

The moving block bootstrap as a convergence diagnostic in Markov chain Monte Carlo ¹

Il moving block bootstrap per diagnosticare la convergenza nelle catene di Markov Monte Carlo

Elisabetta Baldassarri

Rodolfo Rosa

Regione Marche

Dipartimento di Scienze Statistiche

Servizio Sistema Informativo Statistico

“PaoloFortunati” - Università di Bologna

elisabetta-baldassarri@regione.marche.it

rosa@stat.unibo.it

Riassunto: Il moving block bootstrap viene proficuamente impiegato per assegnare l'accuratezza alle stime in serie storiche quando la correlazione tra i dati è relativamente forte, come spesso avviene nelle catene di Markov Monte Carlo. Con esperimenti al computer su un modello di Ising bidimensionale, viene mostrato l'uso del moving block bootstrap anche quale strumento di diagnosi della convergenza.

Keywords: Markov chain Monte Carlo, moving block bootstrap, convergence diagnostics.

1. Introduction

The essential idea of Markov chain Monte Carlo (MCMC) is to generate realizations of a random variable X from an irreducible Markov chain *converging* to (i.e., having as its stationary distributions) the desired distribution of X . Historically, the paper by Metropolis et al. (1953) marked the birth of the Monte Carlo method based on Markov chains in the study of statistical-mechanical systems. Now the statistical applications of MCMC covers a variety of areas, besides physics, ranging from medicine to economics, from archeology to genetics, etc.. In any case, MCMC is most widely used in the framework of Bayesian inference, even though one can find numerous applications in classical statistics too (see, for a fine tutorial, Chib and Greenberg, 1995). Thanks to MCMC methodology, one is able to sample, at least in principle, any density function, regardless of its complexity, in any number of dimensions. However, besides such virtues, some well-known failings are present, due to the very nature of the method: (i) theoretically the “convergence” is reached only asymptotically; (ii) the data generated in the chain are often highly correlated. We will show here that the moving block bootstrap used to assign accuracy to estimates in the presence of considerable correlations between successive realizations, can be exploited also to cast light on the first issue.

The problem of convergence is to assess whether or not stationarity has been achieved, since only if the chain has achieved stationarity averages and variances estimates make

¹Funding from MURST is gratefully acknowledged

sense. In other words, we have to assess that the piece of chain we have generated, and on which we base our estimates, is representative of the whole underlying stationary distribution of such a Markov chain. As stressed by Cowles and Carlin (1996, p. 883), such a notion of convergence refers to something which is not “a single number or even a distribution, but rather a *sample* from a distribution” (authors’ italics). A variety of methods to determine whether stationarity is achieved (*convergence diagnostics*) have been proposed. An old but still currently exploited idea is to execute several runs, all with the same parameters but with different starting values. By comparing the results somehow, one can detect possible dangers and try to circumvent them. This useful prescription was formalized and improved by Gelman and Rubin (1992). On the other hand, it was noticed by Geyer (1992) that one may face cases in which many short chains converge to a unique, but completely wrong distribution. Therefore, it should be more advisable to trust results based on a single run. Often, as in the examples above, different authors give different suggestions, but all grant that there is no single method to deal with convergence, and each tool has its pros and cons. The reader is referred to the review article by Cowles and Carlin (1996) and the chapter 8 of the book by Robert and Casella (1999), in which almost all convergence diagnostics available in literature are discussed and compared.

2. Method

At SIS 1994, Mignani and Rosa, henceforth MR, (1994) showed for the first time how the moving block bootstrap (MBB) works in the Ising model. In a number of following papers, MR (2001, and references therein), through computer simulations, studied the MBB extensively and showed that such a method outperforms other methods based on subseries values in presence of highly correlated realizations in relatively short chains. It was also shown that the MBB may be exploited to performs “statistical experiments” when other approaches are infeasible. In this spirit, we discuss here results of statistical experiments on the same Ising model presented at SIS 1994, focusing on those features concerning with the convergence of the chains. Sketches of the MBB and the Ising model are given below, see MR (2001) for more details.

Basically, the MBB consists in dividing the time series, once stationarity is reached, in overlapping blocks of length l , in such a way that observations more than l apart are nearly statistically independent. The blocks are then resampled with replacement and pasted together to form new series, each with the same length of the original. As in the bootstrap for independent data, we can obtain a suitable number of resampled series, from each of which the statistic of interest is computed and bootstrap measures of accuracy (standard errors, confidence intervals) are derived. A crucial point is the choice of l with respect to the total length of the chain, since we require that the correlation between observations belonging to different block has to “die out”. By plotting the MBB estimate $\hat{\sigma}^*$ of the standard error as a function of l , it appears that $\hat{\sigma}^*$ reaches a plateau when the blocks are iid random variables and, at the same time, inside each block the correlation is retained.

Experiments were performed on a two-dimensional (2D) Ising model with interaction limited to neighboring spins. The 2D Ising model consists of a square regular array of “spin” random variables, s_i , taking on only two orientations: “up” ($s_i = +1$) and “down” ($s_i = -1$). With L sites on an edge, the system configuration is represented by the random vector: $\mathbf{s} = s_1, s_2, \dots, s_N$ where $N = L \times L$ is the total number of spins. The energy of

the system, in absence of external magnetic field, is given by:

$$E = -J \sum_{ij} s_i s_j \quad (1)$$

the indices ij indicate nearest-neighbor pairs and J is a constant controlling type and degree of interaction between site i and j . The application of the Metropolis algorithm consists in selecting one spin s_h ($1 \leq h \leq N$) in the vector $\mathbf{s}_i$, representing the system configurations with energy E_i , computed by (1); the spin direction is reversed ($s_h \rightarrow -s_h$); the new state $\mathbf{s}_j$ is accepted with probability $\alpha = \min\{1, \exp[-(E_j - E_i)/k_B T]\}$ (where k_B is Boltzmann constant and T the temperature). In any case, $\mathbf{s}_j$ or $\mathbf{s}_i$ is counted as a “new state”. The desired properties of the state are stored to compute the statistics of interest.

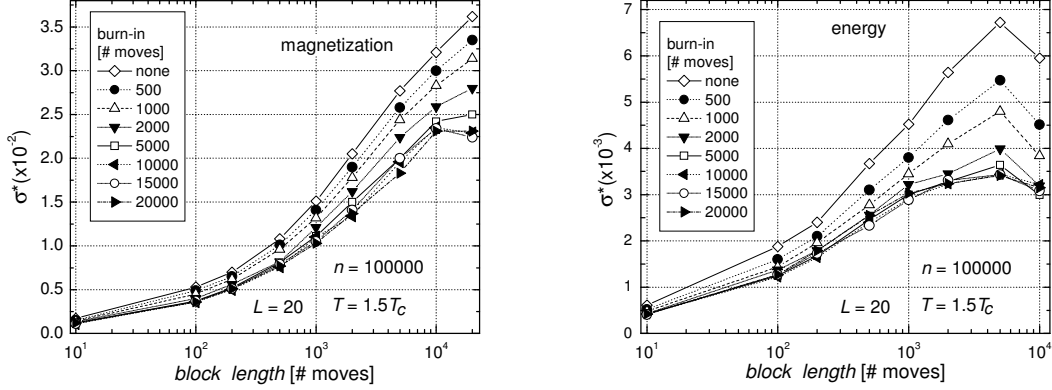
3. Results and discussion

As in MR (1994), computer experiments were performed with $L = 20$, $T = 1.5T_c$ (T_c is the critical temperature above which the spins are almost independent of each other), and periodic boundary conditions. The quantities to be estimated were the magnetization per spin m and the internal energy per spin ϵ , defined as:

$$m = \frac{1}{N} \sum_{h=1}^N s_h \quad \epsilon = \frac{1}{N} E \quad (2)$$

where E is given by (1) and the units of measurement are such that $\epsilon = 1$ and $m = \pm 1$ if $s_h = \pm 1$, all h . To estimate m and ϵ , the chain runs for a number of k interactions or “moves”, i.e. k spin-flip attempts. The output was the sample $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k$, which was used to estimate the quantities of interest. The estimators are of the form $\hat{\mu} = \frac{1}{k-r} \sum_{i=r+1}^k g(\mathbf{s}_i)$, where $\hat{\mu} = \hat{m}$ or $\hat{\epsilon}$ (see (2)). This means that the first r moves were discarded and the averages were done over the remaining $k - r$ moves. Such “burn-in” time r is necessary to avoid, or at least reduce, the effect of starting configurations. We chose the starting configuration with all the spins up, and produced chains with r from 0 (no burn-in) to 20000 moves, in any case we kept $k - r = n = 100000$ moves. To each output we applied the MBB to estimate the standard errors of $\hat{m}$ and $\hat{\epsilon}$. Notice that in MR (1994) the “time” unit was the Monte Carlo step (MCS) per spin, i.e., $\# \text{ moves}/N$, and the chains were much longer. The results of our experiments are shown in the figures below. It appears immediately that the MBB works, i.e. $\hat{\sigma}^*$ reaches a convincing plateau, only when the *whole* chain is at stationarity, that is after 5000–10000 moves. We notice also that the time correlation between configurations decays more slowly for the magnetization than for the energy, since the plateaux begin at $l \approx 10^4$ and $l \approx 10^3$ for the former and latter quantity, respectively. In both cases however the MBB is able to signal that convergence is still not achieved. Of course if n is too small nothing could be said, except that too short chains are perfectly useless. The above results suggest the motivation why the MBB can work as a monitoring convergence technique. The validity of the MBB rests on two prerequisite conditions: under the resampling scheme all the blocks (i) must be independent of each another, as evidenced by the plateau of $\hat{\sigma}^*$ vs l ; (ii) they must share the same probabilistic structure. But, if the latter condition is not satisfied (MBB samples formed by stationary and non-stationary blocks), no signal that the

Figure 1: Moving block bootstrap estimates $\hat{\sigma}^*$ of the standard errors of the magnetization and internal energy, as a function of the block length, for the $L \times L$ Ising model, with $L = 20$ and $T = 1.5T_c$. The burn-in times for different runs are also reported.



former requirement has been fulfilled is apparent. We compared also our proposal with the convergence diagnostic by Gelman and Rubin (1992). We repeated independent runs with different starting configurations and the Gelman and Rubins's method (1992) was applied to these multiple chains. The results confirmed our findings.

In conclusion, exploiting the MBB, is reminiscent somehow, for one side, of methods based on batch means and, for the other one, of method generating replications by way of regeneration. The latter approach, however, suffers of considerable drawbacks (e.g. Brooks and Roberts, 1999), and batch methods do not exploit the whole sample. The MBB uses only one chain, of which it exploits the whole information. Finally, with the MBB one can kill two bird with one stone: assessing accuracy and exploring convergence.

References

- Chib, S., Greenberg, E. (1995) Understanding the Metropolis-Hastings algorithm, *American Statistician*, 49, 327-335.
- Cowles, M. K., Carlin, B. P. (1996) Markov chain Monte Carlo convergence diagnostic: a comparative review, *Journal of the American Statistical Association*, 91, 883-904.
- Gelman, A., Rubin, D. B. (1992) Inference from Iterative Simulation Using Multiple Sequences, *Statistical Science*, 7, 457-511.
- Geyer, C. J. (1992) Practical Markov Chain Monte Carlo, *Statistical Science*, 7, 473-511.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. (1953) Equations of State Calculations by Fast Computing Machine, *Journal of Chemical Physics*, 21, 1087-1091.
- Mignani S., Rosa R. (1994) Il metodo bootstrap 'moving block' per dati dipendenti: un'applicazione ai modelli di Ising, in: *Proceedings, XL conference of the Italian Statistical Society* (S. Remo 6-8 aprile), 465-472.
- Mignani S., Rosa R. (2001) Markov Chain Monte Carlo in Statistical Mechanics: the Problem of Accuracy, *Technometrics*, 43, 347-355.
- Robert, C. P., Casella, G. (1999) *Monte Carlo Statistical Methods*, Springer-Verlag, New York.