Research Paper IC Value : 61.33| SJIF Impact Factor(2018) : 8.003| ISI Impact Factor (2017):1.365 (Dubai)

Volume - 7, Issue- 4, April 2019 | e-ISSN : 2347 - 9671| p- ISSN : 2349 - 0187

EPRA International Journal of Economic and Business Review -Peer Reviewed Journal

FORMULATION OF CUMULATIVE INTERVALS AS A MEASURE OF RELIABILITY FOR THE ASSESSMENT OF SAMPLE MEAN BEHAVIOUR

Peter O. Peter

Shanghai Jiao Tong University, School of Mathematical Sciences, Shanghai, China.

ABSTRACT

We propose a new measure of reliability, called the cumulative mean intervals, that assesses the mean behaviour of a process by computing the probability that the cumulative sample mean will remain below its long-term sample mean with a given tolerance over a period of time. We further derive a lower bound for the measure when the underlying data is independent and identically distributed with a normal distribution. This deduction provides a preliminary basis for parallel extensions to the two limiting case when we compute the probability that the sample mean stays within a given distance from the true mean with no assumptions made on independence and normality.

KEYWORDS: *cumulative mean, probability, reliability, intervals, sample mean.*

1. INTRODUCTION

In this paper, we present a new measure called cumulative mean intervals, that pro- duces more general information on the evolution of the system's mean performance compared to 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 further consider the long-term mean as the sample mean calculated after a long period of time. Given a time series Y_i for $i = 1,2,...$, we define the cumulative mean intervals (CMI) measure for the one-sided case as:

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

where μ is the true mean, k is some initial sample size, and δ is a permissible tolerance. We distinguish the Cumulative Mean Interval (CMI) from the Cumulative Mean Bounds (CMB) discussed under [6] which is the probability that the sample mean stays within a given absolute distance δ from μ on both sides. In this paper, we derive CMI for a large sample size m

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

for some $1 \leq k < m$. The parameter m denotes the number of samples used to calcu- late 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 sam- ple size k. The expression of (1) is the limit of (2) as $m \longrightarrow \infty$ and it defines the probability that the sample mean remains below its true mean μ, with permissible tolerance distance δ, after an initial sample size k. In this paper, 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 allow for dependence and non-normality.

We evaluuate the expression in (2) by structuring a time series od data as a stan- dardized which under some conditions converges to a Brownian Bridge in the limit. When the data is 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 Brow- nian bridge. We leverage boundary-crossing probabilities of Brownian Bridges to derive a lower bound for the values of CMI defined above. The lower bound only occurs because we use a continuous Brownian brigde process for the required calcu- lations,rather than discrete realizations of a standardized time series. The proposed technique is formulated in a spirit similar and motivated by mean bounds discussed under [6].

The layout of this report is as follows; 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 CMI in section 3 and section 4 provides a proof of the main result for the derivation of CMI. Section 5 presents our conclusions.

2. PRELIMINARIES

In this section, we establish the background needed to derive CMI. The work of [5] establishes the quality of this bound and derives the limiting case in (1). The method of standardized time series was introduced in [4] to develop interval estimators for the mean μ using data $Y_1, ..., Y_m$. We assume a known value of the variance σ^2, for which straight forward estimators exist under i.i.d. normal case. A standardized time series is defined in [4] as;

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

[4] shows that under the assumptions of a FCLT, $X(t)$ converges weekly to $B(t)$ as $m \longrightarrow \infty$, 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 1. *For i.i.d. normal data, the points of a $X(\frac{i}{m}), i = 1, ...m$ of a standardized time series have the same joint distribution as the corresponding points $B(\frac{i}{m}), i = 1, ...m$ 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 $\hat{B} = (B(\frac{1}{m}), B(\frac{2}{m}), ...B(\frac{m}{m}))$ has a multivariate normal distribution with $EB(\frac{1}{m}) = 0$ 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 $\hat{X} = (X(\frac{1}{m}), X(\frac{2}{m}), ...X(\frac{m}{m}))$ 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 $(Y_1, Y_2, ..., Y_m)$ and A is a deterministic matrix formulated to yield (3). By inspection of (3), $EX(\frac{1}{m}) = 0$. 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:

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

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

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

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

□

3.Cumulative Mean Intervals

In this section we formulate cumulative mean intervals and derive the probability that the cumulative sample mean oof a performance mmeasure stays below its long-term mean, with tolerance δ, after k samples, when the data are i.i.d. normal. We start with some value $k > 0$ so that there is at least one sample collected to estimate the sample mean. We let δ be the prespecified allowed deviation 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, $\bar{Y}_j$, for $j = k, ...m$ stays within $[-\infty, \bar{Y}_m + \delta]$, and define the probability CMI as in (2).

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 CMI in terms of the standardized time series $X(t)$ and a Brownian bridge $B(t)$ when j is an integer within $k \leq j \leq m$:

$$CMI := P\left(\bigcap_{k\leq j\leq m} \frac{1}{j}\sum_{i=1}^{j} Y_i - \frac{1}{m}\sum_{i=1}^{m} Y_i \leq \delta\right) = P\left(\bigcap_{k\leq j\leq m} \sigma X(\frac{j}{m}) \leq \delta\frac{j}{\sqrt{m}}\right) \quad (4)$$

$$= P\left(\bigcap_{k\leq j\leq m} \sigma B(\frac{j}{m}) \leq \delta\frac{j}{\sqrt{m}}\right) \quad (5)$$

$$\geq P\left(\bigcap_{t\in[\frac{k}{m},1]} \sigma B(t) \leq \delta\sqrt{mt}\right) \equiv CMI_L \quad (6)$$

Proposition 1 allows us to move from (4) to (5) and 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 probabilty 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 $k \leq j \leq m$ where j is restricted to the set of integer values.

Boundry crossing properties of Brownian bridges exists that will allow us to compute (6) exactly. The probability that a Brownian bridge ever leaves two symmetric linear bounds that have non-zero intercepts at $t = 0$ is derived in [1].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 non-zero 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 $\bar{Y}_m$, with tolerance δ, over the range $j = k, ..., m$ has a lower bound*

$$P\left(\bigcap_{k\leq j\leq m} \sigma B(\frac{j}{m}) \leq \delta\frac{j}{\sqrt{m}}\right) \geq CMI_L(\delta, \sigma, k, m),$$

where

$$CMI_L(\delta, \sigma, k, m) = 2\Phi\left(\frac{\delta\sqrt{k}}{\sigma\sqrt{1-\frac{k}{m}}}\right) - 1 \quad (7)$$

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

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

where

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

4.Proof of Theorem 3.1

We wish to compute the following one sided calculation of CMI_L:

$$CMI_L(\delta,\sigma,k,m)=P\left(\bigcap_{t\in[\frac{k}{m},1]}\sigma B(t)\leq\delta\sqrt{mt}\right)$$

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:

$$CMI_L(\delta,\sigma,k,m)=\int_{-\infty}^{\frac{\delta k}{\sqrt{m}}}P\left(\bigcap_{t\in[0,1-\frac{k}{m}]}\sigma B_x^{k/m}(t)\leq\frac{\delta k}{\sqrt{m}}+\delta\sqrt{mt}\right)N\left(x,0,\sigma^2\frac{k}{m}(1-\frac{k}{m})\right)dx \tag{9}$$

The first probability can be evaluated using (6) from [3], with a Brownian bridge starting at x at time 0, ending at 0 at time $1-\frac{k}{m}$, and a linear boundary defined by the intercept $\delta k/\sqrt{m}$ and slope $\delta\sqrt{m}$. Then (9) becomes:

$$CMI_L(\delta,\sigma,k,m)=\int_{-\infty}^{\frac{\delta k}{\sqrt{m}}}\left(1-exp\left(-2\frac{(\frac{\delta k}{\sqrt{m}}-x)(\frac{\delta k}{\sqrt{m}}+\delta\sqrt{m}(1-\frac{k}{m}))}{\sigma^2(1-\frac{k}{m})}\right)\right)N\left(x,0,\sigma^2\frac{k}{m}(1-\frac{k}{m})\right)dx$$

$$=\int_{-\infty}^{\frac{\delta k}{\sqrt{m}}}N\left(x,0,\sigma^2\frac{k}{m}(1-\frac{k}{m})\right)dx \tag{10}$$

$$-\int_{-\infty}^{\frac{\delta k}{\sqrt{m}}}exp\left(-2\frac{(\frac{\delta k}{\sqrt{m}}-x)\delta\sqrt{m}}{\sigma^2(1-\frac{k}{m})}\right)N\left(x,0,\sigma^2\frac{k}{m}(1-\frac{k}{m})\right)dx \tag{11}$$

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

$$exp\left(-2\frac{(\frac{\delta k}{\sqrt{m}}-x)\delta\sqrt{m}}{\sigma^2(1-\frac{k}{m})}\right)\frac{exp\left(\frac{-x^2}{2\sigma^2\frac{k}{m}(1-\frac{k}{m}}\right)}{\sigma\sqrt{2\pi\frac{k}{m}(1-\frac{k}{m})}}$$

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

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

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

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

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 results holds by taking the limit in m. The detailes are outside the scope of this report and are further discussed under [5].

5.CONCLUSION

In this article, we have developed the CMI as a measure of reliability to calculate the probability that the cumulative sample mean stays below its long-term sample mean $Y\ m$, with allowed tolerance ä, 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 quality control measure to estimate the probability that 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 have been developed based on the results derived under [2].

6. ACKNOWLEDGEMENTS

The author would like to thank colleagues and other stakeholders for their valuable comments and suggestions.

7.FUNDING

The author declares that there is no funding or grant support for this study.

REFERENCES

1. *Abundo, M. (2002). Some Conditional Crossing Results of Brownian Motion Over a Piecewise-Linear Boundary. Statistics and Probability Letters, 58 (2), 131–145.*
2. *Atkinson, M. P. & Singham, D. I. (2015). Multidimensional Hitting Time Re- sults for Brownian Bridges with Moving a Hyoerolanar Boundaries. Statistics and Probability Letters, 100, 85–92.*
3. *Scheike, T. H. (1992). A Boundry-Crossing Results for Brownian Motion. Journal of Applied Probability, 29 (2), 448–453.*
4. *Schruben, L. (1982). Confidence Interval Estimation Using Time Series. Opera- tions Research, 36 (6), 1090–1106.*
5. *Singham, D. I. & Atkinson, M. P. (2014). Estimating Cumulative Mean Behavior Using Standardized Time Series.*
6. *Singham, D. I. & Atkinson, M. P. (2015). Construction of Cumulative Mean Bounds for Simulation Output Analysis.*