.92</td>
<td>0.93</td>
<td>0.92</td>
<td>0.93</td>
<td>0.93</td>
<td>0.92</td>
<td>0.94</td>
<td>0.94</td>
</tr>
<tr>
<td>LCT-007-003</td>
<td>0.39</td>
<td>0.56</td>
<td>1.00</td>
<td>0.40</td>
<td>0.39</td>
<td>0.41</td>
<td>0.42</td>
<td>0.46</td>
<td>0.42</td>
<td>0.42</td>
<td>0.39</td>
<td>0.43</td>
<td>0.39</td>
</tr>
<tr>
<td>LCT-039-001</td>
<td>0.98</td>
<td>0.92</td>
<td>0.40</td>
<td>1.00</td>
<td>0.98</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>LCT-039-002</td>
<td>0.98</td>
<td>0.92</td>
<td>0.39</td>
<td>0.98</td>
<td>1.00</td>
<td>0.97</td>
<td>0.97</td>
<td>0.95</td>
<td>0.96</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>LCT-039-003</td>
<td>0.97</td>
<td>0.92</td>
<td>0.41</td>
<td>0.97</td>
<td>0.97</td>
<td>1.00</td>
<td>0.97</td>
<td>0.94</td>
<td>0.95</td>
<td>0.97</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>LCT-039-004</td>
<td>0.97</td>
<td>0.93</td>
<td>0.42</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>1.00</td>
<td>0.96</td>
<td>0.95</td>
<td>0.97</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>LCT-039-005</td>
<td>0.96</td>
<td>0.92</td>
<td>0.46</td>
<td>0.96</td>
<td>0.95</td>
<td>0.94</td>
<td>0.96</td>
<td>1.00</td>
<td>0.95</td>
<td>0.96</td>
<td>0.94</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>LCT-039-006</td>
<td>0.96</td>
<td>0.93</td>
<td>0.46</td>
<td>0.96</td>
<td>0.96</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>1.00</td>
<td>0.96</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
</tr>
<tr>
<td>LCT-039-007</td>
<td>0.97</td>
<td>0.93</td>
<td>0.42</td>
<td>0.98</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>1.00</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>LCT-039-008</td>
<td>0.97</td>
<td>0.92</td>
<td>0.39</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.94</td>
<td>0.95</td>
<td>0.97</td>
<td>1.00</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>LCT-039-009</td>
<td>0.98</td>
<td>0.94</td>
<td>0.43</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>1.00</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>LCT-039-010</td>
<td>0.97</td>
<td>0.94</td>
<td>0.39</td>
<td>0.97</td>
<td>0.96</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
LEU Lattice Results

Conclusions are much the same as for the HEU solutions.

The increase in calculational margin is more modest, which is because there were already high negative bias benchmarks in the initial population prior to applying correlations.

<table>
<thead>
<tr>
<th>Application</th>
<th># (orig.)</th>
<th># (corr.)</th>
<th>$\sum p_i$</th>
<th>$\Delta m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCT-007-001</td>
<td>74</td>
<td>81</td>
<td>34.4</td>
<td>0.0007</td>
</tr>
<tr>
<td>LCT-007-002</td>
<td>47</td>
<td>50</td>
<td>30.3</td>
<td>0.0003</td>
</tr>
<tr>
<td>LCT-007-003</td>
<td>51</td>
<td>51</td>
<td>25.1</td>
<td>0.0000</td>
</tr>
<tr>
<td>LCT-039-001</td>
<td>72</td>
<td>76</td>
<td>34.9</td>
<td>0.0006</td>
</tr>
<tr>
<td>LCT-039-002</td>
<td>70</td>
<td>75</td>
<td>35.0</td>
<td>0.0007</td>
</tr>
<tr>
<td>LCT-039-003</td>
<td>58</td>
<td>70</td>
<td>35.2</td>
<td>0.0009</td>
</tr>
<tr>
<td>LCT-039-004</td>
<td>68</td>
<td>77</td>
<td>35.0</td>
<td>0.0008</td>
</tr>
<tr>
<td>LCT-039-005</td>
<td>53</td>
<td>55</td>
<td>34.1</td>
<td>0.0005</td>
</tr>
<tr>
<td>LCT-039-006</td>
<td>53</td>
<td>62</td>
<td>34.6</td>
<td>0.0005</td>
</tr>
<tr>
<td>LCT-039-007</td>
<td>72</td>
<td>81</td>
<td>34.7</td>
<td>0.0007</td>
</tr>
<tr>
<td>LCT-039-008</td>
<td>65</td>
<td>72</td>
<td>35.2</td>
<td>0.0009</td>
</tr>
<tr>
<td>LCT-039-009</td>
<td>65</td>
<td>73</td>
<td>35.2</td>
<td>0.0009</td>
</tr>
<tr>
<td>LCT-039-010</td>
<td>53</td>
<td>71</td>
<td>35.1</td>
<td>0.0008</td>
</tr>
</tbody>
</table>
The correlations in Marshall’s Scenario A are very high, > 0.95 in many cases.

This causes the cluster of 12 benchmarks to become essentially worth 1 benchmark with most benchmarks being included as redundant most of the time.

The adjusted sample weights are significantly decreased because of the high degree of correlation.
Conclusions

• The uniformly ordered binary decision algorithm has been developed and implemented into a research version of Whisper1.1
  – This addresses the outstanding issue of not accounting for benchmark experiment correlations when estimating the calculational margin

• Results were collected for HEU solutions and LEU lattices
  – The results show the presence of benchmark correlations may significantly increase the calculational margin because Whisper must expand the benchmark population

• This method will lead to lower predicted USLs for some applications
Final Thoughts

• Need to get results for fast spectrum systems
  – Several of the ZPR/ZPPR series have correlation data associated with them, but Whisper1.1 does not have models

• Benchmark correlation data is currently limited and the results will change as more correlations are added
  – There is a need to generate this data for a wider variety of cases
  – Could apply some arbitrary correlation for cases that DICE identifies as potentially correlated but has no quantitative data
  – Would be helpful to have benchmark correlations for different cases quantified during the development of benchmark evaluations

• Allows for the inclusion of different benchmark models from multiple sites
  – Different modeling assumptions may lead to different answer
  – Could all be included and assumed perfectly correlated
Acknowledgments

The author would like to acknowledge the generous support provided by the Nuclear Criticality Safety Program and the technical staff at Los Alamos National Laboratory the facilitating this subcontract that made this research possible.
Discussion

Email contact: bckiedro@umich.edu