Multicanonical MCMC for Sampling Rare Events

Yukito Iba · Nen Saito · Akimasa Kitajima

Received: date / Revised: date

Abstract Multicanonical MCMC (Multicanonical Markov Chain Monte Carlo; Multicanonical Monte Carlo) is discussed as a method of rare event sampling. Starting from a review of the generic framework of importance sampling, multicanonical MCMC is introduced, followed by applications in random matrices, random graphs, and chaotic dynamical systems. Replica exchange MCMC (also known as parallel tempering or Metropolis-coupled MCMC) is also explained as an alternative to multicanonical MCMC. In the last section, multicanonical MCMC is applied to data surrogation; a successful implementation in surrogating time series is shown. In the appendix, calculation of averages in an exponential family, phase coexistence, simulated tempering, parallelization, and multivariate extensions are discussed.

Keywords multicanonical MCMC · Wang-Landau algorithm · replica exchange MCMC · rare event sampling · random matrix · random graph · chaotic dynamical system · exact test · surrogation

1 Introduction

Multicanonical MCMC (Multicanonical Markov Chain Monte Carlo; Multicanonical Monte Carlo) was introduced in statistical physics at the beginning of the 1990s [Berg and Neuhaus (1991, 1992); Berg and Celik (1992)]; it can be

Y. Iba
The Institute of Statistical Mathematics, 10-3 Midori-cho, Tachikawa, Tokyo 190-8562, Japan E-mail: iba@ism.ac.jp

N. Saito
Graduate School of Arts and Sciences The University of Tokyo 3-8-1 Komaba, Meguro-ku Tokyo 153-8902, Japan E-mail: saito@complex.c.u-tokyo.ac.jp

A. Kitajima
Digital Information Services Division, Digital Information Department, National Diet Library, 1-10-1 Nagata-cho, Chiyoda-ku, Tokyo 100-8924 Japan E-mail: a-kitaji@ndl.go.jp
viewed as a variant of umbrella sampling, whose origin is traced back to the 1970s (Torrie and Valleau (1974)). The Wang-Landau algorithm developed in Wang and Landau (2001b,a) provides an effective realization of a similar idea and many of current studies use their way of implementation.

In these studies, multicanonical MCMC is applied to simultaneous sampling from Gibbs distributions of different temperatures; in terms of statistics, it corresponds to sampling from an exponential family. From this viewpoint, a major advantage of multicanonical MCMC is fast mixing in multimodal problems. It often realizes an order of magnitude improvement on the speed of convergence. Some examples in statistical physics are provided by the references in Sec. 3.3.1; see also review articles, Berg (2000), Janke (1998), Landau et al (2004), Higo et al (2012), Iba (2001).

Recent studies, however, provide another look at this algorithm. Multicanonical MCMC enables an efficient way of sampling rare events under a given distribution. Suppose that rare events of $x$ in a high-dimensional sample space are characterized by the value of statistics $\xi(x)$. Then, in some examples, rare events even with probabilities $P(\xi_0 \leq \xi(x)) \approx 10^{-100}$ are sampled within a reasonable computational time. Also, these probabilities are precisely estimated without additional computation.

This novel viewpoint opens the door to a broad application field of multicanonical MCMC, as well as gives a natural and quick introduction to it. Even though some studies have already introduced multicanonical MCMC as a method of rare event sampling (e.g., Driscoll and Maki (2007), Bononi et al (2009)), it will be useful to conduct another survey with a broad perspective and novel applications. An aim of this paper is to provide such an introduction, including recent results by the authors.

Another aim of this paper is to apply multicanonical MCMC to exact tests in statistics. Multicanonical MCMC is useful for sampling from highly constrained systems; it will be explained in this paper in connection with rare event sampling. Hence, it can be naturally applied to MCMC exact tests (Besag and Clifford (1989), Diaconis and Sturmfels (1998)), where constraints among variables make it difficult to construct Markov chains for efficient sampling from null distributions. As an example, we will discuss surrogation of nonlinear time series; yet it can be generalized to the other MCMC exact tests such as sampling from tables with fixed marginals. Results discussed in Sec. 4 are published here for the first time in English.

The rest of this paper is organized as follows: In Sec. 2, multicanonical MCMC is surveyed as a rare event sampling technique. Starting from general issues on rare event sampling, the use of an exponential family with replica exchange MCMC is discussed as an alternative to multicanonical MCMC. Then, the key idea of multicanonical MCMC is introduced, as well as a concise description of the Wang-Landau algorithm. Sec. 3 provides examples of multicanonical rare event sampling, focus-
Multicanonical MCMC for Sampling Rare Events

2 Multicanonical Sampling of Rare Events

2.1 Rare Event Sampling

We first consider general issues in rare event sampling, namely, importance sampling and the use of exponential families; replica exchange MCMC is also explained. For further details on general frameworks and other approaches, see 

5

6

Bucklew (2004); Rubinstein and Kroese (2007); Rubino and Tuffin (2009).

2.1.1 Importance Sampling

Let us assume that the value of a variable $x$ is randomly sampled from the probability distribution $P(x)$ (or the density $p(x)$ where $x$ is a continuous variable); throughout this paper, we assume that $P(x)$ or $p(x)$ is precisely known.

When we specify target statistics $\xi(x)$ as a function of $x$, “rare events” with rare values of statistics $\xi(x)$ are defined as a set $\mathcal{A} = \{x \mid \xi_0 \leq \xi(x)\}$, where the probability $P(\xi_0 \leq \xi(x))$ takes a small value 5; the constant $\xi_0$ controls the rareness of $x$.

In typical examples, $x$ is a high-dimensional vector variable; thus, even when $x$ is a discrete variable, the probability that $x$ takes a specific value is usually very small. Hence, a small value of $P(x)$ itself does not guarantee its “rareness.”

Our problem is to generate samples of $x$ that satisfy $\xi_0 \leq \xi(x)$ and estimates their probability $P(\xi_0 \leq \xi(x))$. By virtue of the current hardware, we can still complete the task by a direct computation, even when the probability $P(\xi_0 \leq \xi(x))$ takes considerably smaller values such as $10^{-12}$ or even $10^{-100}$. However, when the probability of rare events is much smaller, say, $10^{-12}$ or even $10^{-100}$, it is virtually impossible to deal with the problem by naive random sampling from the original distribution $P(x)$.

5 In this paper, quotation mark “” is used for marking non-technical expressions and technical terms in physics, whereas *italic* is utilized for emphasizing other terms.

6 $P(\xi_0 \geq \xi(x))$ is reduced to the case $P(\xi_0 \leq \xi(x))$ by considering $-\xi(x)$, hence, it is not discussed separately. The probability $P(\xi_0 - \delta \leq \xi(x) \leq \xi_0 + \delta)$ is also considered. In this case, we should maintain an adequate value of $\delta$ and/or consider the relative probabilities using the same value of $\delta$ for a proper definition of “rareness.”
A standard solution to this problem is the use of importance sampling techniques, i.e., we generate samples of \( x \) from another distribution \( Q(x) \), which has a larger probability in the tail region defined by \( \xi_0 \leq \xi(x) \). Hereafter, we assume that \( Q(x) \neq 0 \) for the value of \( x \) satisfying \( P(x) \neq 0 \). Using samples \( x^{(i)}, i = 1, \ldots, M \) from \( Q(x) \), the probability under the original distribution \( P(x) \) is estimated as

\[
P(\xi_0 \leq \xi(x)) = \frac{1}{M} \sum_{i=1}^{M} \left[ \frac{P(x^{(i)})}{Q(x^{(i)})} I(\xi_0 \leq \xi(x^{(i)})) \right],
\]

where \( I \) is defined by

\[
I(\xi_0 \leq \xi(x^{(i)})) = \begin{cases} 
1, & \xi_0 \leq \xi(x^{(i)}) \\
0, & \xi_0 > \xi(x^{(i)}).
\end{cases}
\]

The equality in (2) holds in the \( M \to \infty \) limit. An average of arbitrary statistics \( A(x) \) in the tail region \( \xi_0 \leq \xi(x) \) with weights proportional to \( P(x) \) is calculated as

\[
\mathbb{E}[A(x) \mid \xi_0 \leq \xi(x)] = \frac{\frac{1}{M} \sum_{i=1}^{M} A(x^{(i)}) \frac{P(x^{(i)})}{Q(x^{(i)})} I(\xi_0 \leq \xi(x^{(i)}))}{P(\xi_0 \leq \xi(x))},
\]

with \( M \to \infty \).

In the above formulae, we assume that the variable \( x \) takes discrete values; for a continuous variable \( x \), \( P(x) \) and \( Q(x) \) are replaced with densities \( p(x) \) and \( q(x) \), respectively. Hereafter, we use \( P(x) \) and \( Q(x) \) in both cases and substitute the sum \( \sum \) for the integral \( \int \cdots dx \) when there is no ambiguity.

### 2.1.2 Exponential Family

A critical issue in importance sampling is the choice of the distribution \( Q(x) \). Prior to the introduction of MCMC, there was a severe limitation on the choice of \( Q(x) \); this was because efficient generation of samples is possible only for a simple \( Q(x) \). In contrast, MCMC provides much freedom in the selection of \( Q(x) \).

A strategy, which we will discuss in this paper, is to choose \( Q(x) \) in the form

\[
Q(x) = \frac{G(\xi(x)) P(x)}{\sum \delta G(\xi(x)) P(x)},
\]

where \( G(\xi) \) is an appropriate univariate function and \( \sum \) indicates the sum or integrals over the domain of \( x \). A common and fairly universal choice is \( G(\xi) = \exp(\beta \xi) \), which leads to

\[
Q_\beta(x) = \frac{\exp(\beta \xi(x)) P(x)}{\sum \exp(\beta \xi(x)) P(x)}.
\]

When the base measure \( P(x) \) is uniform, \( Q_\beta(x) \) is interpreted as an exponential family with sufficient statistics \( \xi(x) \) and a canonical parameter \( \beta \); it is also regarded as the Gibbs distribution with energy \( -\xi(x) \) and inverse temperature \( \beta \).

\[\text{5} \quad \text{A different approach to combine importance sampling with MCMC is found in \cite{Botev2013}.}\]
2.1.3 Replica Exchange MCMC

Assuming $Q_\beta(x)$ defined by (4), we can sample regions with larger values of $\xi$ by increasing the value of $\beta$. Thus, in principle, MCMC sampling from $Q_\beta(x)$ with a large value of $\beta$ can efficiently generate rare events defined by $\xi_0 \leq \xi(x)$. When $\beta$ increases, however, the set $\mathcal{A}$ of $x$ defined by $\xi_0 \leq \xi(x)$ often almost disconnects, i.e., it consists of multiple “islands” of $x$ separated by regions with tiny values of $Q_\beta(x)$. Such a multimodal property of $Q_\beta(x)$ obviously leads to slow convergence of MCMC.

In many examples, this difficulty is reduced using replica exchange MCMC, which is also known as parallel tempering or Metropolis-coupled MCMC (Kimura and Taki (1991); Geyer (1991); Hukushima and Nemoto (1996); Iba (2001)). In this algorithm, Markov chains with different values of $\beta$ run in parallel; here we assume $K$ chains with $(\beta_1, \beta_2, \ldots, \beta_K)$. Selecting a pair of chains with $\beta_i$ and $\beta_j$ in a regular interval of steps, current values $x^*_i$ and $x^*_j$ of the states of chains are swapped, with the probability $P_{\text{swap}}$ defined by

$$P_{\text{swap}} = \max \left\{ 1, \frac{Q_\beta_i(x^*_j)Q_\beta_j(x^*_i)}{Q_\beta_i(x^*_i)Q_\beta_j(x^*_j)} \right\} = \max \left\{ 1, \exp(\beta_i - \beta_j)(\xi(x^*_j) - \xi(x^*_i)) \right\}.$$  

Note that the combined probability $\prod_{i=1}^{K} Q_i(x_k)$ is a stationary distribution of the Markov chain defined by the combination of the original MCMC and exchange procedure defined above. This property ensures that replica exchange MCMC realizes a proper sampling procedure at each value of $\beta$.

Exchange of states between chains is introduced for facilitating mixing at large values of $\beta$. Owing to these exchanges, states generated at smaller values of $\beta$ successively “propagate” to chains with larger $\beta$ (Fig. 1). This mechanism is similar to that in the simulated annealing algorithm (Kirkpatrick et al. (1983)) for optimization. An essential difference is that replica exchange MCMC utilizes a time-homogeneous Markov chain designed for sampling from each of the given distributions. In contrast, simulated annealing utilizes a time-inhomogeneous chain; at least in principle, it is not suitable for sampling.

The combination of replica exchange MCMC and $Q_\beta(x)$ given by (4) provides a powerful tool for rare event sampling, which is easy to implement on parallel hardware. However, estimation of the probability of rare events $P(\xi_0 \leq \xi(x))$ under the original distribution $P(x)$ requires some additional consideration. Namely, samples at a single value of $\beta$ are usually not enough for computing $P(\xi_0 \leq \xi(x))$ for all values
of $\xi_0$. Hence, the normalizing constant $Z_\beta = \sum_x \exp(\beta \xi(x)) P(x)$ should be estimated for combining the results at different $\beta$.

These difficulties are well treated using samples at multiple values of $\beta$, which are most naturally obtained as outputs of replica exchange MCMC. Here, however, we omit details; essentially, the same problem in statistical physics is known as the estimation of “density of states”; e.g., a rather sophisticated approach, the multiple histogram method, is explained in Newman and Barkema (1999).

2.2 Multicanonical MCMC

Here we explain multicanonical MCMC, which is the main subject of this paper. First, we define a “multicanonical weight” and discuss the behavior of MCMC with this weight. Then, we introduce adaptive MCMC schemes for realizing the multicanonical weight.

2.2.1 Multicanonical Weight

As already explained, when $Q_\beta(x)$ of (4) is used, some additional computation is required for estimating the probabilities of rare events. The situation can be worse in some examples; a region of $\xi$ is virtually not sampled for any choice of the canonical parameter $\beta$. This may not be typical but possible; see Sec. A.2 for further details.

In contrast, multicanonical MCMC has an advantage in that it provides probabilities such as $P(\xi_0 \leq \xi(x))$ directly as outputs of the MCMC simulation and no additional computation is required. Also, the problem of the missing region of $\xi$ can be avoided, at least in some examples. On the other hand, multicanonical MCMC enables fast convergence in multimodal problems, similar to replica exchange MCMC.

To realize these properties, multicanonical MCMC utilizes $G(\xi)$ defined in the following way. First, we assume that an approximation $\tilde{P}(\xi)$ of $P(\xi)$ is given, in which the marginal probability of $\xi$ is defined as $P(\xi') = \sum_{\xi(x) = \xi'} P(x)$, where $\sum_{\xi(x) = \xi'}$ indicates the sum over $x$ that satisfies $\xi(x) = \xi'$. Then, $G(\xi)$ is given by the inverse $1/\tilde{P}(\xi)$ of $\tilde{P}(\xi)$; more precisely, we define

$$G(\xi(x)) = \begin{cases} c \tilde{P}(\xi(x))^{-1} & \text{if } \xi \in [\xi_{\text{min}}, \xi_{\text{max}}] \\ 0 & \text{else} \end{cases}$$

(5)

where $c$ is an arbitrary constant and $[\xi_{\text{min}}, \xi_{\text{max}}]$ is an interval $\xi$ of interest. Note that values of $\xi$ that give $P(\xi) = 0$ should be excluded from the set $[\xi_{\text{min}}, \xi_{\text{max}}]$. Hereafter, we refer to $G(\xi)$ defined in (5) as a “multicanonical weight.” The corresponding $Q(x)$ is defined as $Q(x) = G(\xi(x)) P(x) / C$, where $C = \sum \xi G(\xi(x)) P(x)$ is the normalizing constant; hereafter, the constant $c$ is absorbed in $C$ and omitted from the expressions.

Formally, it is easy to extend these definitions to the cases where $x$ and $\xi$ are continuous variables; discretized versions are, however, used in many actual implementations. See references in Sec. 2.2.4 for extensions to continuous cases.

This $Q(x)$ may also be referred to as a “multicanonical weight” in the space of $x$. 
Fig. 2 “Flat” marginal $Q(\xi)$ realized by (5) is compared to the marginal $Q_\beta(\xi)$ of the exponential family (4). Left: $Q_\beta(\xi)$ with a fixed value of $\beta$. Center: A series of $Q_\beta(\xi)$ with $(\beta_1, \beta_2, \ldots, \beta_7)$ are printed over one another. Right: $Q(\xi)$ realized by a multicanonical weight defined in (5). In some cases, behaviors very different from these are observed; see Sec. 2.4.

At first sight, the choice of $G(\xi(x))$ shown in (5) does not make sense in practice since the distribution $P(\xi)$ is essentially the one that we want to calculate by the algorithm. In some cases, we guess a form of $P(\xi)$ and use it to approximate the multicanonical weight (Körner et al. (2006); Monthus and Garel (2006)), but it is rather exceptional. Nevertheless, we leave this question for a while and discuss the properties of a multicanonical weight.

Let us tentatively assume an ideal case that $\tilde{P}(\xi)$, which appeared in the multicanonical weight defined in (5), is exactly identical with $P(\xi)$. Then, the marginal distribution $Q(\xi)$ defined by $Q(x) = G(\xi(x))P(x)/C$ is uniform in the interval $[\xi_{\text{min}}, \xi_{\text{max}}]$, excluding the values of $\xi$ that give $P(\xi) = 0$. This is because the multicanonical weight is designed for canceling the factor $P(\xi)$, which is confirmed by a direct calculation as

$$Q(\xi') = \frac{1}{C} \sum_x G(\xi(x))P(x)\delta_{\xi', \xi(x)}$$

$$= \frac{1}{C} G(\xi') \sum_x P(x) \delta_{\xi', \xi(x)} = \frac{1}{C} P(\xi')^{-1} P(\xi') = \frac{1}{C},$$

where $\sum_x$ is the sum over the all possible values of $x$ and $\delta_{\xi, \xi'}$ is defined as

$$\delta_{\xi, \xi'} = \begin{cases} 1 & \xi = \xi' \\ 0 & \xi \neq \xi' \end{cases}.$$  

This “flat” distribution $Q(\xi)$ of $\xi$ realized by a multicanonical weight defined in (5) is illustrated in the rightmost panel of Fig. 2. For comparison, $Q(\xi)$ given by an exponential family (4) is shown in the other two panels of Fig. 2.

2.2.2 MCMC Sampling with a Multicanonical Weight

So far, we discuss a rather obvious conclusion, but it is more interesting to consider a MCMC simulation that samples the corresponding distribution $Q(x) = G(\xi(x))P(x)/C$. To uniformly cover the region $[\xi_{\text{min}}, \xi_{\text{max}}]$, the sample path moves
randomly in the region. In other words, the multicanonical weight realizes a random walk on the axis of the target statistics \( \xi \); this walk has a memory because the states \( x \) does not determined uniquely by the value of \( \xi(x) \).

This behavior enables to obtain the desired properties using a single chain, as shown in Fig.3. First, efficient sampling of a tail region with a large value of \( \xi \) is possible if we choose a sufficiently large \( \xi_{\text{max}} \). On the other hand, fast mixing of MCMC is attained if we include a region of \( \xi \) where transitions between local maxima of \( \xi(x) \) are frequently realized\(^{10}\). Therefore MCMC sampling with a multicanonical weight shares an “annealing” property with replica exchange MCMC.

Finally, we confirm how probabilities of rare events are computed under the original distribution \( P(x) \). We assume that \( x^{(i)}, i = 1, \ldots, M \) are samples from \( Q(x) \) defined by (5). Then, the following expression is derived from (1):

\[
P(\xi_{\text{min}} \leq \xi(x) \leq \xi_{\text{max}}) = C \times \frac{1}{M} \sum_{i=1}^{M} \left[ P(\xi(x^{(i)})) I(\xi_{\text{min}} \leq \xi(x^{(i)})) \right],
\]

where \( I \) is defined by (2); we assume the \( M \to \infty \) limit in this and following equalities. Because the values of \( \xi \) are limited in \( \xi_{\text{min}} \leq \xi \leq \xi_{\text{max}} \) by our definition of the multicanonical weight,

\[
P(\xi_{\text{min}} \leq \xi(x) \leq \xi_{\text{max}}) = C \times \frac{1}{M} \sum_{i=1}^{M} P(\xi(x^{(i)})),
\]

also holds. Hence, we arrive at

\[
\frac{P(\xi_{\text{min}} \leq \xi(x) \leq \xi_{\text{max}})}{P(\xi_{\text{min}} \leq \xi(x) \leq \xi_{\text{max}})} = \frac{\sum_{i=1}^{M} \left[ P(\xi(x^{(i)})) I(\xi_{\text{min}} \leq \xi(x^{(i)})) \right]}{\sum_{i=1}^{M} P(\xi(x^{(i)}))}.
\]

The value of the denominator \( P(\xi_{\text{min}} \leq \xi(x) \leq \xi_{\text{max}}) \) becomes almost the unity when the interval \( [\xi_{\text{min}}, \xi_{\text{max}}] \) contains most of the probability mass; otherwise, in some

\(^{10}\) Such a region corresponds to a “high temperature” region in statistical physics, whereas the tail region with rare events corresponds to a “low temperature” region.
cases, we are mainly interested in relative probabilities. Expectation of arbitrary statistics \( A(x) \) in the tail region \( \xi_0 \leq \xi(x) \leq \xi_{\max} \) is also derived from (3) in a similar manner as

\[
\mathbb{E}[A(x) | \xi_0 \leq \xi(x) \leq \xi_{\max}] = \frac{\sum_{i=1}^{M} A(x(i)) \bar{P}(\xi(x(i))) I(\xi_0 \leq \xi(x(i)))}{\sum_{i=1}^{M} \bar{P}(\xi(x(i))) I(\xi_0 \leq \xi(x(i)))}.
\]  

(6)

2.2.3 Entropic Sampling

Now, we return to the following problem: How to estimate the multicanonical weight \( G(\xi) \) in (5) without prior knowledge? The key idea is to use adaptive Monte Carlo; “preliminary runs” of MCMC are repeated to tune the weight \( G(\xi) \) until the marginal distribution \( Q(\xi) \) becomes almost flat in the interval \([\xi_{\min}, \xi_{\max}]\). After tuning the weight, a “production run” is performed, where \( G(\xi) \) is fixed; this run realizes MCMC sampling with a multicanonical weight. Note that virtually any type of MCMC can be used for sampling in both of these stages.

An important point is that \( G(\xi) \) is an univariate function of a scalar variable \( \xi \), while \( Q(x) = G(\xi(x))P(x)/C \) is defined on a high-dimensional space of \( x \); thus, tuning \( G(\xi) \) is much easier than performing a direct adaptation of \( Q(x) \) itself.

To illustrate the principle, we describe a simple method, sometimes known as entropic sampling (Lee (1993b,a)). First, we consider the histogram \( H \) of the values of \( \xi \). It is convenient to introduce a discretized or binned version \( \tilde{\xi}(x) \) of \( \xi \), which takes an integer value \( \tilde{\xi} \in \{1, 2, \ldots, N_b\} \). Then, the histogram of the values of \( \tilde{\xi} \) in the \( k \)th iteration of the preliminary runs is represented by \( \{H^k(\tilde{\xi})\} \), \( \tilde{\xi} = 1, \ldots, N_b \). We define \( \tilde{H} \) as expected counts in each bin of a flat histogram, which is the target of our adaptation. Also, the weight in the \( k \)th iteration is represented by \( \{G^k(\tilde{\xi})\} \), \( \tilde{\xi} = 1, \ldots, N_b \). Now that the adaptation in the \( k \)th step is expressed as a recursion

\[
G^{(k+1)}(\tilde{\xi}) = G^{(k)}(\tilde{\xi}) \times \frac{H + \varepsilon}{H^k(\tilde{\xi}) + \varepsilon}.
\]  

(7)

Here, a constant \( \varepsilon \) is required for eliminating the divergence at \( H^k = 0 \), which is set to a small value, say, unity. The idea behind this recursion is simple – increase the weight if the counts are smaller than \( \tilde{H} \) and decrease the weight if the counts are larger than \( \tilde{H} \).

The tuning stage of the algorithm is formally described as follows. Here we use \( LG(\tilde{\xi}) = \log G(\tilde{\xi}) \) instead of \( G(\tilde{\xi}) \).

1. Initialize \( LG \) and set parameters.
   - Set \( LG(i) = 0 \) for \( i = 1, \ldots, N_b \).
   - Set the maximum number \( K_{\max} \) of iteration.
   - Set the number \( M_{\max} \) of MCMC steps within each iteration.

11 If \( \xi_{\max} < \tilde{\xi} \) or \( \xi < \xi_{\min} \), it is often convenient to define \( \xi_0 = N_b \) or \( \xi = 1 \), respectively. Another way is to reject the value of \( x \) that satisfy \( \xi_{\max} < \xi(x) \) or \( \xi(x) < \xi_{\min} \) within the Metropolis-Hastings algorithm (Schulz et al (2003)).

12 The constant factor \( \tilde{H} \) is not essential in the following argument when we consider relative weights, but we retain it because it clarifies the meaning of formulae.
1. Set the number $M_s$ of MCMC steps between histogram update.
2. Set a regularization parameter $\varepsilon$ (e.g., $\varepsilon = 1$).
3. Set $H = (M_{\text{max}}/M_s)/N_b$.
4. Set the counter $K$ of iteration to 0.

2. Initialize $H$ and $x$.

- Set $H(i) = 0$ for $i = 1, \ldots, N_b$.
- Initialize the state $x$.
- Set the counter $M$ of MCMC trials to 0.

3. Run MCMC.

- Run $M_s$ steps of MCMC with the weight $P(x) \exp[LG(\xi(x))]$.
- Update the histogram $H$.
  - $H(\xi(x^*)) = H(\xi(x^*)) + 1$, where $x^*$ is the current state. ♠
  - $M = M + M_s$.
- If $M < M_{\text{max}}$ go to Step 3

4. Update the histogram $H$.

- $LG(i) = LG(i) + \log[(\varepsilon + H(\xi(x^*))]/(\varepsilon + H(\xi(x^*)))$ for $i = 1, \ldots, N_b$. ♦
- $K = K + 1$.
- Goto Step 2
- If not and $K_{\text{max}} \leq K$, the algorithm fails.

Note that the update formula (7) is included as a step marked with ✶, while the histogram is incremented in the step marked with ♠.

After completing the above procedure, the production run is performed. If the above algorithm fails to converge, we can increase the numbers $M_{\text{max}}$ and/or $K_{\text{max}}$. Another choice is to reduce our requirement and decrease the value of $\xi_0$, which determines the rareness of the obtained events.

The construction of the histogram can be replaced by other density estimation techniques. In the original studies (Berg and Neuhaus (1991); Berg and Celik (1992); Mezei (1987)), $\log G(\xi)$ is represented by a piecewise linear curve, instead of a piecewise constant curve used in entropic sampling; parametric curve fitting is also utilized. Another useful method is kernel density estimation, which is particularly convenient in continuous and/or multivariate $\xi$ cases; it is also used with the Wang-Landau algorithm explained immediately after Zhou et al (2006).

2.2.4 The Wang-Landau algorithm

Entropic sampling is already sufficient for realizing a multicanonical weight in many problems. In current studies, however, the Wang-Landau algorithm (Wang and Landau (2001b,a)) is often utilized, which provides a more efficient strategy to construct a multicanonical weight.

An essential feature of the Wang-Landau algorithm is the use of a time-inhomogeneous chain in the preliminary runs; that is to say, we change the weights after each trial of MCMC moves instead of changing them only at the end of each iteration consisting a fixed number of MCMC steps. This may lead to an “incorrect” MCMC sampling in the preliminary runs, but it causes no problem if we fix the
weights in the final production run, where we compute the required probabilities and expectations.

In the actual implementation, whenever a state \( x \) with \( \tilde{\xi}^* = \tilde{\xi}(x) \) appears, we multiply a constant factor \( C < 1 \) to the value of weight \( G(\tilde{\xi}^*) \); it reduces the weights of already visited values of \( \tilde{\xi} \), whereas it effectively increases the relative weights of other values of \( \tilde{\xi} \). In parallel, we construct the histogram \( H \) of \( \tilde{\xi} \) that appeared in MCMC sampling. After some steps of MCMC, we reach a “sufficiently flat” histogram; then, a step of iterative tuning of the weights is completed.

When we rerun MCMC where the weight is fixed to the values obtained by this procedure, the run usually does not provide a sufficiently flat histogram of \( \tilde{\xi} \). Then, an iterative method is introduced, i.e., we decrease the value of the constant \( C \) and repeat the procedure in the preceding paragraph. A heuristics proposed in the original papers (Wang and Landau (2001b,a)) is to change \( C \) to \( \sqrt{C} \). After each step of the iteration, the histogram \( H \) is cleared, whereas the values of \( G \) are retained.

Again, we stress that any type of MCMC can be used for sampling at each of these stages; we use familiar Metropolis-Hasting algorithms in the examples considered in this paper. As shown in later sections, however, the choice of moves in Metropolis-Hasting algorithms significantly affects the efficiency of the entire algorithm.

Tuning of the weight by the Wang-Landau algorithm is summarized as shown below. Again we use \( LG(\tilde{\xi}) = \log G(\tilde{\xi}) \) in place of \( G(\tilde{\xi}) \); also we define \( LC = -\log C \) (i.e., with a minus sign).

1. Initialize \( LG \) and \( LC \); set other parameters.
   - Set \( LG(i) = 0 \) for \( i = 1, \ldots, N_b \).
   - Set \( LC > 0 \) (e.g., \( LC = -\log(1/e) = 1 \)).
   - Set the maximum number \( K_{\text{max}} \) of the iteration (e.g., \( K_{\text{max}} = 15 \) or 18).
   - Set the number \( M_{\text{max}} \) of MCMC steps within each iteration.
   - Set the counter \( K \) of iteration to 0.

2. Initialize \( H \) and \( x \).
   - If \( K > K_{\text{max}} \), end.
   - Set \( H(i) = 0 \) for \( i = 1, \ldots, N_b \).
   - Initialize the state \( x \).
   - Set the counter \( M \) of MCMC trials to 0.

3. Run MCMC.
   - Run a step of MCMC with the weight \( P(x) \exp(LG(\tilde{\xi}(x))) \).

4. Modify \( LG \) and update the histogram \( H \).
   - \( LG(\tilde{\xi}(x^*)) = LG(\tilde{\xi}(x^*)) − LC \), where \( x^* \) is the current state. ♠
   - \( H(\tilde{\xi}(x^*)) = H(\tilde{\xi}(x^*)) + 1 \), where \( x^* \) is the current state. ♠

5. Check whether \( H \) is “sufficiently flat.”
   - If so, \( LC = LC/2 \), \( K = K + 1 \) and go to Step[2]
   - If not and \( M < M_{\text{max}} \), \( M = M + 1 \) and go to Step[3]
   - If not and \( M_{\text{max}} \leq M \), the algorithm fails.

Note that update ♠ of the weight \( G \) and increment ♠ of the histogram \( H \) are done simultaneously, in contrast to entropic sampling.
After completing the above procedure, the production run is performed. If this algorithm does not converge or the production run using the obtained weights does not give a flat histogram of $\xi$, what can we do? One possibility is to change the criterion that the histogram $H$ is “sufficiently flat” when we make it more strict as well as increase the value of $M_{\text{max}}$, the convergence may be attained with increasing computational time. Increasing the value of $K_{\text{max}}$ may not be effective when we use the original $\sqrt{C}$ rule for modifying $C$ because the value of $C$ becomes very small at large $K$. Another possibility is to relax our requirement on the rareness and decrease the value of $\xi_0$.

The algorithm presented here still contains a number of ad hoc procedures and should be manually adapted to a specific problem. It, however, provides solutions to problems otherwise difficult to treat. On the other hand, many modifications of the algorithm are proposed. Examples of treating continuous variables are seen in Shell et al (2002); Liang (2005); Zhou et al (2006); Atchadé and Liu (2010). Following authors have criticized the $\sqrt{C}$ rule and have proposed modified algorithms: Belardinelli and Pereyra (2007a,b); Liang et al (2007); Zhou and Su (2008); Atchadé and Liu (2010). Convergence of the algorithms is analyzed in Lee et al (2006); Belardinelli and Pereyra (2007a), while rigorous mathematical proofs are discussed in Atchadé and Liu (2010); Jacob and Ryder (2011); Fort et al (2012). Bornn et al (2011) proposed an automatic procedure including adaptation of step and bin size.

3 Examples of Rare Event Sampling by Multicanonical MCMC

Here we discuss two examples of applications of multicanonical MCMC, rare event sampling in random matrices and chaotic dynamical systems. Other applications in physics, engineering, and statistics are briefly surveyed.

3.1 Rare Events in Random Matrices

A pioneering study on rare events in random matrices with multicanonical MCMC is found in Driscoll and Maki (2007), which computes large deviation in growth ratio, a quantity relevant to numerical difficulty of treating matrices. Details of the results in this subsection are discussed in Saito et al (2010) and Saito and Iba (2011).

3.1.1 Largest Eigenvalue

Distributions of the largest eigenvalue $\lambda_{\text{max}}$ of random matrices are of considerable interest in statistics, ecology, cosmology, physics, and engineering. Small deviations are well studied in this problem, whose milestone is the celebrated $N^{1/6}$ law by Tracy and Widom (1994, 1996). Here we are interested in the numerical estimation of

13 Usually, in the Wang-Landau algorithm, this criterion should be much severer than the requirement on flatness of the histogram expected in the final production run; e.g., 1 or 2 percent of deviation for counts in each bin.
large deviations; the present analytical approach to large deviations is limited to specific types of distributions (Dean and Majumdar (2008); Majumdar and Vergassola (2009)). Specifically, the probability \( P(\lambda_{\max} < 0) \) that all eigenvalues are negative is important in many examples, because it is often related to the stability of the corresponding systems (May (1972); Aazami and Easther (2006)).

In Saito et al (2010), multicanonical MCMC is applied to this problem. Rare events whose probability \( P(\lambda_{\max} < 0) \) is as small as \( 10^{-200} \) are successfully sampled for the size \( N \leq 30 \) (or 40) of the matrices. Examples of the results in Saito et al (2010) are shown in Figs. 4 and 5. In Fig. 4 the probability \( P(\lambda_0^{\max} < \lambda_{\max} \leq \lambda_0^{\max} + \delta) \) is plotted against the values of \( \lambda_0^{\max} \) with a small binsize \( \delta \). Fig. 5 shows the probability \( P(\lambda_{\max} < 0) \) that all eigenvalues are negative.

The proposed method is quite general and can be applied to random matrices whose components are sampled from an arbitrary distribution, or even random sparse matrices, for which no analytical solution is available. More on these results are discussed in Saito et al (2010), as well as details of the proposed algorithm.

An important lesson from this example is that we should be careful while choosing the moves in the Metropolis-Hasting algorithm. If we generate candidates using conditional distributions of the original distribution, such as the Gaussian distribution for each component in GOE, the algorithm fails in some cases. This occurs because such a method cannot generate candidates with very large deviations in a component. This difficulty is avoided by the use of a random walk Metropolis algorithm with an adequate step size, as discussed in Saito et al (2010).

The most time-consuming part of proposed algorithm is diagonalization procedure required for each step of MCMC; the Householder method is used here. It can be improved by the use of a more efficient method for calculating the eigenvalue \( \lambda_{\max} \).
Fig. 5 Probabilities $P(\lambda_{\text{max}} < 0)$ obtained by proposed method and simple random sampling method are shown against $N$; latter is available only for small $N$. Left: GOE; curve indicates a quadratic fit to the results with Coulomb gas representation (Dean and Majumdar 2008). Right: an ensemble of matrices whose components are uniformly distributed; curve indicates the probability for GOE with the same variance. [Adapted from N. Saito, Y. Iba, and K. Hukushima, Multicanonical sampling of rare events in random matrices, Physical Review E 82, 031142 (2010), © 2010 American Physical Society]

3.1.2 Random Graphs

Search for rare events in random graphs is also an interesting subject. An undirected graph is represented by the corresponding adjacency matrix, whose components take the values in the set $\{0, 1\}$. For a $k$-regular graph, the maximum eigenvalue takes a fixed value equals to $k$, and hence it is not interesting. On the other hand, the spectral gap $\lambda_{\text{gap}}$, defined as the difference of the maximum and second eigenvalue, represents significant properties of the corresponding graph. Specifically, graphs with larger values of the spectral gap are called Ramanujan graphs or expanders and have applications in network theory; Ramanujan graphs have interesting properties for communications and dynamics on networks (see references in Donetti et al (2006); Saito and Iba (2011)).

In earlier studies, Donetti et al (2005, 2006) optimized the spectral gaps of graphs by simulated annealing; in their algorithm, a pair of edges of the graph is modified in each Metropolis-Hasting step. Using this method, they showed that expanders with interesting structures automatically appear.

Saito and Iba (2011) applied multicanonical MCMC to this problem; they defined Metropolis-Hasting update in a similar way as that in Donetti et al (2005, 2006) and used the Wang-Landau algorithm for realizing multicanonical weights. Examples of the obtained graphs are shown in Fig. 6 while Fig. 7 gives probability $P(\lambda_{\text{gap}}^0 < \lambda_{\text{gap}})$ as a function of $\lambda_{\text{gap}}^0$ and the size $N$ of matrices. See Saito and Iba (2011) for further details.

3.2 Rare Events in Dynamical Systems

Rare events in deterministic dynamical systems are important both in theory and application (Ott (2002); Beck and Schlogl (1995)). An example is a quantitative study on tiny tori embedded in a “chaotic sea” of Hamiltonian dynamical systems, which is a familiar subject in this field. Numerical effort required for uncovering these tiny
structures dramatically increases with the dimension of the system. Therefore, it is natural to introduce MCMC and other stochastic sampling methods to this field. Studies on MCMC search for unstable structures in dynamical systems are found in Sasa and Hayashi (2006); Yanagita and Iba (2009); Geiger and Dellago (2010) and references therein.

Kitajima and Iba (2011) applied multicanonical MCMC to the study of dynamical systems. In the proposed algorithm, a measure of the chaoticity of a trajectory is defined as a function of the initial condition, which corresponds to statistics representing rareness. Then, the Metropolis-Hastings update is defined as follows:

15 Sequential Monte Carlo-like algorithms are also used; see Tailleur and Kurchan (2009); Laffargue et al (2013) and references therein.

16 Here the chaoticity is defined as the number of iterations required for the divergence of perturbed trajectories; the algorithm according to this definition is stable on finite precision machines.
(1) Perturb the initial condition, (2) simulate a fragment of trajectory from the new initial condition, (3) calculate the chaoticity of the trajectory and reject/accept the new initial condition using the current weight. Then, the entire algorithm is defined as multicanonical MCMC with the Wang-Landau algorithm for tuning the weight.

Again, the choice of moves in the Metropolis-Hastings algorithm is important; here we sample a perturbation to the initial conditions from a mixture of uniform densities with different order of widths. This idea, taken from Sweet et al. (2001), seems essential for sampling from fractal-like densities; see Kitajima and Iba (2011) for details.

In Kitajima and Iba (2011), sampling of tiny tori in the chaotic sea of a four-dimensional map

\[
\begin{align*}
  u_{n+1} &= u_n - \frac{K}{2\pi} \sin(2\pi v_n) + \frac{b}{2\pi} \sin(2\pi(y_n + y_n)) \\
  v_{n+1} &= v_n + u_{n+1} \\
  x_{n+1} &= x_n - \frac{K}{2\pi} \sin(2\pi y_n) + \frac{b}{2\pi} \sin(2\pi(v_n + y_n)) \\
  y_{n+1} &= y_n + x_{n+1}
\end{align*}
\]

is studied, where \( K \) and \( b \) are constants that characterize the map. An example of tiny tori found by the proposed method is shown in Fig. 8. In addition, the relative volume of initial conditions that lead to trajectories of the given order of "chaoticity" are successfully estimated by the algorithm; i.e., the proposed method is not only useful for the search but also provides quantitative information on rare events in dynamical systems; see Fig. 2 of Kitajima and Iba (2011).

3.3 Other Applications

The rest of this section briefly describes other fields of applications of multicanonical MCMC.
3.3.1 Statistical Physics

Multicanonical MCMC was originally developed for sampling from Gibbs distributions in statistical physics. Hence, a number of studies in this field have successfully applied it to problems where simple MCMC is virtually disabled by slow mixing. Some typical examples are studies on the Potts and other classical spin models (Berg and Neuhaus (1992); Wang and Landau (2001a); Zhou et al (2006)), spin glass models (Berg and Celik (1992); Wang and Landau (2001a)), and liquid models (Shell et al (2002); Calvo (2002)). Multicanonical MCMC is also used for the study of biomolecules; see Mitsutake et al (2001); Higo et al (2012) for full-atom protein models and Chikenji et al (1999) for lattice protein models. More examples are found in the review articles mentioned in Sec.1.

On the other hand, the use of multicanonical MCMC for other types of rare event sampling in physics is a recent challenge. Hartmann (2002) introduced the idea of rare event sampling by MCMC to the community of physicists. Körner et al (2006) and Monthus and Garel (2006) applied MCMC to the sampling of disorder configurations that gives large deviations in the ground state energies; in these studies, modifications of the Gumbel distribution are used for approximating multicanonical weights. Subsequently, Hukushima and Iba (2008) and Matsuda et al (2008) applied the Wang-Landau algorithm to the study of Griffiths singularities in random magnets, which is known to be sensitive to rare configurations of impurities; Wolfsheimer and Hartmann (2010) discussed RNA secondary structures. In these studies, any prior knowledge on the functional form of $\tilde{P}(\xi)$ is assumed.

3.3.2 Optical Telecommunication and Related fields

Multicanonical MCMC is intensively used for rare event sampling in optical telecommunication and related fields. After a pioneering work by Yevick (2002), a number of applications appeared; see e.g., Holzlöhner and Menyuk (2003), and a recent review, Bononi et al (2009). Sampling of rare noises that cause failures of error correction is discussed in Holzlöhner et al (2005) and Iba and Hukushima (2008), which can be useful for predicting the performance of error-correcting codes.

3.3.3 Statistics

Multicanonical MCMC and related algorithms, specifically, the Wang-Landau algorithm and its generalizations, increasingly attract the attention of statisticians. Liang (2005) introduced the Wang-Landau algorithm to statistics. Atchadé and Liu (2010) and Chopin et al (2012) developed closely related algorithms and tested them in examples of Bayesian inference and model selection. Bornn et al (2011) and Kastner et al (2013) also discussed applications in Bayesian statistics; Kwon and Lee (2008) treated a target tracking problem. Yu et al (2011) (also Liang et al (2010))

17 In terms of physics, it corresponds to sampling of the “quenched disorder,” whereas conventional applications in physics deal with sampling from the Gibbs distribution of thermal disorder.
discussed hypothesis testing using stochastic approximation Monte Carlo. In the following section, we will discuss exact tests and data surrogation as an application field of multicanonical MCMC for constrained systems.

4 Sampling from Constrained Systems and Hypothesis Testing

Sampling from highly constrained systems and combinatorial calculations are discussed here as a variation of the theme of rare event sampling. Exact tests and data surrogation are introduced as an application field of this idea, where efficient sampling from constrained systems is essential.

For general issues on Monte Carlo approximate counting, see Jerrum and Sinclair (1996), Rubinstein and Kroese (2007), and Rubino and Tuffin (2009).

4.1 MCMC Sampling from Constrained Systems

MCMC sampling is difficult when constraints exist among stochastic variables. In such cases, it is often not easy to find a set of Metropolis-Hastings moves that realizes an ergodic Markov chain without violating the constraints. For example, considerable effort is devoted to find ergodic moves for contingency tables with fixed margins and other constraints (Diaconis and Sturmfels (1998); Takemura and Aoki (2004)). Although partial success is obtained using highly sophisticated mathematics, the problem becomes increasingly difficult when problem complexity increases.

Yet another general strategy for dealing with highly constrained systems is an introduction of “soft constraints.” First, given constraints \( f_i(x) = 0, i = 1, \ldots, L \), we define statistics \( \xi(x) \) of the state variables \( x \) that satisfy the following conditions:

1. \( \xi(x) \geq 0 \) and
2. \( \xi(x) = 0 \), if and only if \( x \) satisfy \( f_i(x) = 0 \) for all \( i \).

A simple example of such statistics is

\[
\xi(x) = \sum_{i=1}^{L} c_i |f_i(x)|^\alpha.
\]

Here \( \alpha > 0 \) and \( c_i > 0 \) are arbitrary constants; \( \alpha = 1 \) is usually better than \( \alpha = 2 \) because the maximum value of \( \xi \) is smaller when \( \alpha = 1 \). Then, a finite value of \( \xi \) represents soft constraints, whereas \( \xi = 0 \) corresponds to the original hard constraints. Random sampling of the value of \( x \) usually gives a large value of \( \xi(x) \); hence \( \xi(x) = 0 \) can be regarded as a “rare event.”

At this point, we introduce multicanonical MCMC with target statistics \( \xi \) and samples rare events \( \xi(x) = 0 \) (or, for a continuous variable \( x \), \( \xi(x) \approx 0 \)). Then, after tuning weights with the Wang-Landau algorithm, a production run provides samples that (nearly) satisfy the constraints \( f_i(x) = 0 \) (or \( f_i(x) \approx 0 \)) for all \( i \). Note that a similar strategy can be implemented using a combination of an exponential family with sufficient statistics \( \xi(x) \) and replica exchange MCMC; in this case, a large value of \( \beta \) corresponds to hard constraints.

\[\text{See also Jacobson and Matthews (1998) for an algorithm specialized to Latin squares; it partially utilized a soft constraint strategy.}\]
Some references are as follows: Pinn and Wieczerkowski (1998) introduced replica exchange MCMC with soft constraints to this field; the number of magic squares of size $6 \times 6$ is estimated in their paper; Kitajima and Kikuchi (private communication) extended it to $30 \times 30$ using multicanonical MCMC. Hukushima (2002) estimated the number of N-queen configurations by replica exchange MCMC, while Zhang and Ma (2009) treated N-queen and Latin squares using a hybrid of simulated tempering (Sec. A.3) and the Wang-Landau algorithm; they dealt with Latin squares up to size $100 \times 100$. Fishman (2012) proposed an approach based on soft constraints for counting contingency tables; conventional MCMC is used in his paper.

4.2 Application to Hypothesis Testing

Here we discuss how multicanonical MCMC (and also replica exchange MCMC) can be useful for exact tests and data surrogation; the proposed method is tested with a simple example of time series.

4.2.1 MCMC Exact Tests

MCMC is useful for implementing statistical tests with a complicated null distribution. Particularly important cases occur when the null distribution is a distribution conditioned with a set of statistics $\zeta_i$. In these cases, the null hypothesis is represented as the uniform distribution of $x$ on the set defined by $\zeta_i(x) = \zeta_i^o$, $i = 1, \ldots, L$, where $x$ is a stochastic variable and $\zeta_i^o$ is the value of statistics $\zeta_i$ corresponding to the observed data. For a continuous variable $x$, this condition can be relaxed as

$$|\zeta_i(x) - \zeta_i^o| < \varepsilon_i, \quad i = 1, \ldots, L,$$

where $\varepsilon_i$ is a constant with a small value.

A prototype of such a test is Fisher's exact test of contingency tables (Agresti (1992)), where marginals of the table correspond to $\zeta_i$'s; a number of extended versions exist and MCMC algorithms with complicated Metropolis moves have been developed for them, as mentioned in the previous section. Besag and Clifford (1989) described a test where an Ising model on the square lattice represents the null hypothesis.

In our view, it is natural to introduce the “soft constraint” strategy described in Sec. 4.1 to this problem. When we define the statistics $\xi(x)$ as $\xi(x) = \sum_{i=1}^{L} |\zeta_i(x) - \zeta_i^o|$, it is straightforward to apply multicanonical MCMC for sampling $x$ that satisfies the condition $\zeta_i(x) = \zeta_i^o$, $i = 1, \ldots, L$ or its generalization (8). This strategy is quite general and can be applied to a variety of MCMC hypothesis testing.

19 “Self-avoidingness” of random walk is also well treated by the soft constraint strategy discussed here; see Vorontsov-Velyaminov et al (1996, 2004); Iba et al (1998); Chikenji et al (1999); Shirai and Kikuchi (2012).

20 As mentioned in the previous section, Yu et al (2011); Liang et al (2010) also discussed hypothesis testing with stochastic approximation Monte Carlo, which can be regarded as a version of multicanonical MCMC in this case. They, however, focused on the problem of calculating small p-values; it differs from our idea of using multicanonical MCMC as a sampler from highly constrained systems.
4.2.2 Data Surrogation

In nonlinear dynamics and neural science, statistical tests for time series based on (8) are well developed (Schreiber and Schmitz (2000)). They are called as “data surrogation” methods, and samples from null distributions defined by (8) are called as “surrogates” of the original data. An example of the problem where surrogation is intensively used is testing of statistical properties of neural spike trains (Gruen and Rotter (2010)).

In conventional approaches, surrogates are generated by partial randomization of the original data. For example, if the phase of time series data \( x(t), t = 1, \ldots, N \) is randomized after the complex Fourier transform, then its inverse transform has the same sets of correlation functions

\[
C(x; \tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} x(t)x(t+\tau)
\]

as the original time series and is considered as a surrogate that maintains the value of sufficient statistics \( \zeta(x) = C(x; \tau) \). Although a quick solution is provided in this case, solutions to general cases are only found on a case-by-case basis, and it becomes increasingly difficult as the complexity of the problems increases.

Therefore, Schreiber proposed a general idea of regarding data surrogation as an optimization problem (Schreiber (1998); Schreiber and Schmitz (2000)). According to this idea, generating a surrogate is equivalent to finding a solution of (8), which can be treated by a general-purpose optimization algorithm, e.g., simulated annealing. An application of this idea in neural science is found in Hirata et al (2008).

This was an epoch-making idea in this field; randomization via a clever idea was no longer required, being replaced by a routine procedure at the cost of computational time. However, in data surrogation, we want to generate a sample (or a set of samples) unbiasedly selected from the null distribution defined by (8), and not obtain a sample that satisfies (8). Therefore, applying multicanonical MCMC seems a better choice. Hence, we again arrive at the idea of exact testing with multicanonical MCMC.

4.2.3 Example

Let us illustrate the idea of “multicanonical surrogation” using an example from Schreiber (1998). In this example, the problem is to generate artificial time series \( x = \{x_1, x_2, \ldots, x_N\} \) by permuting the original time series \( x' = \{x'_1, x'_2, \ldots, x'_N\} \) given as observed data. The constraint is to maintain the correlation functions \( C(x; \tau) \)

To be precise, we should assume a periodic boundary condition and change the upper limit of the summation from \( N - \tau \) to \( N \); it is not be a perfect solution for other boundary conditions; see Schreiber (1998) for details.

Results in this subsection (including Figs. 9 and 10) appeared in IEICE Technical Report IBISML2011-7(2011-06) in Japanese, as a report without peer review. It has never been published in English.

A quick solution is present for this problem, but it is not a perfect one unless the periodic boundary condition is assumed, as mentioned in the footnote of the previous subsection.
Multicanonical MCMC for Sampling Rare Events

Fig. 9 Left: Frequency of the occurrence of $\xi$ in the production run of multicanonical MCMC. Horizontal and vertical axes correspond to the observed frequency in the given bins and value of $\xi$, respectively. Right: Probability density of $\xi$. Vertical and horizontal axes correspond to the estimated log-probabilities and value of $\xi$, respectively; a set of bins used in the left panel is also applied in the right panel for defining probabilities. A spike in the rightmost bin corresponds to accumulated probability of larger values of $\xi$.

defined as \((9)\) are nearly equal to the original correlation functions $C(x^o; \tau)$ for $\tau = 1 \ldots T$; here the constant $0 < T < N$ is the maximum of the delay $\tau$, where we expect coincidence of the correlation.

Here the statistics $\xi(x) = \sum_{\tau=1}^{T} |C(x^o; \tau) - C(x; \tau)|$ is used to define multicanonical MCMC. It is zero if and only of $C(x^o; \tau) = C(x; \tau)$ for all $1 \leq \tau \leq T$. Then, Metropolis-Hastings moves are defined by the swap of a randomly selected pair, i.e., a pair $i$ and $j$ is selected by a random number in each step and a new candidate $x^{\text{new}}$ is generated by $x^{\text{new}}_i = x_j$ and $x^{\text{new}}_j = x_i$, without changing other components; here $\{x_i\}$ is assumed to be initialized as a random permutation of $\{x^o_i\}$.

In the following experiment, we consider time series $x^o$ of length $N = 500$ generated by nonlinear observation of a linear AR process $y$, i.e.,

$$x^o_t = y^3_t, \quad y_{t+1} = 0.3y_t + \eta_t, \quad \eta_t \sim N(0, 2.0^2).$$

Multicanonical MCMC is designed for realizing an approximately flat distribution of $\xi$ in the interval $[0, 6847]$, which is divided into 201 bins. The Wang-Landau algorithm with $K_{\text{max}} = 15$ is used to tune the weight; the $\sqrt{C}$ rule is utilized. At each step of iteration, we run MCMC until counts in each bin coincide with the value for the uniform histogram within 2% accuracy. Total number of Metropolis trials are $2.1 \times 10^8$, of which $5.0 \times 10^7$ are used for the final production run.

Results of this experiment are shown in Figs. 9 and 10. In Fig. 9, the distribution of $\xi$ realized in the production run and the estimated log-density of $\xi$ are shown. The former is not quite flat in a non-logarithmic scale, but enough to ensure efficient production of the desired samples. According to the right panel of Fig. 9, the probability of obtaining a sample within the bin $\xi \approx 0$ is estimated to be as small as $10^{-25}$, assuming random sampling of $x$.

In Fig. 10, the quality of the obtained samples is examined. In the left panel, three samples in the last bin $\xi \approx 0$ are shown, which are considerably different from one another. In the right panel, correlation functions $C(x^{(k)}, \tau)$ are calculated for each

\[\text{Here we round the value of } \xi \text{ to } \xi^{\text{max}} \text{ when it exceeds } \xi^{\text{max}} \text{ instead of rejecting the candidate; it causes a spike at the right edge of the density in the right panel of Fig. 9.}\]
of 1171 samples $x^{(k)}$, $k = 1, \ldots, 1171$, fallen in the bin $\xi \approx 0$ and compared to the original $C(x', \tau)$, which shows extremely good agreement.

![Figure 10](image)

**Fig. 10** Left: Surrogate data generated by the proposed method. Uppermost series correspond to the original data and other three are surrogates. Vertical and horizontal axes correspond to $x_t$ and $t$, respectively. Right: Comparison of correlation functions. Vertical and horizontal axes correspond to the value of correlation functions and delay $\tau$, respectively. Line represents $N \cdot C(x', \tau)$, which corresponds to the original data, while black dots represent sets of $N \cdot C(x^{(k)}, \tau)$ obtained from surrogated data; $N = 500$ is the length of time series. Results of 1171 samples are printed over each other; hence, symbols are almost overlapping.

5 Summary and Discussions

In this paper, we discussed rare event sampling using multicanonical MCMC. Two different methods of tuning the weight: entropic sampling and the Wang-Landau algorithm, are explained. Then, examples on random matrices, random graphs, chaotic dynamical systems, and data surrogation are shown. We hope our exposition will be useful for the exploration of further novel applications of multicanonical MCMC.

A Appendix

A.1 Multicanonical MCMC for Exponential Family

We begin this paper with a history of multicanonical MCMC; it was originally developed as a method for sampling from Gibbs distributions, or an exponential family. Here, we briefly discuss how to use multicanonical MCMC for this original purpose.

Note that not all 1171 samples are independent; some additional test is needed for estimating the number of independent samples in our run.
Assume that we want to compute the expectation \( \mathbb{E}_\beta[A(x)] = \sum_i A(x) \exp(-\beta \xi(x))/Z\) of statistics \( A(x) \) from the output \( x^{(i)}; i = 1, \ldots, M \) obtained from multicanonical MCMC that realizes an almost flat marginal of \( \xi(x) \) in a “sufficiently wide” interval \([\xi_{\text{min}}, \xi_{\text{max}}]\). Then, for \( M \to \infty \), the desired expectation is computed by the reweighting formula \(26\):

\[
\mathbb{E}_\beta[A(x)] = \frac{\sum_{i=1}^M A(x^{(i)}) \exp(-\beta \xi(x^{(i)}))}{\sum_{i=1}^M \exp(-\beta \xi(x^{(i)}))}.
\]

It is easy to derive this expression \(27\), considering that the multicanonical weight is proportional to \( P(\xi(x^{(i)})) \). Expression \(10\), however, is quite unusual in the sense that we can use it for a broad range of \( \beta \) where the interval \([\xi_{\text{min}}, \xi_{\text{max}}]\) covers a necessary region. Using this property, multicanonical MCMC simultaneously gives the expectations \( \mathbb{E}_\beta[A(x)] \) for all \( \beta \), through a single production run of a single chain. This is because a multicanonical weight gives a flat distribution of \( \xi \) that has a considerable overlap with the distribution \( \exp(-\beta \xi(x))/Z \) for any value of \( \beta \), which is intuitively understood from the left panel in Fig. 11.

If we consider a similar reweighing that uses outputs of MCMC at \( \beta' \) for computing the expectation at a different \( \beta \), it is practically impossible for a high-dimensional \( x \) unless the difference \( |\beta' - \beta| \) is very small. It is because the overlap of the distributions virtually vanishes as shown in the right panel of Fig. 11 in such cases, the variance of summands on the right hand side of \(10\) drastically increases.

### A.2 First-Order Transition and “Phase Coexistence”

As already mentioned in the main text, there are examples in which a region of \( \xi \) is virtually not realized for any choice of the canonical parameter \( \beta \) of the exponential family with sufficient statistics \( \xi \). The marginal distribution of \( \xi \) has multiple peaks in this region of \( \beta \), as illustrated in Fig. 12. Such examples naturally appear in statistical physics, when we study “phase coexistence” phenomena near first-order phase transitions \(28\). On the other hand, it seems the significance of such phenomena in statistics and engineering has not been fully explored.

In such cases, distributions defined by multicanonical weights are not well approximated by a mixture of the members of the corresponding exponential family; this is easily understood by considering Fig. 12. Hence, the advantage of replica exchange MCMC is limited because the sample path is blocked by the gap of \( \xi \), while multicanonical MCMC can, in principle, do better. Both methods, however, seem to fail in very difficult cases; see [Iba and Takahashi 2005].

---

26 To use this formula for an off-line calculation of the average of \( A(x) \), the values of \( \xi(x) \) and \( A(x) \) should be recorded as pairs in the simulation, like \( (\xi(x^{(i)}), A(x^{(i)})) \), \( i = 1, \ldots, M \).

27 Note that \(10\) becomes \(8\), if we substitute \( R(\xi \leq \xi(x^{(i)})) \) for \( \exp(-\beta \xi(x^{(i)})) \).

28 The ice and water coexist at 0 °C; that is, both of them correspond to the same \( \beta \) but the values of average energy \( -\mathbb{E}(\xi(x)) \) are different.
Marginals of $\xi$ with different values of $\beta$ in the case of phase coexistence; the horizontal axis corresponds to the sufficient statistics $\xi$. These curves are obtained from the 10-states Potts model (Berg and Neuhaus (1992)), which consists of discrete variables $\{x_i\}, x_i \in \{1, \cdots, 10\}$ on a square lattice; they are computed by reweighting of the outputs of a single production run of multicanonical MCMC. If $x = \{x_i\}$ belongs to an “ordered” component, most variables $x_i$ take the same value. In contrast, their values are almost random in the “disordered” component. In the disordered component, $\xi(x)$ takes smaller values, but the number of $x$ that belongs to the component is large; hence, the total probability is comparable in both components.

Random walk of $\beta$ in simulated tempering. The vertical axis corresponds to the value of $\xi$, whereas the horizontal axis for each sub-chart schematically represents a high-dimensional space of $x$. The values of $\beta$ are assumed to increase from left to right; the shading represents the corresponding changes of high probability regions. Note that a stochastic variable $\beta$ is updated by MCMC, retaining the value of the state $x$ at that time; i.e., a separated procedure for changing $\beta$ is required for simulated tempering.

A.3 Simulated Tempering

The “third” method, simulated tempering (Marinari and Parisi (1992); Geyer and Thompson (1995)), or expanded ensemble Monte Carlo (Lyubartsev et al (1992)) is briefly explained here. Practically, we recommend choosing between multicanonical MCMC and replica exchange MCMC. Simulated tempering, however, provides an idea that interpolates these two algorithms and is conceptually important. The idea is simple — MCMC sampling of $(x, \beta)$ from the combined distribution

$$P(x, \beta) = \frac{\exp(\beta \xi(x))}{Z_\beta} \pi(\beta) = \exp(\beta \xi(x) - \log Z_\beta + \log \pi(\beta)),$$

defined as a mixture of the exponential family. Hereafter, we choose a “pseudo prior” $\pi(\beta)$ as an uniform density on $[\beta_{\text{min}}, \beta_{\text{max}}]$, resulting in a random walk of $\beta$ that uniformly covers the interval $[\beta_{\text{min}}, \beta_{\text{max}}]$; see Fig. 13. This behavior is similar to that of multicanonical MCMC, but here $\beta$ is a variable updated in a separate step of MCMC; in case of multicanonical MCMC, a random walk of $\xi(x)$ is induced by the update of the state $x$.

This paper introduced an idea similar to simulated tempering in an even more general framework.
Although this concept is simple, a difficulty arises because the MCMC update of $\beta$ requests the value of $Z_\beta$ as a function of $\beta$, which is unknown in most cases. Hence, we should introduce the estimation of $Z_\beta$ using repeated preliminary runs, which is similar to the weight tuning procedure in multicanonical MCMC. See the above references, as well as Zhang and Ma (2007), which introduced a method like the Wang-Landau algorithm.

A.4 Implementation on Parallel Hardware

Replica exchange MCMC is naturally parallelizable. Then, how can we efficiently implement multicanonical MCMC on parallel hardware? A simple solution is parallelization of the weight tuning stage. That is, a set of preliminary runs is performed in parallel, each of which runs on a CPU; they share a histogram where total numbers of visits to each value of $\xi$ is recorded. Some variants of this idea are discussed in, e.g., Zhan (2008). If we want to go beyond these schemes, something more intricate is required. For example, the range of $\xi$ is divided into a set of intervals and a multicanonical weight is realized in each of them; see an example in Wang and Landau (2001a) and Mitsutake et al (2001), in the latter study, exchange of states between neighboring intervals is incorporated.

A.5 Multivariate Extensions

Multicanonical MCMC samples a high-dimensional $x$, while adaptation of the weight is performed in one-dimensional space of $\xi$. It is possible to introduce a “multivariate multicanonical weight,” which realizes an almost uniform density in a region of two-dimensional ($\xi_1, \xi_2$) or even three-dimensional ($\xi_1, \xi_2, \xi_3$) spaces, where $\xi_k$ is a function of $x$. Examples of such extensions are found in Shteto et al (1997), Iba et al (1998), Chikenji et al (1999), Chikenji and Kikuchi (2000), Zhou et al (2006). Usually, adaptation of weights in a multivariate case is more difficult than that in a univariate case, because of the sparseness of the data collected in the preliminary run. Zhou et al (2006) proposed the use of kernel density estimation for this problem.

Acknowledgements The authors would like to thank Koji Hukushima for helpful discussions and permission for the use of figures in Saito et al (2010). This work was supported by JSPS KAKENHI Grant Number 22500217.

References

Aazami A, Easther R (2006) Cosmology from random multifield potentials. Journal of Cosmology and Astroparticle Physics 3:013
Agresti A (1992) A survey of exact inference for contingency tables. Statistical Science 7(1):131–153
Atchadé YF, Liu JS (2010) The Wang-Landau algorithm in general state spaces: applications and convergence analysis. Statistica Sinica 20(1):209–233
Baumann B (1987) Noncanonical path and surface simulation. Nuclear Physics B 285:391–409
Beck C, Schlögl F (1995) Thermodynamics of Chaotic Systems: An Introduction. Cambridge University Press
Belardinelli R, Pereyra V (2007a) Wang-Landau algorithm: A theoretical analysis of the saturation of the error. The Journal of Chemical Physics 127:184105
Belardinelli RE, Pereyra VD (2007b) Fast algorithm to calculate density of states. Physical Review E 75:046701
Berg BA (2000) Introduction to multicanonical Monte Carlo simulations. Fields Inst Commun 26(1):1–24

\textsuperscript{30} The normalizing constant (partition function) $Z_\beta$ is not required for replica exchange MCMC, because it cancels in the Metropolis-Hastings ratio necessary for deciding accept/reject of the swap of the states between chains; it is an essential advantage of replica exchange MCMC.
Holzloher R, Mahadevan A, Menyuk CR, Morris JM, Zweck J (2005) Evaluation of the very low BER of FEC codes using dual adaptive importance sampling. IEEE Communications Letters 9(2):163–165
Hukushima K (2002) Extended ensemble monte carlo approach to hardly relaxing problems. Computer Physics Communications 147(1):77–82
Hukushima K, Iba Y (2008) A Monte Carlo algorithm for sampling rare events: application to a search for the Griffith singularity. Journal of Physics: Conference Series 95:012005
Hukushima K, Nemoto K (1996) Exchange Monte Carlo method and application to spin glass simulations. Journal of the Physical Society of Japan 65(6):1604–1608
Iba Y (2001) Extended ensemble Monte Carlo. International Journal of Modern Physics C 12(5):623–656
Iba Y, Hukushima K (2008) Testing error correcting codes by multicanonical sampling of rare events. Journal of the Physical Society of Japan 77(10):103801
Iba Y, Takahashi H (2005) Exploration of multi-dimensional density of states by multicanonical Monte Carlo algorithm. Progress of Theoretical Physics Supplement 157:345–348
Iba Y, Chikenji G, Kikuchi M (1998) Simulation of lattice polymers with multi-self-overlap ensemble. Journal of the Physical Society of Japan 67:3327–3330
Jacob PE, Ryder RJ (2011) The Wang-Landau algorithm reaches the flat histogram criterion in finite time. ArXiv e-prints [1110.4028]
Jacobson MT, Matthews P (1996) Generating uniformly distributed random Latin squares. Journal of Combinatorial Designs 4(6):405–437
Janke W (1998) Multicanonical Monte Carlo simulations. Physica A: Statistical and Theoretical Physics 254(1-2):164–178
Jerrum M, Sinclair A (1996) The Markov chain Monte Carlo method: an approach to approximate counting and integration. Approximation algorithms for NP-hard problems pp 482–520
Kastner CA, Braumann A, Man PL, Moshbach S, Brownbridge GP, Akroyd J, Kraft M, Himawan C (2013) Bayesian parameter estimation for a jet-milling model using Metropolis-Hastings and Wang-Landau sampling. Chemical Engineering Science 89:244 – 257
Kimura K, Taki K (1991) Time-homogeneous parallel annealing algorithm. Proceedings of the 13th IMACS World Congress on Computation and Applied Mathematics (IMACS’91) 2:827–828
Kirkpatrick S, Vecchi M, et al (1983) Optimization by simulated annealing. Science 220(4598):671–680
Kitajima A, Iba Y (2011) Multicanonical sampling of rare trajectories in chaotic dynamical systems. Computer Physics Communications 182(1):251–253
Körner M, Katzgraber HG, Hartmann AK (2006) Probing tails of energy distributions using importance-sampling in the disorder with a guiding function. Journal of Statistical Mechanics: Theory and Experiment P04005
Kwon J, Lee KM (2008) Tracking of abrupt motion using Wang-Landau Monte Carlo estimation. In: Proceedings of the 10th European Conference on Computer Vision: Part I, Springer-Verlag, Berlin, Heidelberg, ECCV ’08, pp 387–400
Laffargue T, Nguyen Thu Lam KD, Kurchan J, Tailleur J (2013) Large deviations of Lyapunov exponents. ArXiv e-prints [1302.6258]
Landau D, Tsai SH, Exler M (2004) A new approach to Monte Carlo simulations in statistical physics: Wang-Landau sampling. American Journal of Physics 72:1294–1302
Lee H, Okabe Y, Landau D (2006) Convergence and refinement of the Wang-Landau algorithm. Computer Physics Communications 175(1):36–40
Lee J (1993a) Erratum: “New Monte Carlo algorithm: Entropic sampling”[phys. rev. lett. 72, 211 (1993)]. Physical Review Letters 71:2353
Lee J (1993b) New Monte Carlo algorithm: entropic sampling. Physical Review Letters 71(2):211–214
Liang F (2005) A generalized Wang-Landau algorithm for Monte Carlo computation. Journal of the American Statistical Association 100(472):1311–1327
Liang F, Liu C, Carroll RJ (2007) Stochastic approximation in Monte Carlo computation. Journal of the American Statistical Association 102(477):305–320
Liang F, Liu C, Carroll R (2010) Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples (Wiley Series in Computational Statistics). Wiley
Lyubartsev A, Martynovski A, Shevkonov S, Vorontsov-Velyaminov P (1992) New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. The Journal of Chemical Physics 96(3):1776–1783
Majumdar SN, Vergassola M (2009) Large deviations of the maximum eigenvalue for Wishart and Gaussian random matrices. Physical Review Letters 102(6):060601
Marinari E, Parisi G (1992) Simulated tempering: a new Monte Carlo scheme. Europhysics Letters 19(6):451–458
Matsuda Y, Nishimori H, Hukushima K (2008) Distribution of Lee-Yang zeros and Griffiths singularities in the ±J model of spin glasses. Journal of Physics A Mathematical General 41:324012
May RM (1972) Will a large complex system be stable? Nature 238:413–414
Mezei M (1987) Adaptive umbrella sampling: self-consistent determination of the non-Boltzmann bias. Journal of Computational Physics 68(1):237–248
Mitsutake A, Sugita Y, Okamoto Y (2001) Generalized-ensemble algorithms for molecular simulations of biopolymers. Biopolymers (Peptide Science) 60:96–123
Monthus C, Garel T (2006) Probing the tails of the ground-state energy distribution for the directed polymer in a random medium of dimension d= 1, 2, 3 via a Monte Carlo procedure in the disorder. Physical Review E 74(5):051109
Newman ME, Barkema GT (1999) Monte Carlo methods in statistical physics. Clarendon Press Oxford
Ott E (2002) Chaos in Dynamical Systems. Cambridge University Press
Pinn K, Wiockerowski C (1998) Number of magic squares from parallel tempering Monte Carlo. International Journal of Modern Physics C 9(04):541–546
Rubino G, Tuffin B (eds) (2009) Rare Event Simulation using Monte Carlo Methods. Wiley
Rubinstein RY, Kroese DP (2007) Simulation and the Monte Carlo Method (Wiley Series in Probability and Statistics), 2nd edn. Wiley-Interscience
Saito N, Iba Y (2011) Probability of graphs with large spectral gap by multicanonical Monte Carlo. Computer Physics Communications 182(1):223–225
Saito N, Iba Y, Hukushima K (2010) Multicanonical sampling of rare events in random matrices. Physical Review E 82(3):031142
Sasa Si, Hayashi K (2006) Computation of the Kolmogorov-Sinai entropy using statistical mechanics: Application of an exchange Monte Carlo method. Europhysics Letters 74(1):156
Schreiber T (1998) Constrained randomization of time series data. Physical Review Letters 80(10):2105–2108
Schreiber T, Schmitz A (2000) Surrogate time series. Physica D: Nonlinear Phenomena 142(3-4):346–382
Schulz BJ, Binder K, Muller M, Landau DP (2003) Avoiding boundary effects in Wang-landau sampling. Physical Review E 67(6):067102
Shell MS, Debenedetti PG, Panagiotopoulos AZ (2002) Generalization of the Wang-Landau method for off-lattice simulations. Physical Review E 66:056703
Shirai NC, Kikuchi M (2012) Multicanonical simulation of the Domb-Joyce model and the Go model: new enumeration methods for self-avoiding walks. ArXiv e-prints [1212.2181]
Shiteti J, Linares J, Varret F (1997) Monte Carlo entropic sampling for the study of metastable states and relaxation paths. Physical Review E 56(5):5128–5137
Sweet D, Nusse HE, Yorke JA (2001) Stagger-and-step method: Detecting and computing chaotic saddles in higher dimensions. Physical Review Letters 86:2261–2264
Tailleur J, Kurchan J (2007) Probing rare physical trajectories with Lyapunov weighted dynamics. Nature Physics 3(3):203–207
Takemura A, Aoki S (2004) Some characterizations of minimal Markov basis for sampling from discrete conditional distributions. Annals of the Institute of Statistical Mathematics 56(1):1–17
Torrie GM, Valleau JP (1974) Monte Carlo free energy estimates using non-Boltzmann sampling: Application to the sub-critical Lennard-Jones fluid. Chemical Physics Letters 28(4):578 – 581
Tracy CA, Widom H (1994) Level-spacing distributions and the Airy kernel. Communications in Mathematical Physics 159(1):151–174
Tracy CA, Widom H (1996) On orthogonal and symplectic matrix ensembles. Communications in Mathematical Physics 177(3):727–754
Vorontsov-Velyaminov P, Broukhko A, Kuznetsova T, Lyubartsev A (1996) Free energy calculations by expanded ensemble method for lattice and continuous polymers. The Journal of Physical Chemistry 100(4):1153–1158
Vorontsov-Velyaminov P, Volkov N, Yurchenko A (2004) Entropic sampling of simple polymer models within Wang-Landau algorithm. Journal of Physics A: Mathematical and General 37(5):1573–1588
Wang F, Landau DP (2001) A random walk algorithm to produce a flat histogram. Physical Review E 64(5):056101
Wang F, Landau DP (2001) Efficient, multiple-range random walk algorithm to calculate the density of states. Physical Review Letters 86(10):2050–2053
Wolfsheimer S, Hartmann AK (2010) Minimum-free-energy distribution of RNA secondary structures: Entropic and thermodynamic properties of rare events. Phys Rev E 82:021902
Yanagita T, Iba Y (2009) Exploration of order in chaos using the replica exchange Monte Carlo method. J Stat Mech 2009(02):P02043
Yevick D (2002) Multicanonical communication system modeling - Application to PMD statistics. IEEE Photonics Technology Letters 14(11):1512–1514
Yu K, Liang F, Ciampa J, Chatterjee N (2011) Efficient p-value evaluation for resampling-based tests. Biostatistics 12(3):582–593
Zhan L (2008) A parallel implementation of the Wang–Landau algorithm. Computer Physics Communications 179(5):339–344
Zhang C, Ma J (2007) Simulation via direct computation of partition functions. Physical Review E 76(3):036708
Zhang C, Ma J (2009) Counting solutions for the N-queens and Latin-square problems by Monte Carlo simulations. Physical Review E 79(1):016703
Zhou C, Su J (2008) Optimal modification factor and convergence of the Wang–Landau algorithm. Physical Review E 78:046705
Zhou C, Schulthess TC, Torbrügge S, Landau D (2006) Wang–Landau algorithm for continuous models and joint density of states. Physical Review Letters 96(12):120201