

THE AUTOMATED BLOCKING METHOD

Marius Jonsson, Morten Hjorth-Jensen

August 22, 2017

ABSTRACT

Suppose $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ is the mean of $n = 2^d$ weakly stationary observations X_1, X_2, \dots, X_n . Using $8n$ floating point operations, we automate the blocking method for estimating the variance of \bar{X}_n . We develop the existing mathematics and prove that the rate of change of the estimate is only dependent on the covariance between successive observations. On the basis of this, the behaviour of the method is easily proven. Using the results, we propose a hypothesis test which automates the blocking method. We provide an algorithm and code to show how the automation can be implemented. We finally provide method validation by the means of autoregressive AR(1) and AR(2) stochastic processes as well as validation on physics applications. We find that the convergence rate is exponential as a function of data size, and that the method is significantly more accurate than dependent bootstrapping for the autoregressive models ($p < 0.0001$).

INTRODUCTION

Estimation of the variance of the sample mean is important in all of the sciences. Since 1867 (ref) Chebyshev's inequality told us that the expected value, $\langle X \rangle$ is well approximated by the sample mean \bar{X}_n . However, the variance of the mean is necessary to quantify the uncertainty in the estimate. And for a random sample, this is easily obtained, but for correlated data, the computation is more complicated. However, in this text, we show that if X_1, X_2, \dots, X_n are $n = 2^d$ weakly stationary observations, then the complexity is no larger than that of the sample mean.

We will utilize so called blocking transformations. This refers to forming a new sample of data by taking the mean of every pair of subsequent observations. People has been making such transformations in all areas of mathematics, and (ref) Flydberg and Pedersen made popular a method where blocking transformations reduced the correlations of data, and proposed a way to estimate the variance of the mean. Our develop the mathematics further and propose an automation algorithm which is sufficiently robust that supervision is not required.

The article is structured by «introduction»-, «methods»-, «results and discussion»- and finally a «conclusion and perspectives»-sections.

METHODS

We will agree about the background -Almost surely, - * means conjugate - bracket means expectation - def: subject to k blocking transformations

$$\text{Var}(\langle A \rangle) = \langle A^2 \rangle - \langle A \rangle^2, \quad \text{Cov}(A, B) = \langle AB \rangle - \langle A \rangle \langle B \rangle \quad (1)$$

$$X_i^{(k+1)} = \frac{1}{2} \left(X_{2i-1}^{(k)} + X_{2i}^{(k)} \right) \quad (2)$$

It is known that

THEOREM 1. *If $X_i = \mu + \sum_{j=1}^{2^d} \psi_j w_{i-j}$ is a stationary linear process and there exist $\eta \sigma_w^2 \in \mathbb{R}$ such that for all $0 < i < 2^d$ its fourth moment satisfies $\langle w_i^4 \rangle = \eta \sigma_w^2$ and $k < 2^d$, then $\hat{\rho}_t$ is asymptotically multivariate normal distributed with expected value ρ_t and covariance matrix Σ with elements*

$$\Sigma_{pq} = 2^{-d} \sum_{u=1}^{2^d-1} [\rho(u+p) + \rho(u-p) - 2\rho(p)\rho(u)] [\rho(u+q) + \rho(u-q) - 2\rho(q)\rho(u)].$$

Proof. See Shumway and Stoffer 2011 or the appendix. ■

RESULTS

We first state the results, then prove them in the order they were introduced. Our main result is an automated algorithm of the blocking method, some preliminary results are required.

PRELIMINARY RESULTS

The most counterintuitive of which is the following. It tells us that the correlation between sequential observations contain all the information that exist about the rate of change of the error.

PROPOSITION 1. *Suppose 2^d is the number of observations, and $\gamma_0^{(k)}$ is finite for all $k \in \{0, 1, \dots, d-1\}$. Then the rate of change of the error is*

$$\varepsilon_k - \varepsilon_{k+1} = \frac{\gamma_1^{(k)}}{n^{(k)}} \quad \text{for all} \quad 0 \leq k < d-1. \quad (3)$$

Perhaps not surprisingly, the behaviour of the blocking method follows directly from this proposition. To be precise, we will prove that

COROLLARY 1. *If 2^d is the number of observations, $i, j \in \mathbb{N}$ and $\gamma_0^{(k)}$ is finite for all $k \in \{0, 1, \dots, d-1\}$;*

1. *if there exist $k \in \mathbb{N}$ such that $\gamma_1^{(k)} > 0$ for all $i \leq k \leq j$, then the sequence of errors ε_k is strictly decreasing for all $j \leq k \leq i$.*
2. *if there exist $k \in \mathbb{N}$ such that $\gamma_1^{(k)} \geq 0$ for all $i \leq k \leq j$, then the sequence of errors ε_k is nonstrictly decreasing for all $j \leq k \leq i$.*
3. *if there exist some $i \in \mathbb{N}$ such that $\gamma_1^{(i)} = 0$, and $\gamma_t^{(i)} \geq \gamma_{t+1}^{(i)}$ then the sequence of errors ε_k is constant for all $i \leq k < d-1$.*

COROLLARY 2. *If the blocked variables become independent, for some number of blocking transformation, then the sequence of errors ε_k become constant.*

```

# (1) initialization
n = len(x)
d = log2(n)
mu = mean(x)
s, gamma = zeros(d), zeros(d)

# (2) estimate the auto-covariance and variances
#   for each blocking transformation
for i in range(0,d):
    n = len(x)
    # estimate autocovariance of x
    gamma[i] = n**(-1)*sum( (x[2:n]-mu)*(x[1:(n-1)]-mu) )
    # estimate variance of x
    s[i] = var(x)
    # perform blocking transformation
    x = zeros(n/2)
    for j in range(0, n/2+1):
        x[j] = 0.5*( x[2*j] + x[2*j+1] )

# (3) generate the test observator Z_k from the theorem
Z = zeros(d)
for k in range(0,d):
    Z[k] = (gamma[k]/s[k])**2*(d-k)
Z = flip(cumsum(flip(Z)))

# (4) save vector of 99-percentiles of chi square
q = array([6.634897,9.210340, ... , 50.892181])

# (5) determine the smallest value k such that H_0 is true
for k in range(0,d):
    if(Z[k] < q[k]):
        break

# (6) the answer is ans
n = 2**(d-k)
ans = s[k]/n

```

Figure 1: A simple implementation for python. In step (1) we initialize some variables. n is the length of the observation vector x containing $X_1^0, X_2^0, \dots, X_{2^d}^0$. μ is the mean of x which is required when we for each blocking iteration compute the auto-covariance γ . The array s will contain the variances for each blocking iteration. In step (2), we for each blocking iteration i compute the associated $\gamma[i]$ and $s[i]$ and perform a blocking transformation. In step (3) we cumulatively sum the test observator Z_k . In step (4) we save an array of at least length d which contains all the quantiles from the chi square distribution. In step (5) we perform the hypothesis test to determine the appropriate k and finally (6) compute the answer at the iteration number k .

ALGORITHM

By the proposition, we know that the rate of change of the error is given by $\gamma_1^{(n)}$. That means that if there exist some $k \in \mathbb{N}$ such that for all $n > k$, then $\gamma_1^{(n)}$ is not significantly different from zero, then we stop the algorithm, because iterating further has no benefit (since by the colollary, the estimate cannot get any

better). Actually, readers whom are familiar with the blocking method knows that iterating to far actually induces extra errors since we are not dealing with the $\gamma_i^{(n)}$ s directly, but with estimators $\hat{\gamma}_i^{(n)}$. So stopping the algorithm at this point is optimal. We want to use hypothesis testing to determine k .

THEOREM 2. Suppose $k \in \mathbb{N}$ and we want to test the null hypothesis $H_0 : \langle \gamma_1^{(n)} = 0 \text{ for all } n \geq k \rangle$ against the alternative $H_a : \text{There exist an } n \geq k \text{ such that } \gamma_1^{(n)} \neq 0$. Then

$$Z_k = \sum_{i=k}^{d-1} \hat{\rho}_1^{(i)} n^{(i)} \approx \chi_{d-k}^2$$

is a test observator with upper tailed rejection region.

Suppose we we choose $\alpha = 0.99$ The simplest scheme would be the following:

Add flow chart. As you can imagine the quality of the estimate depends on the number of observations per time constant τ of the correlation structure. Just like any method which attempts to compute the variance of the mean based on the covariances or variance, such as the blocking method, or a brute force estimation

$$\widehat{\text{Var}}(\bar{X}) = 2^{-d} \left[\hat{\gamma}_0^{(0)} + 2 \sum_{t=1}^{2^d-1} \left(1 - \frac{t}{n} \right) \hat{\gamma}_t^{(0)} \right].$$

In fact, we shall see in the validation of the method that the quality of the estimate is what we would expect, it improves exponentially as a function of n/τ .

VALIDATION OF METHOD

PROOF OF CLAIMS

The remaining results develop the mathematics of the blocking method. Let's start with a soft result on which we will build more structure.

PROPOSITION. Suppose $d > 1$ is an integer and X_1, \dots, X_m are $m = 2^d$ random variables from a weakly stationary sample. If we let $\bar{X}_m = (1/m) \sum_{i=1}^m X_i$ denote the mean, and for $t = |i - j|$ we let $\gamma_t^{(k)} = \gamma_{i,j}^{(k)}$ be the autocovariance subject to k blocking-transformations. Then the block transformed variables $X_1^{(k)}, \dots, X_{2^{-k}m}^{(k)}$ are weakly stationary and

$$\text{Var}(\bar{X}_m) - \frac{1}{n^{(k)}} \text{Var}(X_i^{(k)}) = \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}} \right) \gamma_t^{(k)} \quad (4)$$

and

$$\gamma_j^{(k)} = \begin{cases} \frac{1}{2} \gamma_{2j}^{(k-1)} + \frac{1}{2} \gamma_{2j+1}^{(k-1)} & \text{if } j = 0 \\ \frac{1}{4} \gamma_{2j-1}^{(k-1)} + \frac{1}{2} \gamma_{2j}^{(k-1)} + \frac{1}{4} \gamma_{2j+1}^{(k-1)} & \text{if } j > 0 \end{cases} \quad (5)$$

Proof. See `flyvbjerg_error_1989` or the appendix. ■

The right hand side of equation (4) represent the difference between the quantity which we want to estimate, and the estimator which we will use. Therefore, at blocking iteration number k , we will call the right hand side the **error**, and denote it by ε_k . Next let's make a couple of simple observations

LEMMA. Suppose $m = 2^d$, $d > 1$ denotes the number of observations and k denote the number of blocking transformations. Then

$$n^{(k)} = 2^{-k} m \quad \text{and} \quad k < \log_2 m \quad \text{for all} \quad k < d.$$

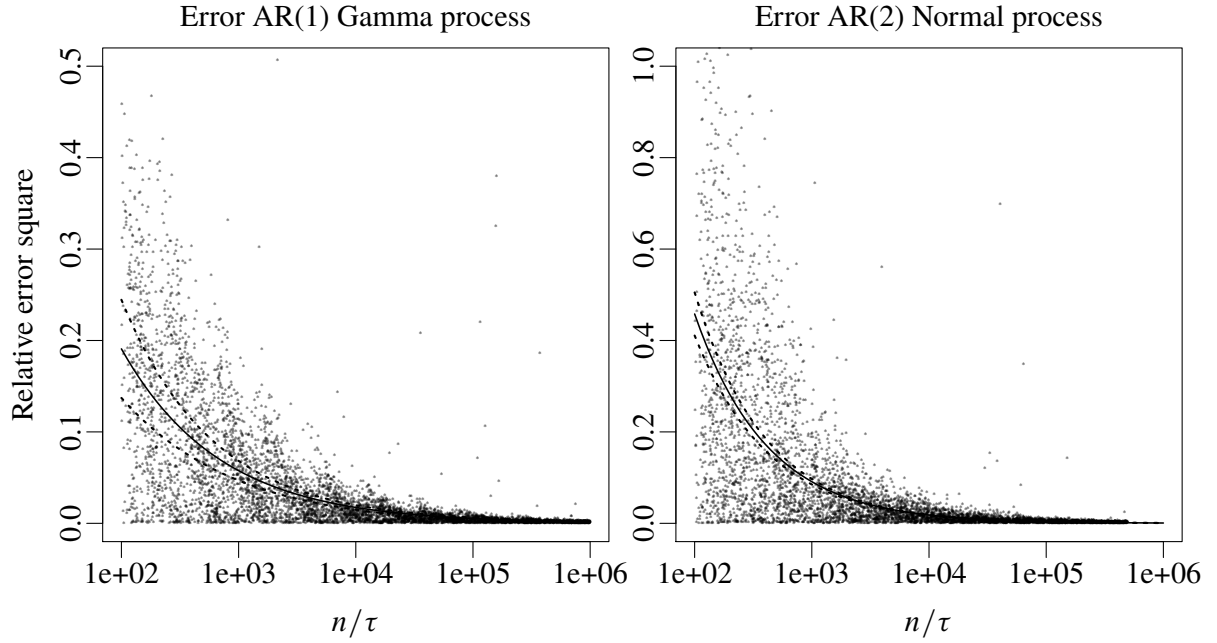


Figure 2: Relative error squared of two autoregressive models versus observations per time autocorrelation time-constant. There is exponential convergence rate for two vastly different correlation structures. AR(1) autocorrelation is positive with exponential decay whilst AR(2) is oscillatory with exponential decay. The processes were distributed vastly different. They were Gamma(1,1) and Standard Normal respectively. The method was insensitive to the distribution of the observations ($p = 0.895$) [sjekk dette]. It is apparant that the first digit of the method was correct for $n = 50\tau$, and two digits correct for $n = 500\tau$. The mean relative errors μ were modelled by gamma regression, $\log \mu = \beta_0 + \beta_1 x$. Deviance explained was 50.65% and 65.39% respectively on 6078 degrees of freedom. Dashed lines give 95% confidence intervals of the means.

Proof. We proof the first part by induction. Suppose $k = 0$. By definition, $n^{(0)} = m$ and so

$$n^{(0)} = m = 2^{-0}m,$$

Suppose now that there exist some $k \in \{0, 1, 2, \dots\}$ such that $n^{(k)} = 2^{-k}m$ and $n^{(k+1)}$ exist, this is possible since $d > 1$ by hypothesis. By construction each blocking transformation halve the number of observations, so $n^{(k+1)} = n^{(k)}/2$ and therefore

$$n^{(k+1)} = \frac{n^{(k)}}{2} = \frac{2^{-k}m}{2} = 2^{-(k+1)}m.$$

To see the last observation is true, just observe that if $k < d$ and $m = 2^d$, then $k < d = d \log_2 2 = \log_2 m$. \blacksquare

PROPOSITION. Suppose 2^d is the number of observations, and $\gamma_0^{(k)}$ is finite for all $k \in \{0, 1, \dots, d-1\}$. Then the rate of change of the error is

$$\varepsilon_k - \varepsilon_{k+1} = \frac{\gamma_1^{(k)}}{n^{(k)}} \quad \text{for all} \quad 0 \leq k < d-1. \quad (6)$$

Proof. We prove formula (6), but first we define a functional to make notation more convenient. We will manipulate sums and will be interested in which function $f(t)$ appear in the terms $\gamma_{f(t)}^{(k)}$ inside the summation. Suppose $a \in \mathbb{Z}$ and define the polynoms $p_a(x) = 2x - a$ and a functional $S_{p_a(j)} : \{p_{-1}, p_0, p_1\} \rightarrow \mathbb{R}$ by:

$$S_{p_a(j)} \equiv \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{p_a(t)}^{(k)}. \quad (7)$$

That means, using $n^{(k+1)} = n^{(k)}/2$, we can rewrite ε_{k+1} as

$$\begin{aligned} \varepsilon_{k+1} &\stackrel{(14)}{=} \frac{2}{n^{(k+1)}} \sum_{t=1}^{n^{(k+1)}-1} \left(1 - \frac{t}{n^{(k+1)}}\right) \gamma_t^{(k+1)} \stackrel{(15)}{=} \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) (\gamma_{2t-1}^{(k)} + 2\gamma_{2t}^{(k)} + \gamma_{2t+1}^{(k)}) \\ &\stackrel{(7)}{=} S_{2j-1} + 2S_{2j} + S_{2j+1} = 2(S_{2j-1} + S_{2j}) + S_{2j+1} - S_{2j-1} \end{aligned} \quad (8)$$

where we added and subtracted S_{2j-1} in the last step. Write out the sums explicitly to obtain:

$$S_{2j-1} = \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} = \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t-1}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} - \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \frac{\gamma_{2t-1}^{(k)}}{n^{(k)}} \quad (9)$$

$$\begin{aligned} S_{2j+1} &= \frac{1}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{2t+1}^{(k)} \\ &= \frac{1}{n^{(k)}} \left[\sum_{t=1}^{n^{(k)}/2-1} \gamma_{2t+1}^{(k)} - \left(\frac{2}{n^{(k)}} \gamma_3^{(k)} + \frac{4}{n^{(k)}} \gamma_5^{(k)} + \dots + \frac{n^{(k)}-2}{n^{(k)}} \gamma_{n^k-1}^{(k)} \right) \right] \\ &= \frac{1}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \left(1 - \frac{2(t-1)}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} = \frac{1}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \left(1 - \frac{2t-1}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} + \frac{1}{n^{(k)}} \sum_{t=2}^{n^{(k)}/2} \frac{\gamma_{2t-1}^{(k)}}{n^{(k)}} \end{aligned}$$

If the last term is subtracted from the first, we obtain after a few manipulations

$$S_{2j+1} - S_{2j-1} = \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} - \frac{1}{n^{(k)}} \gamma_1^{(k)} \quad (10)$$

Lets also investigate the following quantity

$$\begin{aligned} n^{(k)}(S_{2j-1} + S_{2j}) &\stackrel{(9)}{=} \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t-1}{n^{(k)}}\right) \gamma_{2t-1}^{(k)} - \sum_{t=1}^{n^{(k)}/2-1} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} + \sum_{t=1}^{n^{(k)}/2-1} \left(1 - \frac{2t}{n^{(k)}}\right) \gamma_{2t}^{(k)} \\ &= \left[\left(1 + \frac{1}{n^{(k)}}\right) \gamma_1^{(k)} + \left(1 + \frac{2}{n^{(k)}}\right) \gamma_2^{(k)} + \dots + \left(1 + \frac{n^{(k)}-2}{n^{(k)}}\right) \gamma_{n^{(k)}-2}^{(k)} \right] - \sum_{t=1}^{n^{(k)}/2-1} \frac{\gamma_{2t-1}^{(k)}}{n^{(k)}} \\ &= \sum_{t=1}^{n^{(k)}-2} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} - \sum_{t=1}^{n^{(k)}/2-1} \frac{\gamma_{2t-1}^{(k)}}{n^{(k)}} = \sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} - \sum_{t=1}^{n^{(k)}/2} \frac{\gamma_{2t-1}^{(k)}}{n^{(k)}} \end{aligned} \quad (11)$$

In the last equality we added and subtracted $(1 - t/n^{(k)})\gamma_t^{(k)}$ and in addition used that for $t = n^{(k)} - 1$, $(1 - t/n^{(k)})\gamma_t^{(k)} = 1/n^{(k)}\gamma_{n^{(k)}-1}^{(k)}$. This means that if we consider

$$2(S_{2j-1} + S_{2j}) \stackrel{(11)}{=} \frac{2}{n^{(k)}} \left[\sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)} - \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \right] = \varepsilon_k - \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} \quad (12)$$

Now substitute (10) and (12) into (8)

$$\begin{aligned} \varepsilon_{k+1} &\stackrel{(8)}{=} 2(S_{2j-1} + S_{2j}) + S_{2j+1} - S_{2j-1} \\ &\stackrel{(10)(12)}{=} \varepsilon_k - \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} + \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}/2} \frac{1}{n^{(k)}} \gamma_{2t-1}^{(k)} - \frac{1}{n^{(k)}} \gamma_1^{(k)} = \varepsilon_k - \frac{\gamma_1^{(k)}}{n^{(k)}} \end{aligned}$$

Subtract $\varepsilon_{k+1} - \gamma_1^{(k)}/n^{(k)}$ from each side of the equation, and the proposition follows. \blacksquare

The proposition says that the rate of convergence of the method is only dependant on the correlation between sequential observations $X_i^{(k)}$, and not the full correlation structure. But the following two corollaries show that the proposition prove the expected behaviour of the blocking method.

COROLLARY. Suppose 2^d is the number of observations, $i, j \in \mathbb{N}$ and $\gamma_0^{(k)}$ is finite for all $k \in \{0, 1, \dots, d-1\}$;

1. if there exist $k \in \mathbb{N}$ such that $\gamma_1^{(k)} > 0$ for all $i \leq k \leq j$, then the sequence of errors ε_k is strictly decreasing for all $j \leq k \leq i$.
2. if there exist $k \in \mathbb{N}$ such that $\gamma_1^{(k)} \geq 0$ for all $i \leq k \leq j$, then the sequence of errors ε_k is nonstrictly decreasing for all $j \leq k \leq i$.
3. if there exist some $i \in \mathbb{N}$ such that $\gamma_1^{(i)} = 0$, and $\gamma_t^{(i)} \geq \gamma_{t+1}^{(i)}$ then the sequence of errors ε_k is constant for all $i \leq k < d-1$.

Proof. 1. Suppose there exist $k \in \mathbb{N}$ such that $\gamma_1^{(k)} > 0$ for all $i \leq k \leq j$. That means proposition is true. Suppose $u, v \in \mathbb{N}$ are distinct natural numbers such that $i \leq v \leq k \leq u \leq j$ and notice that since $n^{(k)} > 0$ and $\gamma_1^{(k)} > 0$ is true by hypothesis, then the sum of such terms must be positive, and therefore

$$\begin{aligned} 0 < \sum_{k=v}^{u-1} \frac{\gamma_1^{(k)}}{n^{(k)}} &= \frac{\gamma_1^{(v)}}{n^{(v)}} + \frac{\gamma_1^{(v+1)}}{n^{(v+1)}} + \cdots + \frac{\gamma_1^{(u-1)}}{n^{(u-1)}} \stackrel{(6)}{=} (\varepsilon_v - \varepsilon_{v+1}) + (\varepsilon_{v+1} - \varepsilon_{v+2}) + \cdots + (\varepsilon_{u-1} - \varepsilon_u) \\ &= \varepsilon_v - \varepsilon_u, \end{aligned}$$

since every term of the sum cancel except from the first and the last. Now, by adding ε_u to each side of the inequality, the first part is proved.

2. To obtain the second part of the Corollary, replace the strict inequalities with inequalities.

3. To see the last point, suppose there exist some $i \in \mathbb{N}$ such that $\gamma_1^{(i)} = 0$ and $\gamma_t^{(i)} \geq \gamma_{t+1}^{(i)}$. Since proposition is true,

$$\varepsilon_i - \varepsilon_{i+1} = \frac{\gamma_1^{(i)}}{n^{(i)}} = 0$$

We use induction to see that $\varepsilon_{i+b} - \varepsilon_{i+b+1} = 0$ for $b > 0$:

$$\begin{aligned} \varepsilon_{i+1} - \varepsilon_{i+2} &= \frac{\gamma_1^{(i+1)}}{n^{(i+1)}} = \frac{1}{n^{(i+1)}} \frac{1}{4} \left(\gamma_1^{(i)} + 2\gamma_2^{(i)} + \gamma_3^{(i)} \right) \\ &\leq \frac{1}{n^{(i+1)}} \frac{1}{4} \left(\gamma_1^{(i)} + 2\gamma_1^{(i)} + 2\gamma_1^{(i)} \right) = \frac{1}{n^{(i+1)}} \frac{1}{4} (0 + 0 + 0) = 0. \end{aligned}$$

■

COROLLARY. *If the blocked variables become independent, for some number of blocking transformation, then the sequence of errors ε_k become constant.*

Proof. Use the previous corollary, and notice that $\gamma_1^{(k)} = 0$ if X_i, X_{i+1} are independent for all $0 \leq i, j \leq 2^{-k}m$. ■

PROPOSITION. *Suppose X_1, X_2, \dots, X_n are weakly stationary observations, then subject to i blocking transformations,*

$$\left\langle \hat{\gamma}_1^{(i)} - \gamma_1^{(i)} \right\rangle = \text{Var}(\bar{X}_n) \left(1 - \frac{2n^{(i)}}{n^{(i)} - 1} \right) + \frac{1}{n^{(i)}} \frac{2}{n^{(i)} - 1} \sum_{t=0}^{n^{(i)}-1} \gamma_t^{(i)}, \quad \text{for all } i \in \{0, 1, \dots, d-1\}$$

Suppose $X_1^i, X_2^i, \dots, X_n^i$ is uncorrelated, then

$$\left\langle \hat{\gamma}_1^{(i)} \right\rangle = -\text{Var}(\bar{X}_n) \frac{n^i + 1}{n^i - 1} + \frac{2}{n^{(i)}(n^{(i)} - 1)} \gamma_0^{(i)}$$

PROPOSITION 2.

$$\begin{aligned} 4 \sum_{t=1}^{n^{(i)}-1} t \gamma_t^{(i)} &= \sum_{t=1}^{n^{(i-1)}-1} t \gamma_t^{(i-1)} - n_i \gamma_{n_{i-1}-1}^{(i-1)} \\ 2 \sum_{t=1}^{n^{(i)}-1} \gamma_t^{(i)} &= \sum_{t=1}^{n^{(i-1)}-1} \gamma_t^{(i-1)} - \frac{1}{2} \left(\gamma_1^{(i-1)} + \gamma_{n_{i-1}-1}^{(i-1)} \right) \end{aligned}$$

LEMMA 1. Suppose j and k are positive natural number and the sample size $m \geq 2^k(j-1) + 2^{k+1} - 1$, then $k < \log_2 m$ and

$$\gamma_j^{(k)} = 2^{-2k} \left[\gamma_{2^k(j-1)+1}^{(0)} + 2\gamma_{2^k(j-1)+2}^{(0)} + 3\gamma_{2^k(j-1)+3}^{(0)} + \cdots + 2^k \gamma_{2^k(j-1)+2^k}^{(0)} \right. \\ \left. + (2^k - 1)\gamma_{2^k(j-1)+2^k+1}^{(0)} + (2^k - 2)\gamma_{2^k(j-1)+2^k+2}^{(0)} + \cdots + \gamma_{2^k(j-1)+2^{k+1}-1}^{(0)} \right]. \quad (13)$$

Proof. We first show that $k < \log_2 m$. Fix j and k such that $m \geq 2^k(j-1) + 2^{k+1} - 1$, then

$$m \geq \underbrace{2^k(j-1)}_{\geq 0} + \underbrace{2^{k+1} - 1}_{\geq 2^k+1} \geq 2^k+1 \quad \text{only if} \quad \log_2 m \geq \log_2(2^k+1) > \log_2 2^k = k \log_2 2 = k$$

We prove the rest of the lemma by induction. Fix j such that $m \geq 2^1(j-1) + 2^{1+1} - 1$, in particular this ensures that if $k = 1$, then $m \geq 2j + 1$ and therefore $\gamma_{2j+1}^{(0)}$ exists. Define $M = \sup_{k \in \mathbb{N}} \{m \geq 2^k(j-1) + 2^{k+1} - 1\}$. Assume $k = 1$ and write

$$\gamma_j^{(1)} \stackrel{(15)}{=} 2^{-2} \left(\gamma_{2j-1}^{(0)} + \gamma_{2j}^{(0)} + \gamma_{2j+1}^{(0)} \right)$$

Assume now that there exist a positive natural number $k < M$ such that equation (13) is true. This implies $k+1 \leq M$ and hence $\gamma_{2^k 2j+2^k+2^k-1}^{(0)}$ exists, and we can write

$$\gamma_j^{(k+1)} \stackrel{(15)}{=} 2^{-2} (\gamma_{k,2j-1} + 2\gamma_{k,2j} + \gamma_{k,2j+1}) \stackrel{(13)}{=} \\ 2^{-2} 2^{-2k} \left(\gamma_{0,2^k(2j-2)+1} + 2\gamma_{0,2^k(2j-2)+2} + \cdots + 2^k \gamma_{0,2^k(2j-2)+2^k} \right. \\ \left. + (2^k - 1)\gamma_{0,2^k(2j-2)+2^k+1} + \cdots + \gamma_{0,2^k(2j-2)+2^k+2^k-1} \right) + \\ 2^{-2} 2^{-2k} \left(2\gamma_{0,2^k(2j-1)+1} + 4\gamma_{0,2^k(2j-1)+2} + \cdots + 2^k \gamma_{0,2^k(2j-1)+2^k} \right. \\ \left. + 2(2^k - 1)\gamma_{0,2^k(2j-1)+2^k+1} + \cdots + 2\gamma_{0,2^k(2j-1)+2^k+2^k-1} \right) + \\ 2^{-2} 2^{-2k} \left(\gamma_{0,2^k 2j+1} + 2\gamma_{0,2^k 2j+2} + \cdots + 2^k \gamma_{0,2^k 2j+2^k} \right. \\ \left. + (2^k - 1)\gamma_{0,2^k 2j+2^k+1} + \cdots + \gamma_{0,2^k 2j+2^k+2^k-1} \right).$$

By using that $2^{-2} 2^{-2k} = 2^{-2(k+1)}$ and factoring $\gamma_K^{(0)}$ together for all $2^{k+1}(j-1) + 1 \leq K \leq 2^{k+1}(j-1) + 2^{k+2} - 1$, the lemma follows. ■

The next proposition is useful because combined with proposition , it shows how the convergence of the method is uniquely determined by the initial correlation structure.

PROPOSITION 3. Suppose $(\gamma_1^{(k)}, \gamma_2^{(k)}, \dots, \gamma_{n_k}^{(k)})$ denote the correlation structure at blocking iteration number k . Suppose $m \geq 2^{k+1} - 1$, then $k < \log_2 m$ and,

$$2^{2k} \gamma_1^{(k)} = \gamma_1^{(0)} + 2\gamma_2^{(0)} + 3\gamma_3^{(0)} + \cdots + 2^k \gamma_{2^k}^{(0)} + (2^k - 1)\gamma_{2^k+1}^{(0)} + \cdots + \gamma_{2^{k+1}-1}^{(0)}$$

Proof. Use the previous lemma with $j = 1$. ■

Forklar ideen i neste bevis. Vis noen ligninger

LEMMA 2. Bla bla

$$\langle \hat{\sigma}^2 \rangle = \text{Var}(X_i) - \text{Var}(\bar{X}_n)$$

Proof. To come ■

Introduser ideen ved å vise hvordan vi trekker fra hverandre ligningene.

THEOREM 3 (Revised blocking method). *Suppose X_1, \dots, X_n are $n = 2^d$ weakly stationary random variables. Let $\bar{X}_n = (1/n) \sum_{j=1}^n X_j$ denotes their average, $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ and $\gamma_1^{(j)}$ be a the autocovariance of subsequent random variable subject to j blocking transformations, then*

$$\text{Var}(\bar{X}_n) = \frac{1}{n-1} \left(\langle \hat{\sigma}^2 \rangle + \sum_{i=0}^{d-1} 2^i \gamma_1^{(i)} \right).$$

Proof to come. Did not have time to finish it before the exams. Furthermore, the proof appear to be more difficult than first anticipated. If there is interest in it, I can try to work out the details after the exams finish.

This theorem says that after less than $\log_2 m$ blocking iterations, the difference between the variance of the mean and variance of the blocked variables, is essentially zero. This is the case since m is typically many orders of magnitude larger than $\text{Var}(X_i^{(k)})$.

ESTIMATION

Bla bla Forklar hvorfor man gjør det. Vis at den er effektiv

CONCLUSION AND PERSPECTIVES

Bla bla

APPENDIX

PROPOSITION. *Suppose $d > 1$ is an integer and X_1, \dots, X_m are $m = 2^d$ random variables from a weakly stationary sample. If we let $\bar{X} = (1/m) \sum_{i=1}^m X_i$ denote the mean, and for $t = |i - j|$ we let $\gamma_t^{(k)} = \gamma_{i,j}^{(k)}$ be the autocovariance subject to k blocking-transformations. Then the block transformed variables $X_1^{(k)}, \dots, X_{2^{-k}m}^{(k)}$ are weakly stationary,*

$$\text{Var}(\bar{X}_m) - \frac{1}{n^{(k)}} \text{Var}(X_i^{(k)}) = \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}} \right) \gamma_t^{(k)} \quad (14)$$

and

$$\gamma_j^{(k)} = \begin{cases} \frac{1}{2} \gamma_{2j}^{(k-1)} + \frac{1}{2} \gamma_{2j+1}^{(k-1)} & \text{if } j = 0 \\ \frac{1}{4} \gamma_{2j-1}^{(k-1)} + \frac{1}{2} \gamma_{2j}^{(k-1)} + \frac{1}{4} \gamma_{2j+1}^{(k-1)} & \text{if } j > 0 \end{cases} \quad (15)$$

Proof. We first show that $X_1^{(k)}, \dots, X_{2^{-k}m}^{(k)}$ are weakly stationary. We prove each of criteria by induction. By hypothesis, for all $i \in \{1, \dots, m\}$ there exist some $\mu \in \mathbb{R}$ such that $X_i = \mu$ since the sample is weakly stationary. Therefore, for all $i \in \{1, 2, \dots, 2^0 m\}$, $\langle X_i^{(0)} \rangle = \mu$. The induction step follows by

$$\langle X_i^{(k+1)} \rangle \stackrel{(2)}{=} \frac{1}{2} \langle X_{2i-1}^{(k)} + X_{2i}^{(k)} \rangle = \mu$$

For the second criterion, notice that $\gamma_{i,j}^{(0)} = \gamma_{|i-j|}^{(0)}$ since the sample is weakly stationary. To see that for all $n \in \mathbb{N}$ the autocovariance $\gamma_{i,j}$ only depends on i, j through the difference $|i - j|$, perform the induction step by

$$\begin{aligned}\gamma_{u,v}^{(k+1)} &= \text{Cov}\left(X_u^{(k+1)}, X_v^{(k+1)}\right) \stackrel{(2)}{=} \text{Cov}\left(\frac{1}{2}(X_{2u-1}^{(k)} + X_{2u}^{(k)}), \frac{1}{2}(X_{2v-1}^{(k)} + X_{2v}^{(k)})\right) \\ &= \frac{1}{4} \left(\text{Cov}\left(X_{2u}^{(k)}, X_{2v}^{(k)}\right) + \text{Cov}\left(X_{2u-1}^{(k)}, X_{2v-1}^{(k)}\right) + \text{Cov}\left(X_{2u-1}^{(k)}, X_{2v}^{(k)}\right) + \text{Cov}\left(X_{2u}^{(k)}, X_{2v-1}^{(k)}\right) \right) \\ &= \begin{cases} \gamma_0^{(k)}/2 + \gamma_1^{(k)}/2 & \text{if } |u - v| = 0 \\ \gamma_{2|u-v|-1}^{(k)}/4 + \gamma_{2|u-v|}^{(k)}/2 + \gamma_{2|u-v|+1}^{(k)}/4 & \text{if } |u - v| > 0. \end{cases}\end{aligned}$$

This formula proves equation (15) proves that the sample $X_i^{(k)}$ is weakly stationary for all $k \in \{0, 1, \dots, d-1\}$. It remains to prove the formula (14). Suppose we let $\bar{X}_{n^{(k)}}^{(k)}$ denote the mean subject to k blocking transformations, then

$$\begin{aligned}\text{Var}\left(\bar{X}_{n^{(k)}}^{(k)}\right) &\stackrel{(1)}{=} \left\langle \bar{X}_{n^{(k)}}^{(k)2} \right\rangle - \left\langle \bar{X}_{n^{(k)}}^{(k)} \right\rangle^2 \\ &= \frac{1}{(n^{(k)})^2} \left\langle \left(\sum_{i=1}^{n^{(k)}} X_i^{(k)} \right) \left(\sum_{j=1}^{n^{(k)}} X_j^{(k)} \right) \right\rangle - \frac{1}{(n^{(k)})^2} \left\langle \sum_{i=1}^{n^{(k)}} X_i^{(k)} \right\rangle \left\langle \sum_{j=1}^{n^{(k)}} X_j^{(k)} \right\rangle \\ &= \frac{1}{(n^{(k)})^2} \sum_{i=1}^{n^{(k)}} \left[\left\langle X_i^{(k)} X_j^{(k)} \right\rangle - \left\langle X_i^{(k)} \right\rangle \left\langle X_j^{(k)} \right\rangle \right] = \frac{1}{(n^{(k)})^2} \sum_{i=1}^{n^{(k)}} \sum_{j=1}^{n^{(k)}} \gamma_{i,j}^{(k)}\end{aligned}\quad (16)$$

Recall that the sample subject to k blocking transformations is weakly stationary, that means that the autocovariance only depends on its indices i, j though the difference $|i - j| \equiv t$. And since $|i - j| = |j - i|$ it follows that the covariance matrix $\Sigma^{(k)}$ satisfies

$$\Sigma^{(k)} = \begin{bmatrix} \gamma_{1,1}^{(k)} & \gamma_{1,2}^{(k)} & \cdots & \gamma_{1,n^{(k)}}^{(k)} \\ \gamma_{2,1}^{(k)} & \gamma_{2,2}^{(k)} & \cdots & \\ \vdots & & \ddots & \\ \gamma_{n^{(k)},1}^{(k)} & \gamma_{n^{(k)},2}^{(k)} & \cdots & \gamma_{n^{(k)},n^{(k)}}^{(k)} \end{bmatrix} = \begin{bmatrix} \gamma_0^{(k)} & \gamma_1^{(k)} & \cdots & \gamma_{n^{(k)}-1}^{(k)} \\ \gamma_1^{(k)} & \gamma_0^{(k)} & \cdots & \\ \vdots & & \ddots & \\ \gamma_{n^{(k)}-1}^{(k)} & \gamma_{n^{(k)}-2}^{(k)} & \cdots & \gamma_0^{(k)} \end{bmatrix}\quad (17)$$

We want to evaluate the double sum of equation (16), which is the sum of all the terms in $\Sigma^{(k)}$, but since by equation (17), $\Sigma^{(k)}$ is n -diagonal, we can rewrite the double sum as a single sum over each of its diagonal like so:

$$\begin{aligned}\text{Var}\left(\bar{X}_{n^{(k)}}^{(k)}\right) &\stackrel{(16)}{=} \frac{1}{(n^{(k)})^2} \sum_{i=1}^{n^{(k)}} \sum_{j=1}^{n^{(k)}} \gamma_{i,j}^{(k)} \stackrel{(17)}{=} \frac{1}{(n^{(k)})^2} \sum_{t=0}^{n^{(k)}-1} (n^{(k)} - t) \gamma_t^{(k)} \\ &= \frac{\gamma_0^{(k)}}{n^{(k)}} + \frac{2}{n^{(k)}} \sum_{t=1}^{n^{(k)}-1} \left(1 - \frac{t}{n^{(k)}}\right) \gamma_t^{(k)}\end{aligned}$$

That means the only thing which remain is to show is $\text{Var}\left(\bar{X}_{n^{(k)}}^{(k)}\right) = \text{Var}\left(\bar{X}_m\right)$, which follows by induction using

$$\begin{aligned}\bar{X}_{n^{(k+1)}}^{(k+1)} &= \frac{1}{n^{(k+1)}} \sum_{i=1}^{n^{(k+1)}} X_i^{(k+1)} = \frac{1}{n^{(k)}/2} \sum_{i=1}^{n^{(k)}/2} \frac{1}{2} (X_{2i-1}^{(k)} + X_{2i}^{(k)}) \\ &= \left(X_1^{(k)} + X_3^{(k)} + \cdots X_{n^{(k)}-1}^{(k)} + X_2^{(k)} + X_4^{(k)} + \cdots X_{n^{(k)}}^{(k)} \right) = \bar{X}_{n^{(k)}}^{(k)}.\end{aligned}$$

This proves formula (14), and the proposition follows. ■