Three Curious Properties of the Sample Variance and Autocovariance for Stationary Processes with Unknown Mean

by

Donald B. Percival*
Applied Physics Laboratory, HN–10
University of Washington
Seattle, WA 98195
(206)–543–1300
FAX: (206)–543–6785

Abstract – In most books on time series analysis, estimators of the variance and autocovariance for a stationary process are discussed under the assumption that the process mean is known. Here we illustrate that, if the process mean is unknown and hence is estimated by the sample mean, these estimators have some surprising properties.

KEYWORDS: Time series analysis; sample variance; sample autocovariance; stationary processes

*Donald B. Percival is Senior Mathematician, Applied Physics Laboratory, College of Ocean and Fishery Sciences, HN–10, University of Washington, Seattle, WA 98195. The author wishes to thank Brad Bell, Andrew Walden and an anonymous referee for helpful comments that substantially improved the manuscript. The 'zero correlation time' issue arose while the author was working with Mike Gregg and Harvey Seim under Gregg's Office of Naval Research Contract N00014–86–K–0690, which supported the 'Mixing to Mesoscale' University Research Initiative.

1. INTRODUCTION

Let $\{X_t\}$ denote a discrete parameter stationary process with mean $\mu = E\{X_t\}$ and autocovariance sequence (acvs) given by

$$s_\tau = cov\{X_t, X_{t+\tau}\} = E\{(X_t - \mu)(X_{t+\tau} - \mu)\}, \qquad \tau = 0, \pm 1, \pm 2, \ldots$$

(note that s_0 is the variance of X_t). If we are given a time series of length N that can be regarded as a realization of a portion $X_1, \ldots, X_N$ of $\{X_t\}$, the usual estimator of μ is just the sample mean $\bar{X} = \frac{1}{N}\sum_{t=1}^{N} X_t$. Textbooks on time series analysis commonly discuss two estimators of the acvs, namely, the 'unbiased' estimator

$$\tilde{s}_\tau \equiv \frac{1}{N - |\tau|} \sum_{t=1}^{N-|\tau|} (X_t - \bar{X})(X_{t+|\tau|} - \bar{X})$$

and the 'biased' estimator

$$\hat{s}_\tau \equiv \frac{1}{N} \sum_{t=1}^{N-|\tau|} (X_t - \bar{X})(X_{t+|\tau|} - \bar{X}) = \left(1 - \frac{|\tau|}{N}\right) \tilde{s}_\tau,$$

where $|\tau| \le N - 1$. The statistical properties of $\tilde{s}_\tau$ and $\hat{s}_\tau$ are sufficiently complicated that it is convenient to assume that μ is known. Under this assumption, the acvs estimators reduce to

$$\tilde{s}'_\tau \equiv \frac{1}{N - |\tau|} \sum_{t=1}^{N-|\tau|} (X_t - \mu)(X_{t+|\tau|} - \mu) \text{ and } \hat{s}'_\tau \equiv \frac{1}{N} \sum_{t=1}^{N-|\tau|} (X_t - \mu)(X_{t+|\tau|} - \mu).$$

While these 'mean known' estimators are mathematically more convenient to work with, the 'mean adjusted' estimators $\tilde{s}_\tau$ and $\hat{s}_\tau$ are invariably used in practice. The purpose of this note is to point out three properties of $\tilde{s}_\tau$ and $\hat{s}_\tau$ that are surprisingly different from those of the 'mean known' estimators. These 'curiosities' are useful as supplementary material for students in an introductory time series analysis class.

2. WHICH ACVS ESTIMATOR HAS LESS BIAS?

If the process mean μ is known, it is an easy exercise to show that $\tilde{s}'_\tau$ is an unbiased estimator of s_τ, whereas $\hat{s}'_\tau$ is a biased estimator unless $s_\tau = 0$. If μ is unknown, then both $\tilde{s}_\tau$ and $\hat{s}_\tau$ are in general biased estimators, but it is common practice to refer to the former as the 'unbiased' estimator and to the latter as the 'biased' estimator (see, for example, Priestley, 1981, Section 5.3.3). With this convention, it is somewhat surprising that the 'biased' estimator can in fact be less biased than the 'unbiased' estimator for all nonzero lags for which the estimators are defined (the two estimators are identical for $\tau = 0$). To see this, suppose that $\{X_t\}$ is a white noise process; i.e., $s_\tau = 0$ for $\tau \neq 0$. An easy exercise shows that

$$E\{\tilde{s}_\tau\} = -\frac{s_0}{N}, \text{ whereas } E\{\hat{s}_\tau\} = -\left(1 - \frac{|\tau|}{N}\right)\frac{s_0}{N}.$$

That $\hat{s}_\tau$ is less biased than $\tilde{s}_\tau$ when $0 < |\tau| < N-1$ follows from the fact that $E\{\hat{s}_\tau\}$ is closer to 0 than $E\{\tilde{s}_\tau\}$. (A general formula for the bias in $\hat{s}_\tau$ can be obtained from Theorem 6.2.2 of Fuller, 1976.)

3. NEGATIVE AUTOCOVARIANCES AND ZERO CORRELATION TIMES

When the process mean is unknown, the biased estimator of the acvs satisfies the following surprising constraint:

$$\sum_{\tau=-(N-1)}^{N-1} \hat{s}_\tau = 0. \tag{3.1}$$

Proof. Consider an $N \times N$ matrix whose (t,u)th entry is $(X_t - \bar{X})(X_u - \bar{X})$ for $1 \leq t, u \leq N$:

$$\begin{bmatrix} (X_1-\bar{X})(X_1-\bar{X}) & (X_1-\bar{X})(X_2-\bar{X}) & \dots & (X_1-\bar{X})(X_N-\bar{X}) \\ (X_2-\bar{X})(X_1-\bar{X}) & (X_2-\bar{X})(X_2-\bar{X}) & \dots & (X_2-\bar{X})(X_N-\bar{X}) \\ \vdots & \vdots & \ddots & \vdots \\ (X_N-\bar{X})(X_1-\bar{X}) & (X_N-\bar{X})(X_2-\bar{X}) & \dots & (X_N-\bar{X})(X_N-\bar{X}) \end{bmatrix}.$$

The sum of the elements on the main diagonal is $\sum_{t=1}^{N}(X_t - \bar{X})^2 = N\hat{s}_0$. The sum of the elements on the τth super-diagonal is $N\hat{s}_\tau$ for $\tau = 1, \dots, N-1$, and the same is true for the τth sub-diagonal. Since $\hat{s}_{-\tau} = \hat{s}_\tau$, the sum of all the elements in the matrix is $N\sum_{\tau=-(N-1)}^{N-1}\hat{s}_\tau$. However, the sum of the tth row of the matrix is $(X_t - \bar{X})\sum_{u=1}^{N}(X_u - \bar{X})$

– this is identically zero for all t, from which the stated result follows immediately. (A variation on this proof shows that

$$N \sum_{\tau=-(N-1)}^{N-1} \hat{s}'_\tau = \left(\sum_{t=1}^{N} (X_t - \mu) \right)^2 \geq 0 \tag{3.2}$$

when the process mean is known.)

Equation (3.1) has three interesting implications.

[1] Since $\hat{s}_0 > 0$ unless X_t is constant in t (an uninteresting special case), it follows that $\hat{s}_\tau$ *must* be negative for some values of τ, a property that is not necessarily shared by the true acvs.

[2] Property [1] in turns illustrates a limitation in the interpretation of $\hat{s}_\tau / \hat{s}_0$ as a correlation coefficient measuring linear association between $X_{t+\tau}$ and X_t for $t = 1, \ldots, N-\tau$ and $0 \leq \tau < N$. An extreme example is a time series whose tth value is given by bt, where b is a nonzero constant. A scatter plot of $X_{t+\tau}$ versus X_t is perfectly linear (with a slope of unity) for all τ such that $0 \leq \tau \leq N-2$. The usual relationship between scatter plots and correlation coefficients suggests that $\hat{s}_\tau / \hat{s}_0$ should be unity (see, for example, Wright, 1992), whereas in fact it can be close to zero or negative. Figure 1 illustrates this example.

[3] Suppose that $\{X(t)\}$ is a continuous parameter stationary process with unknown mean and autocovariance function $s(\tau)$. A measure of the correlation time (or integral time scale) of this process can be defined as

$$\int_0^\infty \frac{s(\tau)}{s(0)} \, d\tau = \frac{1}{2} \int_{-\infty}^{\infty} \frac{s(\tau)}{s(0)} \, d\tau.$$

If it is finite, this measure is sometimes interpreted in the physical sciences as 'the time needed for any correlation between $X(t)$ and $X(t+\tau)$ to die out' (Yaglom, 1987, p. 113). For a sampling interval $\Delta > 0$, let us now define the discrete parameter process $\{X_t\}$ by letting $X_t = X(t\Delta)$ for $t = 0, \pm 1, \pm 2, \ldots$. If we now obtain a realization of $X_1, \ldots, X_N$, a naive estimator of correlation time would be proportional to $\sum_{\tau=-(N-1)}^{N-1} \hat{s}_\tau$, which Equation (3.1) tells us must always be zero. Clearly this naive

estimator is not very useful! On the other hand, if $E\{X_t\} = E\{X(t)\} = \mu$ is known, we can use $\hat{s}'_\tau$ instead of $\hat{s}_\tau$. Equation (3.2) suggests that the naive estimator might then be of some use.

4. BIAS IN THE SAMPLE VARIANCE

For $\tau = 0$, we have

$$\hat{s}_0 = \tilde{s}_0 = \frac{1}{N}\sum_{t=1}^{N}(X_t - \bar{X})^2 \text{ and } \hat{s}'_0 = \tilde{s}'_0 = \frac{1}{N}\sum_{t=1}^{N}(X_t - \mu)^2,$$

all of which estimate the process variance $\text{var}\{X_t\} = s_0$. It readily follows that $\hat{s}'_0$ is an unbiased estimator of s_0 for all stationary processes. Further, it is straightforward to show that

$$E\{\hat{s}_0\} = s_0 - \text{var}\{\bar{X}\}. \tag{4.1}$$

Since $\text{var}\{\bar{X}\} \geq 0$ and since $\hat{s}_0$ is nonnegative, we have established that

$$0 \leq E\{\hat{s}_0\} \leq s_0 \tag{4.2}$$

(this is a special case of an elegant proof due to David, 1985). Note that this result says that, on the average, we cannot overestimate s_0 using the estimator $\hat{s}_0$. If s_τ converges to 0 as τ gets large, then $\text{var}\{\bar{X}\} \to 0$ as $N \to \infty$ (Corollary 6.1.1.1, Fuller, 1976), so we can state the following asymptotic result:

$$\lim_{N\to\infty} E\{\hat{s}_0\} = s_0.$$

Is it possible to sharpen the lower bound of 0 for $E\{\hat{s}_0\}$ for finite N? The following result says that the answer is 'no.'

Theorem. For every sample size $N \geq 1$ and every $\epsilon > 0$, there exists a stationary process such that

$$\frac{E\{\hat{s}_0\}}{s_0} < \epsilon.$$

Proof. Let $\{X_t\}$ be a stationary first-order autoregressive process, i.e.,

$$X_t = \phi X_{t-1} + e_t,$$

where $|\phi| < 1$, and $\{e_t\}$ is a white noise process with zero mean and unit variance. We have $s_0 = 1/(1-\phi^2)$ and $s_\tau = \phi^{|\tau|} s_0$ (Equation (2.3.5), Fuller, 1976). Using Equation (4.1) and the result

$$\text{var}\,\{\bar{X}\} = \frac{1}{N^2}\sum_{t=1}^{N}\sum_{u=1}^{N} s_{t-u}$$

(Corollary 6.1.1.2, Fuller, 1976), we obtain

$$\frac{E\{\hat{s}_0\}}{s_0} = 1 - \frac{1}{N^2}\sum_{t=1}^{N}\sum_{u=1}^{N}\phi^{|t-u|} \equiv 1 - f(\phi).$$

If $N > 1$, then $f(\cdot)$ is a continuous strictly increasing function of ϕ on the closed interval $[0, 1]$. Since $f(0) = 1/N$ and $f(1) = 1$, we can find a value $\phi < 1$ such that $1 - f(\phi) < \epsilon$, from which we obtain the stated result (the case $N = 1$ is trivial because then $\hat{s}_0 = 0$).

This theorem shows that the effect of not knowning the process mean can be rather large: whereas $\hat{s}_0'$ is an unbiased estimator of s_0 when μ is known, the commonly used $\hat{s}_0$ can severely underestimate s_0 when μ is unknown. Figure 2 illustrates this result.

REFERENCES

David, H. A. (1985) Bias of S^2 Under Dependence. *The American Statistician*, 39, 201.

Fuller, W. A. (1976) *Introduction to Statistical Time Series.* New York: John Wiley & Sons.

Priestley, M. B. (1981) *Spectral Analysis and Time Series.* London: Academic Press.

Wright, T. (1992) Lagrange's Identity Reveals Correlation Coefficient and Straight-Line Connection. *The American Statistician*, 46, 106–107.

Yaglom, A. M. (1987) *Correlation Theory of Stationary and Related Random Functions, Volume I: Basic Results.* New York: Springer-Verlag.

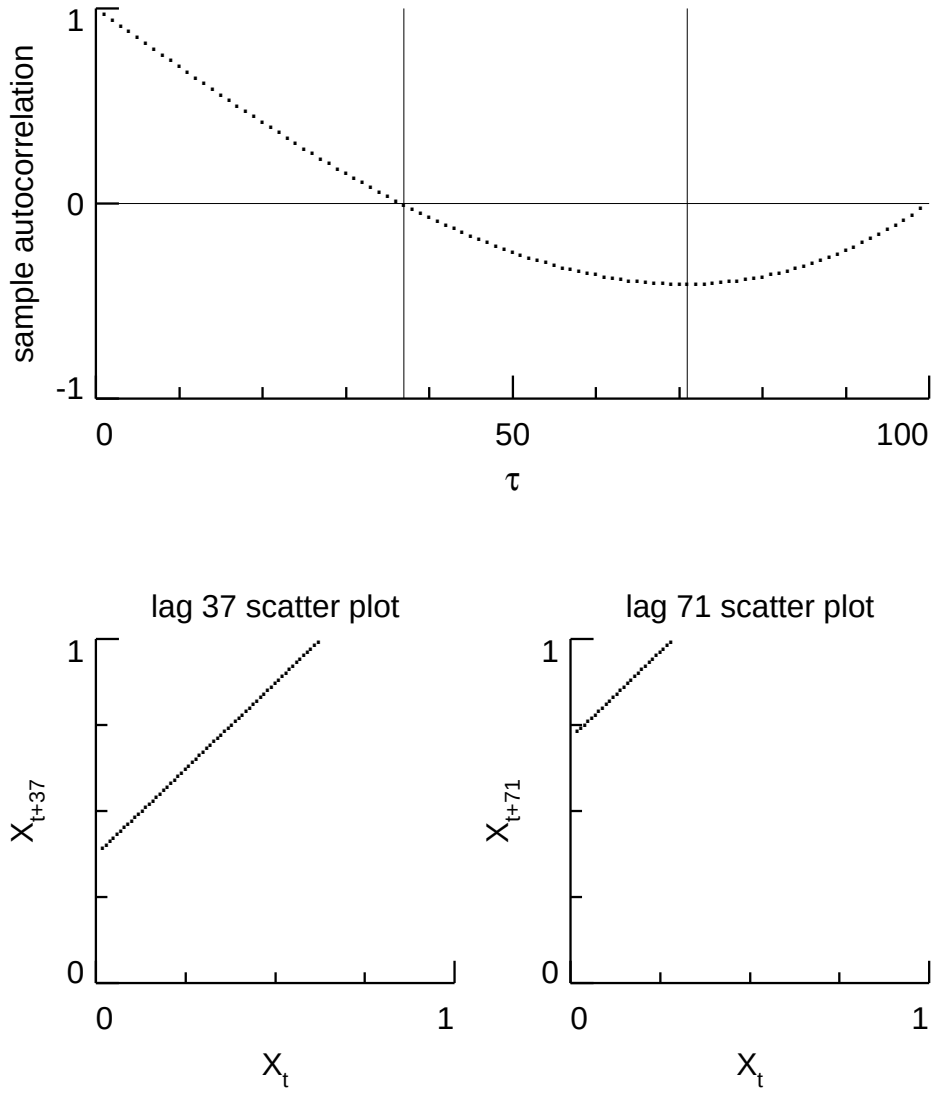

Figure 1. The top plot shows $\hat{s}_\tau/\hat{s}_0$ versus τ for a time series of length $N = 100$ whose tth value is given by $0.01t$. The two thin vertical lines mark the positions $\tau = 37$ and 71, for which $\hat{s}_{37}/\hat{s}_0 = -0.01$ and $\hat{s}_{71}/\hat{s}_0 = -0.41$. The two bottom plots show the scatter plots $X_{t+\tau}$ versus X_t for these two lags, where $t = 1, \ldots, 100 - \tau$.

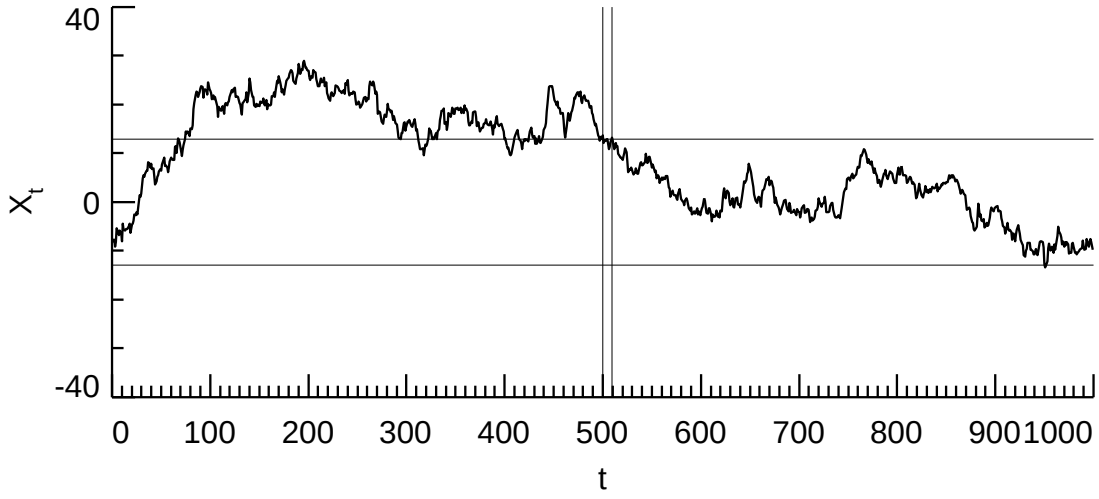

Figure 2. The jagged curve shows a portion of size 1000 of a realization from the stationary zero mean AR(1) process $X_t = 0.997X_{t-1} + e_t$, where $\{e_t\}$ is a white noise process with zero mean and unit variance. The variance of X_t is $s_0 = 166.9$, and the two thin horizontal lines mark $\pm s_0^{1/2} = \pm 12.9$. For sample size $N = 10$, we have $E\{\hat{s}_0/s_0\} = 0.01$; i.e., the sample variance is biased by a factor of 100. The two thin vertical lines enclose one such sample of 10 points, for which $\hat{s}_0 = 0.7$ (well below $s_0 = 166.9$). If we sweep across the time series plotted above and compute $\hat{s}_0$ for all possible 991 subseries of size 10, we find that $\hat{s}_0$ varies from 0.1 to 17.1 and has an average value of 1.8 (again well below the true variance). On the other hand, if we compute $\hat{s}_0'$ based upon the known process mean of zero, we find that $\hat{s}_0'$ varies from 0.6 to 764.0 and has an average value of 184.1, which is rather close to $s_0 = 166.9$.