Draft version December 22, 2022
Typeset using LaTeX **twocolumn** style in AASTeX631

arXiv:2212.10684v1 [astro-ph.IM] 20 Dec 2022

Measures of Variance on Windowed Gaussian Processes

Daeyoung Lee[1] and Charles F. Gammie[1, 2, 3, 4]

[1] *Department of Physics, University of Illinois, 1110 West Green Street, Urbana, IL 61801, USA*
[2] *Department of Astronomy, University of Illinois at Urbana-Champaign, 1002 West Green Street, Urbana, IL 61801, USA*
[3] *NCSA, University of Illinois, 1205 W. Clark St., Urbana, IL 61801, USA*
[4] *Illinois Center for the Advanced Study of the Universe, University of Illinois, 1110 West Green St., Urbana, IL 61801, USA*

ABSTRACT

The variance and fractional variance on a fixed time window (variously known as "rms percent" or "modulation index") are commonly used to characterize the variability of astronomical sources. We summarize properties of this statistic for a Gaussian process.

1. INTRODUCTION

A recent study of millimeter variability of Sgr A* (Wielgus et al. 2022) makes use of the so-called modulation index M on a fixed time window. This statistic, which simply calculates the ratio between the standard deviation and the mean of the light curve on the window, has been used in various contexts (cf. Vaughan et al. (2003), where it is called $F_{\rm var}$, the "fractional rms variability amplitude"). The Event Horizon Telescope used this statistic to compare observed light curves with simulations (Event Horizon Telescope Collaboration et al. 2022), but in the course of that work we were unable to find a clear summary of the effects of the length of the window on M for Gaussian processes. The relevant quantities are easy to derive; in this note we collect them in one place using consistent notation.

In this brief note μ_T and σ_T^2 are the mean and variance (for example of a light curve) measured in a window of finite duration T. Assuming the underlying time series is well modeled as a Gaussian process, we answer the following questions: (1) how is the covariance related to the expected σ_T^2? (2) given the covariance, what is the expected variance in σ_T^2 for a single measurement? (3) how correlated are successive samples of σ_T^2 and, in particular, can they be treated as independent measurements? (4) what is the relationship of the expected σ_T^2 to the structure function? (5) what are the answers to these general questions for the important special case of a damped random walk? (6) what are the effects of sparse and nonuniform sampling? (7) what are the implications for the modulation index? In this discussion we neglect measurement errors. We also assume the observation duration is small compared to the window duration.

2. MEASURED VARIANCE

Let $f(t)$ be a stationary Gaussian random process with mean $\langle f(t)\rangle = \mu$ and unconditional variance $\langle f^2(t)\rangle - \mu^2 = \sigma^2$. Since the process is stationary, we may set the covariance function to be $\langle f(t_1)f(t_2)\rangle - \langle f(t_1)\rangle\langle f(t_2)\rangle \equiv \mathrm{Cov}(t_1,t_2) = \mathrm{Cov}(|t_1-t_2|)$. Here $\langle\,\cdot\,\rangle$ denotes the expectation value.

When observed on a window $[0,T]$, the measured mean and variance are

$$\mu_T \equiv \frac{1}{T}\int_0^T f(t)\,dt, \tag{1}$$

$$\sigma_T^2 \equiv \frac{1}{T}\int_0^T \left(f(t)-\mu_T\right)^2\,dt. \tag{2}$$

Notice that while $\langle\mu_T\rangle = \mu$, $\mu_T \neq \mu$ but is rather a random variable that depends on $f(t)$. Let

$$\rho(t_1,t_2) \equiv \frac{1}{\sigma^2}\mathrm{Cov}(t_1,t_2) \tag{3}$$

be the correlation function of f, and define

$$c(T) \equiv \frac{1}{T^2}\int_0^T\int_0^T \rho(t_1,t_2)\,dt_2\,dt_1. \tag{4}$$

Then

$$\langle\mu_T^2\rangle = \mu^2 + \sigma^2 c(T), \tag{5}$$

$$\langle\sigma_T^2\rangle = \sigma^2(1-c(T)). \tag{6}$$

What is the variance of σ_T^2 for a single measurement? It is helpful to define

$$c_1(T) \equiv \frac{1}{T^2}\iint \rho(t_1,t_2)^2\,dt_2\,dt_1, \tag{7}$$

$$c_2(T) \equiv \frac{1}{T^3}\iiint \rho(t_1,t_2)\rho(t_2,t_3)\,dt_3\,dt_2\,dt_1, \tag{8}$$

$$\begin{aligned} c_3(T) &\equiv \frac{1}{T^4}\left[\iint \rho(t_1,t_2)\,dt_2\,dt_1\right]^2 \\ &= [c(T)]^2, \end{aligned} \tag{9}$$

where all integrals are evaluated from 0 to T. Then the variance of σ_T^2 is

$$\mathrm{Var}(\sigma_T^2) = 2\sigma^4\left[c_1(T) - 2c_2(T) + c_3(T)\right]. \tag{10}$$

As a result of the correlation between points in a time series, the behavior of the measured mean and variance are dependent on both the covariance function and the length of the observation window. Assuming $c(T) \to 0$ as $T \to \infty$, $\langle \sigma_T^2 \rangle$ is smaller than the unconditional variance σ^2 for small T, and approaches it as T increases.

3. CORRELATION BETWEEN DIFFERENT WINDOWS

How independent are measurements of σ_T^2 over successive windows? Consider a second window $[\Delta T, T + \Delta T]$, with

$$\mu_T' \equiv \frac{1}{T} \int_{\Delta T}^{T+\Delta T} f(t)\, dt, \tag{11}$$

$$\sigma_T'^2 \equiv \frac{1}{T} \int_{\Delta T}^{T+\Delta T} \left(f(t) - \mu_T\right)^2\, dt. \tag{12}$$

Since f is stationary and thus $\rho(t_1, t_2) = \rho(|t_2 - t_1|)$, define

$$c'(T) \equiv \frac{1}{T^2} \int_0^T \int_{\Delta T}^{T+\Delta T} \rho(t_1, t_2)\, dt_2\, dt_1, \tag{13}$$

$$c_1'(T) \equiv \frac{1}{T^2} \int_0^T \int_{\Delta T}^{T+\Delta T} \rho(t_1, t_2)^2\, dt_2\, dt_1, \tag{14}$$

$$\begin{aligned} c_2'(T) &\equiv \frac{1}{T^3} \int_0^T \int_0^T \int_{\Delta T}^{T+\Delta T} \rho(t_1, t_3)\rho(t_2, t_3)\, dt_3\, dt_2\, dt_1 \\ &= \frac{1}{T^3} \int_{\Delta T}^{T+\Delta T} \int_{\Delta T}^{T+\Delta T} \int_0^T \rho(t_1, t_3)\rho(t_2, t_3)\, dt_3\, dt_2\, dt_1, \end{aligned} \tag{15}$$

$$\begin{aligned} c_3'(T) &\equiv \frac{1}{T^4} \left[\int_0^T \int_{\Delta T}^{T+\Delta T} \rho(t_1, t_2)\, dt_2\, dt_1 \right]^2 \\ &= [c'(T)]^2. \end{aligned} \tag{16}$$

Then

$$\mathrm{Cov}(\sigma_T^2, \sigma_T'^2) = 2\sigma^4 \left[c_1'(T) - 2c_2'(T) + c_3'(T)\right]. \tag{17}$$

Evidently the measured variance on windows with a small time separation are correlated. The degree of correlation depends on the separation between windows, the length of the windows, and the shape of the covariance function.

4. RELATIONSHIP BETWEEN σ_T^2ND THE STRUCTURE FUNCTION

The structure function at a time lag Δt, here defined as

$$SF(\Delta t) = \frac{1}{T - \Delta t} \int_0^{T-\Delta t} \left[f(t + \Delta t) - f(t)\right]^2\, dt \tag{18}$$

is closely related to σ_T^2, since

$$\langle SF(\Delta t) \rangle = 2\sigma^2 - 2\mathrm{Cov}(\Delta t) \tag{19}$$

and thus

$$\langle \sigma_T^2 \rangle = \frac{1}{T^2} \int_0^T \int_0^T \frac{\langle SF(|t_1 - t_2|) \rangle}{2}\, dt_1\, dt_2. \tag{20}$$

Since $c(y)$ corresponds to the integral of the covariance function over $[0, T] \times [0, T]$, $\langle \sigma_T^2 \rangle$ corresponds to the integral of the structure function over $[0, T] \times [0, T]$.

5. DAMPED RANDOM WALK EXAMPLE

For the damped random walk

$$\mathrm{Cov}(t_1, t_2) = \sigma^2 e^{-|t_1 - t_2|/\tau}, \tag{21}$$

$$SF(\Delta t) = 2\sigma^2 \left(1 - e^{-\Delta t/\tau}\right). \tag{22}$$

Recall that $\Delta t > 0$ is a time lag and is not the same as ΔT, which labels differences between windows. For $y \equiv T/\tau$,

$$c(y) = 2\left(\frac{1}{y} - \frac{1}{y^2}(1 - e^{-y})\right), \tag{23}$$

$$c_1(y) = c(2y), \tag{24}$$

$$c_2(y) = \frac{1}{y^3}\left(2y(2 + e^{-y}) - (7 - e^{-y})(1 - e^{-y})\right), \tag{25}$$

$$c_3(y) = [c(y)]^2, \tag{26}$$

with

$$\langle \sigma_T^2 \rangle = \sigma^2(1 - c(y)), \tag{27}$$

$$\mathrm{Var}(\sigma_T^2) = 2\sigma^4 \left[c(2y) - 2c_2(y) + c^2(y)\right]. \tag{28}$$

For illustrative purposes we will look at the correlation between consecutive windows, i.e. taking $\Delta T = T$. Then,

$$c'(y) = \frac{1}{y^2}(1 - e^{-y})^2, \tag{29}$$

$$c_1'(y) = c'(2y), \tag{30}$$

$$c_2'(y) = \frac{1}{2y^3}(1 - e^{-y})^3(1 + e^{-y}), \tag{31}$$

$$c_3'(y) = [c(y)]^2, \tag{32}$$

with

$$\mathrm{Cov}(\sigma_T^2, \sigma_T'^2) = 2\sigma^4 \left[c'(2y) - 2c_2'(T) + c'^2(y)\right], \tag{33}$$

and the correlation between the two segments is

$$\mathrm{Corr}(\sigma_T^2, \sigma_T'^2) = \frac{\mathrm{Cov}(\sigma_T^2, \sigma_T'^2)}{\mathrm{Var}(\sigma_T^2)}. \tag{34}$$

As $y \to 0$,

$$\langle \sigma_T^2 \rangle \to \sigma^2 \left(\frac{y}{3} - \frac{y^2}{12} + \mathcal{O}(y^3) \right), \tag{35}$$

$$\mathrm{Var}(\sigma_T^2) \to \sigma^4 \left(\frac{4y^2}{45} - \frac{y^3}{15} + \mathcal{O}(y^4) \right), \tag{36}$$

$$\mathrm{Cov}(\sigma_T^2, \sigma_T'^2) \to \sigma^4 \left(\frac{y^4}{72} - \frac{y^5}{36} + \mathcal{O}(y^6) \right), \tag{37}$$

$$\mathrm{Corr}(\sigma_T^2, \sigma_T'^2) \to \frac{5y^2}{32} - \frac{25y^3}{128} + \mathcal{O}(y^4), \tag{38}$$

and as $\frac{1}{y} \to 0$,

$$\langle \sigma_T^2 \rangle \to \sigma^2 \left(1 - \frac{2}{y} + \mathcal{O}(y^{-2}) \right), \tag{39}$$

$$\mathrm{Var}(\sigma_T^2) \to \sigma^4 \left(\frac{2}{y} - \frac{9}{y^2} + \mathcal{O}(y^{-3}) \right), \tag{40}$$

$$\mathrm{Cov}(\sigma_T^2, \sigma_T'^2) \to \sigma^4 \left(\frac{1}{2y^2} - \frac{2}{y^3} + \mathcal{O}(y^{-4}) \right), \tag{41}$$

$$\mathrm{Corr}(\sigma_T^2, \sigma_T'^2) \to \frac{1}{4y} + \frac{1}{8y^2} + \mathcal{O}(y^{-3}). \tag{42}$$

All four quantities are plotted in Figure 1. Evidently the correlation between consecutive windows disappears as $T \to 0$ or $T \to \infty$, and peaks at $T \approx 2\tau$.

6. SPARSE AND NON-UNIFORM SAMPLING

In the previous section, σ_T^2 was evaluated as an integral. For real data, if the observation was uniformly and densely sampled (relative to both the characteristic timescale τ and the length of the observation T), σ_T^2 can be approximated as the variance of the observation. However, data is commonly sampled sparsely and nonuniformly.

If the observation is sampled by $f_i = f(t_i)$, $i = 1, \cdots, N$, then we would measure

$$\mu_T = \frac{1}{N} \sum_{i=1}^{N} f_i, \tag{43}$$

$$\sigma_T^2 = \frac{1}{N} \sum_{i=1}^{N} \left[f_i - \mu_T \right]^2. \tag{44}$$

Then

$$\langle \mu_T \rangle = \mu, \tag{45}$$

$$\langle \mu_T^2 \rangle = \mu^2 + \frac{\sigma^2}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \rho(t_i, t_j), \tag{46}$$

$$\langle \sigma_T^2 \rangle = \sigma^2 \left[1 - \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \rho(t_i, t_j) \right], \tag{47}$$

are the discrete analogues of equations 5 and 6. If the sampling was uniform, the sum would serve as an approximation to the integral in equation 4, but non-uniform sampling will bias our estimate. The nature of this bias depends on the details of the sampling.

One strategy for minimizing bias would be to weight each sample by the gaps Δt_i between observations. However, this can fail if there are large gaps in the data. For example, consider a densely sampled observation across the first and third hour of a three-hour observation, with an hour-long gap in the middle of the observation period. On the first and third segments of observation, we can assume that the measured variance of the segments well-approximates the integrated variance of the underlying process.

For this example, we can look at the effects of two different methods of addressing this gap in our observation: taking the unweighted variance across our samples and weighting each sample by the time gaps in observation before and after the sample.

For a given statistic $g(t)$ calculated from $f(t)$ (such as the point-wise variance, $g(t) = f(t)^2 - \mu_T^2$), the underlying "true" measured value from $t = 0h$ to $t = 3h$ is

$$G = \frac{1}{3} \int_0^3 g(t)\, dt. \tag{48}$$

The first method essentially assumes all observations are uniformly spaced across the entire observation, dilating the function $g(t)$:

$$G_1 = \frac{1}{3} \left[\int_0^{3/2} g(2t/3)\, dt + \int_{3/2}^{3} g(2t/3 + 1)\, dt \right] \tag{49}$$

$$= \frac{1}{2} \left[\int_0^1 g(t)\, dt + \int_2^3 g(t)\, dt \right]. \tag{50}$$

For a quantity like the variance, this gives a consistent underestimate of the true measured value across three hours. On the other hand, the second method instead places undue weight on the endpoints:

$$G_2 = \frac{1}{3} \left[\int_0^1 g(t)\, dt + \frac{1}{2} g(1) + \frac{1}{2} g(2) + \int_1^2 g(t)\, dt \right]. \tag{51}$$

The resulting value of G_2 is strongly dependent on the values $g(1)$ and $g(2)$ at the endpoints of the gap in the observation.

Thus, it is useful to Monte Carlo the analysis on Gaussian processes with the same sampling as the data, to see the effects of the sampling on the measurement of variability. In our applications the variance in the measurement tends to outweigh any effects of sampling.

7. MODULATION INDEX

The modulation index or rms percent is $M_T \equiv \sigma_T / \mu_T$, the standard deviation divided by the mean over a period of observation. For a Gaussian process, the measured variance and the measured mean are uncorrelated, since

$$\begin{aligned}
\left\langle \sigma_T^2 \mu_T \right\rangle &= \left\langle \frac{1}{T} \int_0^T \left[f^2(t) \mu_T - \mu_T^3 \right] \, dt \right\rangle \\
&= \frac{1}{T^2} \iint \left\langle f_1^2 f_2 \right\rangle dt_1 \, dt_2 \\
&\quad - \frac{1}{T^3} \iint \left\langle f_1 f_2 f_3 \right\rangle dt_1 \, dt_2 \, dt_3 \\
&= 0,
\end{aligned}$$

The standard deviation and mean are similarly uncorrelated. However, they are clearly not independent. Thus, rather than deriving an analytic expression for M_T, we will explore it computationally.

Figure 1 shows the mean and variance of M_T for a damped random walk, along with the covariance and correlation of measurements of M_T between consecutive windows. Similar statistics for σ_T^2 are also given. These quantities were averaged over 10^7 independent realizations, with damped random walk parameters $\sigma = 0.24$, $\mu = 2.4$, and $\tau = 1$. Each damped random walk was generated on 2048 uniformly spaced points between 0 and 20, and the statistics were measured as a function of the window length $T \in [0, 10]$, calculated on the first window (or first two, where applicable).

Qualitatively, σ_T^2 and M_T behave similarly. Of particular interest is the correlation between consecutive windows, which peaks on the order of τ but remains fairly low across all window lengths, never rising above 0.1. Thus, as long as σ_T^2 and M_T are measured on independent windows ($\Delta T \geq T$), the correlation between measurements remains small.

This work was supported by NSF grants AST 17-16327 (horizon), OISE 17-43747, and AST 20-34306, and by the Grainger College of Engineering's Willett Chair.

REFERENCES

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al., 2022, ApJL, 930, L16

Wielgus, M., et al., 2022, ApJL, 930, L19

Vaughan, S., Edelson, R., Warwick S., & Uttley, P., 2003, MNRAS, 345, 1271

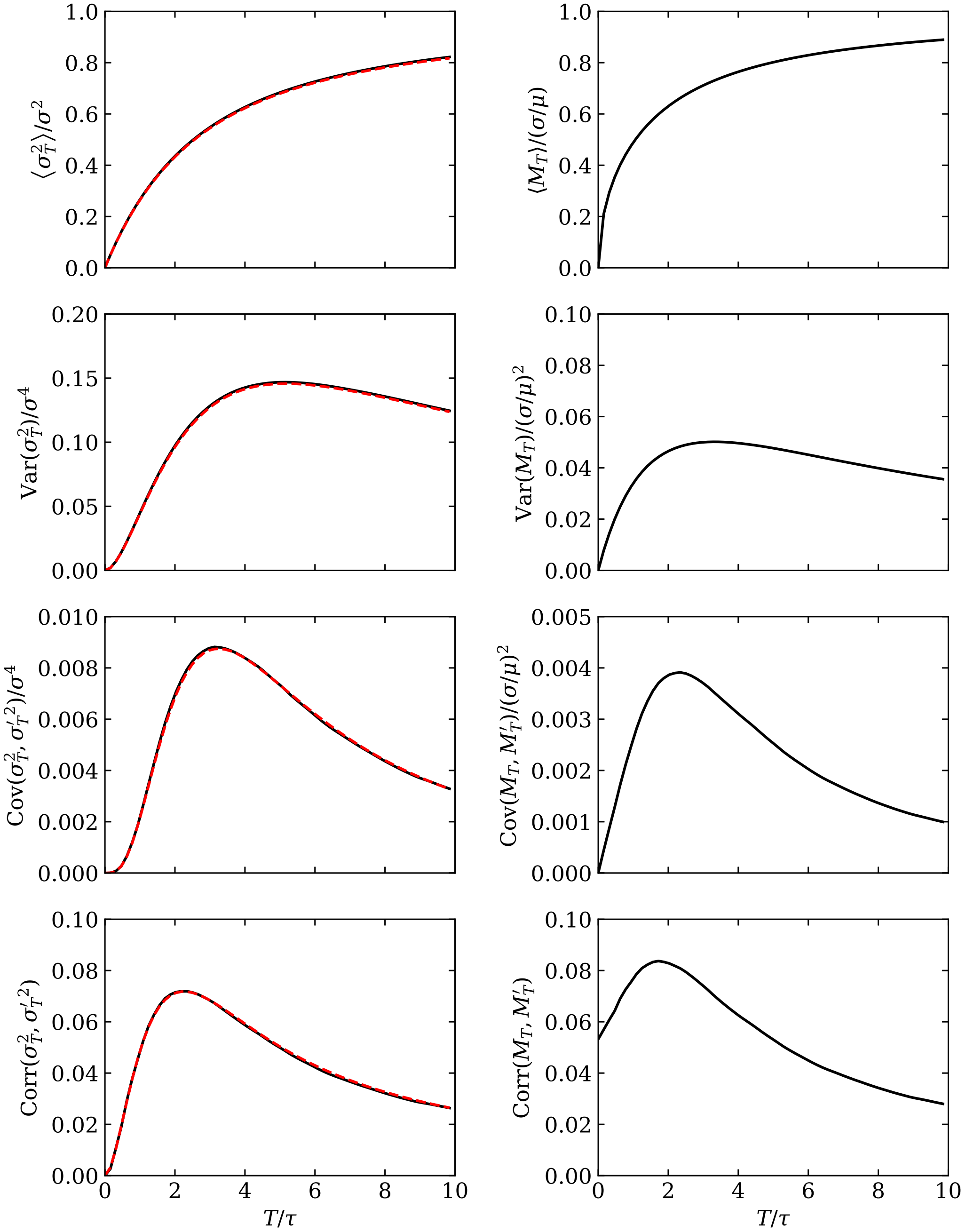

Figure 1. The mean and variance of σ_T^2 and M_T for a damped random walk, along with the covariance and correlation of these measurements on consecutive windows, as functions of the window length T/τ, averaged over 10^7 realizations with $\sigma = 0.24$ and $\mu = 2.4$. Analytic results, where available, are shown in red. The means are normalized by σ^2 and σ/μ, respectively, and the variance and covariance are normalized by σ^4 and $(\sigma/\mu)^2$.