Finding Optimal Window Size to Calculate Integrated Autocorrelation Time for Lattice QCD Ensembles

Ayush Singhal
National Institute of Science Education and Research
Bhubaneswar, Odisha 751005, India
4th year, Integrated M.Sc. Physics
Roll No.: 2011042

This report presents a comprehensive analysis of autocorrelation time and its pivotal role in time series analysis, focusing on lattice Quantum Chromodynamics (QCD). A significant highlight is the implementation of the automatic windowing algorithm, which optimizes the estimation of integrated autocorrelation time (τ_{int}) in Monte Carlo simulations. By intelligently selecting an appropriate window size, the algorithm enhances the precision of τ_{int} calculations, crucial for accurate statistical analyses and efficient algorithm performance in lattice QCD studies.

I. INTRODUCTION

Autocorrelation time (τ) measures the timescale over which a signal or process correlates with itself. It quantifies how quickly values in a time series become uncorrelated as the time lag between them increases. In this report, we discuss the definition, calculation methods, and applications of autocorrelation time in lattice QCD simulations.

A. What is Autocorrelation Time?

Autocorrelation time, often denoted as τ, quantifies the time it takes for a signal or a process to decorrelate with itself. Mathematically, it is defined as the inverse of the decay rate of the autocorrelation function. For a time series $x(t)$, the autocorrelation function $C(t)$ at lag t is given by:

$$C(t) = \frac{\langle x(t)x(0)\rangle - \langle x(t)\rangle\langle x(0)\rangle}{\text{Var}(x)} \tag{1}$$

where $\langle\cdot\rangle$ denotes averaging over time, and $\text{Var}(x)$ is the variance of the time series. The autocorrelation time τ_c is then given by:

$$C(\tau) = e^{-\frac{\tau}{\tau_c}} \tag{2}$$

B. Calculation of Autocorrelation Time

Methods for estimating autocorrelation time include the autocorrelation function and the integrated autocorrelation time. The autocorrelation function $C(t)$ can be calculated directly from a time series $x(t)$ using:

$$C(t) = \frac{1}{N-t}\sum_{n=1}^{N-t}(x(n)-\bar{x})(x(n+t)-\bar{x}) \tag{3}$$

where N is the total number of data points, t is the lag, and $\bar{x}$ is the mean of the time series. The autocorrelation time τ_c can then be determined by fitting an exponential decay curve to $C(t)$.

Integrated autocorrelation time, denoted as τ_{int}, involves summing the autocorrelation function to obtain a direct estimate of the autocorrelation time. It takes into account the entire autocorrelation function rather than just the first few lags. It is defined as:

$$\tau_{\text{int}} = \frac{1}{2} + \sum_{t=1}^{\infty}\frac{C(t)}{C(0)} \tag{4}$$

where $C(t)$ is the autocorrelation function at lag t, and $C(0)$ is the autocorrelation at zero lag. In practice, τ_{int} is often approximated by truncating the sum at a sufficiently large lag T.

To calculate τ_{int} in practice, one typically truncates the sum at a sufficiently large lag T beyond which $C(t)$ becomes negligible. Then, the sum becomes:

$$\tau_{\text{int}} = \frac{1}{2} + \sum_{t=1}^{T}\frac{C(t)}{C(0)} \tag{5}$$

This approximation provides a practical way to estimate the integrated autocorrelation time from the autocorrelation function up to a certain lag T.

The calculation of the integrated autocorrelation time involves summing the autocorrelation function over various lags, normalized by the variance of the time series. This measure provides a comprehensive assessment of the autocorrelation structure of the data and is valuable in determining the effective decorrelation time for statistical analyses and simulations.

II. SU(3) WILSON GAUGE THEORY

SU(3) Wilson gauge theory is a lattice formulation of Quantum Chromodynamics (QCD), describing interactions between quarks and gluons. In lattice QCD, spacetime is discretized onto a lattice, enabling numerical simulations.

A. Mathematical Equations

The Wilson action for SU(3) lattice gauge theory is given by:

$$S_W = \beta \sum_{\text{plaquettes}} \left(1 - \frac{1}{3} \mathrm{Re}\,\mathrm{Tr}(U_p)\right) \tag{6}$$

where β is the gauge coupling constant, and U_p represents the product of link variables around a plaquette.

In lattice QCD, the gauge fields are represented by link variables $U_\mu(x)$, which are SU(3) matrices associated with the links between lattice sites. These link variables are elements of the SU(3) group.

The Wilson loop operator is a gauge-invariant observable in lattice QCD, defined as the product of link variables around a closed loop in space-time. For a rectangular Wilson loop of size $T \times R$, the Wilson loop $W(T, R)$ is given by:

$$W(T, R) = \frac{1}{3} \mathrm{Tr} \left(\prod_{\text{loop}} U_\mu(x) \right) \tag{7}$$

where the product is taken over all links forming the loop.

Monte Carlo simulations generate gauge configurations according to the Boltzmann weight e^{-S_W}, allowing computation of observables such as correlation functions and hadron masses.

These equations form the basis for performing numerical simulations and analyzing results in SU(3) Wilson gauge theory within lattice QCD. They provide a framework for studying the non-perturbative aspects of QCD and understanding the properties of hadrons from first principles.

B. Autocorrelation Time in Lattice QCD

In the context of lattice QCD simulations, autocorrelation time represents the timescale over which successive configurations generated in a Markov chain Monte Carlo (MCMC) simulation become decorrelated. It quantifies the rate at which fluctuations in the lattice gauge fields decorrelate as the Markov chain evolves, affecting the statistical accuracy of observables computed from the configurations.

C. Significance of Autocorrelation Time in Lattice QCD

- Efficient Sampling: Understanding autocorrelation time is crucial for optimizing the performance of MCMC algorithms in lattice QCD simulations. It helps determine appropriate step sizes and decorrelation strategies to enhance the efficiency of sampling configurations.
- Algorithm Optimization: Knowledge of autocorrelation time guides the development and optimization of algorithms tailored for lattice QCD simulations, aiming to minimize computational costs and improve sampling efficiency.
- Error Estimation: Autocorrelation time influences the estimation of statistical errors in lattice QCD calculations. Longer autocorrelation times lead to larger error bars, affecting the precision of computed observables.

III. IMPORTANCE OF LAG IN CALCULATING INTEGRATED AUTOCORRELATION TIME

In the estimation of integrated autocorrelation time (τ_{int}), the choice of lag plays a critical role in determining the accuracy and reliability of the calculation. The lag parameter determines the range of temporal separations over which correlations between successive measurements are examined. Here, we discuss the importance of lag in calculating τ_{int} and its implications for statistical analyses.

A. Balance between Bias and Variance

- **Small Lags:** Using small lags may lead to a biased estimation of τ_{int} as it neglects long-range correlations. Additionally, the autocorrelation function may not have sufficiently decayed at small lags, resulting in an underestimation of τ_{int}.
- **Large Lags:** On the other hand, employing large lags could increase the variance of the estimation. Large lags may involve fewer data points for averaging, leading to greater statistical fluctuations in the computed autocorrelation values.

B. Efficiency of Sampling

- Selecting an appropriate lag allows for efficient sampling of the autocorrelation function. It aims to capture both short-range and long-range correlations adequately, ensuring a robust estimation of τ_{int} without excessive computational costs.
- The lag should be chosen such that it strikes a balance between exploring the autocorrelation structure effectively and minimizing the computational effort required to compute it.

C. Convergence of Autocorrelation

- The autocorrelation function $C(t)$ typically exhibits a decay with increasing lag t. It is crucial to verify that the autocorrelation has sufficiently decayed to near-zero values before truncating the sum used to calculate τ_{int}.

- An inadequate choice of lag may result in an inaccurate assessment of autocorrelation decay, leading to an erroneous determination of τ_{int} and potentially biased statistical analyses.

D. Trade-off with Computational Resources

- The computational resources required for estimating τ_{int} increase with the number of lags considered. Therefore, choosing an optimal lag strikes a balance between the accuracy of the estimation and the computational cost involved.
- Adjusting the lag parameter enables researchers to tailor the analysis to the available computational resources while ensuring reliable results.

IV. AUTOMATIC WINDOWING ALGORITHM

The automatic windowing algorithm as described in [1] and [2]; is designed to select an optimal window size, W or lag, for error estimation in Monte Carlo simulations. It is based on the hypothesis that the autocorrelation time, τ, is approximately a certain factor, S, times the integrated autocorrelation time, τ_{int}. The algorithm automatically chooses a window size, W, based on a hypothesis involving a factor, S, to optimize the estimation of autocorrelation functions. It involves solving a formula to determine an optimal W value, considering the statistical errors and the behavior of the autocorrelation function. The algorithm aims to find a self-consistently close to optimal summation window, ensuring that the chosen value exhibits a plateau in the statistical errors of the autocorrelation function. The default setting for the estimate, S_τ, is 1.5, and the algorithm tries to determine an optimal window size within certain constraints.

The algorithm involves the following steps:

1. **Initial Hypothesis**: Assume $\tau \approx S \cdot \tau_{\text{int}}$, with S typically set between 1 and 2.
2. **Estimator Calculation**: Compute the estimator $\bar{\tau}(W)$ using the formula:
$$\bar{\tau}(W) = \frac{1}{\ln(2\bar{\tau}_{\text{int}}(W) + 1)}$$
If $\bar{\tau}_{\text{int}} \leq \frac{1}{2}$, set $\bar{\tau}(W)$ to a small positive value.
3. **Window Selection**: For $W = 1, 2, \ldots$, calculate:
$$g(W) = \exp\left[-\frac{W}{\bar{\tau}(W)}\right] - \frac{\bar{\tau}(W)}{\sqrt{WN}}$$
The first value of W where $g(W)$ is negative is chosen as the window size.
4. **Plateau Verification**: Ensure that $\bar{\tau}_{\text{int}}(W)$ exhibits a plateau around the chosen W value. If not, adjust S accordingly.

This algorithm helps in determining an optimal summation window for the autocorrelation function, which is crucial for accurate error estimation in Monte Carlo studies. It is particularly useful when dealing with systems and observables where τ and τ_{int} are of the same order. The method is advantageous as it automatically adapts to the autocorrelation times involved and requires minimal manual intervention.

V. EXPERIMENTS AND RESULTS

A. Lattice configuration data

The lattice parameters were:

- Lattice Dimensions: $24 \times 24 \times 24 \times 4$
- Data was taken for ten values of lattice gauge coupling β varied non-linearly from $\beta = 5.6$ till $\beta = 5.9$.
- For each β, 1000 lattice configurations were generated.
- The observable used is the norm of the value of Polyakov loop, and autocorrelations were calculated on them for each value of lattice gauge coupling.

B. Data visualisation

A nice way to visualise the correlation in our generated samples can be done simply by plotting the observable value (norm in our case) for all the configurations for each lattice gauge coupling. In our case, the phase transition lies between $\beta = 5.68$ and $\beta = 5.692$. It is expected that the configurations should become more and more correlated as we move closer to the phase transition lattice gauge coupling and then start to get de-correlated as we move away from the same. We can see in the figures [1-10] how this correlation became prevalent as we changed our parameter (in our case, lattice gauge coupling); this provides us with useful insight into the data.

C. Results obtained from automatic windowing and their corresponding integrated τ values

- Figures [11-20] are plotted for integrated autocorrelation time vs window size used for each value of β.
- As discussed above, the optimal window size was calculated using an automatic windowing algorithm implemented in Python.

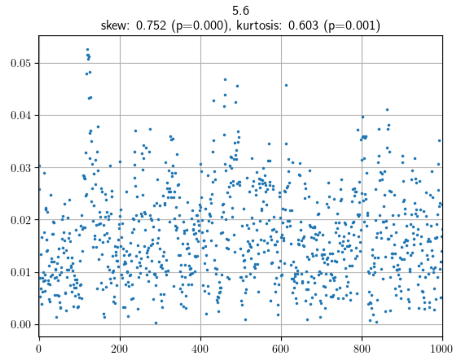

FIG. 1. for $\beta = 5.6$

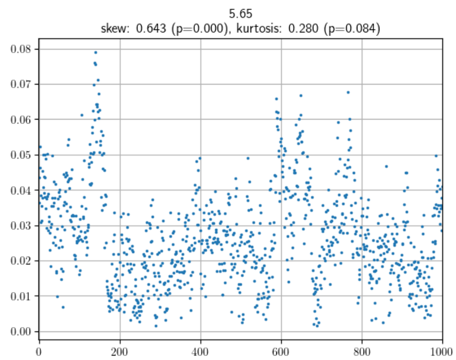

FIG. 2. for $\beta = 5.65$

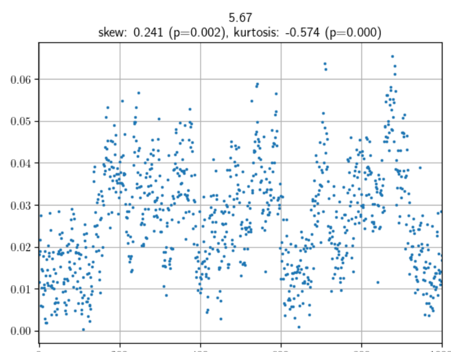

FIG. 3. for $\beta = 5.67$

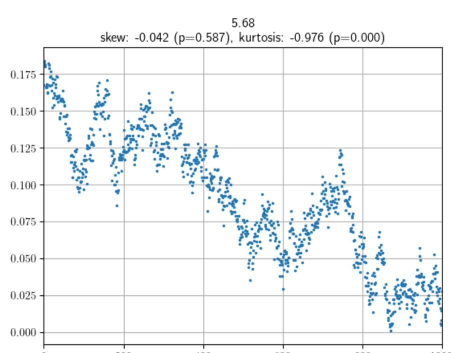

FIG. 4. for $\beta = 5.68$

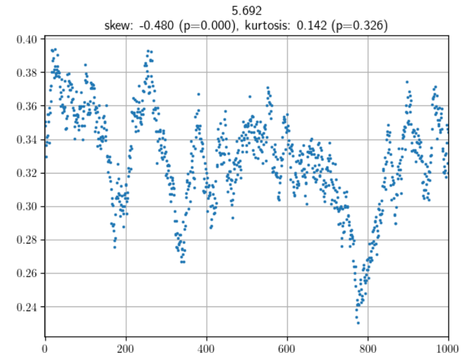

FIG. 5. for $\beta = 5.692$

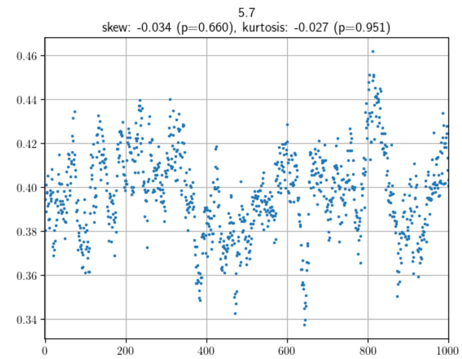

FIG. 6. for $\beta = 5.7$

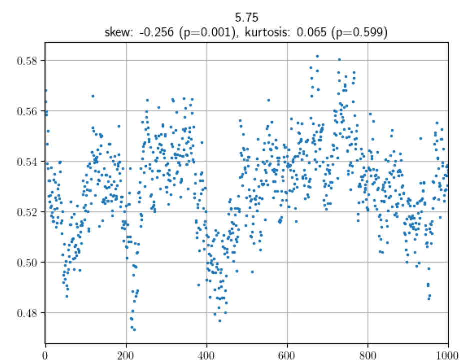

FIG. 7. for $\beta = 5.75$

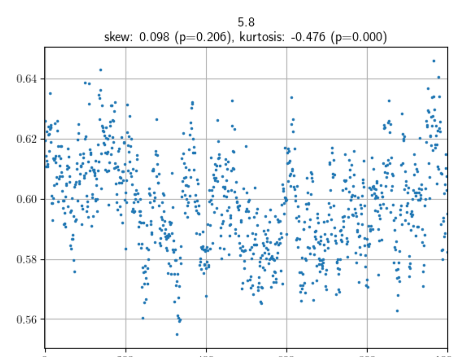

FIG. 8. for $\beta = 5.8$

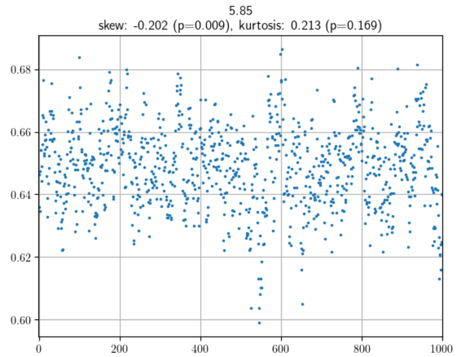

FIG. 9. for $\beta = 5.85$

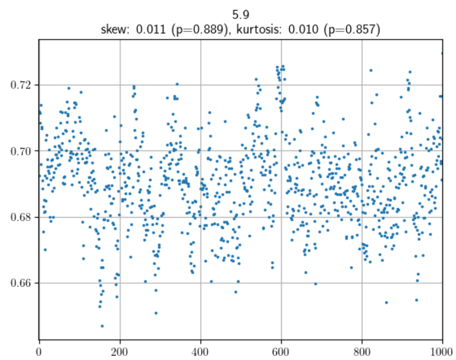

FIG. 10. for $\beta = 5.9$

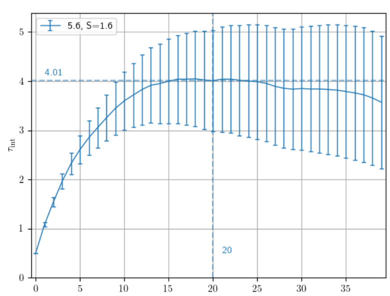

FIG. 11. τ_{int} vs W for $\beta = 5.6$

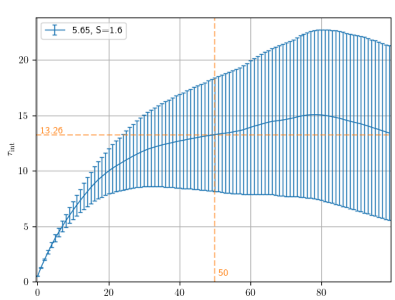

FIG. 12. τ_{int} vs W for $\beta = 5.65$

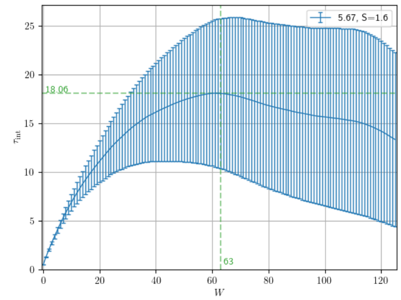

FIG. 13. τ_{int} vs W for $\beta = 5.67$

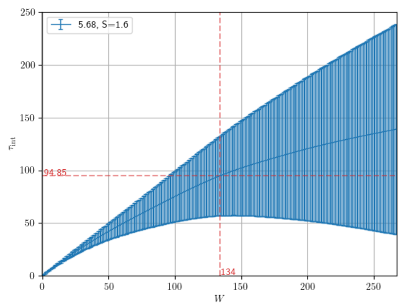

FIG. 14. τ_{int} vs W for $\beta = 5.68$

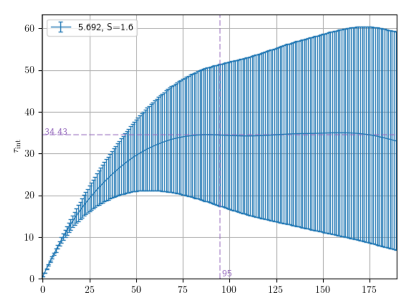

FIG. 15. τ_{int} vs W for $\beta = 5.692$

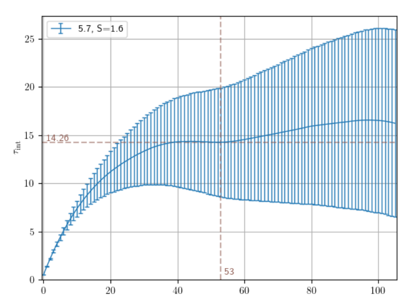

FIG. 16. τ_{int} vs W for $\beta = 5.7$

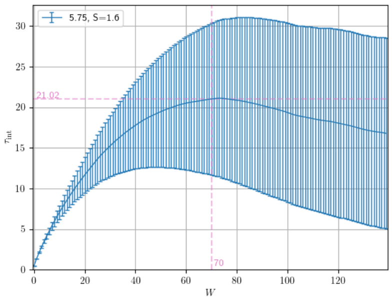

FIG. 17. τ_{int} vs W for $\beta = 5.75$

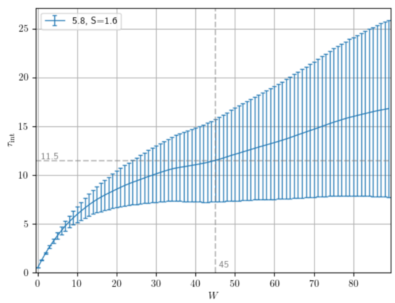

FIG. 18. τ_{int} vs W for $\beta = 5.8$

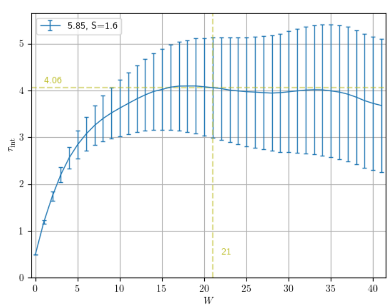

FIG. 19. τ_{int} vs W for $\beta = 5.85$

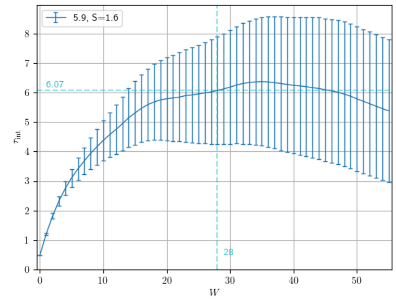

FIG. 20. τ_{int} vs W for $\beta = 5.9$

- The vertical and horizontal dashed line marks the value of optimal window size (or lag) and the corresponding value of τ_{int} respectively.

- We have chosen the value of our hypothesis parameter S as 1.6, as τ_{int} a plateau around the chosen W value for all the lattice gauge coupling values.

VI. CONCLUSION

Autocorrelation time plays a crucial role in analyzing time series data across various fields, including statistics, physics, economics, and more. In lattice Quantum Chromodynamics (QCD), autocorrelation time characterizes the rate at which subsequent configurations in Monte Carlo simulations become uncorrelated. The selection of an appropriate lag or window size is crucial for accurately estimating the integrated autocorrelation time (τ_{int}). A well-chosen lag balances the trade-off between bias and variance, facilitates efficient sampling of the autocorrelation function, ensures convergence of autocorrelation, and optimizes the utilization of computational resources. By considering the importance of lag in calculating τ_{int}, researchers can enhance the reliability and efficiency of statistical analyses in various fields. Autocorrelation time is a fundamental concept in lattice QCD simulations. Understanding and accurately estimating autocorrelation time are crucial for interpreting data and optimizing simulation algorithms.

[1] Wolff, U. (2006). Monte Carlo errors with less errors *arXiv:hep-lat/0306017*. Retrieved from `https://doi.org/10.48550/arXiv.hep-lat/0306017`

[2] Madras, N & Sokal, AD 1988, 'The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk *Journal of Statistical Physics, vol. 50, no. 1-2, pp. 109-186*. Retrieved from `https://doi.org/10.1007/BF01022990`