Determining the Number of Simulations Needed for Stable Results

Rebecca Murphy[1], Jon Harcum[2], Elgin Perry[3],
Breck Sullivan[4], and Peter Tango[4],

[1]UMCES at CBP, [2]Tetra Tech,[3]Statistics Consultant, [4]USGS at the CBP

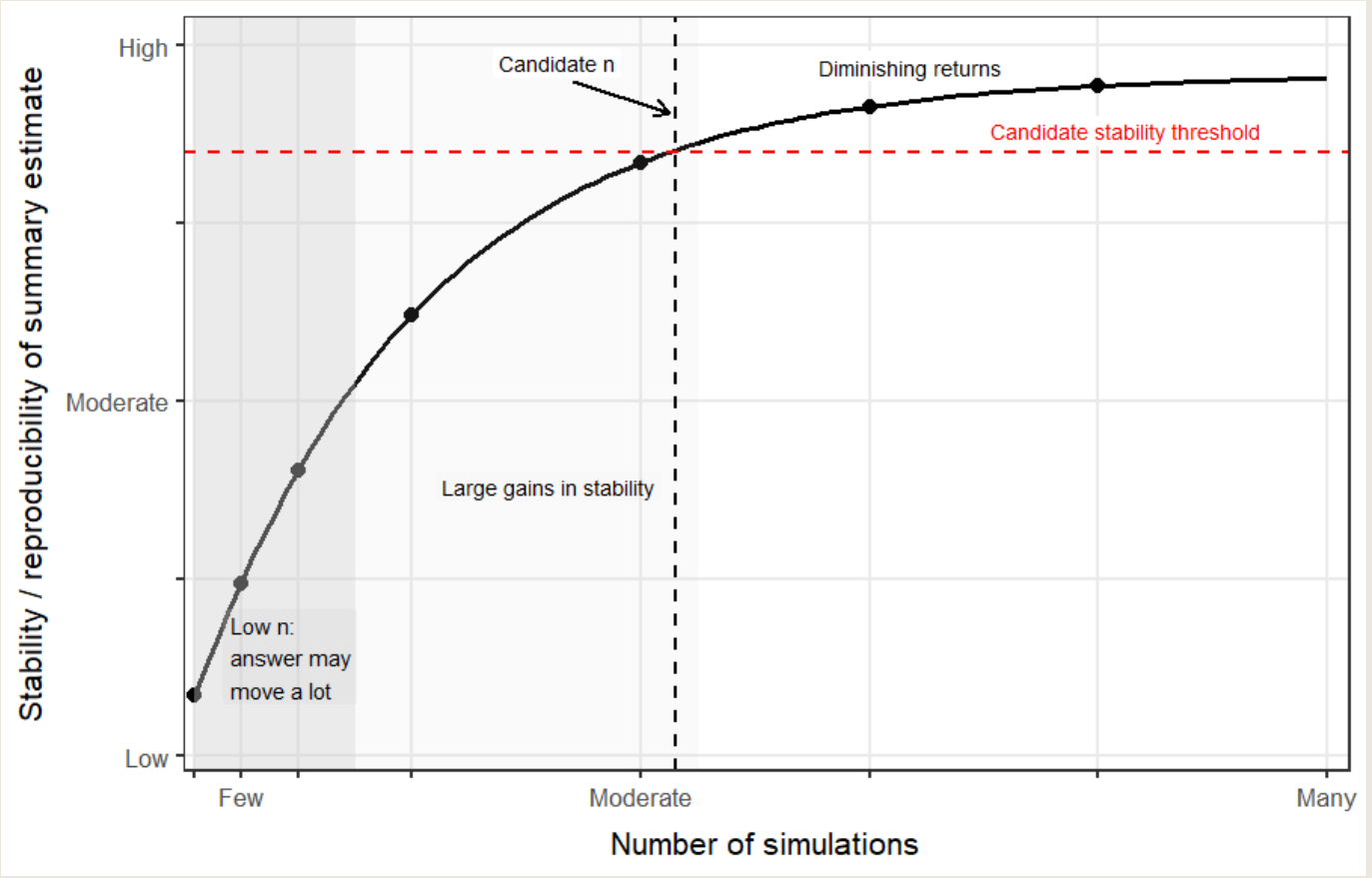
BORG

April 20, 2026

Number of Simulations

How many simulations are enough?

- The goal is to identify the smallest simulation count that yields a scientifically stable answer.
- "Enough" depends on the **endpoint of interest**: central tendency, tail behavior, or the resulting management interpretation.
- In practice, this is a **stability-versus-cost** problem rather than a purely statistical one.
- The key issue is whether additional runs materially change the answer or only add computational expense.

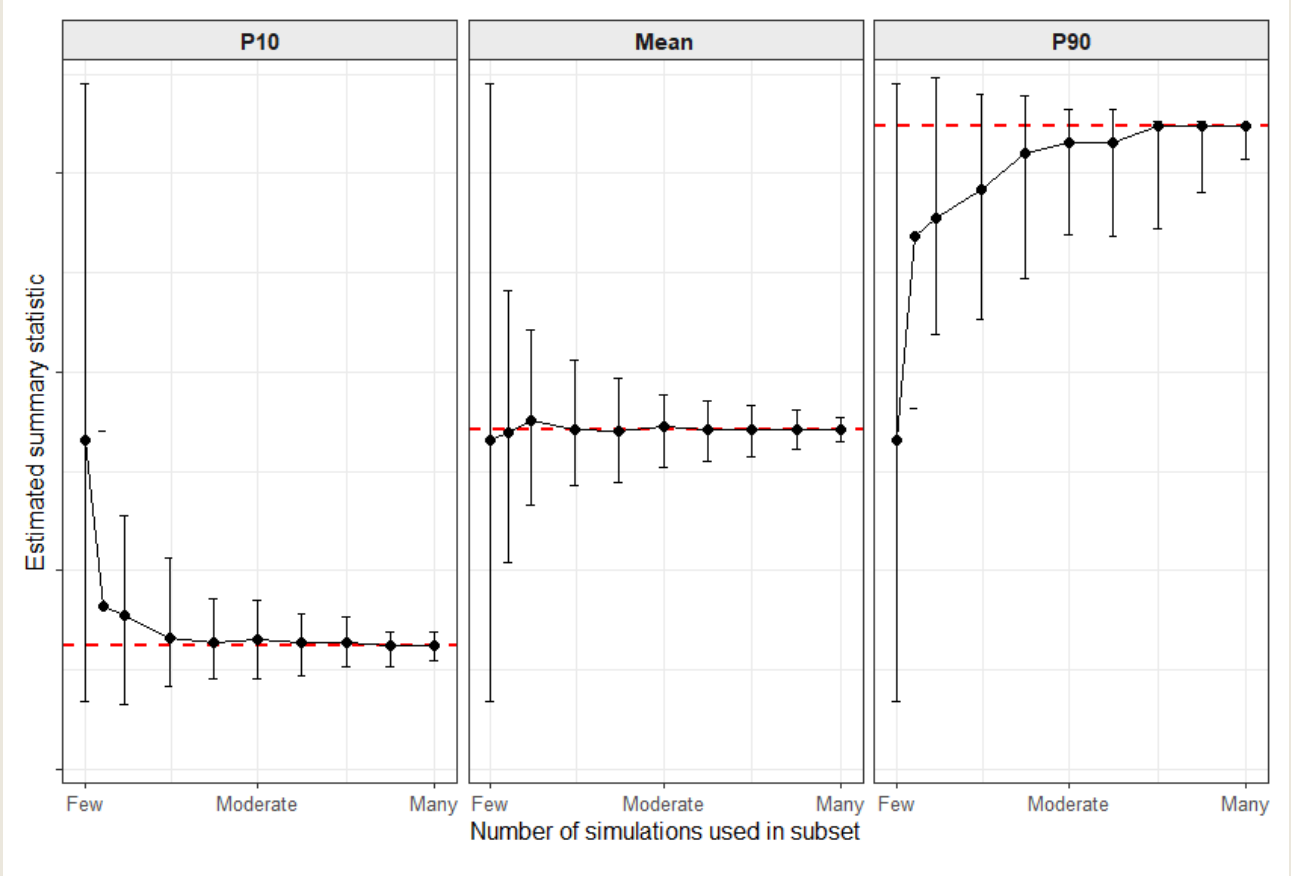


Goal for today: get input on what should be considered in this analysis.

General Approach

A common approach for determining simulation count

- Treat a large simulation set as a working reference distribution.
- Recompute the target summary using smaller subsets of runs.
- Examine two things as subset size increases:
  - *bias relative to the reference*
  - *variability across repeated subsets*
- Ask where the summaries become effectively insensitive to adding more runs.
- The required count will usually differ for:
  - *means versus tails*
  - *well-behaved versus heterogeneous cases*



Questions for the group

- Which quantities are most important to stabilize for the interpolator: mean, lower tail, upper tail? Other (e.g., percent less than)?
- Should adequacy be defined by typical behavior across cases, or by the most variable cases?
- Is it acceptable to use one common simulation count, or should the count depend on class or application?
- At what point does additional precision stop being worth the extra run time?

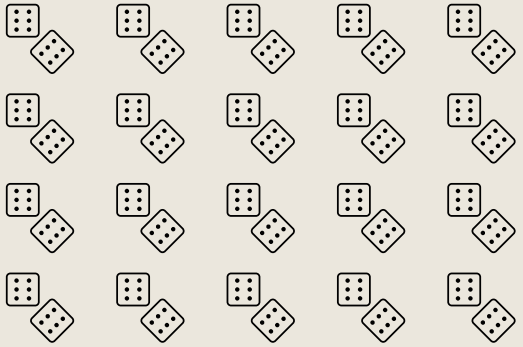
One Experiment

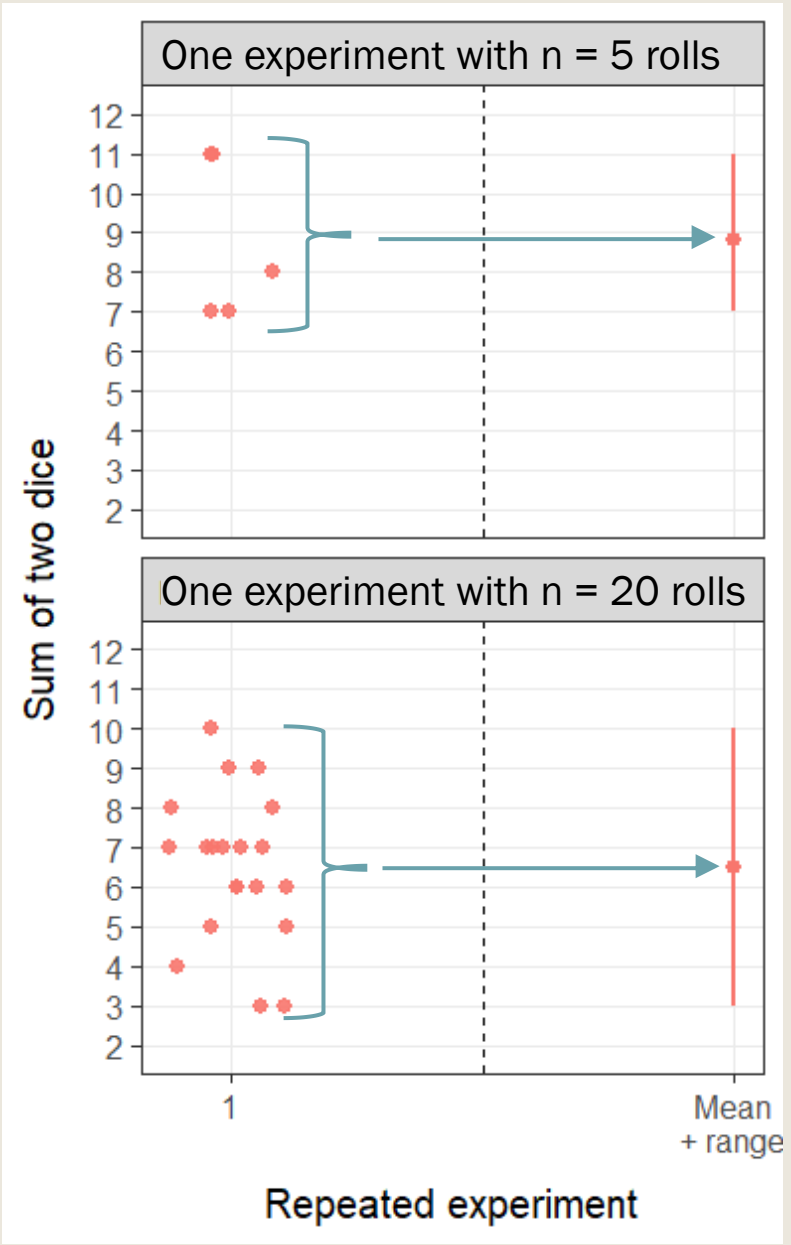
One experiment

- Roll 2 dice, compute sum
- Repeat 5 times

One experiment

- Roll 2 dice, compute sum
- Repeat 20 times





- One experiment produces one summary value.
- Larger n usually gives a more stable summary.

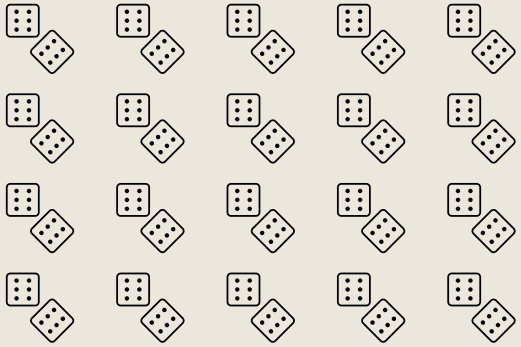
Repeat the full experiment 6 times

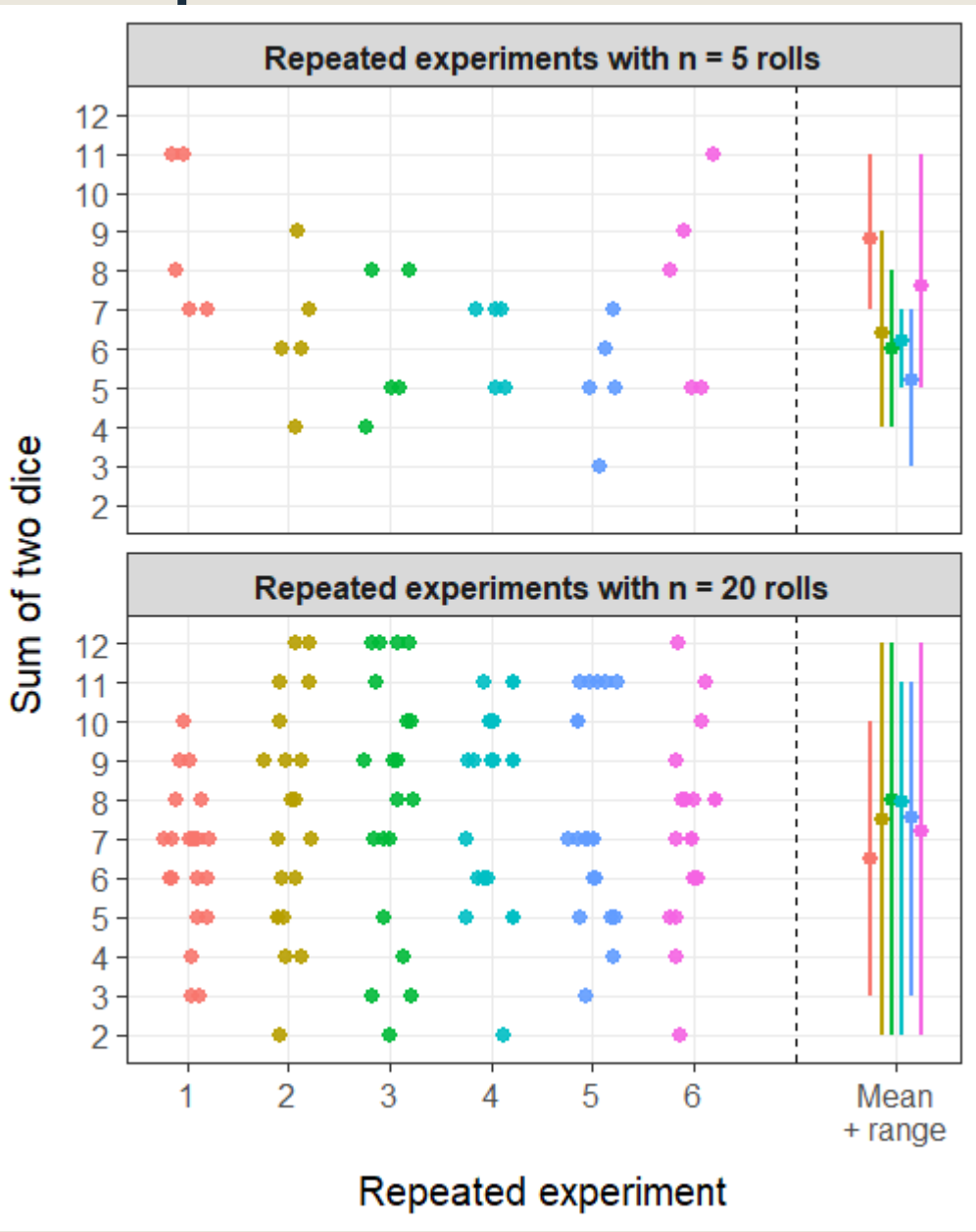
Repeat 6 times

- Roll 2 dice, compute sum
- Repeat 5 times

Repeat 6 times

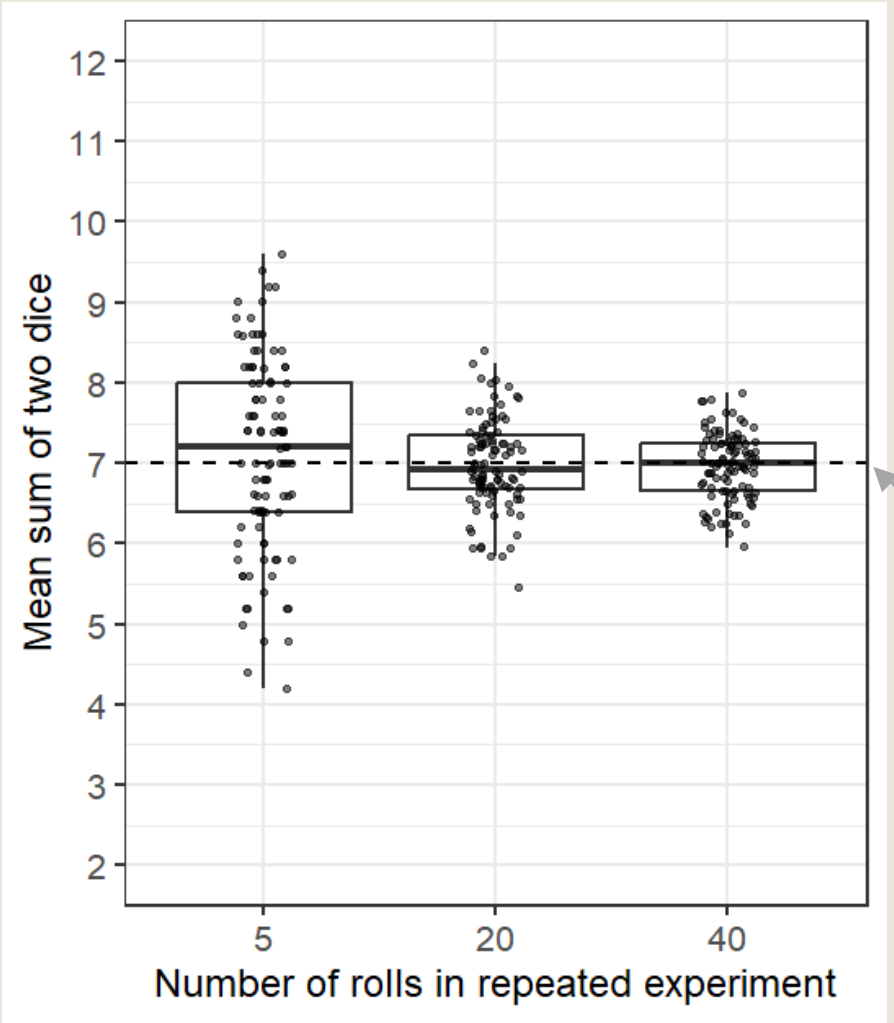
- Roll 2 dice, compute sum
- Repeat 20 times





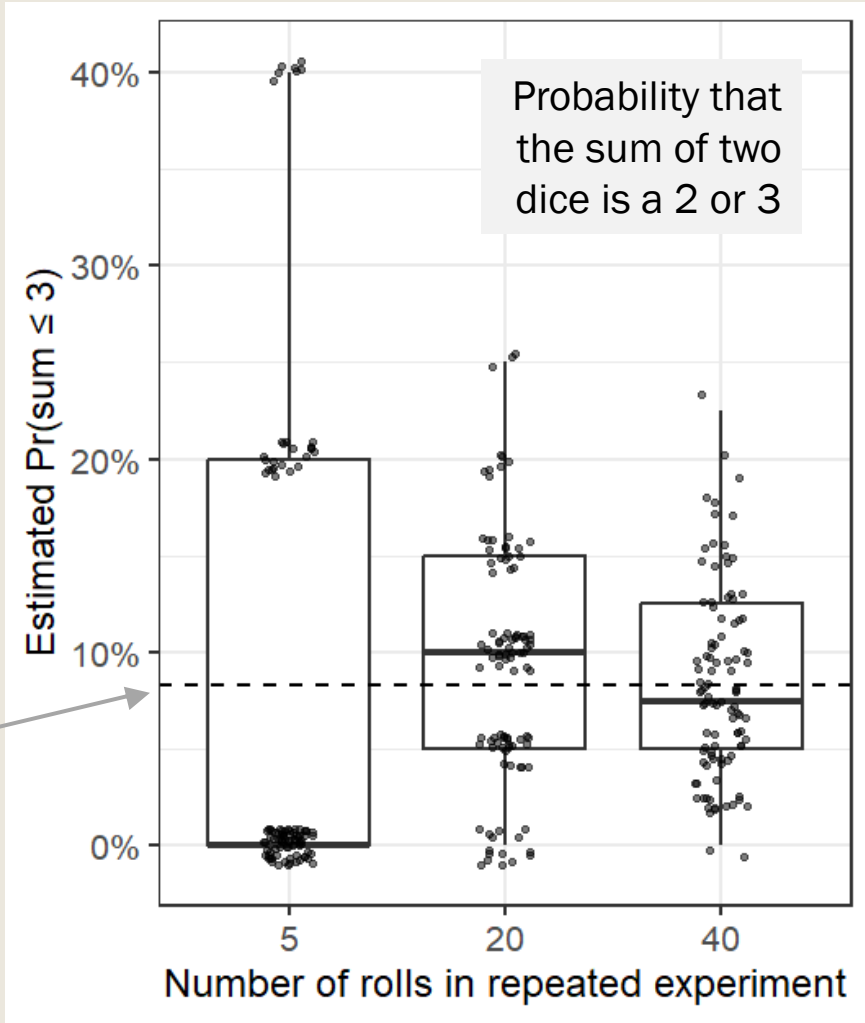
- This is analogous to running a fixed number of simulations and summarizing the result.
- The question is how stable that summary becomes as n increases.

From a Few Repeats to Many Repeats



True mean of 7

Expected Pr = 8.3%



Tail probabilities stabilize more slowly because rare outcomes appear less often with smaller experiments.