2015-05

Construction of cumulative mean bounds for simulation output

Singham, Dashi I.

Monterey, California. Naval Postgraduate School

NPS-OR-15-004

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

CONSTRUCTION OF CUMULATIVE MEAN BOUNDS FOR SIMULATION OUTPUT ANALYSIS

by

Dashi I. Singham
Michael P. Atkinson

May 2015

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704–0188*

1. REPORT DATE *(DD–MM–YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From — To)*
05-30-2015	Technical Report	05-30-2014 to 05-30-2015

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
CONSTRUCTION OF CUMULATIVE MEAN BOUNDS FOR SIMULATION OUTPUT ANALYSIS	**5b. GRANT NUMBER**
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Dashi I. Singham, Michael P. Atkinson	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Operations Research Department Naval Postgraduate School Monterey, CA 93943	NPS-OR-15-004
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

The views expressed in this document are those of the authors and do not reflect the official policy or position of the Department of Defense or the U.S. Government.

14. ABSTRACT

We develop a new measure of reliability, called cumulative mean bounds, that assesses the mean behavior of a process by calculating the probability that the cumulative sample mean will stay below its long-term sample mean, with a given tolerance, over a period of time. In this report, we provide a derivation of a lower bound for the measure when the underlying data are independent and identically distributed with a normal distribution.This derivation provides a preliminary basis for parallel extensions to the two-sided limiting case when we calculate the probability that the sample mean stays within a given distance from the true mean when the assumptions of independence and normality are removed.

15. SUBJECT TERMS

standardized time series, simulation output analysis, quality control measures

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** / **19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	UU	23	Dashi I. Singham 831-656-7836

NSN 7540-01-280-5500

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18

THIS PAGE INTENTIONALLY LEFT BLANK

NAVAL POSTGRADUATE SCHOOL
Monterey, California 93943-5000

The report entitled "Construction of Cumulative Mean Bounds for Simulation Output Analysis" was prepared for public release.

Further distribution of all or part of this report is authorized.

This report was prepared by:

Dashi I. Singham
Research Asst. Professor of Operations Research

Michael P. Atkinson
Assistant Professor of Operations Research

Reviewed by:

Johannes O. Royset
Associate Chairman for Research
Department of Operations Research

Robert F. Dell
Chairman
Department of Operations Research

Released by:

Jeffrey D. Paduan
Dean of Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

We develop a new measure of reliability, called cumulative mean bounds, that assesses the mean behavior of a process by calculating the probability that the cumulative sample mean will stay below its long-term sample mean, with a given tolerance, over a period of time. In this report, we provide a derivation of a lower bound for the measure when the underlying data are independent and identically distributed with a normal distribution. This derivation provides a preliminary basis for parallel extensions to the two-sided limiting case when we calculate the probability that the sample mean stays within a given distance from the true mean when the assumptions of independence and normality are removed.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

We propose a new measure, called "cumulative mean bounds," that produces more general information on the evolution of the system's mean performance than the traditional confidence interval. We present a method for calculating the probability that the sample mean of a time series stays below its true mean on one side, with a given tolerance, over a given period of time. We also consider the "long-term mean," as the sample mean calculated after a long period of time. Given a time series Y_i for $i = 1, 2, \ldots$, we define the cumulative mean bounds (CMB') measure for the one-sided case as:

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

where μ is the true mean, k is some initial sample size, and δ is some allowed tolerance. We distinguish CMB' from CMB, which, in our other papers, is the probability that the sample mean stays within a given absolute distance δ from μ on both sides. In this report, we derive CMB' for a large sample size m:

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

for some $1 \leq k < m$. The parameter m denotes the number of samples used to calculate a long-term mean, and (2) is the probability that the sample mean stays below the long-term mean, with allowed tolerance δ above the mean, after an initial sample size k. The expression (1) is the limit of (2) as $m \to \infty$ and represents the probability that the sample mean stays below its true mean μ, with allowed tolerance distance δ, after an initial sample size k. In this report, we assume that the underlying time series $\{Y_i,\ i \geq 1\}$ consists of data that are independent and identically distributed (i.i.d.) normal, and ongoing work considers the general case when the data meet the assumptions of a functional central limit theorem (FCLT), which would allow for dependence and nonnormality.

We evaluate the expression in (2) by structuring a time series of data as a standardized time series, which, under some conditions, converges to a Brownian bridge in the limit. When the data are i.i.d. normal, we will show that the points of a standardized time series have the same joint distribution as the same time points of a standard Brownian bridge. We leverage boundary-crossing probabilities of Brownian bridges to derive a lower bound for the values of CMB' defined above. The lower bound only occurs because we use a continuous Brownian bridge process for the necessary calculations, rather than discrete realizations of a standardized time series.

Section 2 provides background on standardized time series and derives the joint distribution of points in a standardized time series when the data are i.i.d. normal. We construct the measure CMB' in Section 3. Section 4 provides a proof of the main result for the derivation of CMB'. Section 5 concludes.

II. PRELIMINARIES

In this section, we establish the background needed to derive CMB'. Ongoing work in [1] establishes the quality of this bound and derives the limiting case in (1).

The method of standardized time series was introduced in [2] to develop interval estimators for the mean μ using data $Y_1, \ldots, Y_m$. We assume a known value of the variance σ^2, for which straightforward estimators exist for the i.i.d. normal case.

A standardized time series is defined in [2] as

$$X(t) = \frac{\lfloor mt \rfloor \left(\frac{1}{m} \sum_{i=1}^{m} Y_i - \frac{1}{\lfloor mt \rfloor} \sum_{i=1}^{\lfloor mt \rfloor} Y_i \right)}{\sigma \sqrt{m}}, \quad t \in [0,1]. \tag{3}$$

Schruben shows in [2] that under the assumptions of a FCLT, $X(t)$ converges weakly to $B(t)$ as $m \to \infty$, where $B(t)$ is a standard Brownian bridge over $t \in [0,1]$. In order to use properties of Brownian bridges applied to standardized time series, we require the following result.

Proposition 2.1. For i.i.d. normal data, the points of a $X(\frac{i}{m}), i = 1, \ldots m$ of a standardized time series have the same joint distribution as the corresponding points $B(\frac{i}{m}), i = 1, \ldots m$ of a standard Brownian bridge, which is Gaussian with mean zero and covariance $\frac{i}{m}(1 - \frac{j}{m})$ for $i \leq j$.

Proof. The Brownian bridge $B(t)$, $0 \leq t \leq 1$, is a Gaussian process with $EB(t) = 0$ and $Cov(B(s), B(t)) = s(1-t)$ for $s \leq t$. Thus, the finite dimensional vector $\hat{B} = (B(\frac{1}{m}), B(\frac{2}{m}), \ldots, B(\frac{m}{m}))$ has a multivariate normal distribution with $EB(\frac{i}{m}) = 0$ for all i and $Cov(B(\frac{i}{m}), B(\frac{j}{m})) = \frac{i}{m}(1 - \frac{j}{m})$ for $i \leq j$.

We next turn to the vector $\hat{X} = (X(\frac{1}{m}), X(\frac{2}{m}), \ldots, X(\frac{m}{m}))$ formed from a standardized time series. $\hat{X}$ has a multivariate normal distribution because we can write $\hat{X} = AY$, where Y is the vector of i.i.d. normal data $(Y_1, Y_2, \ldots, Y_m)$ and A is a deterministic matrix (constructed to produce (3)). By inspection of Equation (3), $EX(\frac{i}{m}) = 0$. Thus, to complete the proof, we must show that $Cov(X(\frac{i}{m}),X(\frac{j}{m})) = \frac{i}{m}(1-\frac{j}{m})$ for $i \leq j$ as follows:

$$Cov\left(X(\frac{i}{m}),X(\frac{j}{m})\right) = \frac{ij}{\sigma^2 m}Cov\left(\frac{1}{m}\sum_{\ell=1}^{m}Y_\ell - \frac{1}{i}\sum_{\ell=1}^{i}Y_\ell, \frac{1}{m}\sum_{\ell=1}^{m}Y_\ell - \frac{1}{j}\sum_{\ell=1}^{j}Y_\ell\right)$$
$$= \frac{ij}{\sigma^2 m}\left(\frac{1}{m^2}Var(\sum_{\ell=1}^{m}Y_\ell) - \frac{1}{mj}Cov(\sum_{\ell=1}^{m}Y_\ell, \sum_{\ell=1}^{j}Y_\ell) - \frac{1}{mi}Cov(\sum_{\ell=1}^{i}Y_\ell, \sum_{\ell=1}^{m}Y_\ell) + \frac{1}{ij}Cov(\sum_{\ell=1}^{i}Y_\ell, \sum_{\ell=1}^{j}Y_\ell)\right).$$

Because the Y_i are i.i.d. normal, this simplifies to

$$\frac{ij}{\sigma^2 m}\left(\frac{\sigma^2}{m} - \frac{\sigma^2}{m} - \frac{\sigma^2}{m} + \frac{\sigma^2}{j}\right) = \frac{ij}{\sigma^2 m}\left(\frac{\sigma^2}{j} - \frac{\sigma^2}{m}\right) = \frac{ij}{m}\left(\frac{1}{j} - \frac{1}{m}\right) = \frac{i}{m}\left(1 - \frac{j}{m}\right)$$

$\square$

III. CUMULATIVE MEAN BOUNDS

In this section, we construct cumulative mean bounds and derive the probability that the cumulative sample mean of a performance measure stays below its long-term mean, with tolerance δ, after k samples, when the data are i.i.d. normal. We must start with some value of $k > 0$ for there to be at least one sample collected to estimate the sample mean. Let δ be the prespecified allowed deviation allowed above the long-term mean, which will have implications in quality control applications. First, we will use the long-term average as collected by the data, $\overline{Y}_m$, as our representation for the long-term mean. We wish to calculate the probability that the cumulative mean $\overline{Y}_j$, for $j = k, \ldots, m$ stays within $[-\infty, \overline{Y}_m + \delta]$, and define the probability CMB' as in (2). Figure 1 illustrates cumulative mean bounds in the one-sided case when comparing the cumulative mean performance to the true mean.

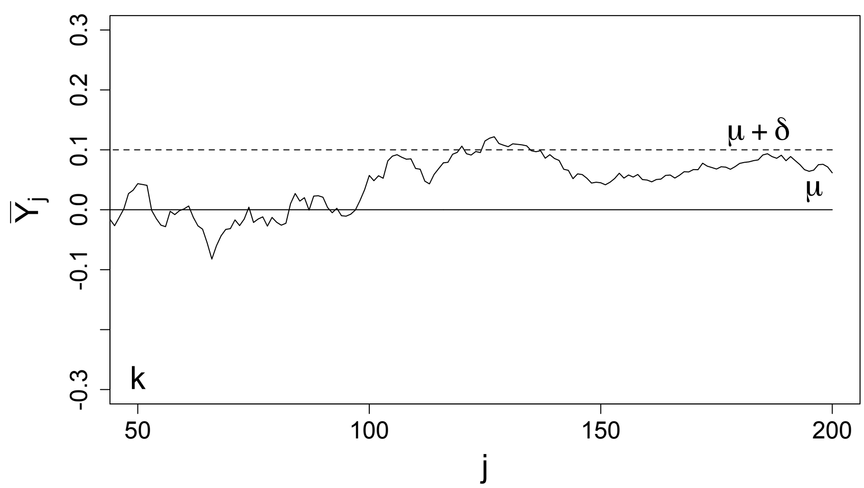

Figure 1: Given an initial sample size k, we evaluate the probability that the cumulative sample mean stays below μ, with allowed tolerance δ.

Using (3), we rewrite CMB' in terms of a standardized time series $X(t)$ and a Brownian bridge $B(t)$ when j is an integer within $k \leq j \leq m$:

$$
\begin{aligned}
CMB' :=& P\left(\bigcap_{k\le j\le m} \frac{1}{j}\sum_{i=1}^{j} Y_i - \frac{1}{m}\sum_{i=1}^{m} Y_i \le \delta\right) = P\left(\bigcap_{k\le j\le m} \sigma X(\frac{j}{m}) \le \delta\frac{j}{\sqrt{m}}\right) && (4)\\
=& P\left(\bigcap_{k\le j\le m} \sigma B(\frac{j}{m}) \le \delta\frac{j}{\sqrt{m}}\right) && (5)\\
\ge& P\left(\bigcap_{t\in[\frac{k}{m},1]} \sigma B(t) \le \delta\sqrt{m}t\right) \equiv CMB'_L. && (6)
\end{aligned}
$$

Proposition 2.1 allows us to move from (4) to (5). To move from (5) to (6), we first set $t = j/m$ to standardize time to lie in $[0,1]$. The lower bound follows because in (6) we evaluate the probability that the Brownian bridge stays within the bounds over all continuous values of $t \in [\frac{k}{m}, 1]$, whereas in (5) we consider only a finite set of discrete points j/m such that $k \le j \le m$ where j is restricted to the set of integer values.

Boundary crossing properties of Brownian bridges exist that will allow us to compute (6) exactly. The probability that a Brownian bridge ever leaves two symmetric linear bounds that have nonzero intercepts at $t = 0$ is derived in [3]. In our case, the slope of these linear bounds is $\pm\delta\sqrt{m}$. Whereas the intercept at $t = 0$ is zero, we start the process at $t = k/m$, which yields a nonzero intercept. In practice, an experiment would require some initial k samples to calculate some estimate of the sample mean. We now present the following result.

Theorem 3.1. Under the assumption that the underlying data are i.i.d. normal, the probability that the sample mean stays below its long-term mean $\overline{Y}_m$, with tolerance δ, over the range $j = k,\ldots,m$ has a lower bound

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

where

$$
CMB'_L(\delta,\sigma,k,m) = 2\Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-\frac{k}{m}}}\right) - 1. \tag{7}
$$

The probability that the sample mean stays below μ, with tolerance δ, for all $j \ge k$, has a lower bound

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

where

$$CMB_L'(\delta,\sigma,k)=2\Phi\left(\frac{\delta\sqrt{k}}{\sigma}\right)-1. \tag{8}$$

THIS PAGE INTENTIONALLY LEFT BLANK

IV. PROOF OF THEOREM 3.1

We wish to compute the following one-sided calculation for CMB'_L:

$$CMB'_L(\delta,\sigma,k,m) = P\left(\bigcap_{t\in[\frac{k}{m},1]} \sigma B(t) \leq \delta\sqrt{m}t\right).$$

We condition on the location of $B(k/m)$, where $B_x^{k/m}$ is a Brownian bridge process that takes value x at time k/m:

$$CMB'_L(\delta,\sigma,k,m) = \int_{-\infty}^{\frac{\delta k}{\sqrt{m}}} P\left(\bigcap_{t\in[0,1-k/m]} \sigma B_x^{k/m}(t) \leq \frac{\delta k}{\sqrt{m}} + \delta\sqrt{m}t\right) N\left(x,0,\sigma^2\frac{k}{m}(1-\frac{k}{m})\right) dx. \tag{9}$$

The first probability can be evaluated using Equation (6) from [4], with a Brownian bridge starting at x at time 0, ending at 0 at time $1-k/m$, and a linear boundary defined by intercept $\delta k/\sqrt{m}$ and slope $\delta\sqrt{m}$. Then (9) becomes:

$$\begin{aligned}
CMB'_L(\delta,\sigma,k,m) &= \int_{-\infty}^{\frac{\delta k}{\sqrt{m}}} \left(1-\exp\left(-2\frac{(\frac{\delta k}{\sqrt{m}}-x)(\frac{\delta k}{\sqrt{m}}+\delta\sqrt{m}(1-\frac{k}{m}))}{\sigma^2(1-\frac{k}{m})}\right)\right) N\left(x,0,\sigma^2\frac{k}{m}(1-\frac{k}{m})\right) dx \\
&= \int_{-\infty}^{\frac{\delta k}{\sqrt{m}}} N\left(x,0,\sigma^2\frac{k}{m}(1-\frac{k}{m})\right) dx && (10) \\
&- \quad \int_{-\infty}^{\frac{\delta k}{\sqrt{m}}} \exp\left(-2\frac{(\frac{\delta k}{\sqrt{m}}-x)\delta\sqrt{m}}{\sigma^2(1-\frac{k}{m})}\right) N\left(x,0,\sigma^2\frac{k}{m}(1-\frac{k}{m})\right) dx && (11)
\end{aligned}$$

and (10) simplifies to $\Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-\frac{k}{m}}}\right)$. The terms inside the integral in (11) are

$$
\begin{aligned}
&\exp\left(-2\frac{(\frac{\delta k}{\sqrt{m}}-x)\delta\sqrt{m}}{\sigma^2(1-\frac{k}{m})}\right)\frac{\exp\left(\frac{-x^2}{2\sigma^2\frac{k}{m}(1-\frac{k}{m})}\right)}{\sigma\sqrt{2\pi\frac{k}{m}(1-\frac{k}{m})}}\\
&=\frac{1}{\sigma\sqrt{2\pi\frac{k}{m}(1-\frac{k}{m})}}\exp\left(-\frac{1}{\sigma^2(1-\frac{k}{m})}\left[2(\frac{\delta k}{\sqrt{m}}-x)\delta\sqrt{m}+\frac{x^2}{\frac{2k}{m}}\right]\right)\\
&=\frac{1}{\sigma\sqrt{2\pi\frac{k}{m}(1-\frac{k}{m})}}\exp\left(-\frac{1}{\sigma^2(1-\frac{k}{m})\frac{2k}{m}}\left[\frac{4k}{m}(\frac{\delta k}{\sqrt{m}}-x)\delta\sqrt{m}+x^2\right]\right).
\end{aligned}
$$

The terms in square brackets above simplify to $x^2-\frac{4kx\delta}{\sqrt{m}}+\frac{4k^2\delta^2}{m}=(x-\frac{2\delta k}{\sqrt{m}})^2$. This implies that (11) simplifies to

$$
\int_{-\infty}^{\frac{\delta k}{\sqrt{m}}}\frac{1}{\sqrt{2\pi\sigma^2\frac{k}{m}(1-\frac{k}{m})}}\exp\left(-\frac{(x-\frac{2\delta k}{\sqrt{m}})^2}{2\sigma^2\frac{k}{m}(1-\frac{k}{m})}\right)dx=\Phi\left(-\frac{\delta\sqrt{k}}{\sigma\sqrt{1-\frac{k}{m}}}\right).
$$

Substituting the various terms back into (10) and (11) we have:

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

To prove the second part of the theorem regarding the probability that the sample average stays below μ, with tolerance δ, we can establish how the result holds by taking the limit in m . The details are outside the scope of this report and are in [1].

V. CONCLUSION

We develop the measure CMB' to calculate the probability that the cumulative sample mean stays below its long-term sample mean $\overline{Y}_m$, with allowed tolerance δ, after an initial sample size k. We rely on properties of standardized time series to perform this calculation. This measure can be used as an alternative to confidence intervals to evaluate the mean performance over time of a system. Additionally, it can be used as a quality control measure to estimate the probability the sample mean will go above a given control limit. Parallel work develops the two-sided case, with fewer restrictions on the data, and allows for further applications. Multidimensional applications will be developed based on results derived in [5].

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

[1] D. Singham and M. Atkinson, “Estimating Cumulative Mean Behavior Using Standardized Time Series,” 2014, Under Review.

[2] L. Schruben, “Confidence Interval Estimation Using Standardized Time Series,” *Operations Research*, vol. 31, no. 6, pp. 1090–1108, 1983.

[3] M. Abundo, “Some Conditional Crossing Results of Brownian Motion Over a Piecewise-Linear Boundary,” *Statistics & Probability Letters*, vol. 58, no. 2, pp. 131–145, 2002.

[4] T. H. Scheike, “A Boundary-Crossing Result for Brownian Motion,” *Journal of Applied Probability*, vol. 29, no. 2, pp. 448–453, 1992.

[5] M. Atkinson and D. Singham, “Multidimensional Hitting Time Results for Brownian Bridges with Moving Hyperplanar Boundaries.” *Statistics and Probability Letters*, vol. 100, pp. 85–92, 2015.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Research Sponsored Programs Office, Code 41
   Naval Postgraduate School
   Monterey, California

4. Richard Mastowski (Technical Editor)..1
   Graduate School of Operational and Information Sciences (GSOIS)
   Naval Postgraduate School
   Monterey, California

5. Dashi I. Singham..2
   Operations Research Department
   Naval Postgraduate School
   Monterey, California

6. Michael P. Atkinson..3
   Operations Research Department
   Naval Postgraduate School
   Monterey, California