The Role of Resampling in Population Annealing

Denis Gessert[1,2], Martin Weigel[1,3]

[1] Centre for Fluid and Complex Systems, Coventry University, Coventry, CV1 5FB, United Kingdom
[2] Institut für Theoretische Physik, Leipzig University, Postfach 100920, D-04009 Leipzig, Germany
[3] Institut für Physik, Technische Universität Chemnitz, D-09107 Chemnitz, Germany

Abstract

Population Annealing (PA) is a population-based algorithm that can be used for equilibrium simulation of thermodynamic systems with a rough free energy landscape. It is known to be more efficient in doing so than standard Markov chain Monte Carlo alone. The algorithm has a number of parameters that can be fine-tuned to improve performance. While there is some theoretical and numerical work regarding most of these parameters, there appears to be a gap in the literature concerning the role of resampling in PA, which this work attempts to bridge.

The $d = 2$ Ising model is used as a benchmarking system for this study. At first various resampling methods are implemented and numerically compared using a GPU PA implementation. In a second part the exact solution of the Ising model is utilized to create an artificial PA setting with effectively infinite Monte Carlo updates at each temperature as well as an infinite population. This allows one to look at resampling in isolation from other parameters.

We identify when resampling choices affect the simulation outcome and obtain some results that are model-independent. Further, we name two resampling methods that appear preferable over the widely used multinomial resampling.

1. Population Annealing Algorithm [1]

Parameters

- population size R
- annealing schedule $\{\beta_k\}$
- # of MC updates at each temperature step
- resampling interval M (usually $= 1$)
- **resampling method**

Algorithm Standard Population Annealing algorithm [1]

```
1: t ← 0
2: Initialize population of R independent replicas at β_0
3: while β_{t+1} < β_max do
4:   t ← t + 1
5:   Calculate Boltzmann weights w_i of the replicas at β_t
6:   if t ≡ 0 (mod M) then resample population according to weights
       (on avg. τ_i = R w_i copies of replica i)
7:   Monte Carlo update of the replicas (θ MC sweeps at β_t)
8:   Measure observables O
9: end while
```

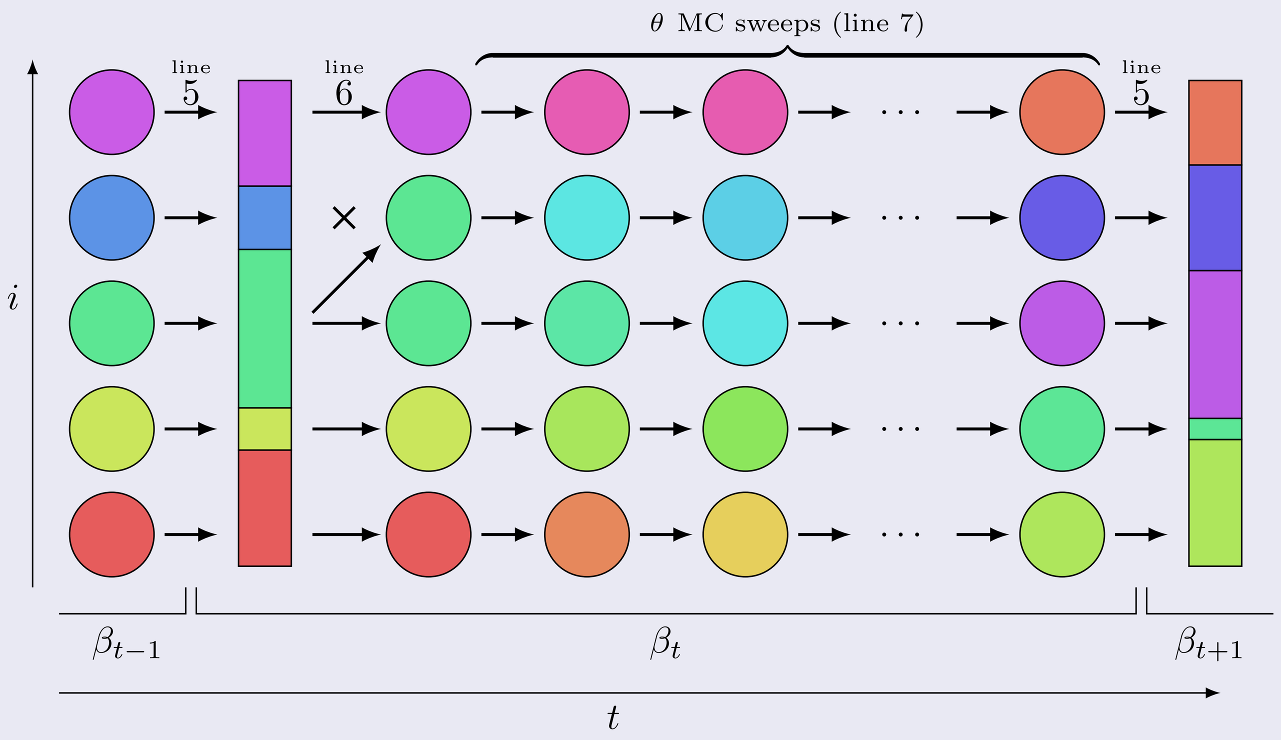
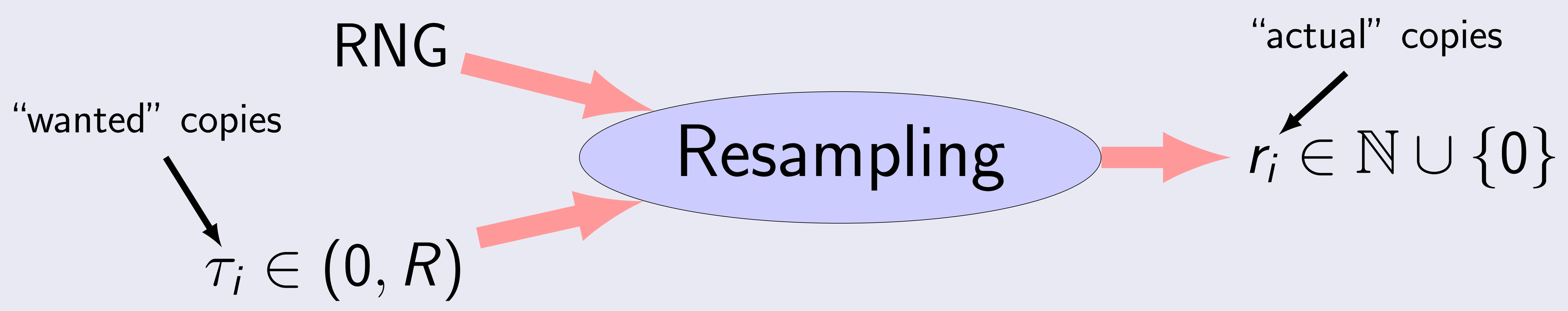


Figure: Schematic of a typical PA simulation. Vertical axis corresponds to R replicas (here five) and horizontal axis to simulation time.

2. Resampling

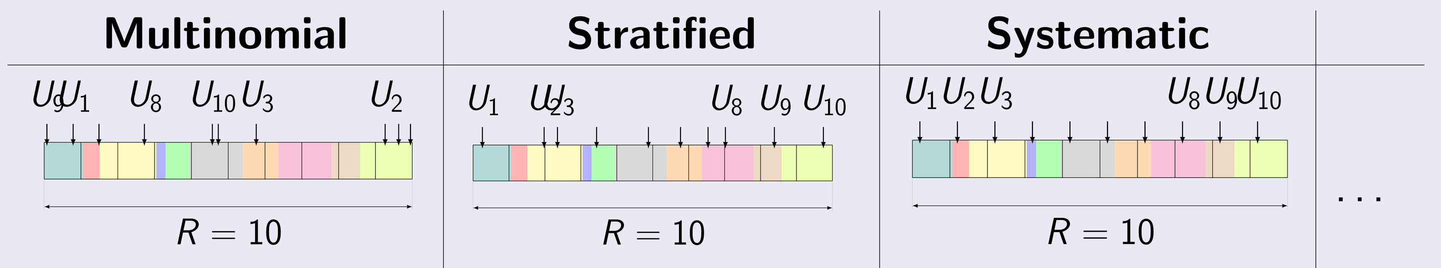


The goal of resampling is to make a number of copies proportional to the statistical weight of individual replicas. As this in general results in a non-integer number of "wanted" copies τ_i, this is not straightforwardly possible. Instead the actual (integer) number of copies, r_i, is chosen through a random resampling process such that the expectation of r_i is τ_i, i.e. $\langle r_i \rangle = \tau_i$.

Clearly, $\langle r_i \rangle = \tau_i$ does not fully specify the resampling process and a number of methods are possible:

	Nearest Integer	Poisson	
With Fluctuating Population Size	$P_{\tau_i}(r_i = k) = \begin{cases} \tau_i - \lfloor \tau_i \rfloor & , k = \lfloor \tau_i \rfloor + 1 \\ 1 - (\tau_i - \lfloor \tau_i \rfloor) & , k = \lfloor \tau_i \rfloor \\ 0 & \text{else} \end{cases}$	$P_{\tau_i}(r_i = k) = \frac{\tau_i^k e^{-\tau_i}}{k!}$	…

With Constant Population Size: Multinomial, Stratified, Systematic …



3. Benchmarking Quantities

In order to answer the question which resampling method is best, we need measures of how well the PA algorithm performs. These are the ones used in this work:

- Bias and statistical error in observables
- Family quantities n_i (resp. N_i): fraction of the population (resp. # of replicas) descending from replica i

$$\rho_t = R \sum_i n_i^2 = \frac{1}{R} \sum_i N_i^2$$

(average family size)

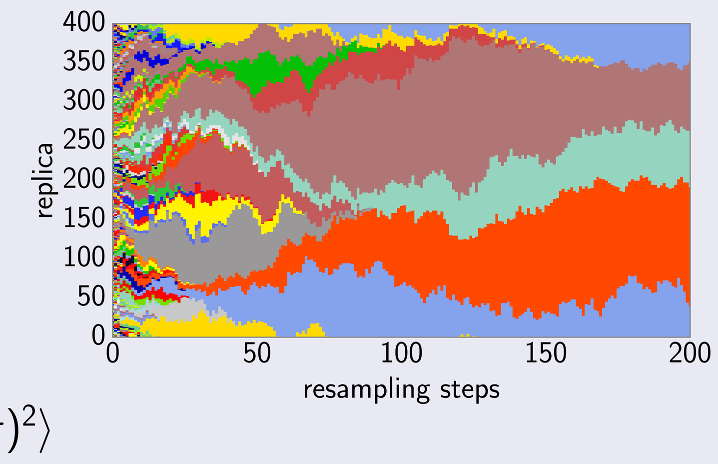


Figure: Pedigree of replicas: Each color corresponds to a family of replicas. The number of surviving families decreases with more and more resampling steps and the average family size increases.

- Sampling variance $SV = \langle (r - \tau)^2 \rangle$

4. Simulation Work

The following data was obtained through PA simulations using the parameters $R = 20{,}000$, $\theta = 5$ and $\Delta\beta = \frac{1}{200}$ and averaged over 5,000 runs (seeds). The implementation is based on the code of Ref. [2].

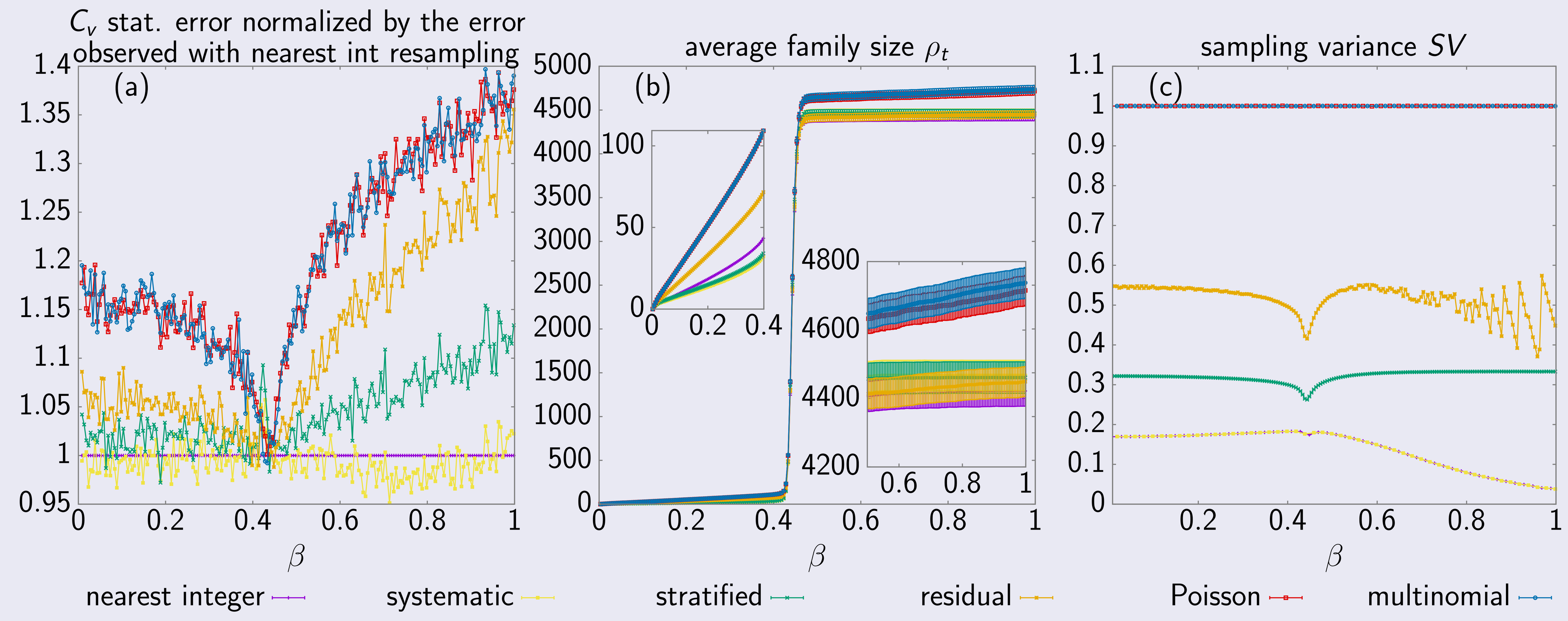


Figure: Measurements of benchmarking quantities as a function of β for various resampling methods. As a rough guide, for all three quantities "The smaller the measurement value the better."

- resampling significantly affects simulation outcome
- away from criticality artificial noise through resampling is of the order of (or larger than) the weight variance (⇒ strong effect)
- at criticality weight variance ≫ sampling variance (⇒ very small effect)

5. Analytical Considerations

Under mild assumptions we can show that asymptotically the average family size is given by the recursion relation

$$\rho_t^{(k+1)} = \rho_t^{(k)} + \underbrace{SV}_{\to \text{from resampling}} + \underbrace{\sigma^2(\{\tau_i\})}_{\to \text{model-dependent and resampling-independent}} .$$

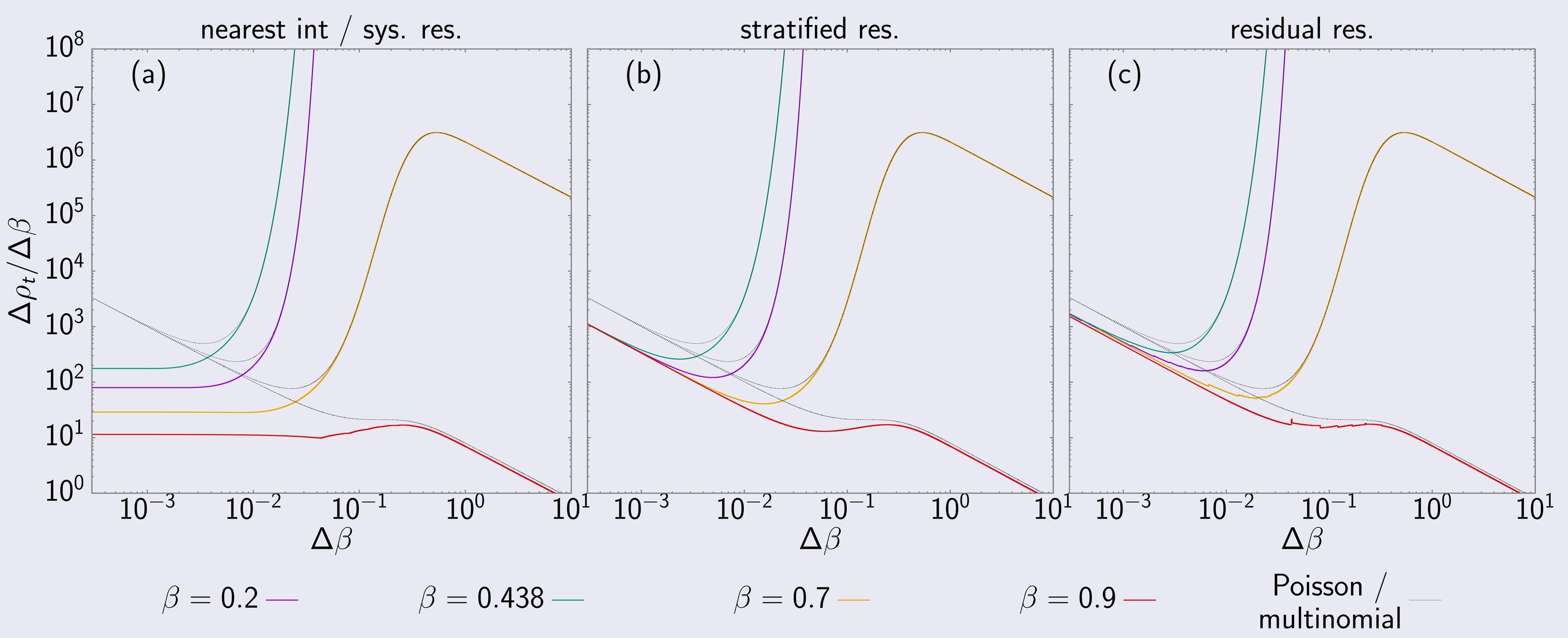


Figure: Increase in average family size per step size as a function of inverse temperature step size for different temperatures and resampling methods.

- family size growth at small (large) steps is governed by sampling variance (weight variance)
- at large $\Delta\beta$ no resampling dependence
- at small $\Delta\beta$: $\Delta\rho_t \approx SV$ and approaches method-specific SV_0 ⇒ divergence if $SV_0 \neq 0$.
- $SV_0 \neq 0$ for nearest int / sys. resampling ⇒ no divergence.
  First order expansion of $\tau(\Delta\beta, E')$ shows that $\Delta\rho_t = SV = \langle |E - \langle E \rangle| \rangle \Delta\beta$ as $\Delta\beta \to 0$.

Conclusion and Outlook

We have demonstrated that

- resampling has a significant effect on the data obtained through population annealing and see a difference in error bars of up to 30% using standard resampling methods, and
- resampling matters mostly when the weight variance of the replicas is of the order of or small compared to sampling variance. Given a constant $\Delta\beta$ this is the case away from criticality or analogously for a fixed β in the case of small $\Delta\beta$'s
- Under most resampling schemes too frequent resampling ("infinitesimal temperature steps") is very disadvantageous.

In a quest of finding the best resampling method nearest integer resampling is a solid choice due to its easy implementation as well as its stability against over-resampling. Systematic resampling is also a good choice (in particular if a constant population size is desired) as it provides the same stability although its implementation is slightly more involved. Follow up questions will include

- adaptive annealing schedule that minimizes ρ_t
- correlation effects that play a role for $\theta < \infty$ that where neglected in the limit $\theta \to \infty$
- extending the notion of resampling cost as $\Delta\rho_t/\Delta\beta$ to also include the number of MC sweeps

References

[1] K. Hukushima and Y. Iba, Population annealing and its application to a spin glass, in AIP Conference Proceedings, Vol. 690, p. 200, AIP, 2003.

[2] L. Y. Barash, M. Weigel, M. Borovský, W. Janke, and L. N. Shchur, Computer Physics Communications **220**, 341 (2017).

Acknowledgments

The project was supported by the Deutsch-Französische Hochschule (DFH-UFA) through the Doctoral College '$\mathbb{L}^4$' under Grant No. CDFA-02-07.