Decision-Based Metrics for Test and Evaluation Experiments

Dashi I. Singham, Ph.D.
Research Associate Professor
Operations Research Department
Naval Postgraduate School
April 2017

Research Question

- Sequential sampling rules can be used to decide the number of experimental replications
- These rules are often made independently of the desired outcome or decision.
- Let D be the decision threshold for making a decision about a system.

How can we use decision criteria to inform sequential sampling rules?

Confidence Intervals

Estimated standard deviation

Estimated mean

$$\left[\bar{x}_n \pm \underbrace{t_{\alpha,n-1}\frac{s_n}{\sqrt{n}}}_{\text{Half-width}}\right]$$

- Confidence intervals represent the uncertainty in the mean performance of a system based on *n* samples
- Often assume normality in the data
- The half-width should be small enough to ensure that the variation in the mean estimate is acceptable

Fixed Sampling Rules – Choosing the sample size

Can use standard deviation estimate

$$n \geq \left(\frac{z_\alpha \sigma}{\delta}\right)^2$$

Desired half-width of confidence interval

- If a variance estimate is available, can calculate ahead of time how many samples should be taken to obtain a confidence interval with a half-width smaller than δ.
- Challenges:
  - hard to choose δ
  - n might be large
  - Variance estimate might not be available

Solution: Sequential Sampling Rules

Absolute precision rules: fix a value of δ and collect samples until the half-width is smaller than δ.

$$n^* = \text{smallest } n \text{ s.t. } t_{\alpha,n-1}\frac{s_n}{\sqrt{n}} \leq \delta$$

Relative precision rules: fix a percentage δ and collect samples until the half-width is within some percent of the sample mean.

$$n^* = \text{smallest } n \text{ s.t. } t_{\alpha,n-1}\frac{s_n}{\sqrt{n}} \leq \delta\bar{x}_n$$

Pros: can stop earlier when desired precision is reached, do not need variance estimate ahead of time

Cons: statistical bias (confidence interval coverage $<1-\alpha$), could still require a large number of samples

Example from Small Arms Testing

Table reproduced from TOP 3-2-045

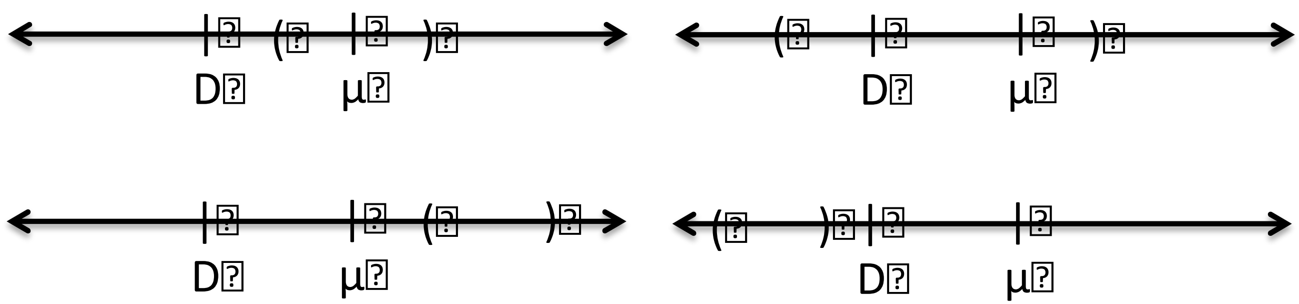
ITEM	MAXIMUM PERMISSIBLE ERROR OF MEASUREMENT*
Brookfield viscometer	$\pm$0.5% full-scale reading.
Cyclic rate recorder	$\pm$1% at rates up to 6000 spm and burst lengths of 100 rounds.
Stargage and airgage	$\pm$0.025 mm.
Thermograph/thermocouples	$\pm$0.6 ^{o}C (1 ^{o}F).
Velocimeter	0.1% or 0.5 m/s (whichever is highest) for bursts to 6000 spm.

*Values can be assumed to represent ±2 standard deviations.

Four Types of Confidence Interval (CI) Results

- Let μ be the true mean system performance.
- Let D be the decision threshold
  - If CI>D, then "Accept" the system as meeting the requirement
  - If CI<D, or includes D, then "Reject" the system as failing to meet the requirement

- Confidence intervals can either
  - correctly include (cover) μ or not
  - correctly determine whether μ is greater than D, or not.



Cover & Correct	Cover & Incorrect
Fail to cover & Correct	Fail to cover & Incorrect

Decision-Based Procedures

- Traditional sequential rules do not incorporate D
- New rules directly incorporate D in the stopping criterion

$$n^* = \text{smallest } n \text{ s.t. } t_{\alpha,n-1} \frac{s_n}{\sqrt{n}} \leq |\bar{x}_n - D|$$

- Results in confidence intervals that usually do not include D
- Will require more samples if the true performance is close to D
  - Ensures precise confidence interval if decision is a close call
- Will end early if true performance is far from D
  - Saves time/replications if the decision is obvious

Conclusions and Implementation

- The expected number of samples required by the procedure depends on how far the unknown μ is from the known requirement D.
- σ=1, 1-α=90%

$\lvert\mu - D\rvert$	Expected no. samples
1.0	5.0
0.5	10.9
0.3	28.6
0.1	263.9
0.05	1095.9
0.0	∞

Confidence interval coverage can be poor for small sample sizes.

Solution: aim for high confidence (99%) known actual confidence may be closer to 95% or 90%.

Conclusions and Future Work

- Developed a new sequential stopping rule that incorporates the decision requirement D.
  - Potentially more efficient in making a decision.
- There may be significant bias associated with sequential rules, along with our proposed modifications. Simulation testing can be used to estimate the bias.
- Looking for ongoing/real test data to estimate the impact of sequential rules.