Cumulative mean bounds for quality control analysis

Singham, Dashi I.

D.I. Singham, M.P. Atkinson, "Cumulative mean bounds for quality control analysis", pp. 541-550. Proceedings of the 2015 Winter Simulation Conference, 6-9 December 2015, (L. Yilmaz, et al., eds.), 10 p.

http://hdl.handle.net/10945/48836

Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

CUMULATIVE MEAN BOUNDS FOR QUALITY CONTROL ANALYSIS

Dashi I. Singham
Michael P. Atkinson

Operations Research Department
Naval Postgraduate School
1411 Cunnigham Road
Monterey, CA 93943, USA

ABSTRACT

We develop an alternative to confidence intervals, called cumulative mean bounds, for estimating the mean behavior of system performance. Cumulative mean bounds assess the probability that the cumulative sample mean stays within a given distance from the unknown true mean and rely on properties of standardized time series. We extend the properties of cumulative mean bounds to estimate the probability that the cumulative sample mean will reach a particular value. This idea can be used to analyze quality control methods for predicting mean simulation output behavior given an initial sample of output.

1 INTRODUCTION

Simulation output analysis research is often devoted to estimating the mean performance of a simulation system. Confidence intervals are one way of assessing the quality of a system by providing a measure of variation in the estimate of mean system performance. We introduce a new measure of reliability for evaluating the mean performance of a system, which considers the path taken by the cumulative sample mean over time. We calculate the probability that the cumulative sample mean will stay within a given distance from the unknown true mean after an initial sample size. This measure is called cumulative mean bounds (*CMB*).

This is different from confidence intervals, which consider the sample mean at one point in time, but there are many similarities between the two measures. Both face a tradeoff between desired confidence in the measure (the probability guarantee), and the precision (the prescribed distance allowed from the true mean). Increasing the number of samples will improve the quality of the results. Confidence intervals give an estimate of the mean and the variability associated with that estimate, and our measure is stronger in that we desire the sample mean to be within a given distance from the true mean over an extended period of time.

In this paper, we show how *CMB* can be used to estimate the probability that the cumulative sample mean will cross a particular threshold value given an initial starting sample. This makes the measure useful in a quality control setting, where we may wish to design a system that will stay within a certain acceptable range that is defined in absolute terms. We imagine a setting in which performance of the system is based on the average performance throughout the entire time period, not just the current time period. Examples of this include reporting mean profitability at the end of a fiscal year, or assessing the daily cost of operating a plant over the long term. While a confidence interval can assess the quality of the long-term mean, *CMB* measures the probability that the mean is likely to be consistently good over time. For example, in investment management, average rates of return are reported regularly. Consistent mean performance can help maintain or attract customers, whereas volatility in the sample mean may reflect poorly, even

978-1-4673-9743-8/15/$31.00

if long-term performance is good. A confidence interval will not directly measure this volatility in the reported mean over time.

As with confidence interval procedures, we require a good variance estimator for our measure to be valid. While confidence intervals rely on approximate normality of the sample mean, *CMB* relies on convergence of a standardized time series formed from the data to a Brownian bridge. The validity of *CMB* for a particular data type depends on the extent to which a standardized time series can be approximated by a Brownian bridge. The quality of this approximation depends on the levels of dependence and non-normality in the data, and standardized time series methods often require large amounts of data. The covariance structure of the standardized time series should approximately match that of its limiting Brownian bridge process. Details on the construction of *CMB* for the independent and normally distributed case are in Singham and Atkinson (2015a). In this paper, we summarize the *CMB* metric and focus on estimating the probability the sample mean will cross a given threshold. In particular, we develop the measure in the case where the underlying data is assumed to be independent and identically distributed (i.i.d.) with a normal distribution so that the standardized time series has the statistical properties of a Brownian bridge.

In many classical CUSUM charting methods, it is assumed that the true mean μ and variance σ^2 of the in-control system are known from extensive experimentation in order to develop methods for testing long-term shifts to a new mean. See Hawkins and Olwell (1998) for more information on CUSUM charts. Kim, Alexopoulos, Goldsman, and Tsui (2006) develop CUSUM charts for detecting changes in the mean of a simulation process without assuming a model structure. Rather than detect shifts in a process, we attempt to predict whether the mean performance will reach a particular level given its variability. We do not assume knowledge of the true mean and variance, but will use estimates to predict the probability that the sample mean will stay within acceptable levels.

We propose the following two-stage experiment. In the first stage, estimates for the mean and variance are collected. This is then used in conjunction with *CMB* to estimate the probability that in the second stage, the cumulative sample mean will stay below a pre-specified control limit, given another initial set of samples. While *CMB* was developed to predict performance relative to the true mean, here we use it to estimate performance relative to a fixed value which could be chosen without knowledge of the true mean. In Section 2 we describe the construction of *CMB*. Section 3 introduces the translation of *CMB* to quality control by relating it to the probability of a false detection of a shift in the mean. We provide numerical results and analysis in Section 4 and Section 5 concludes.

2 CUMULATIVE MEAN BOUNDS

We summarize the construction of *CMB* from Singham and Atkinson (2015a). Suppose we have simulation output $Y_1, Y_2, \ldots Y_m$. Define the cumulative mean after j samples, $j = 1, \ldots, m$ as $\overline{Y}_j = \frac{1}{j}\sum_{i=1}^{j} Y_i$. We are interested in evaluating the probability that after k initial samples, the cumulative sample mean stays within a given distance δ from the long term mean $\overline{Y}_m$, when m is large. Taking m to the limit, *CMB* is the probability that after k initial samples, the cumulative sample mean of these samples stays within a given distance δ from the true mean of the process, μ. Explicitly, we define *CMB* as follows:

$$CMB := P\left(\bigcap_{j\geq k}\left|\mu - \frac{1}{j}\sum_{i=1}^{j} Y_i\right| \leq \delta\right). \quad (1)$$

We note that $k > 0$ is the initial sample size required to avoid a trivial value of *CMB* as zero when $k = 0$, and also gives us an initial sample required in order for the probability guarantee to hold. As k becomes large, we expect the probability that the sample mean stays within distance δ from μ approaches 1. In Singham and Atkinson (2015a), the authors suggest a secondary measure which calculates the probability that the sample mean stays within a given distance from the true mean in one direction, where the sample mean stays below the true mean with an allowed tolerance δ. This is written as

$$CMB' := P\left(\bigcap_{j\geq k}\frac{1}{j}\sum_{i=1}^{j}Y_i - \mu \leq \delta\right) \tag{2}$$

where we calculate the probability the sample mean stays less than μ plus some allowed tolerance δ. In this paper, we focus on this "one-sided" measure to develop a quality control metric, where average system performance should stay below some upper threshold value. The results can be modified to consider a lower threshold value.

To calculate the value of *CMB*, we rely on properties of standardized time series developed in Schruben (1983). Standardized time series based on m samples of observations rescale the data to form a new time series over the range of time $t \in [0,1]$ according to the following function:

$$X(t) = \frac{\lfloor mt\rfloor\left(\frac{1}{m}\sum_{i=1}^{m}Y_i - \frac{1}{\lfloor mt\rfloor}\sum_{i=1}^{\lfloor mt\rfloor}Y_i\right)}{\sigma\sqrt{m}}, \quad t \in [0,1].$$

Let σ^2 be the asymptotic variance constant $\sigma^2 = \lim_{m\to\infty} mVar[\overline{Y}_m]$. To use standardized time series, we require the conditions for a Functional Central Limit Theorem (FCLT) to hold, and Schruben (1983) shows that under these mild conditions the standardized time series will converge to a Brownian bridge as $m \to \infty$. For more details on how standardized time series can be used to analyze simulation output, please see Alexopoulos (2006).

In Singham and Atkinson (2015a) we derive a lower bound for CMB' when the data are i.i.d. normally distributed. We note that for a given value of μ, we can estimate the value of CMB' using simulation. We present the associated theorem for CMB' here.

Theorem 1 Under the assumption that the data is i.i.d. normally distributed, the probability that the sample mean stays within distance δ (on one side) from its long term mean $\overline{Y}_m$ for all $k \leq j \leq m$ has a lower bound

$$CMB'_L(\delta,\sigma,k,m) \leq P\left(\bigcap_{k\leq j\leq m}\frac{1}{j}\sum_{i=1}^{j}Y_i - \frac{1}{m}\sum_{i=1}^{m}Y_i \leq \delta\right) = CMB'(\delta,\sigma,k,m)$$

where

$$CMB'_L(\delta,\sigma,k,m) = 2\Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-\frac{k}{m}}}\right) - 1. \tag{3}$$

Define $\Phi(x)$ as the standard normal cumulative distribution function applied to x. We derive (3) by rewriting (2) as a function of a standardized time series $X(t)$. For the i.i.d. normal case the points $\frac{j}{m}, j = 1,\ldots,m$ of a standardized time series will have the same joint distribution as the same points of a standard Brownian bridge. Details and proofs are given in Singham and Atkinson (2015a). We calculate (3) by deriving boundary-crossing properties of Brownian bridges. The lower bound is the result of assuming a continuous Brownian bridge in calculating boundary-crossing probabilities, as opposed to a standardized time series evaluated at discrete points. In practice the lower bound is fairly tight. In the limit as $m \to \infty$ we can extend Theorem 1 by replacing the long term mean $\overline{Y}_m$ with the true mean μ. This extension produces the following result

$$CMB'_L(\delta,\sigma,k) \leq P\left(\bigcap_{j\geq k}\frac{1}{j}\sum_{i=1}^{j}Y_i - \mu \leq \delta\right) = CMB'(\delta,\sigma,k)$$

where

$$CMB'_L(\delta,\sigma,k) = 2\Phi\left(\frac{\delta\sqrt{k}}{\sigma}\right) - 1. \tag{4}$$

The details extending Theorem 1 in the limiting case to produce (4) appear in Singham and Atkinson (2015b). We will focus on the limiting case in (4) for the remainder of this paper. An issue to consider when using *CMB* in a non-i.i.d. normal setting is the quality of the approximation of a standardized time series by a Brownian bridge. Small sample sizes or highly dependent data may prevent this approximation from being valid, and could result in (4) failing to be a lower bound for finite sample sizes. Singham and Atkinson (2015b) study appropriate limiting conditions for FCLT data results to hold.

As confidence intervals can be defined exactly for i.i.d. normal data, in this paper we develop quality control extensions of CMB' using i.i.d. normal data to develop properties of the measure. The objective is to use CMB' to predict the likelihood of the cumulative sample mean staying within some acceptable range. As stated in (1) and (2), we have defined this acceptable range in the original statement of *CMB* as $[\mu-\delta,\mu+\delta]$, and for CMB' as $[-\infty,\mu+\delta]$ in the one-sided case. Next, we will translate CMB' to estimate the probability the cumulative sample mean will stay within the fixed range $[-\infty,Q]$ where Q is the control limit.

3 QUALITY CONTROL METRICS

We translate information from CMB' into quality control metrics for the sample mean. Information about cumulative mean behavior can be used to predict the probability that a sample mean will stay below a specified level Q given past data samples. We note that we have a different objective than traditional quality control metrics which are designed to detect shifts in the mean as data arrives. Our goal is to estimate the probability that the sample mean will stay below some level Q. This can be related to a probability of a false alarm in a CUSUM framework, where the quality of the measure is evaluated using the average run length until a false alarm occurs.

We derive this related measure by conditioning on the distribution of the true mean given past data to obtain absolute control limits. To do this, we set up a two-stage experiment. In the first stage, we collect samples to estimate the mean and variance. To start, assume that we have k_1 samples with sample mean $\overline{Y}_{k_1}$ collected from the first stage. We would like to calculate the probability that the sample mean will stay below some level Q after k_2 further samples. CMB' gives a probability guarantee for the sample mean staying below its true mean μ with allowed tolerance δ after a given number of samples (in this case, k_2 samples in the second stage). Thus, we use CMB' in the second-stage to estimate the probability that the sample mean stays below Q. Figure 1 shows the second stage control limit and how the value of Q can be related to μ and δ in the CMB' context. In a quality control setting, we would fix μ as the in-control mean value and try to detect when the sample mean has shifted by at least δ to be higher than Q. A high average run length until a detection means the probability of a false alarm is low. *CMB* can be used to estimate the probability a false alarm will occur over a given time period if the true mean remains at μ.

For this paper, we assume the underlying samples Y_i are i.i.d. normal to help establish the properties of the estimation. Many algorithms exist for determining batch sizes required to obtain batched means that are approximately i.i.d. normal. For examples, see Steiger and Wilson (2002), or for one that uses standardized time series, see Alexopoulos, Goldsman, Tang, and Wilson (2013). Assuming the samples are obtained from an i.i.d. normal population, given that the first-stage sample mean is $\overline{Y}_{k_1}$ and the sample variance is $\hat{\sigma}^2_{k_1} = \frac{1}{k_1-1}\sum_{i=1}^{k_1}(Y_i-\overline{Y}_{k_1})^2$, we can approximate the distribution of $\overline{Y}_{k_1}$ as $\overline{Y}_{k_1} \sim \mathcal{N}(\mu,\hat{\sigma}^2_{k_1}/k_1)$, so our uncertainty about the value of μ can be quantified as $\mathcal{N}(\overline{Y}_{k_1},\hat{\sigma}^2_{k_1}/k_1)$. To calculate the probability that the sample mean will stay below Q after another k_2 samples, we condition on possible values of μ

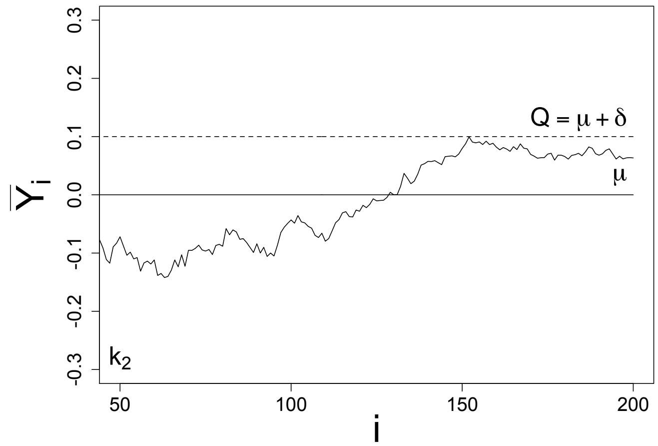

Figure 1: Second stage control limits. We estimate the probability that the cumulative sample mean will stay below Q after k_2 samples. We fix Q, while μ and the corresponding δ are estimated via first-stage sampling.

and use CMB':

$$
\begin{aligned}
CMB'_Q := P\left(\bigcup_{i\geq k_2}(\overline{Y}_i \leq Q)\right) &= CMB'(Q-\mu,\sigma,k_2) && (5)\\
&\approx \int_{-\infty}^{Q} CMB'(Q-x,\hat{\sigma}_{k_1},k_2)\mathcal{N}(x,\overline{Y}_{k_1},\hat{\sigma}^2_{k_1}/k_1)dx && (6)\\
&\geq \int_{-\infty}^{Q} CMB'_L(Q-x,\hat{\sigma}_{k_1},k_2)\mathcal{N}(x,\overline{Y}_{k_1},\hat{\sigma}^2_{k_1}/k_1)dx && (7)\\
&= \int_{-\infty}^{Q}\left(2\Phi\left(\frac{(Q-x)\sqrt{k_2}}{\hat{\sigma}_{k_1}}\right)-1\right)\mathcal{N}(x,\overline{Y}_{k_1},\hat{\sigma}^2_{k_1}/k_1)dx && (8)\\
&:= \widehat{CMB'}_Q.
\end{aligned}
$$

Because we are uncertain about the value of μ, we approximate our uncertainty by conditioning on μ using the distribution obtained from the information in the first stage in (6). As our first stage sample k_1 increases to ∞, our uncertainty about the sample mean decreases and we obtain (5) in the limit. Conditioning on the value of μ as x, we can use CMB'_L to estimate obtain a lower bound for the probability that the cumulative sample mean ever exceeds Q in (7). The lower bound estimate $\widehat{CMB'}_Q$ is then readily available as (8) using numerical integration which is much faster than simulating the value of CMB' exactly in (5) or (6). We evaluate the quality of the approximation and the lower bound in the next section. We need to determine under what conditions the lower bound is actually a lower bound, and we would like the lower bound to be tight for it to be meaningful.

4 RESULTS AND ANALYSIS

With an infinite sample size in stage 1, we would have full knowledge of the distribution of mean and variance of the data and would be able to determine the probability the sample mean ever crossed Q as $CMB'(Q-\mu,\sigma,k_2)$ with k_2 initial samples. Because we approximate the distribution of the data, we assess the quality of this approximation from using a finite sample size k_1.

We explore the tradeoff between investing samples in the first stage (k_1) to obtain a good estimate of μ as opposed to allowing for a larger initial sample size in the second stage (k_2) before the probability guarantee applies. We explore the relationships in two contexts. In the first context, the first stage is a simulation experiment to estimate the mean performance of the system where we can take k_1 to be arbitrarily large. The second stage samples consist of observations of a real process, where we require some initial sample size k_2 for the sample mean to stabilize. In the second context, the second stage has the same sampling cost as the first stage. In this case, we allocate between the two stages using a fixed budget. For both cases, the usage of Section 3 can be summarized as follows:

1. Collect k_1 samples of Y_i data samples.
2. Estimate μ and σ^2 from samples $Y_1, \ldots, Y_{k_1}$ using $\overline{Y}_{k_1}, \hat{\sigma}^2_{k_1}$.
3. Return $\widehat{CMB'}_Q$ using (8).
4. $\widehat{CMB'}_Q$ gives the probability estimate that the sample mean in the second-stage will remain below Q after more than k_2 samples.

4.1 Different Sampling Costs in Each Stage

We experiment with different sample sizes k_1 and k_2 to understand the relationships between the sample sizes and the performance of $\widehat{CMB'}_Q$. We assume that k_1 can be made arbitrarily large as in a simulation context, while k_2 consists of real data samples which are more costly or take time to observe. For example, k_2 could be the number of days needed to observe daily failure rates of a machine to ensure that the mean failure rate will not exceed some level in the future. For all results, we simulate sample paths to estimate the true value of CMB'_Q.

We note that the true value of CMB'_Q increases as k_2 increases, because given a higher starting sample size in the second stage we have a higher probability that the sample mean will stay in a given acceptable range past that point, as extreme fluctuations in the mean are more likely to happen early on. Also, the quality of the lower bound CMB'_L improves as k_2 increases if the data are independent or positively correlated. Thus, we generally expect better results for higher values of k_1 (more information about the likely sample mean) and higher k_2 (the measure CMB' improves).

For the experiments in this subsection, we used $\mu = 0, \sigma = 1, Q = 0.1, m = 10000$, and ran $n = 2000$ replications to generate each table value. Table 1 shows the probability that $\widehat{CMB'}_Q$ is an actual lower bound for CMB'_Q, as there is an approximation made in (6). Table 2 shows the mean values of $\widehat{CMB'}_Q$ as well as the true value of CMB'_Q.

Looking at Table 2 we see that as k_1 increases the mean value of $\widehat{CMB'}_Q$ approaches the true value. For small values of k_1 the mean value of $\widehat{CMB'}_Q$ is biased away from its true value toward 0.5. Table 1 produces a surprising result: the probability that $\widehat{CMB'}_Q$ is a lower bound *decreases* as we increase k_1 for several values of k_2. As we increase k_2 there are two conflicting dynamics that influence whether $\widehat{CMB'}_Q$ will be a lower bound in opposite ways. A larger value of k_1 leads to a reduced variance in the estimate of μ, which decreases the likelihood of a random sample producing a $\widehat{CMB'}_Q$ larger than CMB'_Q. However, we see from Table 2 that as we increase k_1, $\widehat{CMB'}_Q$ is likely to be closer to CMB'_Q, and hence variance in $\widehat{CMB'}_Q$ is more likely to produce values above CMB'_Q. While large k_1 and k_2 produce lower probabilities of being a lower bound, we shall see shortly that any excess of $\widehat{CMB'}_Q$ over CMB'_Q in these situations is small.

Table 3 shows that as k_1 increases, the variance in $\widehat{CMB'}_Q$ decreases, as more information in the mean is obtained in the first stage of sampling. The variance remains relatively constant as k_2 increases.

To further evaluate the distribution of $\widehat{CMB'}_Q$, we plot its cumulative distribution function for different values of k_1 and k_2 in Figure 2. We see that as k_1 increases, the distribution of the estimate tightens around

Table 1: Probability that $\widehat{CMB'}_Q$ is a lower bound for CMB'_Q.

	k_2=20	56	92	128	164	200
k_1=10	0.472	0.578	0.669	0.729	0.776	0.830
100	0.571	0.612	0.643	0.673	0.710	0.744
1000	0.663	0.594	0.603	0.608	0.631	0.649
10000	0.874	0.710	0.669	0.623	0.630	0.625
100000	1.000	0.965	0.891	0.833	0.826	0.761

Table 2: Mean value of $\widehat{CMB'}_Q$.

	k_2=20	56	92	128	164	200
k_1=10	0.434	0.493	0.509	0.526	0.525	0.525
100	0.358	0.486	0.544	0.588	0.609	0.626
1000	0.339	0.525	0.625	0.692	0.739	0.768
10000	0.345	0.544	0.657	0.738	0.793	0.835
100000	0.345	0.545	0.662	0.742	0.799	0.842
True	0.382	0.571	0.681	0.756	0.812	0.852

Table 3: Variance of $\widehat{CMB'}_Q$.

	k_2=20	56	92	128	164	200
k_1=10	0.077	0.088	0.087	0.085	0.086	0.085
100	0.043	0.061	0.066	0.068	0.069	0.072
1000	0.010	0.020	0.022	0.023	0.022	0.023
10000	0.001	0.002	0.002	0.002	0.002	0.002
100000	0.000	0.000	0.000	0.000	0.000	0.000

the true value, but the estimates may have a higher chance of failing to be a lower bound. As k_2 increases, the true value of CMB'_Q increases leading to higher values of $\widehat{CMB'}_Q$.

4.2 Allocation Within a Fixed Budget

The motivation for the prior section was that we are able to simulate many samples in stage 1 but have limited flexibility in the number of samples in stage 2 because these could be samples collected from physical experimentation. We now explore a situation where we have a fixed budget $B = k_1 + k_2$ where samples in both stages are equally costly and we need to allocate some samples to estimation of the first stage in k_1 and the rest to the second stage initial sample size. We analyze the same metrics from before, but fix the total number of samples. To further evaluate the quality of the lower bound, we estimate the conditional mean shortfall ($E[\widehat{CMB'}_Q - CMB'_Q | \widehat{CMB'}_Q < CMB'_Q]$) and conditional mean overage ($E[\widehat{CMB'}_Q - CMB'_Q | \widehat{CMB'}_Q > CMB'_Q]$) of the estimator. Table 4 shows the results by varying the allocation of a budget of 1000 samples when $\mu = 0, \sigma = 1$, and $Q = 0.1$.

The results for most of the metrics are generally monotone in the direction we expect. The true value of CMB'_Q decreases as k_1 increases (k_2 decreases) and the probability of $\widehat{CMB'}_Q$ being a lower bound decreases as k_1 increases (k_2 decreases) as the true value of CMB'_Q is decreasing. The variance decreases as k_1 increases, as does the expected shortfall, while the expected overshot increases as k_1 increases and k_2 decreases. The increase in overshoot occurs primarily because CMB'_Q is so close to 1 for most columns, that there is a relatively small cap on what the overshoot can be except for the last columns as k_2 decreases.

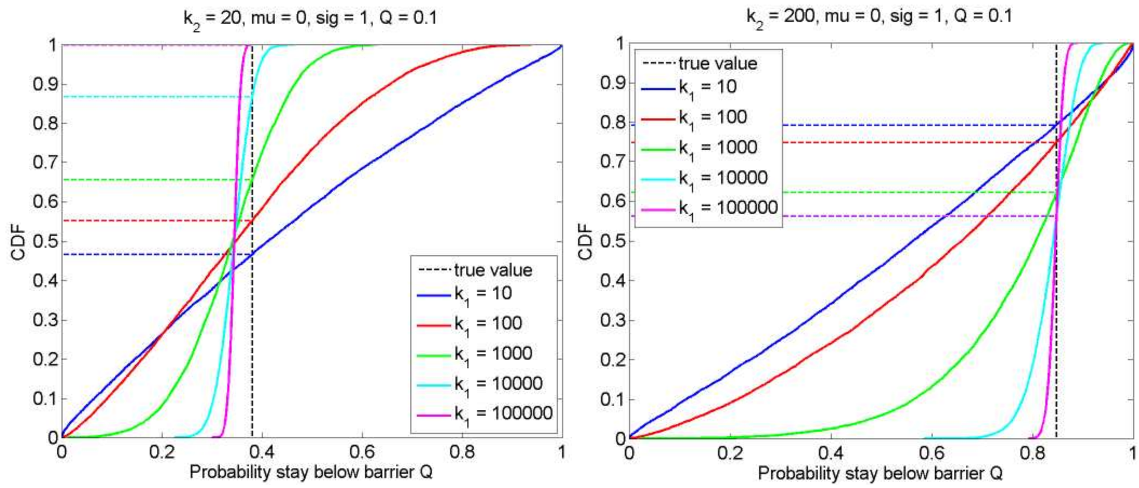

Figure 2: Cumulative distribution function of estimates $\widehat{CMB'}_Q$ for different values of k_1 using a low value of $k_2 = 20$ (left) and high value of $k_2 = 200$ (right). The dashed vertical line represents the true value of CMB'_Q.

Table 4: Allocation combination results, used $\mu = 0, \sigma = 1, Q = 0.1, nreps = 10000$, $m = 10000$, $B = 1000$.

	k_1=10%	20%	30%	40%	50%	60%	70%	80%	90%
True CMB	0.997	0.995	0.992	0.986	0.975	0.957	0.921	0.852	0.700
Probability Lower Bound	0.980	0.950	0.915	0.875	0.815	0.769	0.720	0.659	0.614
Mean	0.696	0.762	0.793	0.814	0.828	0.822	0.802	0.756	0.637
Variance	0.067	0.052	0.044	0.038	0.033	0.030	0.028	0.028	0.025
Expected Shortfall	-0.307	-0.246	-0.218	-0.197	-0.184	-0.182	-0.180	-0.176	-0.157
Expected Overshot	0.001	0.003	0.005	0.008	0.014	0.022	0.037	0.060	0.087

The one metric for which the relationship is non-monotonic is the mean value of the estimator $\widehat{CMB'}_Q$. The mean increases and then decreases, appearing to be its highest at around a 50% allocation to each stage. Even though k_2 is high when k_1 is low, and hence CMB'_Q should be higher for smaller k_1, the large variation in our estimate of μ caused by small k_1 leads to a larger probability that the estimated value of the true mean is actually greater than the boundary Q, which produces a lower estimate of $\widehat{CMB'}_Q$ and drags down the mean calculation in Table 4. As k_1 increases, the uncertainty around the mean decreases and it is much less likely that the possible mean will exceed Q and deflate $\widehat{CMB'}_Q$. But if k_1 is too large, then k_2 is small and the the value of CMB'_Q decreases and hence so too must $\widehat{CMB'}_Q$.

We next decrease Q to impose a more strict limit on the cumulative mean. In Table 5 we set $Q = 0.05$ which is even closer to the true mean 0. The results are similar to Table 4. The highest mean estimate $\widehat{CMB'}_Q$ occurs when 30% of the samples are allocated to the first stage.

Table 6 shows the results for an even more strict upper bound when $Q = 0.01$. In this case, the mean value of $\widehat{CMB'}_Q$ is larger than the true value of CMB'_Q in all cases, though the average overage decreases as k_1 increases and k_2 decreases. In this case, the true value of CMB'_Q is quite low to begin with, and it seems unlikely the sample mean will stay within this strict range, in which case the quality of the estimate $\widehat{CMB'}_Q$ is less useful anyway.

Table 5: Allocation combination results, used $\mu = 0, \sigma = 1, Q = 0.05, nreps = 10000, m = 10000, B = 1000$.

	k_1=10%	20%	30%	40%	50%	60%	70%	80%	90%
True CMB	0.906	0.877	0.845	0.807	0.763	0.707	0.636	0.543	0.406
Probability Lower Bound	0.868	0.816	0.768	0.726	0.695	0.643	0.612	0.577	0.566
Mean	0.563	0.585	0.595	0.594	0.585	0.570	0.540	0.482	0.375
Variance	0.079	0.077	0.074	0.070	0.067	0.063	0.057	0.050	0.032
Expected Shortfall	-0.403	-0.373	-0.348	-0.328	-0.306	-0.283	-0.248	-0.219	-0.161
Expected Overshot	0.050	0.063	0.078	0.093	0.113	0.128	0.144	0.155	0.139

Table 6: Allocation combination results, used $\mu = 0, \sigma = 1, Q = 0.01, nreps = 10000, m = 10000, B = 1000$.

	k_1=10%	20%	30%	40%	50%	60%	70%	80%	90%
True CMB	0.260	0.245	0.230	0.213	0.195	0.176	0.153	0.127	0.091
Probability Lower Bound	0.318	0.321	0.334	0.330	0.331	0.340	0.337	0.341	0.353
Mean	0.453	0.433	0.402	0.388	0.359	0.328	0.292	0.248	0.179
Variance	0.081	0.078	0.073	0.071	0.063	0.057	0.048	0.037	0.021
Expected Shortfall	-0.134	-0.127	-0.120	-0.113	-0.103	-0.092	-0.083	-0.068	-0.048
Expected Overshot	0.345	0.335	0.319	0.316	0.296	0.278	0.251	0.218	0.163

Finally, we consider a generous upper bound Q, where $Q = 0.2$. In this case, the true value of CMB'_Q is much higher. We note that a higher allocation to k_1 leads to the highest mean estimate at 50-60% allocation. Overall the estimator performs well for all allocations and across all metrics.

Table 7: Allocation combination results, used $\mu = 0, \sigma = 1, Q = 0.2, nreps = 10000, m = 10000, B = 1000$.

	k_1=10%	20%	30%	40%	50%	60%	70%	80%	90%
True CMB	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.996	0.959
Probability Lower Bound	1.000	1.000	1.000	1.000	1.000	0.915	0.857	0.756	0.676
Mean	0.887	0.954	0.977	0.986	0.991	0.991	0.989	0.979	0.930
Variance	0.025	0.008	0.003	0.002	0.001	0.000	0.001	0.001	0.003
Expected Shortfall	-0.113	-0.046	-0.023	-0.014	-0.009	-0.009	-0.012	-0.023	-0.051
Expected Overshot	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.015

For practical purposes, it is critical to obtain a good estimate of the variance for the estimate to be reasonable. Higher values of k_1 result in better estimators for μ and better precision in $\widehat{CMB'}_Q$, while higher values of k_2 lead to larger values of CMB'_Q. If it is likely that Q is close to μ, then even larger values of k_1 will be needed to reduce the uncertainty in μ and larger values of k_2 will be needed to obtain a reasonable value of CMB'_Q. If Q is far from μ, as in Table 7, then the results are less sensitive to the values of k_1 and k_2.

5 CONCLUSIONS

We present a new way of evaluating mean system performance called cumulative mean bounds. As an alternative to confidence intervals, we consider the probability that the cumulative system mean will stay within a given distance from its true mean. We translate this measure to estimate the probability that the cumulative sample mean will stay below a specific value after a given time period. This can be used as a quality control measure to relate the mean behavior of the system to some externally imposed limit. In order to achieve this translation, we require an estimate of the mean and variance based on some initial samples in the first stage of the experiment. We then require further samples in a second stage before the probability guarantee can be applied. We analyze the tradeoff between these two stages of sampling. We find that whereas a better probability guarantee can be found by increasing samples in the second stage, it is important to have an adequate number of samples in the first stage to estimate the mean.

REFERENCES

Alexopoulos, C. 2006. "A Comprehensive Review of Methods for Simulation Output Analysis". In *Proceedings of the 2006 Winter Simulation Conference*, edited by L. P. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, 168–178. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Alexopoulos, C., D. Goldsman, P. Tang, and J. R. Wilson. 2013. "A Sequential Procedure for Estimating the Steady-State Mean Using Standardized Time Series". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 613–622. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Hawkins, D., and D. Olwell. 1998. *Cumulative Sum Charts and Charting for Quality Improvement*. Springer.

Kim, S.-H., C. Alexopoulos, D. Goldsman, and K.-L. Tsui. 2006. "A new model-free CUSUM procedure for autocorrelated processes". Technical report, Georgia Institute of Technology.

Schruben, L. 1983. "Confidence Interval Estimation Using Standardized Time Series". *Operations Research* 31 (6): 1090–1108.

Singham, D., and M. Atkinson. 2015a. "Construction of Cumulative Mean Bounds for Simulation Output Analysis". Technical Report NPS-OR-15-004, Naval Postgraduate School.

Singham, D., and M. Atkinson. 2015b. "Estimating Cumulative Mean Behavior Using Standardized Time Series". Under Review.

Steiger, N., and J. Wilson. 2002. "An Improved Batch Means Procedure for Simulation Output Analysis". *Management Science* 48 (12): 1569–1586.

AUTHOR BIOGRAPHIES

DASHI SINGHAM is a Research Assistant Professor at the Naval Postgraduate School. This research was conducted during her time as a Visiting Assistant Professor at the University of California at Berkeley. Her main research area is simulation output analysis. She received a Ph.D. in Industrial Engineering & Operations Research and a Masters in Statistics from the University of California at Berkeley, and a B.S.E. from Princeton University in Operations Research & Financial Engineering. Her email address is dsingham@nps.edu.

MICHAEL ATKINSON is an Associate Professor of Operations Research at the Naval Postgraduate School. His research focuses on applying stochastic models to study homeland security and military applications. He received his Ph.D in Computational and Mathematical Engineering and his M.S. in Statistics from Stanford University. His e-mail address is mpatkins@nps.edu.