Population Annealing
An approach to finite-temperature calculation

Yukito Iba *
The Institute of Statistical Mathematics,
Minato-ku, Tokyo, Minami-Azabu,
iba@ism.ac.jp

1 Background - Improved Monte Carlo Algorithms

In the last decade of the 20th century, Dynamic Monte Carlo algorithms, which have been essential tools in statistical physics, are proved to be also useful in statistical and information sciences [1, 2]. Now, these methods are considered as indispensable part of undergraduate/graduate courses in statistics and engineering. Note that the ability of sampling from high dimensional distributions and calculating averages is also important in statistical science, for example, for the sampling from posterior distributions.

This change increases the attention to an inherent difficulty of these algorithms, that is, Markov chains used in these algorithms mix very slowly in difficult problems. It causes annoying bias and inaccuracy of the outputs. Many efforts have been devoted to overcome this difficulty. "Simulated Annealing", which is a method for optimization, or equivalently, a method for finding a ground state of the system, does not resolve the difficulty at finite temperatures.

In 1990s (or late of 1980s),"Extended Ensemble" type algorithms are introduced, which are useful for finite temperature calculations. In these algorithms, ensembles defined as compositions or extensions of given distributions are simulated. Typical examples are Multicanonical algorithm, Simulated Tempering, and Exchange Monte Carlo (Parallel Tempering) algorithms. Now they become popular in various fields in statistical physics

*The main part of the study is done in collaboration with Koji Hukushima (Univ. Tokyo).

and statistical information processing. An interdisciplinary review on this subject is found in [3].

These methods based on improved ensembles have an advantage of great flexibility and generality compared with other approaches to overcome the slow mixing problem, for example, algorithms with cluster updates or improved dynamics. But there are still possibilities of more flexible usage or extensions. Specifically, we are interested in the following aspects:

1. The extension of an ensemble is usually done in the space of the temperature, or in the space of the energy. In principle, however, we can use any parameter, parameters, and physical quantities to construct extended ensembles, when it is useful to facilitate the relaxation and/or to realize efficient measurement of required quantities. This idea leads to the development of special purpose ensembles and multivariate extensions.

2. It will be interesting to extend the idea of Extended Ensembles and combine it with "Sequential" or "Population" Monte Carlo algorithms [6, 9, 10], whose studies are reactivated [1] in the fields of statistical information processing and (bio)polymer studies.

The direction **1** is investigated by the author and other researchers in this several years(see [3], for a comprehensive review including historical notes). For example, a method based on a two-dimensional extension in (energy, self-overlap) space is proved to be very efficient for simulating lattice proteins [4, 5]. We are now planning some studies on frustrated spin systems and/or error-correcting codes by the principle of **1**.

There are a few possibilities in the direction **2**. In this paper, we discuss one of them and present a way to modify Simulated Annealing to an algorithm for finite-temperature calculation. The proposed method, Population Annealing, uses multiple copies of the original system to represent distributions, which we call "particles" or "replicas". It is, however, essentially different from Exchange Monte Carlo (Parallel Tempering). In Population Annealing, the temperature is common for all replicas and gradually decreases throughout the simulation. It does not up and down stochastically like the one in Simulated Tempering. Instead, we introduce weights of the replicas and a resampling procedure that split/remove the replicas (particles)

[1] Population Monte Carlo has long history, whose most famous example in physics is *diffusion Monte Carlo* for quantum systems.

using the weights, which is commonly used in population Monte Carlo methods [6]. Despite the name of "Annealing", it enables us correct computation of canonical averages in the limit of infinite number of replicas.

2 Population Annealing

Here we give an implementation of Population Annealing for canonical distributions. In principle, Population Annealing can be used with any family of distributions parameterized by a parameter, which plays the role of β in a canonical distribution.

Let us represent the state of the system by x and consider a family of canonical distributions parameterized by the inverse temperature β

$$P_\beta(x) = \frac{\exp(-\beta E(x))}{Z_\beta}. \tag{1}$$

Consider K copies of the system x (replicas or paricles), which are randomly initialized. We will indicate them by $\{x^k\}, k = 1 \dots K$. These replicas $\{x^k\}$ are simulated in a parallel manner as described in the following description of the algorithm. We also associate weight W^k to each replica k, which is initialized by $W_0^k := 1$. Starting from sufficiently high temperature $1/\beta_0$, we repeat the following procedures with decreasing sequences of temperatures $\{1/\beta_i\}$.

1. Calculate weights of the replicas:

   $$\triangle\beta_i := \beta_i - \beta_{i-1}.$$

   For each k,

   $$w^k := \exp(-\triangle\beta_i E(x^k))$$
   $$W_i^k := W_{i-1}^k w^k.$$

2. Split/Remove (Resampling) of the replicas:
   If $k \equiv 0 \,(\bmod M)$, then the following procedure is performed.
   For all k, set

   $$P^k := \frac{W_i^k}{\sum_k W_i^k}$$
   $$x_{old}^k := x^k.$$

   Then, set the new value of x^k to x_{old}^k with the probability P^k.

Multiple selection of a replica x^k_{old} is possible. Thus, a replica with a small weight W^k is removed with a high probability, while a replica with a large weight tends to have multiple "descendants" . After the resampling, all the weights are re-initialized to the unity.

$$W^k := 1$$

3. Monte Carlo update of the replicas:
   Simulate each replica x^k independently with $\beta = \beta_i$ using the energy $E(x)$ in *finite* number of Monte Carlo steps (usually, a small number of steps is preferable, for example, 1 MCS). Sample average A^k_i of any desired physical quantity $A(x)$ is calculated.

   Any dynamical Monte Carlo algorithm which can sample from the canonical distribution (1) with $\beta = \beta_i$ can be used for the update.

4. Calculate averages:
   Canonical average of the quantity A is calculated by

$$\langle A \rangle_{\beta_i} = \sum_k A^k_i W^k_i.$$

   This part of the algorithm can be done in an off-line manner, when we store $\{A^k_i\}$ and $\{W^k_i\}$.

5. Set $i := i + 1$.

The interval M of the split/remove procedure should be properly chosen. In general, we can perform the procedure **2** with unequal intervals. But we should be careful to on-line adaptive implementation, because it can introduce systematic bias.

The algorithm without the steps **1** and **2** reduces *simulated annealing*, with which canonical averages at finite temperatures are not correctly reproduced. On the other hand, the one with the step **1** is formally correct even without the step **2**. Its efficiency is, however, severely affected by the increase of the variance of weights $\{W^k\}$. It seems not to work well in a complex and large-scale system.

The algorithm without the step **2** essentially the same as an algorithm proposed by Jarzynski [8]. He discussed it from the viewpoint of non-equibirium thermodynamics. Neal [7] discussed a similar idea "Annealed Importance Sampling" in a more transparent manner. From this viewpoint, the proposed algorithm, "Population Annealing", are regarded as an extension of Jarzynski's algorithm or "Annealed Importance Sampling".

"Population Annealing" is also understood as a special case of generalized Sequential Monte Carlo or population Monte Carlo (see, [6] and [9, 10]). Another related study is found in [11], although the algorithm proposed by them is more complicated and covered by a (somewhat superfluous) quantum mechanical analogy. They did not treat random spin systems and combinatorial problems with discrete variables, which are main subject of our study.

3 Application to Spin Glass

We have tested the validity and efficiency of the proposed algorithm with a ferromagnetic Ising model, an Ising spin glass model, and estimation of the number of magic squares of given sizes. Here I will discuss results from a collaborated work with Koji Hukushima [12]. The model is a three-dimensional Ising spin glass model with Gaussian coupling in a magnetic field. All the results shown here are obtained from a single sample (bond realization) of $N = 8^3$.

In Fig. 1, the proposed algorithm is compared to conventional algorithms, i.e., simulated annealing with different cooling rates, an algorithm without resampling (without the step **2**), and Exchange Monte Carlo. Simulated annealing gives incorrect canonical averages at low temperatures even with very slow annealing speed, while the output of algorithm without resampling shows large variance. The proposed method gives good results, which coincides with the results of Exchange Monte Carlo. That is, in this example, both of the steps **1** and **2** are crucial for the efficient calculation of canonical averages.

In Fig. 2, an example of the "pedigree", i.e., the graph of ancestor - descendants relations is shown. For a replica in the population at the lowest temperature, its "ancestor" in the highest temperature $1/\beta_0$ is identified, and then all the descendants of it are illustrated in the figure with a color (or gray level). Repeating this procedure, we can visualize the branching process induced by the resampling process **2**. Note that the descendants of an ancestor are not necessarily in the same state, because Monte Carlo updates in the step **3** changes the states. The data shown in Fig. 2 is taken from a toy-simulation with a small number of replicas.

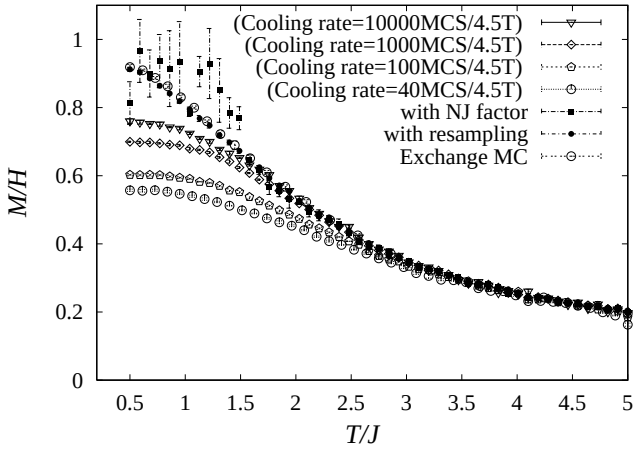

Figure 1: Comparison of the proposed algorithm (Population Annealing), Simulated Annealing with different cooling rates, an algorithm without resampling (with NJ factor, Neal-Jarzynski algorithm), and Exchange Monte Carlo. The magnetization at each temperature is shown in the vertical axis. Gaussian coupling, $N = 8^3$, $H/J = 0.1$.

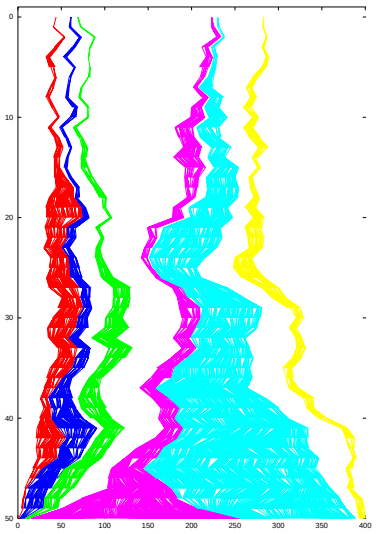

Figure 2: A pedigree of the replicas (particles). The uppermost line corresponds to the highest temperature, while the bottom line corresponds to the lowest temperature.

References

[1] Iba,Y., "Bayes-toukei to Toukei-Butsuri" (Bayesian Statistics and Statistical Physics), a book in Japanese, Iwanami, to appear in this summer.

[2] Gilks et al. (eds.), Markov chain Monte Carlo in Practice, Chapman and Hall (1996).

[3] Iba,Y., Extended Ensemble Monte Carlo, International Journal of Modern Physics, C12 (2001) 623-656.

[4] Iba, Y., Chikenji,G., and Kikuchi, M., Simulation of Lattice Polymers with Multi-Self-Overlap Ensemble, J. Phys. Soc. Japan, 67 (1998) 3327-3330.

[5] Chikenji,G., Kikuchi,M., and Iba,Y., Multi-Self-Overlap Ensemble for Protein folding: Ground state search and Thermodynamics, Physical Review Letters, 83 (1999) 1886-1889.

[6] Iba,Y., Population Monte Carlo algorithms, Transactions of the Japanese Society for Artificial Intelligence, 16 (2001) 279-286, e-print: cond-mat/0008226.

[7] Neal, R. M., Annealed Importance Sampling, Statistics and Computing, 11 (2001) 125-139.

[8] Jarzynski, C., Equilibrium Free Energies from Nonequilibrium Processes Phys. Rev. E 56, (1997) 5018.

[9] Doucet et al. (eds.), Sequential Monte Carlo in Practice, Springer (2001).

[10] Liu, J.S., Monte Carlo Sterategies in Scientific Computing, Springer, 2001.

[11] Cho, A. E., Doll, J. D., and Freeman, D. L., The Construction of Double-Ended Classical Trajectories, Chem. Phys. Lett., 229 (1994) 218.

[12] Hukushima, K. and Iba, Y., in preperation. Some of the results in this paper are presented at the annual meeting of Physical Society of Japan (2003).