

Contents

1	How the code works	1
2	Workflow	1
3	Testing first set of parameters	2
3.1	Probability distribution	2
4	Appendix	3

1 How the code works

For plotting the probability distribution as seen in the later evaluation part, you have to run the file "prob_distr.py" as is. The input of the functions used for plotting the histograms can be changed regarding the input data and the parameters of the bootstrap-method. Finally, at the end of the code (but not being plotted when you run the file) there is the code which we used for visualizing our original replicas.

2 Workflow

...

Probability distribution

A simple probability distribution for our calculated data of 500 mean values (500 replicas with one mean per replica, each for magnetization and energy) can easily be calculated with numpy's histogram module and then be plotted with matplotlib's bar module. These histograms for our original 500 replicas can be seen in the appendix. For this, we also tested the expectation of a Gaussian distribution for our histograms via fitting a Gaussian curve to the data and then displaying both overlayed, which worked very well. (This also showed the importance of the further down explained mean error as a standard deviation.) Moreover, as we want to implement specific errors for each bar into our distribution, we can fundamentally choose between two procedures:

1. We can compute a lot more replicas for the same given parameters to then calculate the overall mean and statistical error (with given equations in lecture) for each bar of the histogram. This would be an exact method. However, in order to not have to compute hundreds and thousands of more replicas we chose a resampling procedure.
2. The bootstrap-method can be used to estimate a real sampling distribution by just resampling our given data. We implemented the method in the following way: first, we take a certain number (function input "size_small_samples") of elements from the given array and append these values into a new array until we reach the original length of the given data; in our case 500. This will be done a specific number of times, e.g. N-times (function input "numb_samples"). Also using the original data, at the end we get (N+1) "independent" arrays of same size; just as we would have computed 500 replicas and their means (N+1)-times explicitly. This procedure is of course an approximation, but we are also working with statistical changes from one replica to another, which underlines the bootstrap-method being used. As already explained for method 1, we can split our newly computed arrays of replica-means into specific

bars of the final histogram (function input "numb_bars"). For each bar, we calculate the overall mean with its statistical error, as explained in the lecture. These values can then be used for the final histogram with error bars. The histogram plot was afterwards refurbished for better visibility and evaluation of the distributions.

An important note: as expected for statistical progresses, the outcome of our bootstrap-method is highly dependent on the chosen parameters of sampling number and chosen random seed. These parameters were tested extensively with hindsight on the error of the overall mean via the bootstrap method being criterion for statistical σ -deviation. The parameters used for our final calculations via bootstrap-method are: random.seed=5, numb_bars=50, numb_boot_samples=100, size_small_samples=5.

3 Testing first set of parameters

3.1 Probability distribution

We now evaluate the probability distributions of magnetization and energy. The procedure used for calculating our data and resampling sets was already explained in the workflow. Thus, both of the histograms can be seen in the figures below. All parameters for the bootstrap-method and the plotting itself, except the x-axis and therefore x-size of the bars, were kept the same for both figures.

One can observe a relatively good Gaussian distribution for both histograms, as expected for a statistical deviation of replicas (viable for calculating replicas and bootstrapping them). Both the beforehand and via bootstrap calculated mean are nearly identical (maximum ± 0.2 for magnetization). Whereas the bootstrap errors are around 50-times larger, stemming from the statistical resampling nature of the bootstrap-method. Still, just from looking at both plots, one could argue, that the errors of the bootstrap mean are approximately in a 1σ -deviation.

Furthermore, the energy means are more densely distributed around the overall mean than the magnetization means, also visible in different sizes of errors. This can also be seen in figure 3 in the appendix, where the same energy histogram is plotted with the x-axis being in the same size limits as for the magnetization.

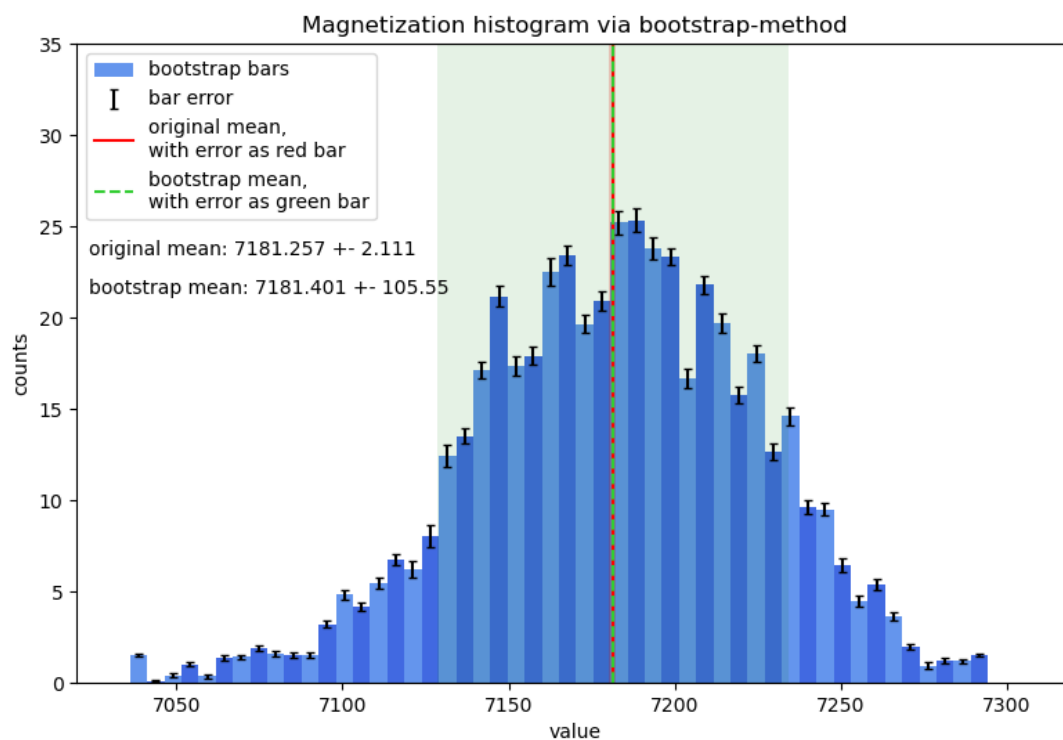


Figure 1: Magnetization histogram via bootstrap-method. With statistical error bars, mean of the original- and bootstrap-distribution, respective statistical error added as Δx bars (view values in figure).

4 Appendix

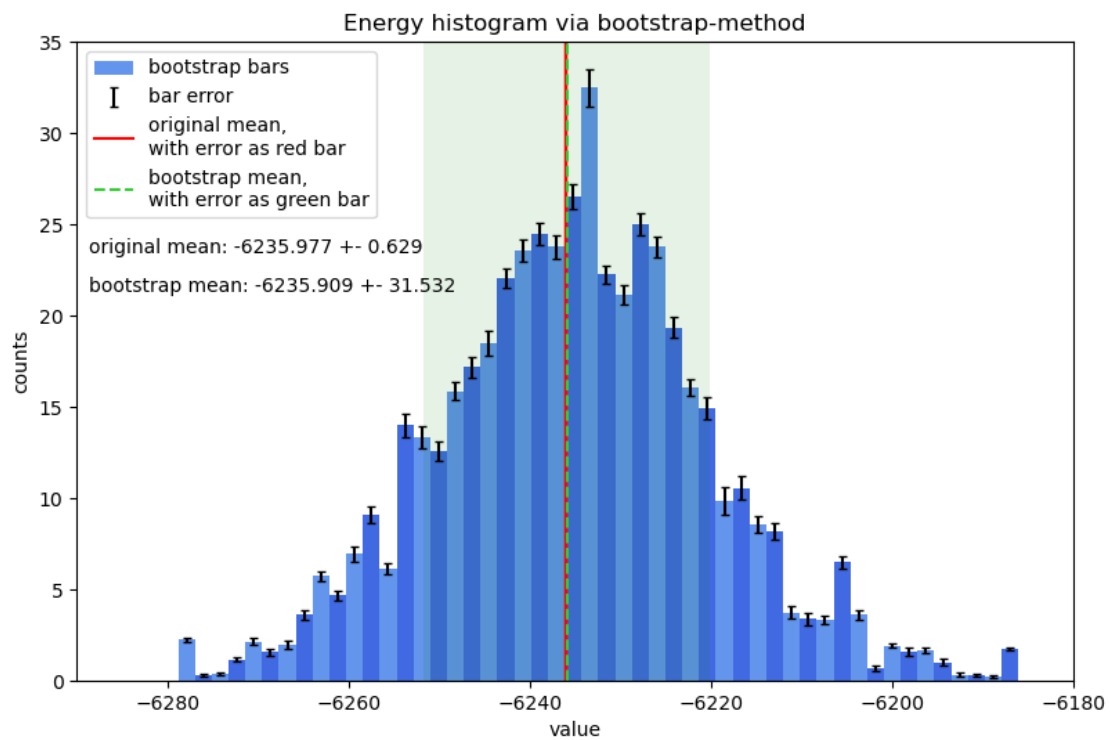


Figure 2: Energy histogram via bootstrap-method. With statistical error bars, mean of the original- and bootstrap-distribution, respective statistical error added as Δx bars (view values in figure).

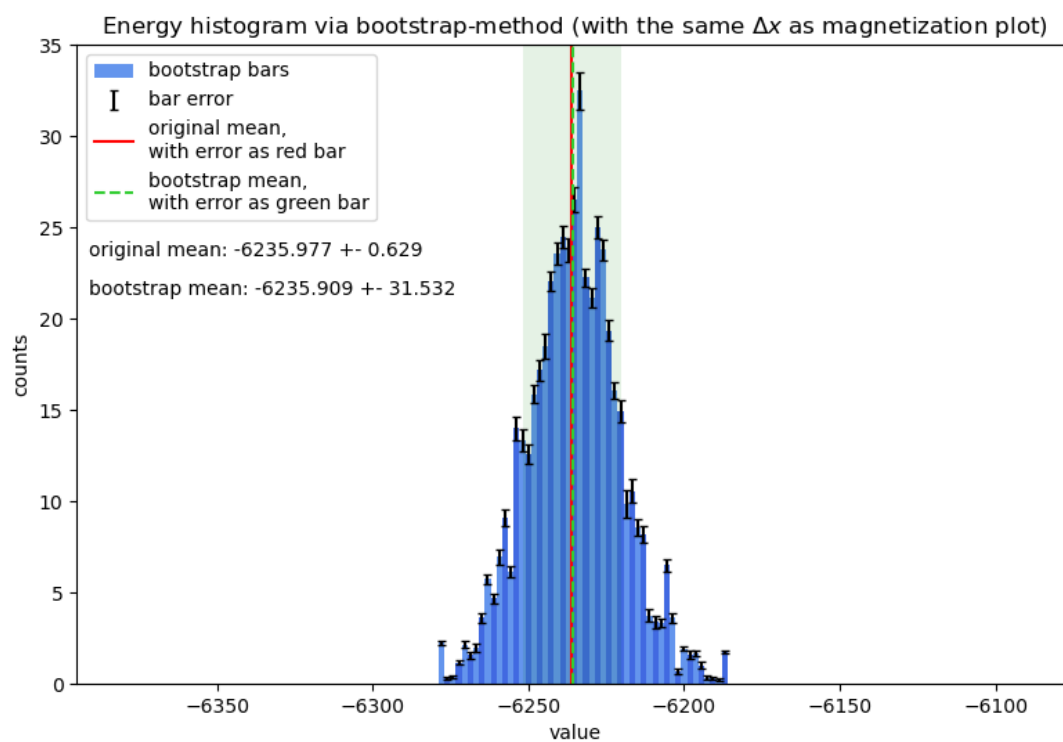


Figure 3: To visualize the narrower distribution of the energy in regards to magnetization: same energy histogram via bootstrap-method as seen before, now with the x-axis being in the same size limits as for the magnetization. With statistical error bars, mean of the original- and bootstrap-distribution, respective statistical error added as Δx bars (view values in figure).

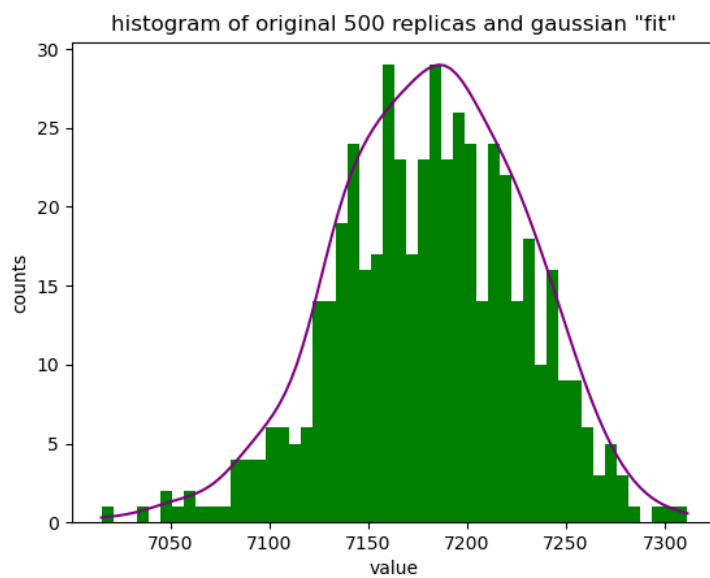


Figure 4: Magnetization: original data histogram and Gaussian "fit"

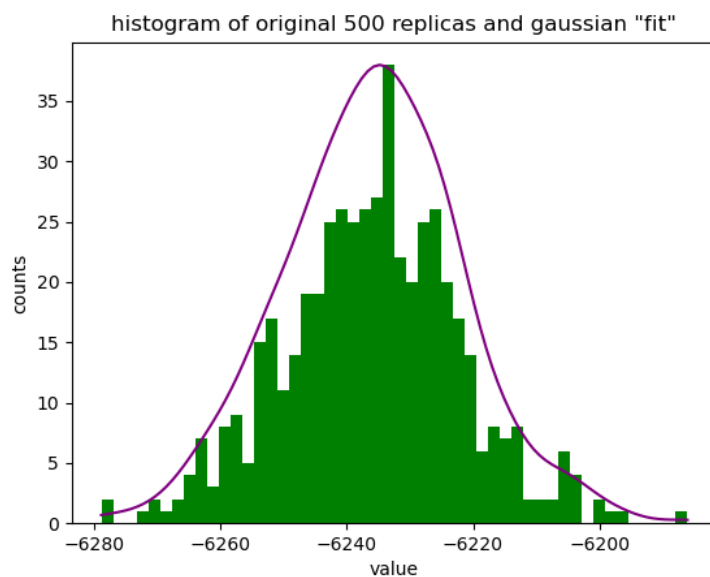


Figure 5: Energy: original data histogram and Gaussian "fit"