Proceedings of the 1986 Winter Simulation Conference
J. Wilson, J. Henriksen, S. Roberts (eds.)

BEHAVIOR OF SAMPLE MEANS AND NONPARAMETRIC TIME SERIES ESTIMATION

Emanuel Parzen
Department of Statistics
Texas A&M University
College Station, TX 77843, U.S.A.

ABSTRACT

The behavior of sample means, which needs to be understood by all applied statisticians and users of simulation methods, can be considered to be the most basic question of both classical and modern probability and statistics. This theory, and its implications for practice, will be surveyed (in the methodology session on Time Series Analysis of Sample Means) by three statisticians who are experts on time series analysis.

This paper consists of five sections discussing: notation; spectral density classification of memory type of a time series; equivalent degrees of freedom of asymptotic confidence intervals for the mean; sample Fourier transforms and sample spectral density; sample Brownian Bridge functionals and standardized time series.

1. NOTATION AND INTRODUCTION

The reader of these papers on time series analysis should be warned that time series analysts are far from agreeing on a standard notation to use for the basic concepts of the field. My notation is chosen not to be different but because it seems optimal according to my philosophy of notation.

$Y(1),\ldots,Y(T)$: a sample of size T, considered to be observations of a random variable Y, indexed by $t=1,2,\ldots,T$.

$F(y)=\text{Prob}[Y\le y]$, $-\infty<y<\infty$: distribution function of Y.

$Q(u)$, $0\le u\le 1$: quantile function of Y, also denoted $Q(u;Y)$, defined by

$$Q(u) = F^{-1}(u) = \inf\{y: F(y)\ge u\},\ 0\le u\le 1,$$

$F^{\sim}(y)$: sample distribution function of Y, defined by $F^{\sim}(y)$=fraction of sample $\le y$.

$Q^{\sim}(u)=F^{\sim -1}(u)$: sample quantile function of Y.

MY, mean of Y: usually denoted $E[Y]$

$$MY=\int_{-\infty}^{\infty} y\, dF(y) = \int_0^1 Q(u)\, du$$

$Z(t)=Y(t)-MY$: fluctuation series used to represent $Y(t)$ as a sum $Y(t)=MY+Z(t)$ of an unknown mean value MY to be estimated and a zero mean series $Z(t)$ called the error series.

$MY^{\sim}$, sample mean of Y: under a general set of conditions (which we do not state in detail) an asymptotically efficient estimator of MY is the sample mean, denoted $MY^{\sim}$ or $\overline{Y}$, defined by

$$MY^{\sim}= \int_{-\infty}^{\infty} y\, dF^{\sim}(y) = \int_0^1 Q^{\sim}(u)\, du$$

$$= (1/T) \sum_{t=1}^{T} Y(t)$$

Using the Central Limit Theorem, one can derive the asymptotic distribution of $MY^{\sim}$. In developing expressions for the distribution of $MY^{\sim}$ to be used to form confidence intervals for MY one could use the sample variance $VAR^{\sim}[Y]$ and sample standard deviation $DS^{\sim}[Y]$ defined by

$$VAR^{\sim}[Y] = (1/T) \sum_{t=1}^{T} \{Y(t)-MY^{\sim}\}^2$$

$$DS^{\sim}[Y]=\{VAR^{\sim}[Y]\}^{0.5}$$

We assume that $Y(1),\ldots,Y(T)$ are identically distributed as Y but may not be independent. However they are a sample from a time series which is covariance stationary in the sense that there is a function $R(v)$, $v=0,\pm 1,\pm 2,\ldots$, such that

$$COV[Y(s),\ Y(t)] = R(t-s)\ .$$

In the study of time series, the effect of the marginal distribution of Y(t) can be separated from the bivariate dependence of Y(t) and Y(t+v) by defining the correlation function

$$rho(v) = R(v)/R(0) = CORR[Y(t),\ Y(t+v)].$$

The Fourier transforms of R(v) and rho(v) are denoted $S(\omega)$, $0 \le \omega \le 1$, and $f(\omega)$, $0 \le \omega \le 1$ and are called the power spectrum and spectral density of the time series. Two basic definitions are

$$f(\omega) = \sum_{v=-\infty}^{\infty} \exp(2\pi i v\omega)\ rho(v),\ 0 \le \omega \le 1\ ,$$

$$S(\omega) = R(0)f(\omega)\ .$$

We interpret ω as representing frequency and its reciprocal $1/\omega$ represents period; thus a peak (local maximum) in the spectral density at frequency $\omega = 1/12$ represents the presence in the time series of a sinusoidal component or disturbed periodicity of period 12. Note that a time series Y(t) has period P if Y(t+P) = Y(t) for all t.

An AR(1) time series $Z(T) - \rho Z(t-1) = e(t)$ where e(t) are independent $N(0,\sigma^2)$ and $|\rho|<1$ has $VAR[Y] = \sigma^2(1-\rho^2)$, $rho(v) = \rho^{|v|}$,

$$f(0) = (1+\rho)/(1-\rho),$$

$$f(\omega) = (1-\rho^2)/(1+\rho^2 - 2\rho \cos 2\pi\omega).$$

When Y(1),...,Y(T) are a sample from a stationary time series, the sample mean (under suitable conditions called "mixing") is asymptotically normal. To find suitable formulas for its asymptotic variance we write

$$VAR[MY^{\sim}] = (1/T^2) \sum_{s,t=1}^{T} COV[Y(s),\ Y(t)]$$

$$T\ VAR[MY^{\sim}] = VAR[Y] \sum_{|v|<T} (1-|v/T|)\ rho(v)$$

As $T\to\infty$, assuming $\sum_{v=-\infty}^{\infty} |rho(v)|<\infty$,

$$T\ VAR[MY^{\sim}] \to VAR[Y]\ f(0)$$

$$\sqrt{T}\{MY-MY^{\sim}\}/\{VAR[Y]f(0)\}^{0.5} \to N(0,1)$$

The time series analyst is interested in estimating the spectral density function to help identify models for the time series. In various applications, one only seeks to estimate the value of the spectral density at zero frquency because its value is required in other formulas. The infinite sum

$$f(0) = \sum_{v=-\infty}^{\infty} rho(v)$$

cannot be estimated by merely replacing each rho(v) by an estimator. I would like to emphasize that the observation that this infinite sum is the value at zero frequency of the spectral density function is very important and useful because it provides a variety of methods for estimating f(0).

Correlations are estimated by sample correlations defined by

$$rho^{\sim}(v) = R^{\sim}(v)/R^{\sim}(0),$$

in terms of the sample covariance function (defined for v=0,1,...,T-1)

$$R^{\sim}(v) = (1/T) \sum_{t=1}^{T-v} \{Y(t)-MY^{\sim}\}\{Y(t+v)-MY^{\sim}\},$$

The variances of $rho^{\sim}(v)$ decrease to zero, as $T\to\infty$, as 1/T; the sum of M values of $rho^{\sim}(v)$ converge to zero as M/T and does not converge to zero when the limit of M/T is positive. The theoretical problem of spectral estimation can be regarded as how to choose M as a function of T so that M/T tends to 0 while M tends to ∞.

2. SPECTRAL DENSITY CLASSIFICATION OF MEMORY TYPE OF A TIME SERIES

To estimate the spectral density one must identify the memory type of the time series in the spectral density domain. We define a time series to be:

1. No memory if $f(\omega) = 1$ for all ω;
2. Short memory if there exist positive finite constants c and C such that

   $0<c<f(\omega)<C<\infty$ for all ω;
3. Long memory if it has a zero or an infinity.

At a frequency ω_0 at which $f(\cdot)$ is zero or infinite we seek to model the rate of approach to 0 or ∞ by representing $f(\omega)$ for ω near ω_0 by

$$f(\omega) = (\omega-\omega_0)^{-\delta}L(\omega)$$

where $L(\omega)$ is slowly-varying or log-like function, and δ is called the index of regular variation at ω_0.

A diagnostic statistic for memory is the spectral dynamic range (SDR) and its logarithm to base e (LSDR):

$$SDR = \max_{0\leq\omega\leq 1} f(\omega) - \min_{0\leq\omega\leq 1} f(\omega), \quad LSDR = \log SDR$$

The goal of the concept of memory is best illustrated by considering an AR(1) time series $Y(t) - \rho\, Y(t-1) = e(t)$ where $e(\cdot)$ is independent normal$(0,\sigma^2)$ series.

For an AR(1), $SDR = (1+|\rho|)/(1-|\rho|)$. A table of LSDR corresponding to different values of ρ gives us a guide to how to assign memory types:

ρ	.05	.15	.25	.35	.45	.55	.65	.75	.85	.95
LSDR	.2	.6	1.	1.5	1.5	1.9	2.5	3.1	3.9	7.3

Based on this table and empirical experience we might regard LDSR<1 as very short memory and LSDLR>7 as very long memory.

When one simulates time series in order to study the behavior of sample means, insight into the numbers obtained is provided by understanding: (1) the distribution of Y, especially the type of its departure from normality, and (2) the correlation structure of the time series Y(t), especially the type of its departure form independence (no memory) as measured by f(0), the value at zero frequency of the spectral density function. An initial way to empirically study the role of these effects on the behavior of the sample mean is by simulating an AR(1) process whose marginal distributions are exponential (a technical report by Will Alexander is in preparaton).

Let us examine from the point of view of time series memory type the time series model considered by Titus (1985):

$$Y(t) = MY + Z(t),\ MY = 10,\ e(t) = a(t) - 1$$

where a(t) are independent exponential with mean 1, Z(t) obeys the model

$$Z(t)-.5\, Z(t-1)-.3\, Z(t-2)-.2\, Z(t-3)=e(t),$$

initial values $Z(1)=Z(2)=Z(3)=-3$. The memory type of the observed time series Y(t) is long memory since $f(0) = 22.8$ and its log spectral dynamic range equals 6. Results of simulations of this time series model should be interpreted as conclusions about the behavior of the sample mean when computed from long memory time series.

3. EQUIVALENT DEGREES OF FREEDOM OF ASYMPTOTIC CONFIDENCE INTERVALS FOR THE MEAN

One of the great contributions of statistical theory in the first half of the 20th century was the development of small sample statistical methods based on t distributions, chi-squared distributions, and F distributions. For a no-memory time series exact (rather than asymptotic) confidence intervals for MY are obtained by using the fact:

$$\sqrt{T}\ \{MY^{\sim}-MY\}/\{c\ VAR^{\sim}[Y]\}^{0.5} = t_{T-1}$$

where t_k denotes Student's t distribution with k degrees of freedom, and $c = T/(T-1)$.

We can attain an approximate exactness by using an approximation to t_k (see Gaver and Kafadar (1984)); an example of an approximation (see Parzen (1985)) is an asymptotic formula with correction factor h:

$$hk \log\{1 + (1/k)t_k^2\} \to Z^2$$

where Z^2 obeys chi-squared distribution with 1 degree freedom, and we define

$$h=h(k) = (k-1)^2/k(k-1.5).$$

We write symbolically

$$t_k^2 = \{\exp(Z^2/h(k)k)-1\}k$$

This is a relation between distributions of random variables which is stated more precisely in terms of quantile functions; let $Q(u;X)$ denote the quantile function of a random variable X. We argue that approximately

$Q(1-(u/2);t_k)=[\{\exp(Q(1-u;Z^2)/h(k)k)-1\}k]^{0.5}$.

For u = 0.05, $Q(.95;Z^2) = 3.84146$. The approximate and exact values of $Q(.975;\ t_k)$ are given in Table I for $k\geq 6$.

Spectral density estimators f^(0) can be used to form confidence intervals for the mean, to take account of the (possibly severe) effects of dependence, by using an approximate distribution:

$$\sqrt{T}\ (MY\tilde{}\ - MY)/\{VAR\tilde{}[Y]\ f\hat{}(0)\}^{0.5} =$$
$$= t_k = [\{\exp(Z^2/h(k)k)-1\}k]^{0.5}$$

for a suitable value of k. One calls k the *equivalent degrees of freedom*. Research continues on suitable formulas for, and interpretation of, k. The next section suggests a formula for f^(0) which illustrates this approach to describing the behavior of sample means in a manner suitable for forming confidence intervals for MY.

TABLE I

Values of $Q(0.975;t_k)$

k	Exact	Approximate
6	2.447	2.445
7	2.365	2.365
8	2.306	2.306
9	2.262	2.262
10	2.228	2.228
11	2.201	2.201
12	2.179	2.179
13	2.160	2.160
14	2.145	2.145
15	2.131	2.132
16	2.120	2.120
17	2.110	2.110
18	2.101	2.101
19	2.093	2.093
20	2.086	2.086
21	2.080	2.080
22	2.074	2.074
23	2.069	2.069
24	2.064	2.064
25	2.060	2.060
26	2.056	2.056
27	2.052	2.052
28	2.048	2.048
29	2.045	2.045
30	2.042	2.042
40	2.021	2.021
120	1.980	1.980

4. SAMPLE FOURIER TRANSFORMS AND SAMPLE SPECTRAL DENSITY

To a time series sample Y(t), t=1,...,T, one can compute (by the Fast Fourier Transform) for k=0,1,...,T-1

$$YFOUR(k) = (1/T) \sum_{t=1}^{T} Y(t) \exp(2\pi ikt/T) .$$

The sample mean and variance can be expressed

$$MY\tilde{} = YFOUR(0)$$

$$VAR\tilde{}[Y] = \sum_{k=1}^{T-1} |YFOUR(k)|^2$$

Note that YFOUR(k) are complex valued, obeying YFOUR(T-k)=YFOUR*(k) where YFOUR*(k) denotes the complex conjugate of YFOUR(k).

The random variables YFOUR(k), k=0,1,...,[T/2] are asymptotically uncorrelated for a stationary time series. The sample spectrum S~(ω) and sample spectral density f~(ω) are defined at ω=k/T, k=0,1,...,T-1, by

$$S\tilde{}(k/n) = T|YFOUR(k)|^2$$

$$f\tilde{}(k/T) = S\tilde{}(k/T)/VAR\tilde{}[Y] \quad .$$

One can show that

$$VAR\tilde{}[Y] = (1/T) \sum_{k=1}^{T-1} S\tilde{}(k/T)$$

$$1 = (1/T) \sum_{k=1}^{T-1} f\tilde{}(k/T)$$

Sample spectral densities are very wiggly and need to be smoothed. For white noise (random sample) the random variables f~ (k/n) are asymptotically independent exponentially distributed with mean 1, and their optimal smoothing yields an estimated spectral density f^(ω) = 1. For short memory time series one forms estimators f^(ω) by suitable averages of f~(k/T) for k/T in suitable neighborhoods of ω.

Thus it is natural to estimate S(0) by S^(0) of the form (first suggested by Albert Einstein in 1914 in a paper only recently discovered)

$$S\hat{}(0) = (1/m) \sum_{k=1}^{m} S\tilde{}(k/T)$$

Asymptotically

2m S^(0)/S(0) → chi-squared distribution, 2m degrees freedom

$$\sqrt{T}\ \{MY^{\sim}-MY\}/\{S^{\wedge}(0)\}^{0.5} \rightarrow \text{t distribution, 2m degrees freedom}$$

One can form a confidence interval for MY from the foregoing statistic, which can be written in terms of estimators of the spectral density (rather than the spectrum) by

$$\sqrt{T}\ \{MY^{\sim}-MY\}/\{VAR^{\sim}[Y]\ f^{\wedge}(0)\}^{0.5} \rightarrow t_{2m}$$

One can regard this statistic as being of standardized time series type (discussed in the next section) and also as standardization by an estimator of the spectrum at zero frequency if one lets m tend to ∞ as T tends to ∞.

One approach to choosing m in practice is to choose it to be as large as is compatible with the hypothesis that $f^{\sim}(1/T),\ldots,f^{\sim}(m/T)$ are identically distributed (using a test such as Bartlett's test for equality of variances). One expects m to be small when time series memory is long.

5. SAMPLE BROWNIAN BRIDGE FUNCTIONALS AND STANDARDIZED TIME SERIES

When the sample mean $MY^{\sim}$ of a stationary time series is asymptotically normally distributed, its asymptotic distribution can be described by writing

$$\sqrt{T}\ \{MY^{\sim} - MY\} \rightarrow \sqrt{S(0)}\ W(1)$$

where W(1) is a N(0,1) random variable. We use the notation W(1) to introduce the role of the Weiner process W(u), $0 \leq u \leq 1$, and the Brownian Bridge process B(u), $0 \leq u \leq 1$.

We define W(u) to be a zero mean Gaussian process with covariance kernel $E[W(s)W(t)] = \min(s,t)$. We define B(u) to be a zero mean Gaussian process with covariance kernel

$$E[B(s)B(t)] = \min(s,t) - st.$$

Equivalently one can represent $B(u) = W(u) - uW(1)$.

An important role in the empirical analysis of time series (and in understanding the theory of standardized time series introduced into simulation studies by Schruben (1983)) is played by the <u>sample Brownian Bridge</u> of a time series:

$$B^{\sim}(u) = \sum_{t \leq [Tu]} \{Y(t)-MY^{\sim}\}/\sqrt{T}\ ,\quad 0 \leq u \leq 1\ .$$

Under suitable mixing conditions one can show weak convergence of the stochastic processes

$$\{B^{\sim}(u),\ 0 \leq u \leq 1\} \rightarrow \{\sqrt{S(0)}\ B(u),\ 0 \leq u \leq 1\}$$

An important graphical tool for analyzing a time series is a plot of $B^{\sim}(u)$, $0<u<1$. An important feature of the plot is its range

$$R^{\sim} = \max_{0<u<1} B^{\sim}(u) - \min_{0<u<1} B^{\sim}(u)\quad .$$

Let $(S^{\sim})^2 = VAR^{\sim}[Y] = R^{\sim}(0)$ denote the sample variance. The sample R/S statistic is the ratio $R^{\sim}/S^{\sim}$; Mandelbrot (1973) emphasizes plots of $\log R^{\sim} - \log S^{\sim}$ versus $\log T$ as a diagnostic tool for measuring the Hurst exponent of the time series which is a measure of its "long memory" nature [see Parzen (1986)].

$R^{\sim}$ is an example of a functional of the sample Brownian Bridge $B^{\sim}(\cdot)$; its asymptotic distribution obeys $R^{\sim} \rightarrow \sqrt{S(0)}\ R$, defining

$$R = \max_{0 \leq u \leq 1} B(u) - \min_{0 \leq u \leq 1} B(u)$$

To form a confidence interval for the population mean MY from the sample mean $MY^{\sim}$, without estimating S(0), the approach of the method of standardized time series is to form a random variable whose asymptotic distribution does not depend on S(0); an example of such a random variable is

$$\sqrt{T}\ \{MY^{\sim} - MY\}/R^{\sim} \rightarrow W(1)/R$$

One can find an explicit formula for the random variable on the right hand side which can be related to the theory of the Kuiper statistic in the theory of nonparametric inference:

$$\mathrm{Prob}[R>x] = 2 \sum_{k=1}^{\infty} (4k^2x^2-1)\exp(-2k^2x^2),$$

$$\mathrm{Prob}[R>1.75] = .05,\ \mathrm{Prob}[R>2.0) = .01.$$

Glynn and Iglehart (1985) have shown that alternatively one can form standardized time series type statistics such that the limit random variable has a distribution which is Student's distribution with k degrees of freedom for a suitable value of k. They show that standardized time series methods of forming confidence intervals for MY are asymptotically larger than intervals obtained by a method which consistently estimates the asymptotic variance by estimating S(0), the spectrum at zero frequency. The connections between spectral density estimation, construction of confidence intervals for the mean MY, and standardized time series can perhaps be clarified by studying statistics which are simultaneously a standardized time series method and an estimated spectral density standardization. We believe that an example of such a statistic was introduced in the preceding section, namely

$$T(MY^{\sim}-MY)^2/(1/m) \sum_{k=1}^{m} S^{\sim}(k/T)$$

where one chooses m as a function of T so that $m \to \infty$ and $m/T \to 0$ as $t \to \infty$.

ACKNOWLEDGEMENT

Our research is supported by U.S. Army Research Office and Office of Naval Research. I would like to thank Professors H. J. Newton and Ruey S. Tsay for their help in organizing the session on Time Series Analysis of Sample Means at the December 1986 Winter Simulation Conference.

REFERENCES

Gaver, D. P. and Kafadar, K. (1984). "A Retrievable Recipe for Inverse t," American Statistician, 38, 308-311.

Glynn, Peter W. and Iglehart, Donald L. (1985). "The theory of standardized time series," Technical report No. 32, Stanford Operations Research Department.

Mandelbrot, B. (1973). "Statistical methodology for non-periodic cycles: from the covariance to R/S analysis." Review of Economics and Social Measurement, 259-290.

Newton, H. J. (1987). Timeslab, Wadsworth: Belmont.

Parzen, E. (1985). "Quantile-information-functional statistical inference and unification of discrete and continuous data analysis," Proceedings 30th Army Conference on Design of Experiments, 213-226.

Parzen, E. (1986). "Quantile spectral analysis and long memory time series, Journal Applied Probability, 23A, 41-54.

Schruben, L. (1983). "Confidence interval estimation using standardized time series," Operations Research 31, 1090-1108.

Titus, Birney D. (1985). "Modified confidence intervals for the mean of an autoregressive process," Technical report 34, Stanford Operations Research Department.

AUTHOR'S BIOGRAPHY

EMANUEL PARZEN is Distinguished Professor of Statistics at Texas A&M University. His degrees are in Mathematics (A.B. 1949 Harvard; Ph.D. 1953 U. California, Berkeley). He was at SUNY Buffalo (1970-1978) and at Stanford (1956-1970). His fields of research include Scientific Statistical Science (development of methods usable by applied scientists which are not a bag of tricks but are based on unified approach which links diverse fields of statistics); Time Series Analysis (time domain, spectral domain, reproducing kernels); Signal Processing; Quantile, Information, Functional Statistical Inference approach to Data Modeling.

He is a fellow of International Statistical Institute, American Statistical Association, Institute of Mathematical Statistics, Royal Statistical Society, AAAS, and member of many societies in the mathematical sciences.

Emanuel Parzen
Department of Statistics
Texas A&M University
College Station, Texas 77843, U.S.A.