Tutorial 6

Statistical analysis of time series

Peter Košovan*, Marcello Sega

May 4, 2011
ICP, University of Stuttgart

We return back to analysis of simulation data, error estimation and identifying the equilibration period. We re-use some simulation programs from previous semester and focus mainly on data processing and obtaining reliable and unbiased estimates of the observables. We compare several methods of estimating correlation times in terms of efficiency and accuracy as well as computational cost.

1 Correlation functions and correlation times

In the last tutorial of the previous semester, we have been estimating correlation times of observables generated as a time series. Our estimates were done in a "quick and dirty" way which relies on assumptions about the properties of the correlation functions. In the current tutorial we revisit the same problem again. We will compute the correlation functions and correlation times explicitly, and compare different methods. This should provide more insight into reliability of the estimation methods.

In general, an autocorrelation function of an observable $\mathcal{O}$ can be defined as

$$C(\tau) = \langle\, ((\mathcal{O}(t) - \langle \mathcal{O} \rangle\,) \cdot (\mathcal{O}(t+\tau) - \langle \mathcal{O} \rangle)\,\rangle \tag{1}$$

considering that $\langle \mathcal{O}(t) - \langle \mathcal{O} \rangle \rangle = 0$ it simplifies to

$$C(\tau) = \langle \mathcal{O}(t) \cdot \mathcal{O}(t+\tau) \rangle - \langle \mathcal{O} \rangle^2 \tag{2}$$

Where $\langle \cdot \rangle$ refers to an ensemble average. It may be useful to normalize the autocorrelation function such as

$$\Gamma(\tau) = C(\tau)/C(0)$$

*kosovan@icp.uni-stuttgart.de

There are two important limits: $\Gamma(0) = 1$ and $\Gamma(\infty) = 0$. Asymptotically for large τ, one might expect the autocorrelations to decay exponentially

$$\Gamma(\tau) \sim \exp\left(-\frac{\tau}{\tau_{\text{exp}}}\right) \tag{3}$$

where τ_{exp} is the exponential correlation time. In a more general case, $\Gamma(\tau)$ can be composed of several exponential functions

$$\Gamma(\tau) = \sum_{i=0}^{i_{\max}} A_i \exp\left(-\frac{\tau}{\tau_i}\right)$$

or it does not have to be exponential at all. In the exponential case, the integrated autocorrelation time is a weighted average of the individual correlation times

$$\tau_{\text{int}} = \frac{1}{2} + \int_0^\infty \Gamma(\tau)\mathrm{d}\tau = \frac{1}{2} + \sum_{i=0}^{i_{\max}} A_i \tau_i \tag{4}$$

It is possible to compute $\Gamma(\tau)$ from a simulation. Then we can perform numerical integration and obtain an estimator for τ_{int}:

$$\overline{\tau_{\text{int}}}(t) = \frac{1}{2} + \int_0^t \Gamma(\tau)\mathrm{d}\tau \tag{5}$$

Where Δt is the time interval between two consecutive points for which we have computed $\Gamma(\tau)$. Performing numerical integration using the parallelogram law, we get

$$\overline{\tau_{\text{int}}}(t) = \frac{1}{2}\left(1 + j\Delta t \sum_{0<j\Delta t<t} \big(\Gamma((j-1)\Delta t) + \Gamma((j)\Delta t)\big)\right) \tag{6}$$

Provided that we have enough data, $\tau_{\text{int}}(t)$ should converge to a constant value. The value of $\overline{\tau_{\text{int}}}$ can then be determined self-consistently from the value of $\overline{\tau_{\text{int}}}(t)$ at $t \geq 6\,\overline{\tau_{\text{int}}}(t)$ (see also [2]). One has to check, however, that there is a plateau in $\overline{\tau_{\text{int}}}(t)$ around this value.

1.1 Some technical issues in computing correlations

The simplest algorithm to compute $\Gamma(\tau)$ is to consider all possible pairs of values. For a given set of correlated consecutive values $\mathcal{O}_i$, we compute the estimator $bar\Gamma(j\Delta t)$ as follows

$$\overline{\Gamma}(j\Delta t) = \frac{1}{\overline{\mathcal{O}^2} - (\overline{\mathcal{O}})^2} \frac{1}{(N-j)} \sum_{i=0}^{N-j} \left(\mathcal{O}_i \cdot \mathcal{O}_{i+j} - \langle\mathcal{O}\rangle^2\right) \tag{7}$$

Recall the distinction between the (true) ensemble average $\langle\mathcal{O}\rangle$ and an estimator of its value $\overline{\mathcal{O}}$. In principle one can in this way compute $\overline{\Gamma}(j\Delta t)$ up to $j = N - 1$, but the number of operations required to do so is proportional to $N(N-1)$ and therefore the computational cost grows steeply.

Another problem is that the error of estimating $\overline{\Gamma}(j\Delta t)$ also grows very quickly with j and soon the noise becomes bigger than the measured value. To obtain reliable estimates of $\overline{\Gamma}(j\Delta t)$ for higher values of j, one needs much more data than to obtain a reliable estimate of $\langle\mathcal{O}\rangle$. For $j\Delta t > \tau_{\text{int}}$ the estimate of $\overline{\Gamma}(j\Delta t)$ is a sum of many positive and negative fluctuations which eventually cancel to give zero. Therefore it is very sensitive to correctly excluding the equilibration period. If a part of equilibration period is included in computing the estimate, $\overline{\Gamma}(j\Delta t > \tau_{\text{int}})$ diverges rather than converging to zero. Even stronger effect can be observed in computing $\overline{\tau_{\text{int}}}(t)$ as we sum up all individual errors in $\overline{\Gamma}(j\Delta t)$. Typically, after reaching a plateau, $\overline{\tau_{\text{int}}}(t)$ diverges quickly due to the errors.

2 Simulation programs

To save time and keep our attention focused on physics and statistics, we avoid programming in this tutorial. Instead, will re-use the programs from the last semester and provide one additional to compute autocorrelation functions. A brief help on the use of all the programs can be obtained upon calling them with no arguments.

Apart from earlier programs, a new program for autocorrelations is provided. It computes the estimator for $\overline{\Gamma}(j\Delta t)$ and $\overline{\tau_{\text{int}}}(t)$ according to formulas 7 and 6 respectively. It also computes the self-consistent estimate of $\bar{\tau}_{\text{int}}$ but it does not check if there is a plateau in $\bar{\tau}_{\text{int}}(t)$ around this value. As usual, it is compiled by typing

```
make
```

and it takes the following arguments

```
autocorr <filename> <n_lines> <column> <tmin> <tau_max> <dt>
```

For convenience, you can get a short help when invoking the program without any arguments. Of course, you are encouraged to read the code and understand how it works.

Homework 1: Ising – follow the recipe (4 points)

In the first task, we take the simple ising model and relatively well-behaved quantities. Well behaved in a sense that their correlations decay exponentially and relatively fast.

1. Perform the simulation of the ising model at temperature $T = 2.5$ with the following parameters: `n_sweeps=100000`, `lattice_size=50`, `dump_frequency=10000`. The output file `magnetization.dat` contains three columns: MC step, magnetization and energy, respectively.

2. Use the `analyze` routine from the ising tutorial to measure the averages and Estimate the correlation time of magnetization (column 2) using the binning method. Take `tmin=10000` and different numbers of blocks as $N_{\text{blocks}} = 4 + 2^n,\ n \in \mathbb{N}$. Try also lower values of `tmin` and observe what they do to the result.

3. Plot the estimated correlation time as a function of the number of blocks and as a function of block size. Log-scale for N would be convenient!

4. Based on the plot, try to estimate the correct value of τ_{int} and discuss the possible confidence interval. Keep in mind that to obtain a reliable estimate of τ_{int}, you need to satisfy two contradicting requirements: $N \gg 1$ and $k \gg \tau_{\text{int}}$ where k is the block size. If you are unsure about the location of the plateau, you may try more values of N in the range where you expect the plateau to be.

5. Using the `autocorr` routine, compute the autocorrelation function and $\tau_{\text{int}}(t)$ for both the magnetization and energy. Take `tau_max=100`. Plot the autocorrelation function and $\tau_{\text{int}}(t)$ and discuss their shapes. *Remember that useful information is contained before $t \approx 10\,\tau_{\text{int}}$. The tail is just noise!*

6. Estimate the correlation time by identifying the plateau in $\tau_{\text{int}}(t)$.

7. Estimate the correlation time by fitting the initial part of $\Gamma(\tau)$ (before it reaches zero) using equation 3.

8. Put estimates from different methods in a table. Where available, include the error or try to obtain it as an educated guess. Discuss if the quality of different methods.

9. Repeat the same procedure for energy density (column 3 of magnetization.dat). Notice the energy density is not as well-behaved quantity as magnetization.

Homework 2: Explore the realm of ising correlations (2 points)

1. Try to measure how the correlation time for the magnetization depends on temperature.

2. Perform simulations also for $T = 2.35, 2.4, 2.7, 3.0$. Be aware that *each* temperature requires different length of the simulation and equilibration!

3. Plot $\Gamma(\tau)$ for different temperatures in one plot.

4. Plot $\tau_{\text{int}}(t)$ for different temperatures in one plot.

5. Estimate τ_{int} from your results.

6. Plot $\log(\overline{\tau_{\text{int}}})$ against $\log(|T - T_{\text{c}}|)$ where $T_{\text{c}} = 2/\log(1+\sqrt{2})$ is the critical temperature of the 2D ising model.

Homework 3: Correlations in a LJ liquid (4 points)

1. Perform a simulation of a LJ liquid in the Langevin thermostat at $T = 1.2$, $\rho = 0.5$, $N = 256$, $\Delta t = 0.01$ and $\gamma = 1.0$. Again, you are responsible for determining the necessary simulation time and discarding the equilibration.
   Hint: try starting with `ns=10000` timesteps to get the first result quickly and then make your simulation longer until you find the plot of $\bar{\Gamma}(\tau)$ satisfactory.

2. Analyze average system temperature in a similar fashion as you analyzed magnetization in Homework 1.

3. Estimate correlation times of T at more values of γ: 0.5 0.7, 1.4 and 2.0.

4. Plot the estimated correlation times and (including errors) as a function of γ^{-1} and discuss the plot.

Optional Homework: error of the error (2 points)

As it was mentioned in the introduction, errors in $\overline{\Gamma}(\tau)$ grow very quickly with increasing τ. Therefore, one would like to go one step further and obtain some error estimates also for the autocorrelation function.

1. Modify the `autocorr` code so that it uses binging method to compute error estimates for individual values of $\overline{\Gamma}(\tau)$.

2. Use the error propagation formula [1] to use the errors of $\overline{\Gamma}(\tau)$ and obtain error estimates for $\overline{\tau_{\text{int}}}(t)$. For simplicity, neglect the covariance term.

3. Apply the program to analyze the results of Homework 1.

4. Plot how the absolute and relative errors of $\overline{\Gamma}(\tau)$ and $\overline{\tau_{\text{int}}}(t)$ grow with time.

5. Compare your error estimates to the self-consistent estimates from `autocorr` and those from fitting the exponential to the initial part of $\bar{\Gamma}(\tau)$. If desired, try running a longer simulation for better accuracy.

Additional suggestions for playing around (no points)

1. Try what happens to $\bar{\Gamma}(\tau)$ if you forget to exclude the equilibration or a part of it.

2. Try to see how the correlation time increases with the lattice size in the Ising model.

3. Follow how the correlation times change with system size in the LJ liquid. Note qualitative differences between different quatintties (T, p, E_{kin}).

References

[1] http://mathworld.wolfram.com/ErrorPropagation.html

[2] http://www.fz-juelich.de/nic-series/volume10/janke2.pdf