Reducing Simulation Replications for Future Combat System Analysis

22 October 2004

Purpose

- **Describe an alternative method of determining, and possibly reducing, the simulation replications required to determine differences among alternatives.**

Agenda

- **Coefficient of variation**
- **Bootstrapping**
- **Methodology description**
- **Application of methodology to FCS data sets**
- **Insights and summary**

Background

- **“Accepted” standard is to execute 21 replications in CASTFOREM.**
- **Due to the long run times of CASTFOREM for the FCS KPP analysis, TRAC-WSMR Deputy Director requested investigation of replications required and comparison of alternatives.**
- **Subsequent research by TRAC-Monterey, in conjunction with TRAC-WSMR, resulted in:**
  - **Coefficient of variation (CV) as a tool to determine replications required.**
  - **Bootstrap to reduce replications.**
  - **Comparing alternatives using bootstrap confidence intervals.**

Coefficient of Variation

- **Defined as the standard deviation divided by mean.**
- **A statistical measure of the deviation of a variable from its mean.**
- **No units associated with this measure.**
- **A smaller value is better and implies less variability.**
- **The data does not have to be normally distributed.**
- **A data set with a higher CV will have a larger confidence interval than a data set with a smaller CV.**

Bootstrap

- **A bootstrap sample $x^* = (x^*_1, x^*_2, x^*_3, \ldots, x^*_n)$ is obtained by randomly sampling *n* times, with replacement, from the original data points $x_1, x_2, x_3, \ldots, x_n$. The corresponding measure of interest (e.g., mean or median) is taken.**
- **For example, assume we have seven data points of (3, 9, 8, 5, 6, 1, 10) and its mean is 6. One bootstrap sample of these seven data points might be (6, 6, 1, 8, 1, 8, 10) and its mean is 5.714.**
- **A total of 1000 bootstrap samples are done. Above is only one example. This procedure is done rapidly (within seconds) using a computer.**
- **A bias-corrected and accelerated bootstrap confidence interval (BCa) is calculated (via computer) and can be used to compare alternatives.**

Data Sets

- **TRAC-Monterey previously gained insights on MOE data characteristics from WSMR's NVESD STAMP effort.**
- **Requested and received from WSMR FCS MOE data consisting of 36 MOE each with four alternatives (11 replications per MOE and alternative combination).**
- **Thus, we were provided 144 data sets to determine the potential of the CV and bootstrap.**

Methodology

- **This methodology was used for each of the 144 data sets.**
- **The mean and median were calculated for all 11 replications.**
- **The CV, test for normality, mean 90% BCa, and median 90% BCa were calculated for the first five replications, then for the first six replications, then for the first seven replications,…, then for all 11 replications.**
- **Insights for applicable CV measures, effect of non-normal data, and assessing alternatives with the BCa were gained.**

Methodology Example

- **The first data set included the 11 data points of (279, 287, 356, 297, 302, 291, 294, 288, 286, 352, 306).**
- **The true mean of these 11 replications is 303.5 and true median is 294. The true 90% confidence interval (using parametric statistics) of the 11 replications is (289.2, 317.7), but note the data is non-normal.**
- **1000 bootstrap samples were taken for each of the number of replications.**
- **Note that a different 1000 bootstrap samples can yield slightly different numbers, but these differences are negligible.**
- **Note the BCa has a slightly wider confidence interval, but it does not require normality assumptions.**

Number of Replications	Normal	CV	Mean 90% BCa	Median 90% BCa
5	No	0.087	(287.8, 330.4)	(279, 302)
6	No	0.087	(289.2, 327.2)	(283, 302)
7	No	0.087	(290.7, 324.5)	(279, 297)
8	No	0.087	(290, 318.1)	(287, 297)
9	No	0.087	(289.2, 316.1)	(286, 294)
10	No	0.086	(292.7, 322.8)	(287, 299.5)
11	No	0.086	(293.1, 319.8)	(287, 302)

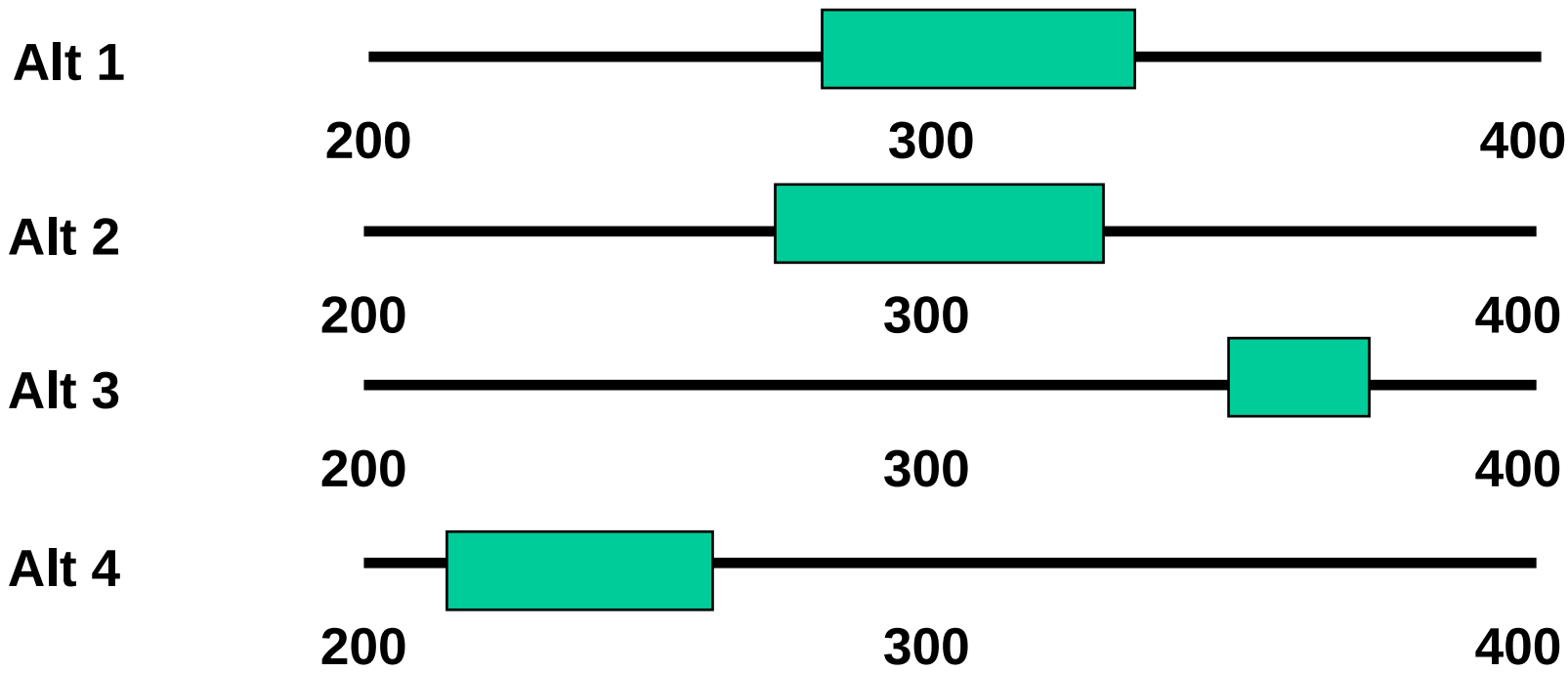
Methodology Example (continued)

- **The previous slide was for one MOE on one of the four alternatives. For this specific MOE, the other three alternatives were examined.**
- **When a CV < .20 was achieved, the procedure was terminated at that number of replications.**
- **For this particular MOE, after five replications, each alternative had a CV < .20.**
- **"Regardless of the cost per replication, we (Law & Kelton) recommend always making at least three to five replications of a stochastic simulation to assess the variability of the X_j's."**

	True Mean	True Median	Replications	Normal	CV	Mean 90% BCa	Median 90% BCa
Alternative 2	316.3	314	5	Yes	0.131	(276.6, 318.6)	(258, 316)
Alternative 3	349.9	361	5	Yes	0.066	(337.2, 361.6)	(326, 364)
Alternative 4	235.4	241	5	Yes	0.109	(218.2, 264)	(200, 261)

Comparing Alternatives

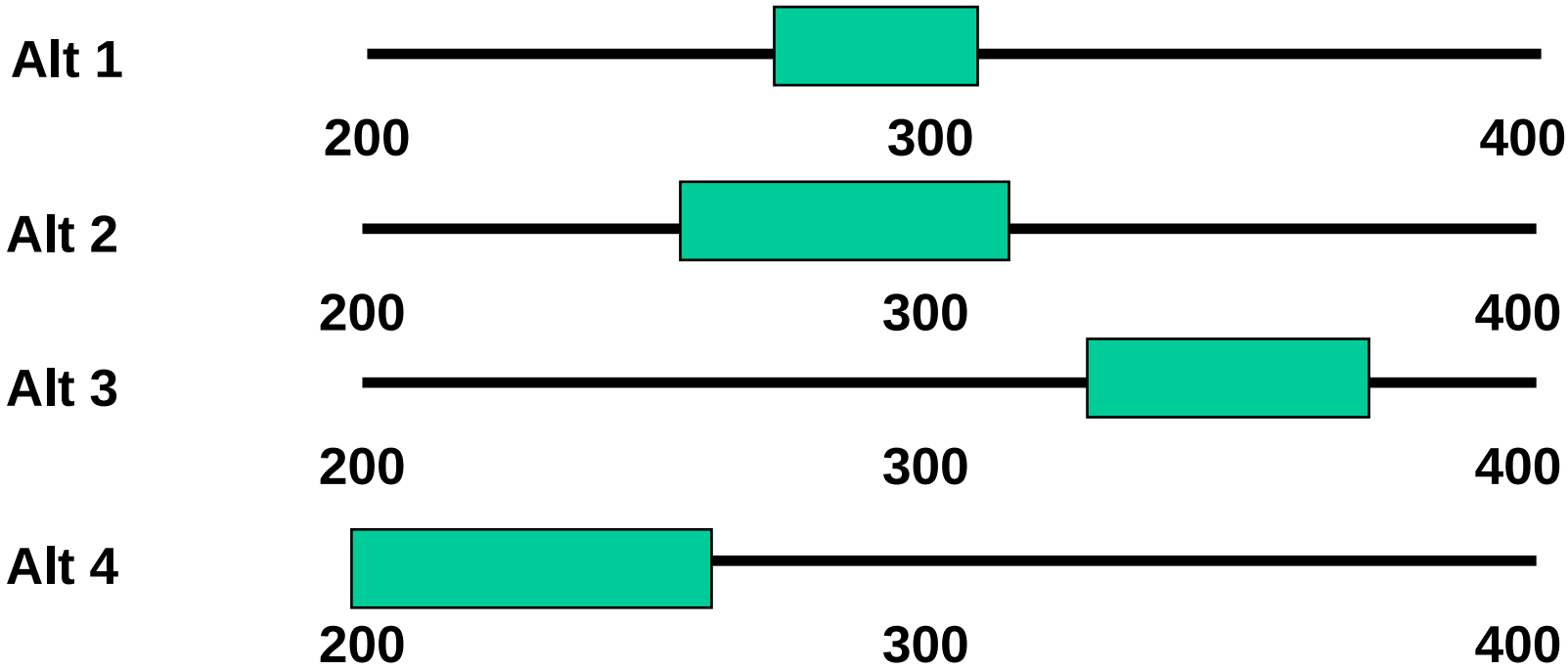
- **Use the mean 90% BCa (green rectangles) to compare the MOE for the four alternatives. Experience shows that a 90% BCa provides a robust confidence interval.**



- **If "bigger is better," then Alt 3 > Alt 1 = Alt 2 > Alt 4. This is obtained by examining if significant overlap of the green bands occurs between alternatives. Also, insight gained that Alt 3 has the least variability.**

Comparing Alternatives (continued)

- **Similarly, use the median 90% BCa (green rectangles) to compare the MOE for the four alternatives.**



- **Similar to the mean, Alt 3 > Alt 1 = Alt 2 > Alt 4 by examining if significant overlap of the green bands occurs between alternatives.**

Analysis of Remaining MOE's

- **This procedure was executed for each of the remaining 35 MOE's.**
- **These results can be presented upon request.**
- **The purpose of this work was not to check the work of TRAC-WSMR, but to determine the merits and implementation insights of the CV and bootstrap methods.**

Methodology Insights

- **Although both the mean and median can be used, the mean appears to be sufficient. There was not a significant difference when comparing alternatives on whether the data was normally distributed or not (assessed using the Kolmogorov-Smirnov Test).**
- **The CV was less than .10 in almost 80% of the 144 data sets after five replications. When the bootstrap procedure was done on these five replications, the resulting mean 90% BCa included the true mean in all cases.**
- **The CV was less than .20 in over 86% of the 144 data sets after eight replications. When the bootstrap procedure was done on these eight replications, the resulting mean 90% BCa included the true mean in all cases.**

Methodology Insights (continued)

- **Approximately 14% of the CV values were greater than .20 (some greater than .70), but after eight replications, the resulting mean 90% BCa included the true mean in all cases.**
- **If the CV is high for a particular MOE in one alternative, it is high for all of the alternatives for that MOE.**
- **The CV does not significantly change from 5-11 replications. For example, from our MOE example for alternative 1, the CV after five replications was .087 and after 11 replications, the CV was .086.**
- **The magnitude of the MOE value does not effect the CV (unitless). For example, the MOE example for Alternative 1 had values ranging from 279 to 352 and had a CV of .087. Another MOE we examined had values ranging from .79 to .831 and had a CV of .017.**

Summary

- **If you have the available resources, then there is nothing that substitutes for the actual data obtained from executing the simulation.**
- **The CV value (especially when paired with a “picture” of the data) appears to be a good measure to determine how many replications are required and does not require normality assumptions.**
- **If the FCS KPP simulation runs do require significant resources (mainly time), the bootstrap appears to offer good results after five replications when compared to the 11 replications.**
- **The 90% mean BCa is an excellent analytical and visual tool to show where differences between alternatives exist.**