Adv. Appl. Probab. **54,** 536–569 (2022)
*doi:*10.1017/apr.2021.40

NORMAL APPROXIMATION FOR FUNCTIONS OF HIDDEN MARKOV MODELS

CHRISTIAN HOUDRÉ,* *Georgia Institute of Technology*
GEORGE KERCHEV,** *Université du Luxembourg*

Abstract

The generalized perturbative approach is an all-purpose variant of Stein's method used to obtain rates of normal approximation. Originally developed for functions of independent random variables, this method is here extended to functions of the realization of a hidden Markov model. In this dependent setting, rates of convergence are provided in some applications, leading, in each instance, to an extra log-factor vis-à-vis the rate in the independent case.

Keywords: Stein's method; Markov chains; generalized perturbative approach; normal approximation; stochastic geometry

2020 Mathematics Subject Classification: Primary 60F05
Secondary 60K35; 60D05

1. Introduction

Let $X = (X_1, \ldots, X_n)$ be a random vector with coordinates in a Polish space E, and let $f : E^n \to \mathbb{R}$ be a measurable function such that $f(X)$, for n large, is square-integrable. For a large class of such functions f it is expected that as n grows without bound, $f(X)$ behaves like a normal random variable. To quantify such estimates one is interested in bounding the distance between $f(X)$ and the normal random variable $\mathcal{N} \sim N(m_f, \sigma_f^2)$ where $m_f = \mathbb{E}[f(X)]$ and $\sigma_f^2 = Var(f(X))$. Two such distances of interest are the Kolmogorov distance

$$d_K(f(X), \mathcal{N}) := \sup_{t\in\mathbb{R}} |\mathbb{P}(f(X) \le t) - \mathbb{P}(\mathcal{N} \le t)|$$

and the Wasserstein distance

$$d_W(f(X), \mathcal{N}) := \sup_{h} |\mathbb{E}[h(f(X))] - \mathbb{E}[h(\mathcal{N})]|,$$

where this last supremum is taken over real-valued Lipschitz functions h such that $|h(x) - h(y)| \le |x - y|$ for all $x, y \in \mathbb{R}$.

For the case where the components of X are independent random variables, upper bounds on $d_W(f(X), \mathcal{N})$ were first obtained in [2], and these were extended to $d_K(f(X), \mathcal{N})$ in [14]. Both results rely on a class of difference operators that will be described in Section 2.

Received 9 October 2020; revision received 15 July 2021.
* Postal address: School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332-0160, USA. Email address: houdre@math.gatech.edu
** Postal address: Université du Luxembourg, Unité de Recherche en Mathématiques, Maison du Nombre, 6 Avenue de la Fonte, L-4364 Esch-sur-Alzette, Grand Duché du Luxembourg. Email address: gkerchev@gmail.com

Very few results address the (weakly) dependent case, and in the present work we provide estimates on $d_K(f(X), \mathcal{N})$ and $d_W(f(X), \mathcal{N})$ when X is generated by a hidden Markov model. Such models are of interest because of their many applications in fields such as computational biology and speech recognition; see e.g. [7]. Recall that a hidden Markov model (Z, X) consists of a Markov chain $Z = (Z_1, \ldots, Z_n)$ which emits the observed variables $X = (X_1, \ldots, X_n)$. The possible states in Z are each associated with a distribution on the values of X. In other words the observation X is a mixture model where the choice of the mixture component for each observation depends on the component of the previous observation. The mixture components are given by the sequence Z. Note also that given Z, X is a Markov chain.

The content of the paper is as follows. Section 2 contains a short overview of results on normal approximation in the independent setting and introduces a simple transformation involving independent and identically distributed (i.i.d.) random variables that allows us to adapt these estimates to the hidden Markov model. Moreover, further quantitative bounds are provided for the special case when f is a Lipschitz function with respect to the Hamming distance. Applications to variants of the ones analyzed in [2] and [14] are developed in Sections 3 and 4, leading to an extra log-factor in the various rates obtained there. The majority of the more technical computations are carried out and presented in Section 5.

2. Main results

For a comprehensive review of Stein's method we refer the reader to [4]. The exchangeable pairs approach was outlined in [2]. We now recall below a few of its main points.

Let $W := f(X)$. Originally in [2], and then in [14], various bounds on the distance between W and the normal distribution were obtained through a variant of Stein's method. As is well known, Stein's method is a way to obtain normal approximations based on the observation that the standard normal distribution $\mathcal{N}$ is the only centered and unit-variance distribution that satisfies

$$\mathbb{E}\big[g'(\mathcal{N})\big] = \mathbb{E}\big[\mathcal{N} g(\mathcal{N})\big]$$

for all absolutely continuous g with almost-everywhere (a.e.) derivative g' such that $\mathbb{E}|g'(\mathcal{N})| < \infty$ (see [4]), and for the random variable W, $|\mathbb{E}[Wg(W) - g'(W)]|$ can be thought of as a distance measuring the proximity of W to $\mathcal{N}$. In particular, for the Kolmogorov distance, the solutions g_t to the differential equation

$$\mathbb{P}(W \le t) - \mathbb{P}(\mathcal{N} \le t) = g_t'(W) - Wg_t(W)$$

are absolutely continuous with a.e. derivative such that $\mathbb{E}\big|g_t'(\mathcal{N})\big| < \infty$ (see [4]). Then,

$$d_K(W, \mathcal{N}) = \sup_{t \in \mathbb{R}} \big|\mathbb{E}\big[g_t'(W) - Wg_t(W)\big]\big|. \tag{1}$$

Further properties of the solutions g_t allow for upper bounds on $\mathbb{E}\big[g_t'(W) - Wg_t(W)\big]$ using difference operators associated with W that were introduced in [2] (see [14]). This is called the *generalized perturbative approach* in [3], and we describe it next. First, we recall the perturbations used to bound the right-hand side of (1) in [2, 14]. Let $X' = \big(X_1', \ldots, X_n'\big)$ be an independent copy of X, and let $W' = f(X')$. Then (W, W') is an exchangeable pair, since it has the same joint distribution as (W', W). A perturbation $W^A = f^A(X) := f\big(X^A\big)$ of W is defined through the change X^A of X as follows:

$$X_i^A = \begin{cases} X_i' & \text{if } i \in A, \\ X_i & \text{if } i \notin A, \end{cases}$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

for any $A \subseteq [n] := \{1, \dots, n\}$, including $A = \emptyset$. With these definitions, still following [2], difference operators are defined for any $\emptyset \subseteq A \subseteq [n]$ and $i \notin A$, as follows:

$$\Delta_i f^A = f\big(X^A\big) - f\Big(X^{A\cup\{i\}}\Big).$$

Moreover, set

$$T_A(f) := \sum_{j\notin A} \Delta_j f(X) \Delta_j f\big(X^A\big),$$
$$T'_A(f) := \sum_{j\notin A} \Delta_j f(X) \big|\Delta_j f\big(X^A\big)\big|,$$

and for $k_{n,A} = 1/\binom{n}{|A|}(n - |A|)$, set

$$T_n(f) := \sum_{\emptyset\subseteq A\subsetneq[n]} k_{n,A} T_A(f),$$
$$T'_n(f) := \sum_{\emptyset\subseteq A\subsetneq[n]} k_{n,A} T'_A(f).$$

Now, for $W = f(X_1, \dots, X_n)$ such that $\mathbb{E}[W] = 0$, $0 < \sigma^2 = \mathbb{E}\big[W^2\big] < \infty$, and assuming all the expectations below are finite, [2, Theorem 2.2] gives the bound

$$d_W\big(\sigma^{-1}W, \mathcal{N}\big) \le \frac{1}{\sigma^2}\sqrt{Var\Big(\mathbb{E}\big[T_n(f)|X\big]\Big)} + \frac{1}{2\sigma^3}\sum_{j=1}^{n}\mathbb{E}\big|\Delta_j f(X)\big|^3, \tag{2}$$

while [14, Theorem 4.2] yields

$$d_K\big(\sigma^{-1}W, \mathcal{N}\big) \le \frac{1}{\sigma^2}\sqrt{Var\Big(\mathbb{E}\big[T_n(f)|X\big]\Big)} + \frac{1}{\sigma^2}\sqrt{Var\Big(\mathbb{E}\big[T'_n(f)|X\big]\Big)}$$
$$+ \frac{1}{4\sigma^3}\sum_{j=1}^{n}\sqrt{\mathbb{E}\big|\Delta_j f\big|^6} + \frac{\sqrt{2\pi}}{16\sigma^3}\sum_{j=1}^{n}\mathbb{E}\big|\Delta_j f(X)\big|^3, \tag{3}$$

where in both cases $\mathcal{N}$ is now a standard normal random variable.

Our main abstract result generalizes (2) and (3) to the case when X is generated by a hidden Markov model. It is as follows.

Proposition 2.1. *Let (Z, X) be a hidden Markov model with Z an aperiodic time-homogeneous and irreducible Markov chain with finite state space $\mathcal{S}$, and X taking values in a non-empty finite $\mathcal{A}$. Let $W := f(X_1, \dots, X_n)$ with $\mathbb{E}[W] = 0$ and $0 < \sigma^2 = \mathbb{E}\big[W^2\big] < \infty$. Then there exist a finite sequence of independent random variables $R = \big(R_0, R_1, \dots, R_{|\mathcal{S}|(n-1)}\big)$, with R_i taking values in $\mathcal{S}\times\mathcal{A}$ for $i = 0, \dots, |\mathcal{S}|(n-1)$, and a measurable function $h : (\mathcal{S}\times\mathcal{A})^{|\mathcal{S}|(n-1)+1} \longrightarrow \mathbb{R}$ such that $h\big(R_0, \dots, R_{|\mathcal{S}|(n-1)}\big)$ and $f(X_1, \dots, X_n)$ are identically distributed. Therefore,*

$$d_W\big(\sigma^{-1}W, \mathcal{N}\big) \le \frac{1}{\sigma^2}\sqrt{Var\Big(\mathbb{E}\big[T_{|R|}(h)|R\big]\Big)} + \frac{1}{2\sigma^3}\sum_{i=0}^{|\mathcal{S}|(n-1)}\mathbb{E}\big|\Delta_i h(R)\big|^3 \tag{4}$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

and

$$d_K\left(\sigma^{-1}W, \mathcal{N}\right) \leq \frac{1}{\sigma^2}\sqrt{Var\left(\mathbb{E}\left[T_{|R|}(h)|R\right]\right)} + \frac{1}{\sigma^2}\sqrt{Var\left(\mathbb{E}\left[T'_{|R|}(h)|R\right]\right)} + \frac{1}{4\sigma^3}\sum_{j=0}^{|R|-1}\sqrt{\mathbb{E}\left|\Delta_j h(R)\right|^6} + \frac{\sqrt{2\pi}}{16\sigma^3}\sum_{j=0}^{|R|-1}\mathbb{E}\left|\Delta_j h(R)\right|^3, \tag{5}$$

where $\mathcal{N}$ is a standard normal random variable.

At first glance the above results might appear to be simple corollaries to (2) and (3). Indeed, as is well known, every Markov chain (in a Polish space) admits a representation via i.i.d. random variables $U_1, \ldots, U_n$, uniformly distributed on (0, 1), and the inverse distribution function. Therefore, $f(X_1, \ldots, X_n) \stackrel{d}{=} h(U_1, \ldots, U_n)$, for some function h, where, as usual, $\stackrel{d}{=}$ indicates equality in distribution. However, providing quantitative estimates for $\mathbb{E}|\Delta_j h(U_1, \ldots, U_n)|$ via f seems to be out of reach, since passing from f to h involves the 'unknown' inverse distribution function. For this reason, we develop for our analysis a more amenable, although more restrictive, choice of i.i.d. random variables, which is described intuitively in the next paragraph and then again in greater detail in Section 2.1.

Consider $R = \left(R_0, \ldots, R_{|\mathcal{S}|(n-1)}\right)$ as stacks of independent random variables on the $|\mathcal{S}|$ possible states of the hidden chain that determine the next step in the process, with R_0 specifying the initial state. Each R_i takes values in $\mathcal{S} \times \mathcal{A}$ and is distributed according to the transition probability from the present hidden state. Then, one has $f\left(X_1, \ldots, X_n\right) \stackrel{d}{=} h\left(R_0, \ldots, R_{|\mathcal{S}|(n-1)}\right)$, for $h = f \circ \gamma$, where the function γ translates between R and X. This construction is carried out in more detail in the next section. Further note that when $(X_i)_{i\geq 1}$ is a sequence of independent random variables, the hidden chain in the model consists of a single state, and then the function γ is the identity function and so $h = f$.

In order for Proposition 2.1 to be meaningful, a further quantitative study of the terms in the upper bounds is necessary. It turns out that the variance terms determine the order of decay; see Remark 5.2. Nevertheless, we obtain additional quantitative estimates for all the terms involved; the proofs are presented in Section 5.

Proposition 2.2. *With the notation as above, let f be Lipschitz with respect to the Hamming distance, i.e., let*

$$|f(x) - f(y)| \leq c\sum_{i=1}^{n}\mathbf{1}_{x_i \neq y_i}$$

for every $x, y \in \mathcal{A}^n$ and where $c > 0$. Then, for any $r > 0$,

$$\mathbb{E}|\Delta_i h(R)|^r \leq C_1(\ln n)^r,$$
$$\mathbb{E}|h(R) - \mathbb{E}[h(R)]|^r \leq C_2 n^{r/2}(\ln n)^r,$$

for n large enough and $C_1, C_2 > 0$ depending on r and the parameters of the model.

Let R' and R'' be independent copies of R, and let $\widetilde{\mathbf{R}}$ be the random set of recombinations of R, R', and R''. The set $\widetilde{\mathbf{R}}$ consists of $3^{|R|}$ random vectors of size $|R|$:

$$\widetilde{\mathbf{R}} := \left\{Z = \left(Z_0, \ldots, Z_{|R|-1}\right) : Z_i \in \left\{R_i, R'_i, R''_i\right\}\right\}.$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Let

$$
\begin{aligned}
B_{|R|}(h) &:= \sup_{Y,Y',Z,Z' \in \widetilde{\mathbf{R}}} \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\, \Delta_{j,k}h(Y')\neq 0} \big|\Delta_j h(Z)\big|^2 \big|\Delta_k h\big(Z'\big)\big|^2\Big], \\
B^{(k)}_{|R|}(h) &:= \sup_{Y,Z,Z' \in \widetilde{\mathbf{R}}} \mathbb{E}\big|\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0} \Delta_i h(Z)^2 \Delta_j h\big(Z'\big)^2\big|, \\
B^{(j)}_{|R|}(h) &:= \sup_{Y,Z,Z' \in \widetilde{\mathbf{R}}} \mathbb{E}\big|\mathbf{1}_{\Delta_{i,k}h(Y)\neq 0} \Delta_i h(Z)^2 \Delta_k h\big(Z'\big)^2\big|.
\end{aligned} \tag{6}
$$

Then the following general bound for the variance terms holds.

Proposition 2.3. *With the notation as above and for $U = T_{|R|}(h)$ or $U = T'_{|R|}(h)$, we have*

$$
\sqrt{Var\big(\mathbb{E}[U|R]\big)} \le \frac{1}{\sqrt{2}} \sum_{\emptyset \subseteq A \subsetneq [|R|]} k_{|R|,A} \Bigg(\sum_{i=0}^{|R|-1} \sum_{j,k \notin A} \Big(\mathbf{1}_{i=j=k} \mathbb{E}|\Delta_i h(R)|^4 + \mathbf{1}_{i\neq j\neq k} B_{|R|}(h) \\
+ \big(\mathbf{1}_{i\neq j=k} + \mathbf{1}_{i=k\neq j}\big) B^{(k)}_{|R|}(h) + \big(\mathbf{1}_{i\neq j=k} + \mathbf{1}_{i=j\neq k}\big) B^{(j)}_{|R|}(h) \Big) \Bigg)^{1/2},
$$

where $[|R|] := \{1, \dots, |R|\}$.

Note that in Proposition 2.3, the underlying function f is not assumed to be Lipschitz. Moreover, the function h is not symmetric, and therefore the expression above cannot be simplified further, in contrast to the similar results in [14]. The proofs of Propositions 2.2 and 2.3 are technical and are delayed to Sections 5.1 and 5.2 respectively.

2.1. Additional details on the construction of *R*

Let (Z, X) be a hidden Markov model with Z an aperiodic time-homogeneous and irreducible Markov chain on a finite state space $\mathcal{S}$, and with X taking values in an alphabet $\mathcal{A}$. Let P be the transition matrix of the hidden chain, and let Q be the $|\mathcal{S}| \times |\mathcal{A}|$ probability matrix for the observations; i.e., Q_{ij} is the probability of seeing output j if the latent chain is in state i. Let the initial distribution of the hidden chain be μ. Then

$$
\begin{aligned}
&\mathbb{P}\big((Z_1, \dots, Z_n; X_1, \dots, X_n) = (z_1, \dots, z_n; x_1, \dots, x_n)\big) \\
&\quad = \mu(z_1) Q_{z_1,x_1} P_{z_1,z_2} \dots P_{z_{n-1},z_n} Q_{z_n,x_n}.
\end{aligned}
$$

Next we introduce a sequence of independent random variables $R_0, \dots, R_{|\mathcal{S}|(n-1)}$ taking values in $\mathcal{S} \times \mathcal{A}$ and a function γ such that $\gamma\big(R_0, \dots, R_{|\mathcal{S}|(n-1)}\big) = \big(Z_1, \dots, Z_n; X_1, \dots, X_n\big)$. For any $s, s' \in \mathcal{S}$, $x \in \mathcal{A}$, and $i \in \{0, \dots, n-1\}$, let

$$
\begin{aligned}
\mathbb{P}\big(R_0 = (s, x)\big) &= \mu(s) Q_{s,x}, \\
\mathbb{P}\big(R_{i|\mathcal{S}|+s'} = (s, x)\big) &= P_{s',s} Q_{s,x}.
\end{aligned}
$$

The random variables R_i are well defined since $\sum_x Q_{s,x} = 1$ for any $s \in \mathcal{S}$ and $\sum_s P_{s',s} = \sum_s \mu(s) = 1$ for any $s' \in \mathcal{S}$. One can think of the variables R_i as a set of instructions indicating where the hidden Markov model goes next. The function γ reconstructs the realization

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

$(Z_i, X_i)_{i\geq 1}$ sequentially from the sequence $(R_i)_{i\geq 0}$. In particular, γ captures the following relations:

$$\begin{aligned}(Z_1, X_1) &= R_0,\\ (Z_{i+1}, X_{i+1}) &= R_{i|\mathcal{S}|+s} \quad \text{if } Z_i = s \text{ for } i \geq 1.\end{aligned}$$

One can also think of the sequence $(R_i)_{i\geq 0}$ as $|\mathcal{S}|$ stacks of random variables on the $|\mathcal{S}|$ possible states of the latent Markov chain, and the values being rules for the next step in the model. Note that only one variable on the ith level of the stack will be used to determine the $(i+1)$th hidden and observed pair. Furthermore, the distribution of the random variables R_i for $i \geq 1$ encodes the transition and output probabilities in the P and Q matrices of the original model.

Thus one can write $f(X_1, \ldots, X_n) = h(R_0, \ldots, R_{|\mathcal{S}|(n-1)})$, for $h := f \circ \gamma$, where the function γ does the translation from $(R_i)_{i\geq 0}$ to $(Z_i, X_i)_{i\geq 1}$ as described above.

Let $R' = (R'_0, \ldots, R'_{|\mathcal{S}|(n-1)})$ be an independent copy of R. Let $A \subseteq \{0, 1, \ldots, |\mathcal{S}|(n-1)\}$, and let the change R^A of R be defined as follows:

$$R_i^A = \begin{cases} R'_i & \text{if } i \in A,\\ R_i & \text{if } i \notin A,\end{cases} \tag{7}$$

where, as before, when $A = \{j\}$ we write R^j instead of $R^{\{j\}}$.

Recall that the 'discrete derivative' of h with a perturbation A is

$$\Delta_i h^A = h(R^A) - h(R^{A\cup\{i\}}).$$

Then (4) and (5) follow from (2) and (3), respectively, since when (Z, X) is a hidden Markov model one writes

$$W = f(X_1, \ldots, X_n) \stackrel{d}{=} h(R_0, \ldots, R_{|\mathcal{S}|(n-1)}),$$

where the sequence $(R_i)_{i\geq 0}$ is a sequence of independent random variables.

Remark 2.1. (i) The idea of using stacks of independent random variables to represent a hidden Markov model is somewhat reminiscent of Wilson's cycle popping algorithm for generating a random directed spanning tree; see [17]. The algorithm has also been related to loop-erased random walks in [9].

(ii) If S consists of a single state, making the hidden chain redundant, there is a single stack of instructions. This corresponds to the independent setting of [2] and [14], and then γ is just the identity function.

(iii) The same approach, via the use of instructions, is also applicable when $\mathcal{A}$ and $\mathcal{S}$ are infinite countable. The $Q_{s,x}$ no longer form a finite matrix, but the same definition holds as long as $\sum_{x\in\mathcal{A}} Q_{s,x} = 1$ for all $s \in \mathcal{S}$. We need a countably infinite number of independent instructions to encode $(Z_i, X_i)_{1\leq i\leq n}$. In particular, let R_0 and $(R_{i,s})_{1\leq i\leq n, s\in\mathcal{S}}$ be such that

$$\begin{aligned}\mathbb{P}(R_0 = (s, x)) &= \mu(s)Q_{s,x},\\ \mathbb{P}(R_{i,s'} = (s, x)) &= P_{s',s}Q_{s,x}.\end{aligned}$$

Then the function γ reconstructs $(Z_i, X_i)_{1\leq i\leq n}$ from R_0 and $(R_{i,s})_{1\leq i\leq n, s\in\mathcal{S}}$ via

$$\begin{aligned}(Z_1, X_1) &= R_0,\\ (Z_{i+1}, X_{i+1}) &= R_{i,s} \quad \text{if } Z_i = s \text{ for } i \geq 1.\end{aligned}$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

(iv) It is possible to obtain bounds on the various terms involved in Proposition 2.2 and Proposition 2.3 in the case when $|\mathcal{S}|$ is a function of n. Assuming that there exists a general deterministic upper bound $g_r(n)$ on $\mathbb{E}|\Delta_i h|^r$, one can bound the non-variance terms in Proposition 2.2 by $C|\mathcal{S}|\left(\sqrt{g_6(n)}+g_3(n)\right)/\sigma^3$. The variance terms, using Proposition 2.3, will then be bounded by $C\sqrt{|\mathcal{S}|g_4(n)+|\mathcal{S}|^3(A)}/\sigma^2$, where A is a complicated expression that depends on the particular problem as well as on $\mathbb{E}|\Delta_i h|^r$ and $|\mathcal{S}|$.

The next crucial part in the analysis is the key result we use everywhere: if a change in an instruction propagates X levels (a random variable), then $\mathbb{P}(X > K) \le (1-\epsilon)^K$, for some absolute $0 < \epsilon < 1$. If $|\mathcal{S}|$ is finite, this holds under some standard assumptions on the model. If $|\mathcal{S}|$ grows with n, then under some minor additional restrictions in the hidden Markov model, we have $\mathbb{P}(X > K) \le (1 - 1/|\mathcal{S}|)^K$. Then K can be chosen to be at most a small power of n (we take $K = \ln n$ in the finite case, which explains the logarithmic factors in our bounds). For meaningful bounds, $|\mathcal{S}|$ could grow at most like $\ln n$. However, much more fine-tuning is necessary in the actual proof.

For some recent results (different from ours) on normal approximation for functions of general Markov chains we refer the reader to [5].

3. Covering process

Although our framework was initially motivated by [11] and the problem of finding a normal approximation result for the length of the longest common subsequences in dependent random words, some applications to stochastic geometry are presented below. Our methodology can be applied to other related settings, in particular to the variant of the occupancy problem introduced in the recent article [10] (see Remark 4.1).

Let $(K, \mathcal{K})$ be the space of compact subsets of $\mathbb{R}^d$, endowed with the hit-and-miss topology. Let E_n be a cube of volume n, and let $C_1, \ldots, C_n$ be random variables in E_n, called *germs*. In the i.i.d. setting of [14], each C_i is sampled uniformly and independently in E_n; i.e., if $T \subset E_n$ with measure $|T|$, then

$$\mathbb{P}(C_i \in T) = \frac{|T|}{n},$$

for all $i \in \{1, \ldots, n\}$.

Here, we consider $C_1, \ldots, C_n$, generated by a hidden Markov model in the following way. Let $Z_1, \ldots, Z_n$ be an aperiodic irreducible Markov chain on a finite state space $\mathcal{S}$. Each $s \in \mathcal{S}$ is associated with a measure m_s on E_n. Then for each measurable $T \subseteq E_n$,

$$\mathbb{P}\big(C_i \in T | Z_i = s\big) = m_s(T).$$

Assume that there are constants $0 < c_m \le c_M$ such that for any $s \in \mathcal{S}$ and measurable $T \subseteq E_n$,

$$\frac{c_m|T|}{n} \le m_s(T) \le \frac{c_M|T|}{n}.$$

Note that $c_m = c_M = 1$ recovers the setting of [14].

Let $K_1, \ldots, K_n$ be compact sets (*grains*) with $Vol(K_i) \in (V_1, V_2)$ (absolute constants) for $i = 1, \ldots, n$. Let $X_i = C_i + K_i$ for $i = 1, \ldots, n$ be the *germ–grain* process. Consider the closed set formed by the union of the germs translated by the grains

$$F_n = \left(\bigcup_{k=1}^{n} X_K\right) \cap E_n.$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

We are interested in the volume covered by F_n,

$$f_V(X_1, \ldots, X_n) = Vol(F_n),$$

and the number of isolated grains,

$$f_I(X_1, \ldots, X_n) = \#\{k : X_k \cap X_j \cap E_n = \emptyset, k \neq j\}.$$

Theorem 3.1. *Let $\mathcal{N}$ be a standard normal random variable. Then, for all $n \in \mathbb{N}$,*

$$d_K\left(\frac{f_V - \mathbb{E}f_V}{\sqrt{Var\, f_V}}, \mathcal{N}\right) \leq C\left(\frac{n(\ln n)^3}{\sqrt{Var(f_V)^3}} + \frac{n^{1/2}(\ln n)^4}{Var(f_V)}\right), \tag{8}$$

$$d_K\left(\frac{f_I - \mathbb{E}f_I}{\sqrt{Var\, f_I}}, \mathcal{N}\right) \leq C\left(\frac{n(\ln n)^3}{\sqrt{Var(f_I)^3}} + \frac{n^{1/2}(\ln n)^4}{Var(f_I)}\right), \tag{9}$$

for some constant $C > 0$ independent of n.

The study of the order of growth of $Varf_I$ and $Varf_V$ is not really part of the scope of the present paper. In the independent case, there are constants $0 < c_V \leq C_V$ such that $c_V n \leq Varf_V \leq C_V n$ and $c_V n \leq Varf_I \leq C_V n$ for n sufficiently large (see [13, Theorem 4.4]). In our dependent setting, a variance lower bound of order n will thus provide a rate of order $(\log n)^4/\sqrt{n}$.

The proofs of the normal approximations for f_V and f_I are carried out in Sections 3.1 and 3.2. Many of the more technical computations are carried out in Section 5.

3.1. Normal approximation for f_V

Write $f_V(X_1, \ldots, X_n) = h(R_0, \ldots, R_{|\mathcal{S}|(n-1)})$ for a set of instructions R defined as in Section 2.1. The volume of each grain is bounded by V_2, so f_V is Lipschitz with respect to the Hamming distance, with constant V_2. Proposition 2.1 holds, and from Proposition 2.2, the non-variance terms in the bounds in Proposition 2.1 are bounded by $C(\ln n)^3/\sqrt{n}$. Here and below, C is a constant, independent of n, which can vary from line to line. Indeed, for instance,

$$\begin{aligned}\frac{1}{4\sigma^3}\sum_{j=0}^{|R|-1}\sqrt{\mathbb{E}|\Delta_j h(R)|^6} \leq\ & CVar(f_V)^{-3/2}(|\mathcal{S}|(n-1)+1)(\ln n)^3 \\ \leq\ & Cn(\ln n)^3/Var(f_V)^{3/2}. \end{aligned}\tag{10}$$

To analyze the bound on the variance terms given by Proposition 2.3, first note that

$$\sum_{i=0}^{|R|-1}\sum_{j,k\notin A}\mathbf{1}_{i=j=k}\mathbb{E}|\Delta_i h(R)|^4 \leq Cn(\ln n)^4,$$

using Proposition 2.2. The other terms are bounded as follows.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Proposition 3.1. *Let $A \subsetneq [|R|]$, and let $B_{|R|}(h)$, $B^k_{|R|}(h)$, and $B^j_{|R|}(h)$ be as in (6). Then*

$$\sum_{i=0}^{|R|-1} \sum_{j,k\notin A} \mathbf{1}_{i\neq j\neq k} B_{|R|}(h) \leq Cn(\ln n)^8, \tag{11}$$

$$\sum_{i=0}^{|R|-1} \sum_{j,k\notin A} \left(\mathbf{1}_{i\neq j=k} + \mathbf{1}_{i=k\neq j}\right) B^{(k)}_{|R|}(h) \leq Cn(\ln n)^4, \tag{12}$$

$$\sum_{i=0}^{|R|-1} \sum_{j,k\notin A} \left(\mathbf{1}_{i\neq j=k} + \mathbf{1}_{i=j\neq k}\right) B^{(j)}_{|R|}(h) \leq Cn(\ln n)^4, \tag{13}$$

for some constant $C > 0$ that does not depend on n.

Proof. See Section 5.3 for the proof of the first bound. The others follow similarly. □

The bound on the variance terms in Proposition 2.3 becomes

$$\begin{aligned}\sqrt{Var(\mathbb{E}[U|R])} \leq& \frac{1}{\sqrt{2}} \sum_{A\subsetneq[|R|]} k_{|R|,A}\left(Cn(\ln n)^4 + Cn(\ln n)^8 + 2Cn(\ln n)^4\right)^{1/2} \\ \leq& C\sqrt{n}(\ln n)^4. \end{aligned} \tag{14}$$

Then (8) follows from (14), (10), and Proposition 2.1.

3.2. Normal approximation for f_I

The proof of (9) is more involved since the function f_I is not Lipschitz. Abusing notation, write $f_I(X_1, \ldots, X_n) = h(R_0, \ldots, R_{|\mathcal{S}|(n-1)})$ for a set of instructions R as in Section 2.1. Proposition 2.1 holds, and, as in our analysis for f_V, we proceed by estimating the non-variance terms in the bounds. The following holds.

Proposition 3.2. *For any $t = 1, 2, \ldots$ and $i \in \{0, \ldots, |\mathcal{S}|(n-1)\}$,*

$$\mathbb{E}|\Delta_i h|^t \leq C(\ln n)^t, \tag{15}$$

where $C = C(t) > 0$.

Proof. See Section 5.4. The approach is similar to the one employed for Proposition 2.2 and uses a graph representation. □

Therefore, for the non-variance term in Proposition 2.1, we have

$$\frac{1}{4\sigma^3} \sum_{j=0}^{|R|-1} \sqrt{\mathbb{E}|\Delta_j h(R)|^6} + \frac{\sqrt{2\pi}}{16\sigma^3} \sum_{j=0}^{|R|-1} \mathbb{E}|\Delta_j h(R)|^3 \leq Cn \left(\frac{\ln n}{\sqrt{Var(f_I)}}\right)^3. \tag{16}$$

We are left to analyze the bound on the variance terms given by Proposition 2.3. First, note that using Proposition 3.2,

$$\sum_{i=0}^{|R|-1} \sum_{j,k\notin A} \mathbf{1}_{i=j=k} \mathbb{E}|\Delta_i h(R)|^4 \leq Cn(\ln n)^4.$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Proposition 3.3. *Let $A \subsetneq [|R|]$, and let $B_{|R|}(h)$, $B^k_{|R|}(h)$, and $B^j_{|R|}(h)$ be as in (6). Then the bounds (11), (12), and (13) hold in this setting as well.*

Proof. See Section 5.5. □

The bound on the variance terms in Proposition 2.3 becomes

$$\begin{aligned}\sqrt{Var(\mathbb{E}[U|R])} \leq & \frac{1}{\sqrt{2}} \sum_{A \subsetneq [|R|]} k_{|R|,A} \Big(Cn(\ln n)^4 + Cn(\ln n)^8 + 2Cn(\ln n)^4 \Big)^{1/2} \\ \leq & C\sqrt{n}(\ln n)^4. \end{aligned} \tag{17}$$

Then (9) follows from (17), (16), and Proposition 2.1.

4. Set approximation with random tessellations

Let $K \subseteq [0,1]^d$ be compact, and let X be a finite collection of points in K. The Voronoi reconstruction, or the Voronoi approximation, of K based on X is given by

$$K^X := \big\{ y \in \mathbb{R}^d : \text{ the closest point to } y \text{ in } X \text{ lies in } K \big\}.$$

For $x \in [0,1]^d$, denote by $V(x;X)$ the Voronoi cell with nucleus x within X, given by

$$V(x;X) := \big\{ y \in [0,1]^d : ||y - x|| \leq ||y - x'|| \text{ for any } x' \in (X,x) \big\},$$

where $(X,x) = X \cup \{x\}$, and where, as usual, $||\cdot||$ is the Euclidean norm in $\mathbb{R}^d$. The volume approximation of interest is

$$\varphi(X) := Vol\big(K^X\big) = \sum_i \mathbf{1}_{X_i \in K} Vol\big(V\big(X_i;X\big)\big).$$

In [14], $X = (X_1, \ldots, X_n)$ is a vector of n i.i.d. random variables uniformly distributed on $[0,1]^d$. Here, we consider $X_1, \ldots, X_n$ generated by a hidden Markov model in the following way. Let $Z_1, \ldots, Z_n$ be an aperiodic irreducible Markov chain on a finite state space $\mathcal{S}$. Each $s \in \mathcal{S}$ is associated with a measure m_s on $[0,1]^d$. Then for each measurable $T \subseteq [0,1]^d$,

$$\mathbb{P}\big(X_i \in T | Z_i = s\big) = m_s(T).$$

Assume, moreover, that there are constants $0 < c_m \leq c_M$ such that for any $s \in \mathcal{S}$ and measurable $T \subseteq [0,1]^n$,

$$c_m \frac{|T|}{n} \leq m_s(T) \leq c_M \frac{|T|}{n}.$$

Recall the notions of Lebesgue boundary of K given by

$$\partial K := \{x \in [0,1]^d : Vol(B(x,\epsilon) \cap K) > 0 \text{ and } Vol\big(B(x,\epsilon) \cap K^c\big) > 0, \text{ for any } \epsilon > 0\}$$

and

$$\partial K^r := \big\{ x : d(x, \partial K) \leq r \big\}, \; \partial K^r_+ := K^c \cap \partial K^r,$$

where $d(x,A)$ is the Euclidean distance from $x \in \mathbb{R}^d$ to $A \subseteq \mathbb{R}^d$.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Now, for $\beta > 0$, let

$$\gamma(K, r, \beta) := \int_{\partial K_+^r} \left(\frac{Vol(B(x, \beta r) \cap K)}{r^d} \right)^2 dx.$$

Next, recall that K is said to satisfy the (weak) rolling ball condition if

$$\gamma(K, \beta) := \liminf_{r>0} Vol\big(\partial K^r\big)^{-1} \big(\gamma(K, r, \beta) + \gamma\big(K^c, r, \beta\big)\big) > 0. \tag{18}$$

Our main result is as follows.

Theorem 4.1. *Let $K \subseteq [0, 1]^d$ satisfy the rolling ball condition. Moreover, assume that there exist $S_-(K)$, $S_+(K)$, $\alpha > 0$ such that*

$$S_+(K) r^\alpha \le Vol\big(\partial K^r\big) \le S_+(K) r^\alpha \quad \text{for every } r > 0.$$

Then, for $n \ge 1$,

$$d_K \left(\frac{\varphi(X) - \mathbb{E}\varphi(X)}{\sqrt{Var(\varphi(X))}}, \mathcal{N} \right) \le C \frac{(\ln n)^3}{n^{1/2-\alpha/d}}, \tag{19}$$

where $C > 0$ is a constant not depending on n.

As in [14], we split Theorem 4.1 into two results. The first one establishes a central limit theorem.

Proposition 4.1. *Let $0 < \sigma^2 = Var(\varphi(X))$. Assume that $Vol(\partial K^r) \le S_+(K) r^\alpha$ for some $S_+(K)$, $\alpha > 0$. Then, for $n \ge 1$,*

$$d_K \left(\frac{\varphi(X) - \mathbb{E}\varphi(X)}{\sigma}, \mathcal{N} \right) \le C \left(\frac{(\ln n)^2}{\sigma^2 n^{3/2+\alpha/2d}} + \frac{(\ln n)^3}{\sigma^3 n^{2+\alpha/2d}} \right), \tag{20}$$

where $C > 0$ is a constant not depending on n.

The second result introduces bounds on the variance under some additional assumptions.

Proposition 4.2. *Let $K \subseteq [0, 1]^d$ satisfy the rolling ball condition. Moreover, assume that there exist $S_-(K)$, $S_+(K)$, $\alpha > 0$ such that*

$$S_+(K) r^\alpha \le Vol(\partial K^r) \le S_+(K) r^\alpha \quad \text{for every } r > 0.$$

Then, for n sufficiently large,

$$C_d^- S_-(K) \gamma(K) \le \frac{Var(\varphi(X))}{n^{-1-\alpha/d}} \le C_d^+ S_+(K) \tag{21}$$

for some $C_d^-, C_d^+ > 0$.

It is clear that Theorem 4.1 will be proved once Proposition 4.1 and Proposition 4.2 are established.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

4.1. Proof of Proposition 4.1

Again, as before, we introduce a set of instructions R and a function h such that $h(R) = \varphi(X)$. We apply Proposition 2.1, and the initial step is to bound $\mathbb{E}|\Delta_i h(R)|^r$, where $r > 0$. In fact, the following holds.

Proposition 4.3. *Under the assumptions of Proposition 4.1,*

$$\mathbb{E}|\Delta_i h(R)|^r \le c_{d,r,\alpha} S_+(K)(\ln n)^r n^{-r-\alpha/d}, \tag{22}$$

where $c_{d,r,\alpha}$ depends on the parameters of the model and the dimension d, as well as on r and α. Moreover, for $n, q \ge 1$,

$$\mathbb{E}|\varphi(X) - \mathbb{E}\varphi(X)|^r \le C_{d,r,\alpha} S_+(K)(\ln n)^r n^{-r/2-\alpha/d} \tag{23}$$

for some $C_{d,r,\alpha} > 0$.

Before presenting the proof, we introduce some notation. Recall that $x, y \in [0, 1]^d$ are said to be Voronoi neighbors within the set X if $V(x; X) \cap V(y; X) \neq \emptyset$. In general, the Voronoi distance $d_V(x, y; X)$ between x and y within X is given by the smallest $k \ge 1$ such that there exist $x = x_0, x_1 \in X, \ldots, x_{k-1} \in X, x_k = y$ such that x_i, x_{i+1} are Voronoi neighbors for $i = 0, \ldots, k-1$.

Denote by $v(x, y; X) = Vol\big(V(y; X) \cap V(x; (y, X))\big)$ the volume that $V(y; X)$ loses when x is added to X. Then, for $x \notin X$,

$$\varphi(X, x) - \varphi(X) = \mathbf{1}_{x\in K} \sum_{y\in X\cap K^c} v(x, y; X) - \mathbf{1}_{x\in K^c} \sum_{y\in X\cap K} v(x, y; X).$$

Let $R_k(x; X)$ be the distance from x to the farthest point in the cell of a kth-order Voronoi neighbor in X; i.e., for $X = (X_1, \ldots, X_n)$,

$$R_k(x; X) = \sup\big\{||y - x|| : y \in V\big(X_i; X\big),\, d_V\big(x, X_i; X\big) \le k\big\},$$

with $R(x; X) := R_1(x; X)$. If x does not have kth-order neighbors, take $R_k(x; X) = \sqrt{d}$. Then

$$Vol(V(x; X)) \le \kappa_d R(x; X)^d,$$

where $\kappa_d = \pi^{d/2}/\Gamma(d/2 + 1)$ is the volume of the unit ball in $\mathbb{R}^d$.

Proof of Proposition 4.3. The proof relies on two results. First, we have the following technical lemma which is established in Section 5.6.

Lemma 4.1. *Assume there exist $S_+(K), \alpha > 0$ such that $Vol(\partial K^r) \le S_+(K) r^\alpha$ for all $r > 0$. Let*

$$U_k(i) = \mathbf{1}_{d\left(X_i, \partial K\right) \le R_k\left(X_i; X\right)} R_k\big(X_i; X\big)^d.$$

Then, for some $c_{d,qd+\alpha,k} > 0$,

$$\mathbb{E}U_k^q(i) \le S_+(K) c_{d,qd+\alpha,k} n^{-q-\alpha/d}$$

for all $n \ge 1$, $q \ge 1$.

Second, within our framework, we have the following version of [14, Proposition 6.4] where $S(R)$ is the original set of points generated by R and $S(R^i)$ is the set of points generated after the change in the instruction R_i.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Proposition 4.4. *(i) If, for every $s \in S(R) \setminus S(R^i)$, the set $R_1(s, S(R))$, which contains s and all its neighbors, is either entirely in K or entirely in K^c, then $\Delta_i h(R) = 0$. A similar result holds for $s \in S(R^i) \setminus S(R)$ and the set $R_1(s, S(R^i))$.*

(ii) Assume $|i-j|$ is large enough so that $(S(R^i) \setminus S(R)) \cup (S(R^j) \setminus S(R)) = S(R^{ij}) \setminus S(R)$, where $S(R^{ij})$ is the set of points generated after the changes in both R_i and R_j. Suppose that for every $s_1 \in S(R^i) \Delta S(R)$ and $s_2 \in S(R^j) \Delta S(R)$, at least one of the following holds:

1. *$d_V(s_1, s_2; S(R^{ij}) \cap S(R)) \geq 2$, or*
2. *$d_V(s_1, \partial K; S(R^{ij}) \cap S(R)) \geq 2$ and $d_V(s_2, \partial K; S(R^{ij}) \cap S(R)) \cap S(R)) \geq 2$.*

Then $\Delta_{i,j} h(R) = 0$.

Now, write

$$|\Delta_i h(R)| \leq \sum_{s \in S(R) \setminus S(R^i)} \mathbf{1}_{d_{S(R)}(s, \partial K) \leq R_1(s; S(R))} k_d R_1(s; S(R))^d + \sum_{s \in S(R^i) \setminus S(R)} \mathbf{1}_{d_{S(R^i)}(s, \partial K) \leq R_1(s; S(R^i))} k_d R_1(s; S(R^i))^d.$$

As before, for some $T > 0$, there exist an event E and $\epsilon > 0$ such that, conditioned on E, $|S(R^i) \setminus S(R)| = |S(R) \setminus S(R^i)| \leq T$ and $\mathbb{P}(E^c) \leq (1-\epsilon)^T$. Then, from Lemma 4.1, there exist $S_+(K)$, $\alpha > 0$ such that

$$\mathbb{E}|\Delta_i h(R)|^r \leq c_{d,r,\alpha}(1-\epsilon)^T + c_{d,r,\alpha} S_+(K) T^r n^{-r-\alpha/d},$$

where $c_{d,r,\alpha}$ depends on the parameters of the model and the dimension d, as well as on r and α. If $T = c \ln n$ for a suitable $c > 0$, then

$$\mathbb{E}|\Delta_i h(R)|^r \leq c_{d,r,\alpha} S_+(K) (\ln n)^r n^{-r-\alpha/d},$$

as desired. An application of the rth-moment Efron–Stein inequality (see [12, 16]) then yields (23). □

For the non-variance term in Theorem 2.1, we have

$$\frac{1}{4\sigma^3} \sum_{j=0}^{|R|-1} \sqrt{\mathbb{E}|\Delta_j h(R)|^6} + \frac{\sqrt{2\pi}}{16\sigma^3} \sum_{j=0}^{|R|-1} \mathbb{E}|\Delta_j h(R)|^3 \leq C\sigma^{-3} (\ln n)^3 n^{-2-\alpha/2d}. \tag{24}$$

To analyze the bound on the variance terms given by Proposition 2.3, first note that

$$\sum_{i=0}^{|R|-1} \sum_{j,k \notin A} \mathbf{1}_{i=j=k} \mathbb{E}|\Delta_i h(R)|^4 \leq C(\ln n)^4 n^{-3-\alpha/d},$$

again using Proposition 4.3. The other terms are bounded as follows.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Proposition 4.5. *Let $A \subsetneq [|R|]$, and let $B_{|R|}(h)$, $B^k_{|R|}(h)$ and $B^j_{|R|}(h)$ be as in (6). Then, for $\epsilon > 0$,*

$$\sum_{i=0}^{|R|-1} \sum_{j,k \notin A} \mathbf{1}_{i \neq j \neq k} B_{|R|}(h) \leq C_\epsilon \left(n^{-3-2\alpha/d} (\ln n)^{10+4\epsilon} \right),$$

$$\sum_{i=0}^{|R|-1} \sum_{j,k \notin A} \left(\mathbf{1}_{i \neq j = k} + \mathbf{1}_{i = k \neq j} \right) B^{(k)}_{|R|}(h) \leq C_\epsilon \left(n^{-3-2\alpha/d} (\ln n)^{10+4\epsilon} \right),$$

$$\sum_{i=0}^{|R|-1} \sum_{j,k \notin A} \left(\mathbf{1}_{i \neq j = k} + \mathbf{1}_{i = j \neq k} \right) B^{(j)}_{|R|}(h) \leq C_\epsilon \left(n^{-3-2\alpha/d} (\ln n)^{10+4\epsilon} \right),$$

for some constant $C_\epsilon > 0$ that does not depend on n.

Proof. See Section 5.7 for the proof of the first bound. The others follow similarly. □

The bounds on the variance terms in Proposition 2.3 become

$$\begin{aligned} \sqrt{Var(\mathbb{E}[U|R])} \leq & \frac{1}{\sqrt{2}} \sum_{A \subsetneq [|R|]} k_{|R|,A} \left(C \frac{(\ln n)^4}{n^{3+\alpha/d}} + C_\epsilon \frac{(\ln n)^{10+4\epsilon}}{n^{3+2\alpha/d}} \right)^{1/2} \\ \leq & C \frac{(\ln n)^2}{n^{3/2+\alpha/2d}}. \end{aligned} \tag{25}$$

Then (20) follows from (25), (24), and Proposition 2.1.

4.2. Proof of Proposition 4.2

The upper bound follows immediately from Proposition 4.3.

Recall the following result ([14, Corollary 2.4]) concerning the variance. Let $X := (X_1, \ldots, X_n) \in E^n$, where E is a Polish space. If X' is an independent copy of X, and $f : E^n \to \mathbb{R}$ is measurable, with $\mathbb{E}\left[f(X)^2\right] < \infty$,

$$Var(f(X)) \geq \sum_{i=1}^{n} \mathbb{E}\left[\left(\mathbb{E}\left[\Delta_i f\left(X', X\right) | X \right] \right)^2 \right]. \tag{26}$$

In our setting we take $f = \varphi$. Unlike in [14], the function φ is not symmetric, and the right-hand side of (26) cannot be simplified. The lower bound provided by [14, Corollary 2.4] recovers the correct order of growth of the variance, which is enough for our purposes. However, more precise generic results are also available; see e.g. [1].

Let H be the realization of the hidden chain for X. By the law of total variance, $Var(\varphi(X)) \geq Var(\varphi(X)|H)$. Let X' be an independent copy of X, given H. Note that, given H, $(X_i)_{i=1,\ldots,n}$ and $\left(X'_i\right)_{i=1,\ldots,n}$ are independent random variables which are *not* identically distributed.

Applying (26) to $\varphi(X|H)$, we obtain

$$Var(\varphi(X)|H) \geq \sum_{i=1}^{n} \mathbb{E}^H_{X'_i} \left(\mathbb{E}^H_X \left[\varphi\left(X^i\right) - \varphi(X) \right] \right)^2,$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

where $X^i = \big(X_1, \ldots, X_{i-1}, X_i', X_{i+1}, \ldots, X_n\big)$, and $\mathbb{E}^H$ indicates that H is given. (To simplify notation in what follows we drop the H.) The difference from the proof in [14] is that now the variables are no longer identically distributed. Write

$$\mathbb{E}_X\big[\varphi\big(X^i\big) - \varphi(X)\big] = \mathbb{E}_X\big[\varphi\big(X^i\big) - \varphi\big(X^{(i)}\big)\big] - \mathbb{E}_X\big[\varphi(X) - \varphi\big(X^{(i)}\big)\big],$$

where $X^{(i)} = \big(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n\big)$. By Lemma 4.1,

$$\mathbb{E}_X\big[\varphi(X) - \varphi\big(X^{(i)}\big)\big] \le c_{d,\alpha} n^{-1-\alpha/d}. \tag{27}$$

We are left to study $\mathbb{E}\big[\varphi\big(X^i\big) - \varphi\big(X^{(i)}\big)\big]$. Recall that

$$\begin{aligned}\varphi\big(X^i\big) - \varphi\big(X^{(i)}\big) =& \mathbf{1}_{\{X_i' \in K\}} \sum_{j \ne i} \mathbf{1}_{\{X_j \in K^C\}} v\big(X_i', X_j; X^{(i,j)}\big) \\ &- \mathbf{1}_{\{X_i' \in K^C\}} \sum_{j \ne i} \mathbf{1}_{\{X_j \in K\}} v\big(X_i', X_j; X^{(i,j)}\big).\end{aligned}$$

Now, for the case $X_i' \in K^C$ (the other case being equivalent), we have

$$\begin{aligned}&\left|\mathbb{E}_{X, X_i'}\left[- \mathbf{1}_{\{X_i' \in K^C\}} \sum_{j \ne i} \mathbf{1}_{\{X_j \in K\}} v\Big(X_i', X_j; X^{(i,j)}\Big)\right]\right| \\ &\quad \ge \quad \mathbb{E}_{X_i'}\left[\mathbf{1}_{\{X_i' \in \partial K_+^{n^{-1/d}}\}} \sum_{j \ne i} \mathbb{E}_X\Big[\mathbf{1}_{\{X_j \in K\}} v\Big(X_i', X_j; X^{(i,j)}\Big)\Big]\right],\end{aligned}$$

since $v\big(X_i', X_j; X^{(i,j)}\big) \ge 0$. Then

$$\begin{aligned}&\mathbb{E}_X\Big[\mathbf{1}_{\{X_j \in K\}} v\big(x, X_j; X^{(i,j)}\big)\Big]\Big] \\ &\ge \mathbb{E}_{X^{(i,j)}}\left[c_1 \int_{y \in K} v\big(x, y; X^{(i,j)}\big) dy\right] \\ &\ge c_1 Vol\big(B\big(x, \beta n^{-1/d}\big) \cap K\big) \inf_{y\,:\,||x-y|| \le \beta n^{-1/d}} \mathbb{E}_{X^{(i,j)}}\Big[v\big(x, y; X^{(i,j)}\big)\Big],\end{aligned}$$

using the independence after conditioning on H and the properties of the model. We want to find an event that implies that $v\big(x, y; X^{(i,j)}\big) \ge cn^{-1}$. One instance is when no point of $X^{(i,j)}$ falls in $B\big(y, 6\beta n^{-1/d}\big)$. Indeed, then $B\big(y, 3\beta n^{-1/d}\big) \subset V\big(y, X^{(i,j)}\big)$. The distance between y and x is less than $\beta n^{-1/d}$, and so there is $z \in B\big(y, 3\beta n^{-1/d}\big)$, namely $z = x + \beta n^{-1/d}(x-y)/||x-y||$, such that

$$B\big(z, \beta n^{-1/d}\big) \subset V\big(x, \big(X^{(i,j)}, y\big)\big) \subset B\big(y, 3\beta n^{-1/d}\big) \subset V\big(y; X^{(i,j)}\big).$$

Then $v\big(x, y; X^{(i,j)}\big) \ge Vol\big(B\big(z, \beta n^{-1/d}\big) = \kappa_d \beta^d n^{-1}$. Finally,

$$\begin{aligned}&\inf_{y\,:\,||x-y|| \le \beta n^{-1/d}} \mathbb{E}_{X^{(i,j)}}\big[v\big(x, y; X^{(i,j)}\big)\big] \\ &\qquad \ge \kappa_d \beta^d n^{-1} \mathbb{P}\Big(X^{(i,j)} \cap B\Big(y, 6\beta n^{-1/d}\Big) = \emptyset\Big) \\ &\qquad \ge \kappa_d \beta^d n^{-1} \big(1 - c_2 \beta^d n^{-1}\big)^n \\ &\qquad \ge c_{d,\beta} n^{-1},\end{aligned}$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

for some $c_{d,\beta} > 0$ depending on the parameters of the model, the dimension d, and β. Then

$$\mathbb{E}_X\Big[\mathbf{1}_{\{X_j \in K\}} v\big(x, X_j; X^{(i,j)}\big)\Big]\Big] \geq c_{d,\beta} Vol\big(B\big(x, \beta n^{-1/d}\big)\big) n^{-1}.$$

Therefore, by the very definition of $\gamma(K, r, \beta)$ and since the case $X_i' \in K$ is symmetric,

$$\begin{aligned}\mathbb{E}_{X_i'}\mathbb{E}_X\Big[\big(\varphi\big(X^i\big) - \varphi\big(X^{(i)}\big)^2\Big] \geq & c_{d,\beta}\Bigg(c_1 \int_{\partial K_+^{n^{-1/d}}} Vol\big(B\big(x, \beta n^{-1/d}\big) \cap K\big)^2 dx \\ & + c_1 \int_{\partial K_-^{n^{-1/d}}} Vol\big(B\big(x, \beta n^{-1/d}\big) \cap K^c\big)^2 dx\Bigg) \\ = & c_{d,\beta}\big(n^{-2}\gamma\big(K, n^{-1/d}, \beta\big) + n^{-2}\gamma\big(K^c, n^{-1/d}, \beta\big)\big).\end{aligned}$$

If the rolling ball condition (18) and the lower bound on $\partial K^{n^{-1/d}}$ both hold, then

$$\mathbb{E}_{X_i'}\mathbb{E}_X\Big[\big(\varphi\big(X^i\big) - \varphi\big(X^{(i)}\big)^2\Big] \geq c_{d,\beta} S_-(K)\gamma(K, \beta) n^{-2-\alpha/d},$$

which dominates the contribution (27) from $\mathbb{E}\big[\varphi(X) - \varphi\big(X^{(i)}\big)\big]$. Therefore, finally,

$$Var(\varphi(X)) \geq c_{d,\beta}^- S_-(K)\gamma(K, \beta) n^{-1-\alpha/d},$$

as desired.

Remark 4.1. Let us expand a bit on another potential application of our generic framework, namely the occupancy problem as studied in [10]. To set up the notation, $(Z_1, \ldots, Z_n)$ is an aperiodic, irreducible, and time-homogeneous (hidden) Markov chain that transitions between different alphabets. To each alphabet is associated a distribution over the collection of all possible letters, giving rise to the observed letters $(X_1, \ldots, X_n)$. We assume that the number of alphabets is finite but that the number of total letters is $\lfloor \alpha n \rfloor$, for some fixed $\alpha > 0$. One studies $W := f(X_1, \ldots, X_n)$—the number of letters that have not appeared among the $X_1, \ldots, X_n$. Then an analysis as in the proof of Theorem 3.1 leads to the following:

$$d_K\left(\frac{W - \mathbb{E}W}{\sqrt{Var(W)}}, \mathcal{N}\right) \leq C\left(\frac{n(\ln n)^3}{\sqrt{Var(W)^3}} + \frac{n^{1/2}(\ln n)^4}{Var(W)}\right),$$

where $Var(W)$ is a function of n, $\mathcal{N}$ is the standard normal distribution, and $C > 0$ is a constant depending on the parameters of the model but not on n. As mentioned at the beginning of the section, the study of the precise order of growth of the variance of W, in our dependent framework, is not within the scope of the current paper. For the i.i.d. case one can show (see e.g. [8]) that $Var(W) \sim \big(\alpha e^{-1/\alpha} - (1+\alpha)e^{-2/\alpha}\big)n$ as $n \to \infty$.

5. Technical results

5.1. Bounds on terms involving $\Delta_i h$

Recall the setting of Proposition 2.2 and Proposition 2.3. Let (Z, X) be a hidden Markov model and let the latent chain Z be irreducible and aperiodic, with finite state space $\mathcal{S}$. Assume that Z is started at the stationary distribution.

We start by establishing a technical result regarding Z.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

First, note that there exist $K \geq 1$ and $\epsilon \in (0, 1)$ such that

$$\mathbb{P}\big(Z_n = s, Z_{n+K} = s'\big) \geq \epsilon,$$

and thus,

$$\mathbb{P}\big(Z_{n+K} = s'\big) \geq \epsilon, \quad \mathbb{P}\big(Z_{n+K} = s' | Z_n = s\big) \geq \epsilon \tag{28}$$

for all $n \geq 1$ and $s, s' \in \mathcal{S}$.

Lemma 5.1. *Let $K \geq 1$ and $\epsilon \in (0, 1)$ be as in (28), and let $(Z_i)_{i\geq 1}$ be an irreducible and aperiodic Markov chain with finite state space $\mathcal{S}$. Then*

$$\mathbb{P}\left(Z_{j+K} \neq s_1, Z_{j+2K} \neq s_2, \ldots, Z_{j+tK} \neq s_t\right) \leq (1-\epsilon)^t \tag{29}$$

for any $t \geq 1$, $j \geq 1$, and $(s_1, \ldots, s_t) \in \mathcal{S}^t$.

Proof. We show (29) by induction. The case $t = 1$ follows from (28). Next, for $(s_1, \ldots, s_{t+1}) \in \mathcal{S}^{t+1}$,

$$\begin{aligned}
&\mathbb{P}\Big(Z_{j+K} \neq s_1, Z_{j+2K} \neq s_2, \ldots, Z_{j+(t+1)K} \neq s_{t+1}\Big) \\
&= \sum_{s'_1 \neq s_1, \ldots, s'_{n+1} \neq s_{t+1}} \mathbb{P}\Big(Z_{j+K} = s'_1, \ldots, Z_{n+1} = s'_{t+1}\Big) \\
&= \sum_{s'_1 \neq s_1, \ldots, s'_{n+1} \neq s_{t+1}} \mathbb{P}\Big(Z_{j+(t+1)K} = s'_{t+1} | Z_{j+K} = s'_1, \ldots, Z_{j+tK} = s'_t\Big) \\
&\qquad \cdot \mathbb{P}\Big(Z_1 = s'_1, \ldots, Z_{j+tK} = s'_t\Big) \\
&= \sum_{s'_1 \neq s_1, \ldots, s'_{n+1} \neq s_{t+1}} \mathbb{P}\Big(Z_{j+(t+1)K} = s'_{t+1} | Z_{j+tK} = s'_t\Big) \mathbb{P}\Big(Z_{j+K} = s'_1, \ldots, Z_{j+tK} = s'_t\Big) \\
&= \sum_{s'_1 \neq s_1, \ldots, s'_t \neq s_t} \mathbb{P}\Big(Z_{j+(t+1)K} \neq s_{t+1} | Z_{j+tK} = s'_t\Big) \mathbb{P}\Big(Z_{j+K} = s'_1, \ldots, Z_{j+tK} = s'_t\Big) \\
&\leq (1-\epsilon) \sum_{s'_1 \neq s_1, \ldots, s'_t \neq s_n} \mathbb{P}\Big(Z_{j+K} = s'_1, \ldots, Z_{j+tK} = s'_t\Big) \\
&= (1-\epsilon) \mathbb{P}\Big(Z_{j+K} \neq s_1, \ldots, Z_{j+tK} \neq s_t\Big) \\
&\leq (1-\epsilon)^{t+1},
\end{aligned}$$

where we have used the Markov property, (28), and finally the induction hypothesis. This suffices for the proof of (29), and thus the proof of the lemma is complete. □

Let $f : \mathcal{A}^n \to \mathbb{R}$ be Lipschitz, i.e., be such that $|f(x) - f(y)| \leq c \sum_{i=1}^n \mathbf{1}_{x_i \neq y_i}$ for every $x, y \in \mathcal{A}^n$, where $c > 0$. Let $R = \big(R_0, \ldots, R_{|\mathcal{S}|(n-1)}\big)$ be a vector of independent random variables, and let h be the function such that

$$f\big(X_1, \ldots, X_n\big) \stackrel{d}{=} h\big(R_0, \ldots, R_{|\mathcal{S}|(n-1)}\big).$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Let R' be an independent copy of R. The next result provides a tail inequality that is key for Proposition 2.2.

Proposition 5.1. *Let $K > 0$ and $\epsilon > 0$ be as in (28). Then*

$$\mathbb{P}(|\Delta_i h(R)| \geq cx) \leq C(1-\epsilon)^{x/K} \tag{30}$$

for any $x \in \mathbb{N}$, where $C > 0$ depends on the parameters of the model but neither on n nor on x.

Proof. The sequence of instructions $R^i := \left(R_0, \ldots, R'_i, \ldots, R_{|\mathcal{S}(n-1)|}\right)$ may give rise to a different realization (Z', X') of the hidden Markov model, as compared to (Z, X), the one generated by R. The two models are not independent. In particular, if instruction R_i determines $\left(Z_j, X_j\right)$ and R'_i determines $\left(Z'_j, X'_j\right)$, then $\left(Z_k, X_k\right) = \left(Z'_k, X'_k\right)$ for $k < j$. Let s be the smallest nonnegative integer (possibly $s = \infty$) such that $Z_{j+s} = Z'_{j+s}$. Then for any $k > j + s$, $\left(Z_k, X_k\right) = \left(Z'_k, X'_k\right)$ as well. Finally, if $k \in \{j, \ldots, j+s-1\}$, the pairs $\left(Z_k, X_k\right)$ and $\left(Z'_k, X'_k\right)$ are independent. We show next that for $K \geq 1$ as in (28), and any $t \in \mathbb{N}$,

$$\mathbb{P}(s \geq tK) \leq (1-\epsilon)^t. \tag{31}$$

Indeed,

$$\begin{aligned}\mathbb{P}(s > tK) \leq & \quad \mathbb{P}\left(Z_{j+K} \neq Z'_{j+K}, Z_{j+2K} \neq Z'_{j+2K}, \ldots, Z_{j+tK} \neq Z'_{j+tK}\right) \\ = & \sum_{(s_1,\ldots,s_t)\in\mathcal{S}^t} \mathbb{P}\left(Z_{j+K} \neq s_1, Z'_{j+K} = s_1, \ldots, Z_{j+tK} \neq s_t, Z'_{j+tK} = s_t\right).\end{aligned}$$

By independence,

$$\begin{aligned}&\mathbb{P}\left(Z_{j+K} \neq s_1, Z'_{j+K} = s_1, \ldots, Z_{j+tK} \neq s_t, Z'_{j+tK} = s_t\right) \\ = & \quad \mathbb{P}\Big(Z_{j+K} \neq s_1, \ldots, Z_{j+tK} \neq s_t\Big)\mathbb{P}\Big(Z'_{j+K} = s_1, \ldots, Z'_{j+tK} = s_t\Big),\end{aligned}$$

and thus by Lemma 5.1

$$\begin{aligned}\mathbb{P}(s > tK) \leq & \sum_{(s_1,\ldots,s_t)} (1-\epsilon)^t \mathbb{P}\Big(Z'_{j+K} = s_1, \ldots, Z'_{j+tK} = s_t\Big) \\ \leq & \quad (1-\epsilon)^t,\end{aligned}$$

as desired.

Let $E(t)$ be the event

$$E(t) := \left\{X_{j+K} \neq X'_{j+K}, X_{j+2K} \neq X'_{j+2K}, \ldots, X_{j+tK} \neq X'_{j+tK}\right\}.$$

Note that $\mathbb{P}(E(t)) \leq \mathbb{P}(s \geq tK) \leq (1-\epsilon)^t$. In particular, if $|h(R) - h(R^i)| \geq cx$, where $c > 0$ is the Lipschitz constant of the associated function f, then $s \geq x$, as there are at least x positions k such that $X_k \neq X'_k$. Thus,

$$\begin{aligned}\mathbb{P}\Big(|h(R) - h\big(R^i\big)| \geq cx\Big) \leq & \quad \mathbb{P}(E(\lfloor x/K \rfloor)) \\ \leq & \quad C(1-\epsilon)^{x/K},\end{aligned} \tag{32}$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

where $C > 0$ depends on the parameters of the model but not on x. This suffices for the proof of (30). □

We now turn to the proof of Proposition 2.2.

Proof of Proposition 2.2. Let E_t be the event that $|h(R) - h(R^i)| \geq tK$. Then

$$\mathbb{E}|h(R) - h\big(R^i\big)|^r = \mathbb{E}|h(R) - h\big(R^i\big)|^r \mathbf{1}_{E_t} + \mathbb{E}|h(R) - h\big(R^i\big)|^r \mathbf{1}_{E_t^c}.$$

Recall that $|g(x)| \leq cn$ for all $x \in \mathcal{A}^n$, and then $|h(R) - h(R^i)| \leq 2cn$. Using (32),

$$\begin{aligned}\mathbb{E}|h(R) - h(R^i)|^r &\leq (2cn)^r \mathbb{P}(E_t) + (ctK)^r \mathbb{P}(E_t^c) \\ &\leq (2cn)^r (1-\epsilon)^t + (ctK)^r. \end{aligned} \tag{33}$$

Let $t = -r \ln n / (\ln (1-\epsilon)) > 0$. Then

$$\mathbb{E}|h(R) - h\big(R^i\big)|^r \leq (2c)^r + \left(-\frac{crK}{\ln (1-\epsilon)}\right)^r (\ln n)^r. \tag{34}$$

The order of the bound is optimal for t such that

$$(1-\epsilon)^t \leq \left(\frac{\ln n}{n}\right)^r, \tag{35}$$

or

$$t \geq -\frac{r(\ln n - \ln (\ln n))}{\ln (1-\epsilon)};$$

it follows that

$$\mathbb{E}|h(R) - h\big(R^i\big)|^r \leq (2c)^r + \left(-\frac{crK}{\ln (1-\epsilon)}\right)^r (\ln n - \ln (\ln n))^r,$$

and the right-hand side has the same order of growth as (34).

If the growth order of $(1-\epsilon)^t$ is larger than the one in (35), the bound on the second term in (33) is of larger order as well.

Then $\mathbb{E}|\Delta_i h(R)|^r \leq C_1 (\ln n)^r$, for $C_1 > 0$ depending on the parameters of the model and r. The first part of Proposition 2.2 is established.

For the upper bound on the central moments of $f(X)$, recall the following generalizations of the Efron–Stein inequality (see [12, 16]): for $r \geq 2$,

$$\big(\mathbb{E}|h(R) - \mathbb{E}h(R)|^r\big)^{1/r} \leq \frac{r-1}{2^{1/r}} \left(\sum_{i=0}^{|R|-1} \big(\mathbb{E}|h(R) - h\big(R^i\big)|^r\big)^{2/r}\right)^{1/2},$$

and for $r \in (0, 2)$,

$$\big(\mathbb{E}|h(R) - \mathbb{E}h(R)|^r\big)^{1/r} \leq \frac{1}{\sqrt{2}} \left(\sum_{i=0}^{|R|-1} \mathbb{E}\big|h(R) - h\big(R^i\big)\big|^2\right)^{1/2}.$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Then, for all $r > 0$,

$$\begin{aligned}\mathbb{E}|h(R) - \mathbb{E}h(R)|^r &\le \left(\max\left\{\frac{1}{\sqrt{2}}, \frac{r-1}{2^{1/r}}\right\}\right)^r \left((|\mathcal{S}|(n-1)+1)C(\ln n)^2\right)^{r/2} \\ &\le C_2 n^{r/2}(\ln n)^r,\end{aligned}$$

where $C_2 > 0$ is a function of $|\mathcal{S}|$ and r. □

Remark 5.1. (i) Recall that in the independent setting, there is a single stack, or equivalently, the state space of the latent chain consists of a single element. Then for s as defined in the first paragraph of the above proof, $\mathbb{P}(s > 1) = 0$. Thus we can take $tK = 2$, and since $\mathbb{P}(E_t) \le \mathbb{P}(s \ge tk) = 0$, (33) becomes

$$\mathbb{E}|h(R) - h\big(R^i\big)|^r \le (2c)^r,$$

which recovers the independent case.
(ii) Note that the bound on the central moments also follows from using an exponential bounded difference inequality for Markov chains proved by Paulin [15]. This holds for the general case when X is a Markov chain (not necessarily time-homogeneous) taking values in a Polish space $\Lambda = \Lambda_1 \times \cdots \times \Lambda_n$, with mixing time τ_{min}. Then, for any $t \ge 0$,

$$\mathbb{P}(|f(X) - \mathbb{E}[f(X)]| \ge t) \le 2\exp\left(\frac{-2t^2}{||c^*||^2\tau_{min}}\right),$$

where f is such that

$$|f(x) - f(y)| \le \sum_{i=1}^{n} c_i \mathbf{1}_{x_i \ne y_i},$$

for any $x, y \in \mathbb{R}^n$ and some $c^* = (c_1, \ldots, c_n) \in \mathbb{R}^n$, and where $||c^*||^2 = \sum_{i=1}^n c_i^2$.

5.2. Proof of Proposition 2.3

In this section we no longer require the underlying function f to be Lipschitz.

Let $U := \sum_{\emptyset \subseteq A \subsetneq [|R|]} k_{|R|,A} U_A/2$ for a general family of square-integrable random variables $U_A(R, R')$. From [2, Lemma 4.4],

$$\begin{aligned}\sqrt{Var(\mathbb{E}[U|R])} &\le \frac{1}{2} \sum_{\emptyset \subseteq A \subsetneq [|R|]} \sqrt{Var(\mathbb{E}[U_A|R])} \\ &\le \frac{1}{2} \sum_{\emptyset \subseteq A \subsetneq [|R|]} \sqrt{\mathbb{E}[Var(U_A|R')]}.\end{aligned}$$

As in [14], this inequality will be used for both $U_A = T_A(h)$ and $U_A = T'_A(h)$. A major difference from the setting in [14, Section 5] is that the function h is not symmetric; i.e., if σ is a permutation of $\{0, \ldots, |\mathcal{S}|(n-1)\}$, it is not necessarily the case that $h\big(R_0, \ldots, R_{|\mathcal{S}|(n-1)}\big) = h\big(R_{\sigma(0)}, \ldots, R_{\sigma(|R|(n-1))}\big)$. Indeed, each variable in R is associated with a transition at a particular step and from a particular state. Fix $A \subsetneq [|R|]$ and let $\tilde{R}$ be another independent copy of R. Introduce the substitution operator

$$\tilde{S}_i(R) = \big(R_0, \ldots, \tilde{R}_i, \ldots, R_{|R|}\big).$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Recall that from the Efron–Stein inequality,

$$Var\big(U_A|R'\big) \le \frac{1}{2}\sum_{i=0}^{|R|-1}\mathbb{E}\big[\big(\tilde{\Delta}_i U_A(R)\big)^2|R'\big],$$

where $\tilde{\Delta}_i U_A(R) = U_A\big(\tilde{S}_i(R)\big) - U_A(R)$.

Then,

$$\sqrt{Var(\mathbb{E}[U|R])} \le \frac{1}{\sqrt{8}}\sum_{\emptyset\subseteq A\subsetneq[|R|]} k_{|R|,A}\sqrt{\sum_{i=0}^{|R|-1}\mathbb{E}\big[\tilde{\Delta}_i U_A\big]^2}. \tag{36}$$

Recall also that $U_A = \sum_{j\notin A}\Delta_j h(R) a(\Delta_j h(X^A))$, where the function a is either the identity, or $a(\cdot) = |\cdot|$. Then

$$\sum_{i=0}^{|R|-1}\mathbb{E}\big[\tilde{\Delta}_i U_A\big]^2 = \sum_{i=0}^{|R|-1}\sum_{j,k\notin A}\mathbb{E}\big[\big|\tilde{\Delta}_i\big(\Delta_j h(R) a\big(\Delta_j h\big(R^A\big)\big)\big)\big| \times \big|\tilde{\Delta}_i\big(\Delta_k h(R) a\big(\Delta_k h\big(R^A\big)\big)\big)\big|\big]. \tag{37}$$

Fix $0 \le i \le |R|-1$, and note that for $j \notin A$,

$$\begin{aligned}&\tilde{\Delta}_i\big(\Delta_j h(R) - a\big(\Delta_j h\big(R^A\big)\big)\big)\\ &\quad = \tilde{\Delta}_i\big(\Delta_j h(R)\big) a\big(\Delta_j h\big(R^A\big)\big) + \Delta_j h\big(\tilde{S}_i(R)\big)\tilde{\Delta}_i\big(a\big(\Delta_j h\big(R^A\big)\big)\big).\end{aligned} \tag{38}$$

Then, using $|\tilde{\Delta}_i a(\cdot)| \le |\tilde{\Delta}_i(\cdot)|$, the summands in (37) are bounded by

$$4\sup_{Y,Y',Z,Z'}\mathbb{E}\big|\tilde{\Delta}_i\big(\Delta_j h(Y)\big)\Delta_j h\big(Y'\big)\tilde{\Delta}_i\big(\Delta_k h(Z)\big)\Delta_k h\big(Z'\big)\big|, \tag{39}$$

where Y, Y', Z, Z' are recombinations of $R, R', \tilde{R}$; i.e., $Y_i \in \big\{R_i, R'_i, \tilde{R}_i\big\}$, for $i \in [0, |R|-1]$.

Next, as in [14], we bound each type of summand appearing in (37).

If $i = j = k$, and using $\tilde{\Delta}_i\big(\Delta_i(\cdot)\big) = \Delta_i(\cdot)$, (39) is bounded by

$$4\sup_{Y,Y',Z,Z'}\mathbb{E}|\Delta_i h(Y)\Delta_i h\big(Y'\big)\Delta_i h(Z)\Delta_i h\big(Z'\big)| \le 4\mathbb{E}|\Delta_i h(R)|^4.$$

If $i \ne j \ne k$, we can switch $\tilde{R}_i$ and R'_i, and Y is still a recombination. Then (39) is equal to

$$\begin{aligned}&4\sup_{Y,Y',Z,Z'}\mathbb{E}\big[\Delta_i\big(\Delta_j h(Y)\big)\Delta_j h\big(Y'\big)\Delta_i\big(\Delta_k h(Z)\big)\Delta_k h\big(Z'\big)\big]\\ &\le\ 4\sup_{Y,Y',Z,Z'}\mathbb{E}\big[\mathbf{1}_{\Delta_{i,j}h(Y)\ne 0}\big(|\Delta_j h(Y)| + |\Delta_j h\big(Y^i\big)|\big)|\Delta_j h\big(Y'\big)|\\ &\quad\times\mathbf{1}_{\Delta_{i,k}h(Z)\ne 0}\big(|\Delta_k h(Z)| + |\Delta_k h\big(Z^i\big)|\big)|\Delta_k h\big(Z'\big)|\big]\\ &\le\ 16\sup_{Y,Y',Z,Z'}\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\ne 0,\,\Delta_{j,k}h\left(Y'\right)\ne 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\Big],\end{aligned} \tag{40}$$

where the last step follows from the Cauchy–Schwarz inequality.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

If $i \neq j = k$, (39) is equal to

$$\begin{aligned}
&4 \sup_{Y,Y',Z,Z'} \mathbb{E}|\tilde{\Delta}_i\big(\Delta_j\big(h(Y)\big)\Delta_j h\big(Y'\big)\tilde{\Delta}_i\big(\Delta_j\big(h(Z)\big)\Delta_j h\big(Z'\big)|\\
&= 4 \sup_{Y,Z} \mathbb{E}|\tilde{\Delta}_i\big(\Delta_j\big(h(Y)\big)^2\Delta_j h(Z)^2|\\
&= 4 \sup_{Y,Z} \mathbb{E}|\Delta_j\big(\Delta_i\big(h(Y)\big)^2\Delta_j h(Z)^2|\\
&\leq 16 \sup_{Y,Z,Z'} \mathbb{E}|\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\Delta_i h(Z)^2\Delta_j h\big(Z'\big)^2|, \qquad (41)
\end{aligned}$$

where we have exchanged $\tilde{R}_i$ and R'_i and used the Cauchy–Schwarz inequality as in (40).

Similarly, if $i = j \neq k$, the bound is

$$\begin{aligned}
&4 \sup_{Y,Y',Z,Z'} \mathbb{E}|\tilde{\Delta}_i\big(\Delta_i\big(h(Y)\big)\Delta_i h\big(Y'\big)\tilde{\Delta}_i\big(\Delta_k\big(h(Z)\big)\Delta_k h\big(Z'\big)|\\
&= 4 \sup_{Y,Y',Z,Z'} \mathbb{E}|\Delta_i h(Y)\Delta_i h\big(Y'\big)\Delta_i\big(\Delta_k\big(h(Z)\big)\Delta_k h\big(Z'\big)|\\
&= 4 \sup_{Y,Z,Z'} \mathbb{E}|\Delta_i h(Y)^2\Delta_i\big(\Delta_k\big(h(Z)\big)\Delta_k h\big(Z'\big)|\\
&\leq 8 \sup_{Y,Z,Z'} \mathbb{E}|\mathbf{1}_{\Delta_{i,k}h(Y)\neq 0}\Delta_i h(Z)^2\Delta_k h\big(Z'\big)^2|. \qquad (42)
\end{aligned}$$

Finally, if $i = k \neq j$, the bound is by symmetry

$$\begin{aligned}
&4 \sup_{Y,Y',Z,Z'} \mathbb{E}|\tilde{\Delta}_i(\Delta_j(h(Y))\Delta_j h(Y')\tilde{\Delta}_i(\Delta_i(h(Z))\Delta_i h\big(Z'\big)|\\
&\leq 8 \sup_{Y,Z,Z'} \mathbb{E}|\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\Delta_i h(Z)^2\Delta_j h\big(Z'\big)^2|. \qquad (43)
\end{aligned}$$

Combining (40), (41), (42), and (43) in (37), we finally get

$$\begin{aligned}
&\sum_{i=0}^{|R|-1} \mathbb{E}\big[\tilde{\Delta}_i U_A\big]^2\\
&\leq 16 \sum_{i=0}^{|R|-1} \sum_{j,k\notin A} \Big(\mathbf{1}_{i=j=k}\mathbb{E}|\Delta_i h(R)|^4 + \mathbf{1}_{i\neq j\neq k}B_{|R|}(h)\\
&\qquad + \big(\mathbf{1}_{i\neq j=k} + \mathbf{1}_{i=k\neq j}\big)B^{(k)}_{|R|}(h) + \big(\mathbf{1}_{i\neq j=k} + \mathbf{1}_{i=j\neq k}\big)B^{(j)}_{|R|}(h)\Big),
\end{aligned}$$

where

$$\begin{aligned}
B_{|R|}(h) &:= \sup_{Y,Y',Z,Z'} \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\,\Delta_{j,k}h\left(Y'\right)\neq 0}\big|\Delta_j h(Z)\big|^2\big|\Delta_k h\big(Z'\big)|^2\Big],\\
B^{(k)}_{|R|}(h) &:= \sup_{Y,Z,Z'} \mathbb{E}\big|\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\Delta_i h(Z)^2\Delta_j h\big(Z'\big)^2\big|,\\
B^{(j)}_{|R|}(h) &:= \sup_{Y,Z,Z'} \mathbb{E}\big|\mathbf{1}_{\Delta_{i,k}h(Y)\neq 0}\Delta_i h(Z)^2\Delta_k h\big(Z'\big)^2\big|.
\end{aligned}$$

This suffices for the proof of Proposition 2.3.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Remark 5.2. As observed in [6], the terms involving $\Delta_i h(R)$ in (4) and (5) can be removed, leaving only the variance terms. Here is a different way to establish this fact for the particular framework we consider in Section 3. Recall that the expressions on the right-hand sides of (4) and (5) are bounds on terms of the form

$$\mathbb{E}\big|g_t'(W) - g_t'(W)T\big| + \big|\mathbb{E}\big[g_t(W)W - g_t'(W)T\big]\big|,$$

where $|g_t'| \le 1$ and $|g_t(W)W - g_t'(W)| = |\mathbf{1}_{W\le t} - \mathbb{P}(\mathcal{N} \le t)| \le 1$ (see [14] and [3]). First, note that

$$\big|g_t'(W) - g_t'(W)T\big| \ge \big|g_t'(W)T\big| - 1$$

and

$$1 \ge \big|g_t(W)W - g_t'(W)\big| \ge \big|g_t(W)W\big| - 1.$$

Then, by the triangle inequality and the above,

$$\big|g_t(W)W - g_t'(W)T\big| \le \big|g_t(W)W\big| + \big|g_t'(W)T\big| \le \big|g_t'(W) - g_t'(W)T\big| + 3.$$

Let $\mathbb{E}\big|g_t'(W) - g_t'(W)T\big| \le f(n)$ for some function f, with $f(n) \to \infty$ and such that for $\sigma^2 = \sigma^2(n)$, $f(n)/\sigma^2 \to 0$. Then

$$\big|\mathbb{E}\big[g_t(W)W - g_t'(W)T\big]\big|/\sigma^3 \le Cf(n)/\sigma^2,$$

for some constant $C > 0$ that does not depend on n. Therefore, the asymptotic behavior of the bounds in (4) and (5) is given by the terms corresponding to $\mathbb{E}|g_t'(W) - g_t'(W)T|$, i.e., the terms involving the variance. This modification of the method is also valid in our framework and would 'improve' our results. However, it does not have a significant effect on the rates obtained in our applications in Section 3, and so we will not pursue it any further here.

5.3. Proof of Proposition 3.1

Recall that

$$B_{|R|}(h) := \sup_{Y,Y',Z,Z'} \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\ne 0,\,\Delta_{j,k}h\left(Y'\right)\ne 0}|\Delta_j h(Z)|^2|\Delta_k h\left(Z'\right)|^2\Big], \tag{44}$$

where the supremum is taken over recombinations of R and its independent copies R' and R''.

Let E be the event that at least one of the perturbations of the instructions in (44) yields a difference in more than K points. By Proposition 5.1, there is $\epsilon > 0$ such that $\mathbb{P}(E) \le (1-\epsilon)^K$. Then, by the Lipschitz property of h,

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\ne 0,\,\Delta_{j,k}h(Y')\ne 0}|\Delta_j h(Z)|^2|\Delta_k h\left(Z'\right)|^2\Big] \\
&\quad= \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\ne 0,\,\Delta_{j,k}h(Y')\ne 0}|\Delta_j h(Z)|^2|\Delta_k h\left(Z'\right)|^2\mathbf{1}_E\Big] \\
&\qquad+ \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\ne 0,\,\Delta_{j,k}h(Y')\ne 0}|\Delta_j h(Z)|^2|\Delta_k h\left(Z'\right)|^2\mathbf{1}_{E^c}\Big] \\
&\quad\le \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\ne 0,\,\Delta_{j,k}h(Y')\ne 0}|\Delta_j h(Z)|^2|\Delta_k h\left(Z'\right)|^2\mathbf{1}_{E^c}\Big] + Cn^4(1-\epsilon)^K \\
&\quad\le CK^4\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\ne 0,\,\Delta_{j,k}h(Y')\ne 0}\mathbf{1}_{E^c}\Big] + Cn^4(1-\epsilon)^K.
\end{aligned} \tag{45}$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

If $S(Y)$ is the set of points generated by the instructions Y and $S(Y^i)$—the set of points generated by Y after the perturbation of Y_i—let

$$S_1 := S(Y)\Delta S\big(Y^i\big),$$

where Δ is the symmetric difference operator. Similarly, let

$$\begin{aligned} S_2 &:= S(Y)\Delta S\big(Y^j\big),\\ S_3 &:= S(Y')\Delta S\big(\big(Y'\big)^i\big),\\ S_4 &:= S(Y')\Delta S\big(\big(Y'\big)^j\big). \end{aligned}$$

Note that, conditioned on E^c, $|S_i| \le 2K$ for $i = 1, 2, 3, 4$. Furthermore, if $s_1 \cap s_2 = \emptyset$ for all $(s_1, s_2) \in (S_1, S_2)$, then $\Delta_{i,j}h(Y) = 0$. Then

$$\mathbf{1}_{\Delta_{i,j}h(Y)} \le \sum_{(s_1,s_2)\in(S_1,S_2)} \mathbf{1}_{s_1\cap s_2\neq\emptyset}.$$

This bound is meaningful if the sets S_1 and S_2 are disjoint sets of random variables. Conditioned on E^c, this is the case if $|i-j| \ge |\mathcal{S}|K$. We introduce events E_1, E_2, and E_3 corresponding to 0, 1, or 2 of the conditions $\{|i-j| \ge |\mathcal{S}|K, |j-k| \ge |\mathcal{S}|K\}$ holding, respectively. The events E_1, E_2, and E_3 are deterministic. Then we have

$$\mathbb{E}\big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_{E^c}\big] = \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_{E^c}\big(\mathbf{1}_{E_1}+\mathbf{1}_{E_2}+\mathbf{1}_{E_3}\big)\Big].$$

First we use the trivial bound $\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\Delta_{j,k}h(Y')\neq 0} \le 1$ to get

$$\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_{E^c}\mathbf{1}_{E_1}\Big] \le \mathbf{1}_{E_1}. \tag{46}$$

Then, for the term with $\mathbf{1}_{E_3}$,

$$\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_{E^c}\mathbf{1}_{E_3}\Big] \le \mathbf{1}_{E_3}\mathbb{E}\left[\sum_{(s_1,s_2)\in(S_1,S_2)}\sum_{(s_3,s_4)\in(S_3,S_4)}\mathbf{1}_{s_1\cap s_2\neq\emptyset,s_3\cap s_4\neq\emptyset}\right].$$

To bound $\mathbb{E}\big[\mathbf{1}_{s_1\cap s_2\neq\emptyset,s_3\cap s_4\neq\emptyset}\big]$, we condition on s_2, s_3, and the values of all hidden variables H. Then, since S_1 and S_4 are disjoint, we have independence:

$$\begin{aligned} \mathbb{E}\big[\mathbf{1}_{s_1\cap s_2\neq\emptyset,s_3\cap s_4\neq\emptyset}\big] &= \mathbb{E}[\mathbb{E}[\mathbf{1}_{s_1\cap s_2\neq\emptyset,s_3\cap s_4\neq\emptyset}|s_2, s_3, H]]\\ &= \mathbb{E}\big[\mathbb{E}\big[\mathbf{1}_{s_1\cap s_2\neq\emptyset}|s_2, s_3, H\big]\mathbb{E}\big[\mathbf{1}_{s_1\cap s_2\neq\emptyset}|s_2, s_3, H\big]\big]\\ &\le \left(\frac{c_M V_2}{n}\right)^2. \end{aligned}$$

Therefore,

$$\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_{E^c}\mathbf{1}_{E_3}\Big] \le \mathbf{1}_{E_3}CK^4/n^2 \tag{47}$$

for some $C > 0$ independent of K and n, where we have used that $|S_i| \le 2K$ for $i = 1, 2, 3, 4$.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Finally, for the term with E_2, we may assume that $|i-j| \geq |\mathcal{S}|K$, since the case $|j-k| \geq |\mathcal{S}|K$ is identical. Write, using the trivial bound on $\mathbf{1}_{\Delta_{j,k}h(Y')} \neq 0$,

$$\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\,\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_{E^c}\mathbf{1}_{E_2}\Big] \leq \mathbf{1}_{E_3}\mathbb{E}\left[\sum_{(s_1,s_2)\in(S_1,S_2)} \mathbf{1}_{s_1\cap s_2\neq\emptyset}\right].$$

Next, as before,

$$\mathbb{E}\big[\mathbf{1}_{s_1\cap s_2\neq\emptyset}\big] = \mathbb{E}\big[\mathbb{E}[\mathbf{1}_{s_1\cap s_2\neq\emptyset}|s_2, H]\big] \leq \frac{c_M V_2}{n}.$$

Then

$$\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\,\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_{E^c}\mathbf{1}_{E_2}\Big] \leq \mathbf{1}_{E_2} CK^2/n. \tag{48}$$

Combining (45), (46), (48), and (47), we get the following bound on (44):

$$B_{|R|}(h) \leq C\Big(\mathbf{1}_{E_1}K^4 + \mathbf{1}_{E_2}K^6/n + \mathbf{1}_{E_3}K^8/n^2 + n^4(1-\epsilon)^K\Big).$$

Then,

$$\begin{aligned}
&\sum_{i=0}^{|R|-1}\sum_{j,k\notin A} \mathbf{1}_{i\neq j\neq k} B_{|R|}(h)\\
&\quad\leq C\Big(nK^5 + n^2K^7/n + n^3K^8/n^2 + n^7(1-\epsilon)^K\Big)\\
&\quad\leq Cn(\ln n)^8,
\end{aligned}$$

when we choose $K = c\ln n$ for a suitable $c > 0$, independent of n.

5.4. Proof of Proposition 3.2

As in the proof of Proposition 5.1, the sequence of instructions R^i may give rise to a different realization (Z', X'). Indeed, if the instruction R_i determines (Z_j, X_j) and R'_i determines (Z'_j, X'_j), it is possible that $(Z_j, X_j) \neq (Z'_j, X'_j)$. Let $s \geq 0$ be the smallest integer (possibly $s = \infty$) such that $Z_{j+s} = Z'_{j+s}$. Then, as in (31), there is $\epsilon > 0$ such that for $K \in \mathbb{N}$,

$$\mathbb{P}(s \geq K) \leq (1-\epsilon)^K.$$

Fix K, and let E be the event corresponding to $\{s \geq K\}$. Using the trivial bound $|h(R)| \leq n$, and thus $|\Delta_i h(R)| \leq 2n$, we have

$$\begin{aligned}
\mathbb{E}|\Delta_i h|^t = &\quad \mathbb{E}\big[|\Delta_i h|^t\mathbf{1}_E\big] + \mathbb{E}\big[|\Delta_i h|^t\mathbf{1}_{E^c}\big]\\
\leq &\quad (2n)^t(1-\epsilon)^K + \mathbb{E}\big[|\Delta_i h|^t\mathbf{1}_{E^c}\big].
\end{aligned} \tag{49}$$

Let $S(R)$ be the set of points generated by the sequence of instructions R, and let $S(R^j)$ be the points generated by R after the perturbation of R_j. Set $S = S(R)\Delta S(R^j)$ for the symmetric difference and $S^c = S(R) \cap S(R^j)$. Note that E^c implies that $|\mathcal{S}| \leq 2K$. Furthermore,

$$|\Delta_i h| \leq \sum_{s\in S}\sum_{x\in S^c} \mathbf{1}_{s\cap x\neq\emptyset},$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

and

$$|\Delta_i h|^t \leq \sum_{(s_1,\ldots,s_t)\in S^t} \sum_{(x_1,\ldots,x_t)\in (S^c)^t} \prod_{j,\ell=1}^{t} \mathbf{1}_{s_j\cap x_\ell\neq\emptyset},$$

To estimate (49), we need to evaluate

$$\mathbb{E}\left[\prod_{j,\ell=1}^{t} \mathbf{1}_{s_j\cap x_\ell\neq\emptyset}\right],$$

and to do so we proceed as in [14] by studying the shape of the relations of $(s_j, x_\ell)_{j,\ell\in\{1,\ldots,t\}}$.

Identify the set $(s_j, x_\ell)_{j,\ell\in\{1,\ldots,t\}}$ with the edges of the graph G whose vertices correspond to $(s_j)_{j\in\{1,\ldots,t\}}$ and $(x_\ell)_{\ell\in\{1,\ldots,t\}}$. In particular, if $s_{j_1} = s_{j_2}$ for some $j_1 \neq j_2$, we identify them with the same point in the graph G. Conditioned on the realization of the hidden chain Z, we have independence. Then, if G is a tree, fix a root and condition recursively on vertices at different distances from the root. By the restrictions on the volume of the grain and the sampling distribution,

$$\mathbb{E}\left[\prod_{j,\ell=1}^{t} \mathbf{1}_{s_j\cap x_\ell\neq\emptyset}\middle| Z = z^n\right] \leq \left(\frac{c_M V_2}{n}\right)^{|E(G)|},$$

where $|E(G)|$ is the number of edges in the graph G. Furthermore,

$$\mathbb{E}\left[\prod_{j,\ell=1}^{t} \mathbf{1}_{s_j\cap x_\ell\neq\emptyset}\right] \leq \left(\frac{c_M V_2}{n}\right)^{|E(G)|}.$$

Note that the same result holds if G is a graph without cycles, i.e., a collection of disjoint trees. In general, G might have cycles. Let T be a subgraph of G that contains no cycles. Then

$$\prod_{j,\ell=1}^{t} \mathbf{1}_{s_j\cap x_\ell\neq\emptyset} \leq \prod_{e=(e_1,e_2)\in E(T)} \mathbf{1}_{e_1\cap e_2\neq\emptyset},$$

where the product on the right-hand side runs over the edges $e = (e_1, e_2)$ of the graph T with $e_1 \in S$ and $e_2 \in S^c$. Let $|\mathcal{S}|$ be the number of distinct vertices in $(s_1, \ldots, s_t)$, and similarly let $|x|$ be the number for $(x_1, \ldots, x_t)$. The graph G is complete bipartite with $|S| + |x|$ vertices. We can find a subgraph T of G, also with $|S| + |x|$ vertices and no cycles. Then

$$\begin{aligned}
\mathbb{E}[|\Delta_i h|^t \mathbf{1}_{E^c}] \leq\ & \mathbb{E}\left[\mathbf{1}_E^c \sum_{(s_1,\ldots,s_t)\in S^t} \sum_{(x_1,\ldots,x_t)\in (S^c)^t} \prod_{j,\ell=1}^{t} \mathbf{1}_{s_j\cap x_\ell\neq\emptyset}\right] \\
=\ & \mathbb{E}\left[\mathbf{1}_E^c \sum_{a,b=1}^{t} \sum_{\substack{(s_1,\ldots,s_t)\in S^t,\\ |\mathcal{S}|=a}} \sum_{\substack{(x_1,\ldots,x_t)\in (S^c)^t,\\ |x|=b}} \prod_{j,\ell=1}^{t} \mathbf{1}_{s_j\cap x_\ell\neq\emptyset}\right] \\
\leq\ & \mathbb{E}\left[\mathbf{1}_E^c \sum_{a,b=1}^{t} C_t |S|^a |S^c|^b \left(\frac{c_M V_2}{n}\right)^{a+b-1}\right] \\
\leq\ & C_t K^r,
\end{aligned}$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

where $C_t > 0$ is a constant depending on t, and where we have used that $|\mathcal{S}| \le 2K$ and $|S^c| \le 2n$.

Let $K = c\ln n$ for a suitable $c > 0$; then (49) implies (15) as desired.

5.5. Proof of Proposition 3.3

As before, let E be the event that all perturbations of instructions in (44) propagate at most K levels. We have that $\mathbb{P}(E^c) \le (1-\epsilon)^K$ for some $\epsilon \in (0,1)$. Using the trivial bound $|h(Y)| \le n$,

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\,\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\Big]\\
&\quad=\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\,\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\Big]\\
&\qquad+\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\,\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_{E^c}\Big]\\
&\quad\le\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\,\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\Big]+4n^4(1-\epsilon)^K. \qquad (50)
\end{aligned}$$

Let $S(Y^i)$ be the set of points generated by the sequence of instructions Y after the perturbation of Y_i. Let S be the set of all points in the expectation above, and furthermore let

$$\begin{aligned}
&S_1 := S(Y)\Delta S\big(Y^i\big), \quad S_2 := S(Y)\Delta S\big(Y^j\big),\\
&S_3 := S\big(Y'\big)\Delta S\big(\big(Y'\big)^j\big), \quad S_4 := S\big(Y'\big)\Delta S\big(\big(Y'\big)^k\big),\\
&S_5 := S(Z)\Delta S\big(Z^j\big), \quad S_6 := S\big(Z'\big)\Delta S\big(Z^k\big),
\end{aligned}$$

where Δ is the symmetric difference operator. Conditioned on E, $|S_i| \le 2K$ for $i = 1,\dots,6$ and $|\mathcal{S}| \le 10n$.

Conditioned on E, if $j - i \le |\mathcal{S}|K$, the perturbation in i might be propagating past the position corresponding to the instruction j, leading to difficulties in the analysis of $\Delta_{i,j}h(Y)$. This is why we condition further on the events E_1, E_2, E_3 corresponding respectively to 0, 1, or 2 of the conditions $\{|i-j| \ge |\mathcal{S}|K, |j-k| \ge |\mathcal{S}|K\}$ holding true. Note that E_1, E_2, and E_3 are deterministic.

If E_1 holds, we use the trivial bound $\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\,\Delta_{j,k}h(Y')\neq 0} \le 1$, which leads to

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\,\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{E_1}\Big]\\
&\le\ \mathbb{E}\Big[|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{E_1}\Big]\\
&\le\ \mathbf{1}_{E_1}CK^4, \qquad (51)
\end{aligned}$$

using the Cauchy–Schwarz inequality.

Conditioned on E_3, the sets S_1, $S_2 \cup S_3$, and S_4 are pairwise disjoint. Next, in similarity to an argument presented in [14], if $s_1 \cap s = \emptyset$ and $s_2 \cap s = \emptyset$ for all $(s_1, s_2, s) \in (S_1, S_2, S)$, then $\Delta_{i,j}h(Y) = 0$. Therefore,

$$\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0} \le \sum_{\substack{s_1\in S_1\\ s_2\in S_2}}\sum_{s\in S}\mathbf{1}_{s_1\cap s\neq\emptyset,\,s_2\cap s\neq\emptyset},$$

and also

$$\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0} \le \sum_{\substack{s_3\in S_3\\ s_4\in S_4}}\sum_{s\in S}\mathbf{1}_{s_3\cap s\neq\emptyset,\,s_4\cap s\neq\emptyset}.$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Furthermore,

$$|\Delta_j h(Z)| \le \sum_{s_5 \in S_5} \sum_{s \in S} \mathbf{1}_{s_5 \cap s \neq \emptyset}$$

and

$$|\Delta_k h(Z')| \le \sum_{s_6 \in S} \sum_{s \in S} \mathbf{1}_{s_6 \cap s \neq \emptyset}.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0, \Delta_{j,k}h(Y')\neq 0} |\Delta_j h(Z)|^2 |\Delta_k h(Z')|^2 \mathbf{1}_E \mathbf{1}_{E_3}\big] \\
&\le \mathbb{E}\Bigg[\Bigg(\sum_{\substack{(s_1,s_2,s_3,s_4)\in(S_1,S_2,S_3,S_4)\\(s',s'')\in S^2}} \mathbf{1}_{\substack{s_1\cap s'\neq\emptyset, s_2\cap s'\neq\emptyset,\\ s_3\cap s''\neq\emptyset, s_4\cap s''\neq\emptyset}}\Bigg) \\
&\qquad \cdot \Bigg(\sum_{s_5\in S_5}\sum_{s\in S}\mathbf{1}_{s_5\cap s\neq\emptyset}\Bigg)^2 \Bigg(\sum_{s_6\in S_6}\sum_{s\in S}\mathbf{1}_{s_6\cap s\neq\emptyset}\Bigg)^2 \mathbf{1}_E\mathbf{1}_{E_3}\Bigg] \\
&\le \mathbb{E}\Bigg[\sum_{\substack{(s_1,\dots,s_4)\in(S_1,\dots,S_4)\\(s_5,\dots,s_8)\in S_{56}^4}} \sum_{\substack{(s',s'')\in S^2\\(s'_5,\dots,s'_8)\in S^4}} \mathbf{1}_{\substack{s_1\cap s'\neq\emptyset, s_2\cap s'\neq\emptyset,\\ s_3\cap s''\neq\emptyset, s_4\cap s''\neq\emptyset}} \prod_{a,b=5}^{8} \mathbf{1}_{s_a\cap s'_b\neq\emptyset} \mathbf{1}_E\mathbf{1}_{E_3}\Bigg], \qquad (52)
\end{aligned}$$

where $S_{56} = S_5 \cup S_6$ and $|S_{56}| \le 4K$, conditioned on E.

To evaluate the summand expression we use the graph representation. Let E_ℓ be the event that there are ℓ distinct points among $s', s'', s'_5, \dots, s'_8$, different from $s_1, \dots, s_8$. Note that $\ell \in [0, 6]$. Conditioned on E_ℓ, we can find a subgraph with no cycles and $\ell + 2$ edges, of the graph with edges $\{\{s_1, s'\}, \{s_2, s'\}, \{s_3, s''\}, \{s_4, s''\}\} \cup \{\{s_a, s'_b\} : a, b \in [5, 8]\}$. Indeed, note that there are at least 3 different points among $s_1, \dots, s_4$. Next, if there are x points present among s', s'' and $\ell - x$ points among $s'_5, \dots, s'_8$, we can find a subgraph with no cycles with at least $\ell - x$ edges among $\{\{s_a, s'_b\} : a, b \in [5, 8]\}$ and $x + 2$ edges among $\{\{s_1, s'\}, \{s_2, s'\}, \{s_3, s''\}, \{s_4, s''\}\}$.

Then, if we further condition on the values of the hidden variables H, we get, by independence,

$$\mathbb{E}\left[\mathbf{1}_{\substack{s_1\cap s'\neq\emptyset, s_2\cap s'\neq\emptyset,\\ s_3\cap s''\neq\emptyset, s_4\cap s''\neq\emptyset}} \prod_{a,b=5}^{8} \mathbf{1}_{s_a\cap s'_b\neq\emptyset} \mathbf{1}_E\mathbf{1}_{E_3}\mathbf{1}_{E_\ell} \,\middle|\, H\right] \le \mathbf{1}_{E_3} \left(\frac{c_M V_2}{n}\right)^{\ell+2}.$$

Then (52) is further bounded by

$$\mathbf{1}_{E_3} \sum_{\ell=0}^{6} (4K)^8 \binom{6}{\ell} (10n)^\ell \left(\frac{c_M V_2}{n}\right)^{\ell+2} \le \mathbf{1}_{E_3} C K^8 n^{-2}, \qquad (53)$$

for some $C > 0$ independent of n and K.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Finally, assume that E_2 holds and that $|i-j| \geq |\mathcal{S}|K$. The case $|j-k| \geq |\mathcal{S}|K$ is identical. As above, using the trivial bound $\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0} \leq 1$,

$$\begin{aligned}
&\mathbb{E}\big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0,\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{E_2}\big]\\
&\leq \mathbb{E}\big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{E_2}\big]\\
&\leq \mathbb{E}\left[\left(\sum_{\substack{(s_1,s_2)\in(S_1,S_2)\\ s'\in S}} \mathbf{1}_{s_1\cap s'\neq\emptyset,s_2\cap s'\neq\emptyset}\right)\left(\sum_{s_5\in S_5}\sum_{s\in S}\mathbf{1}_{s_5\cap s\neq\emptyset}\right)^2\left(\sum_{s_6\in S_6}\sum_{s\in S}\mathbf{1}_{s_6\cap s\neq\emptyset}\right)^2\mathbf{1}_E\mathbf{1}_{E_2}\right]\\
&\leq \mathbb{E}\left[\sum_{\substack{(s_1s_2)\in(S_1,S_2)\\ (s_5,\ldots,s_8)\in S_{56}^4}}\sum_{\substack{s'\in S\\ \left(s'_5,\ldots,s'_8\right)\in S^4}}\mathbf{1}_{s_1\cap s'\neq\emptyset,s_2\cap s'\neq\emptyset}\prod_{a,b=5}^{8}\mathbf{1}_{s_a\cap s'_b\neq\emptyset}\mathbf{1}_E\mathbf{1}_{E_2}\right]. \qquad (54)
\end{aligned}$$

If we condition on E_ℓ and the values of the hidden variables H, we get

$$\mathbb{E}\left[\mathbf{1}_{s_1\cap s'\neq\emptyset,s_2\cap s'\neq\emptyset}\prod_{a,b=5}^{8}\mathbf{1}_{s_a\cap s'_b\neq\emptyset}\mathbf{1}_E\mathbf{1}_{E_2}\mathbf{1}_{E_\ell}|H\right]\leq \mathbf{1}_{E_2}\left(\frac{c_M V_2}{n}\right)^{\ell+1},$$

since in this case s_1 and s_2 are distinct and we can find a subgraph with $\ell+1$ edges and no cycles.

Then (54) is bounded by

$$\mathbf{1}_{E_2}\sum_{\ell=0}^{6}(4K)^6\binom{6}{\ell}(10n)^\ell\left(\frac{c_M V_2}{n}\right)^{\ell+1}\leq \mathbf{1}_{E_2}CK^6n^{-1}, \qquad (55)$$

for some $C>0$.

We get the following bound on $B_{|R|}(h)$ using (50), (51), (55), and (53):

$$B_{|R|}(h)\leq C\Big(\mathbf{1}_{E_1}K^4+\mathbf{1}_{E_2}K^6/n+\mathbf{1}_{E_3}K^8/n^2+n^4(1-\epsilon)^K\Big).$$

Then,

$$\begin{aligned}
&\sum_{i=0}^{|R|-1}\sum_{j,k\notin A}\mathbf{1}_{i\neq j\neq k}B_{|R|}(h)\\
&\quad\leq C\Big(nK^6+n^2K^7/n+n^3K^8/n^2+n^7(1-\epsilon)^K\Big)\\
&\quad\leq Cn(\ln n)^8,
\end{aligned}$$

where we have chosen $K=c\ln n$ for a suitable $c>0$, independent of n.

5.6. Proof of Lemma 4.1

To simplify computations, we introduce the process X' defined as

$$X'=\bigcup_{m\in\mathbb{Z}^d}(X+m).$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Unlike in the independent setting in [14], here the law of X' is invariant only under integer-valued translations. Note that, almost surely, X' has exactly n points in any cube $[t, t+1]^d$, where $t \in \mathbb{R}$. Let $T_x = \{[y, y+1]^d : y \in \mathbb{R}^d, x \in [y, y+1]^d\}$. Define $\overline{R_k}(x; X)$ as

$$\overline{R_k}(x; X) := \sup_{T \in T_x} R_k(x; X' \cap T).$$

Note that if $x \in [0, 1]^d$, then $[0, 1]^d \in T_x$ and so $\overline{R_k}(x; X') \geq R_k(x; X)$. When the X_i are sampled independently and uniformly, as in [14], it is the case that $\overline{R_k}(x; X')$ does not depend on the position of x. However, in the hidden Markov model case we need to find a further bound on $\overline{R_k}(x; X')$.

For that purpose, consider the cube $K_0 := [-1/2, 1/2]^d$ of volume 1 centered at $\mathbf{0} \in \mathbb{R}^d$. Let B_A be the open ball of $\mathbb{R}^d$ centered at $\mathbf{0}$ and of volume $A < 1$, to be chosen later. Next, let $\tilde{X} = (0, \tilde{X}_1, \ldots, \tilde{X}_{n-1})$ be such that $\tilde{X}_i \in K_0$ for all $i = 1, \ldots, n-1$. Furthermore, for any Lebesgue-measurable $T \subseteq K_0$, set

$$\mathbb{P}(\tilde{X}_i \in T) = c_m |T \cap B_A| + c_M |T \cap B_A^c|$$

for all $i \in 1, \ldots, n-1$, where $|\cdot|$ now denotes the Lebesgue measure of the corresponding sets. If $A = (c_M - 1)/(c_M - c_m)$, then the above is a well-defined positive measure on K_0. From the restrictions of the hidden Markov model, if $\tilde{R}_k = R_k(0; \tilde{X})$,

$$\overline{R_k}(x; X) \leq \tilde{R}_k.$$

Indeed, $\tilde{R}_k$ represents the worst-case scenario where the remaining points of X are least likely to be distributed in the volume closest to x.

Then,

$$\mathbb{E}U_k^q(i) \leq \mathbb{E}_{X_i, \tilde{X}}\left[\mathbf{1}_{d(X_i; \partial K) \leq \tilde{R}_k} \tilde{R}_k^{qd}\right] \leq S_+(K)\mathbb{E}_{\tilde{X}}\left[\tilde{R}_k^{qd+\alpha}\right], \tag{56}$$

where we have used the upper bound on $Vol(\partial K^r)$.

To estimate $\mathbb{E}\left[\tilde{R}_k^{qd+\alpha}\right]$, note that if $\tilde{R}_k \geq r$, there will be an open ball of radius $r/2k$ in K_0 containing no points of $\tilde{X}$. Moreover, there will be $s_d \in (0, 1)$, depending only on the dimension d, such that every ball of radius $2k$ contains a cube of side length $s_d r/k$ of the form $[g - s_d r/2k, g + s_d r/2k]$, where $g \in (s_d r/k)\mathbb{Z}^d$. Then, if $s_d r/k < 1$,

$$\begin{aligned}
\mathbb{P}(\tilde{R}_k \geq r) \leq & \mathbb{P}(\exists g \in (s_d r/k)\mathbb{Z}^d : \tilde{X} \cap [g - s_d r/2k, g + s_d r/2k] = \mathbf{0}) \\
\leq & \#\{g : g \in (s_d r/k)\mathbb{Z}^d \cap [-r, r]^d\} \mathbb{P}(\tilde{X} \cap [-s_d r/2k, s_d r/2k] = \mathbf{0}) \\
\leq & \frac{k^d}{(s_d)^d}\left(1 - c_m (s_d r/k)^d\right)^{n-1}.
\end{aligned}$$

If, on the other hand, $s_d r/k \geq 1$, then $\tilde{X} \cap [g - s_d r/2k, g + s_d r/2k] = \tilde{X}$ and $\mathbb{P}(\tilde{R}_k \geq r) = 0$. Using $1 - x \leq e^{-x}$, for any $u > 0$ we have

$$\begin{aligned}
\mathbb{E}\left[\tilde{R}(0, \tilde{X})^u\right] = & \int_0^\infty \mathbb{P}\left(\tilde{R}(0, \tilde{X}) \geq r^{1/u}\right) dr \\
\leq & c_{d,k} \int_0^\infty \left(1 - c_m \left(s_d r^{1/u}/k\right)^d\right)^{n-1} dr \\
\leq & c_{d,k} \int_0^\infty \exp\left(-c_m (n-1)\left(s_d r^{1/u}/k\right)^d\right) dr \\
\leq & c_{d,k,u} (n-1)^{u/d} \int_0^\infty \exp\left(-r^{d/u}\right) dr.
\end{aligned}$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Applying the above in (56) yields

$$\mathbb{E}U_k^q(i) \le c_{d,k,qd+\alpha} S_+(K) n^{-q-\alpha/d},$$

where $c_{d,k,qd+\alpha} > 0$ depends only on the parameters of the transition probabilities of the hidden chain and on d, k, and $qd+\alpha$, but neither on n nor on i.

5.7. Proof of Proposition 4.5

We analyze

$$B_{|R|}(h) := \sup_{Y,Y',Z,Z'} \mathbb{E}\big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\big], \tag{57}$$

where as before the supremum is taken over recombinations Y, Y', Z, Z' of R, R', R''. Let E be the event that all perturbations of the instructions in (57) propagate at most T levels. There is $\epsilon > 0$, depending only on the parameters of the models, such that $\mathbb{P}(E^c) \le (1-\epsilon)^T$.

As before, conditioned on E, if $|j-i| \le |\mathcal{S}|K$, the perturbation in i might be propagating past the position corresponding to the instruction j, leading to difficulties in the analysis of $\Delta_{i,j}h(Y)$. This is the reason for conditioning further on the events E_1, E_2, E_3 corresponding respectively to 0, 1, or 2 of the conditions $\{|i-j| \ge |\mathcal{S}|K, |j-k| \ge |\mathcal{S}|K\}$ holding. Note that E_1, E_2, and E_3 are deterministic.

In this setting, we also study the event that all Voronoi cells are small. For that purpose, as in [14], we introduce the event $\Omega_n(X)$, given by

$$\Omega_n(X) := \left(\max_{1\le j\le n} R\big(X_j; X\big) \le n^{-1/d}\rho_n\right),$$

where $\rho_n = (\ln n)^{1/d+\epsilon'}$ for ϵ' sufficiently small. Then, after conditioning on the realization of the hidden chain, a proof as in [14, Lemma 6.8] leads to

$$n^\eta(1 - \mathbb{P}(\Omega_n(X))) \to 0 \tag{58}$$

as $n \to \infty$, for all $\eta > 0$.

We now estimate $B_{|R|}(h)$. Write

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\Big] \\
&= \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_{E^c}\Big] \\
&\quad + \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{\Omega_n^c}\Big] \\
&\quad + \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{\Omega_n}\mathbf{1}_{E_1}\Big] \\
&\quad + \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{\Omega_n}\mathbf{1}_{E_2}\Big] \\
&\quad + \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{\Omega_n}\mathbf{1}_{E_3}\Big].
\end{aligned} \tag{59}$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Using $|\Delta_j h(Z)|, |\Delta_k h(Z')| \le 1$, we get that the first two terms in (59) are bounded by $\mathbb{P}(E^c) + \mathbb{P}(\Omega_n^c)$. Next,

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h(Z')|^2\mathbf{1}_E\mathbf{1}_{\Omega_n}\mathbf{1}_{E_1}\Big]\\
&\le \mathbf{1}_{E_1}\mathbb{E}\Big[|\Delta_j h(Z)|^2|\Delta_k h(Z')|^2\mathbf{1}_E\mathbf{1}_{\Omega_n}\Big]\\
&\le C\mathbf{1}_{E_1}T^4 n^{-4-2\alpha/d}\rho_n^{4d},
\end{aligned} \tag{60}$$

where we have used the Cauchy–Schwarz inequality.

Next, define as before

$$\begin{aligned}
S_1 &:= S(Y)\Delta S(Y^i), \quad S_2 := S(Y)\Delta S(Y^j),\\
S_3 &:= S(Y')\Delta S((Y')^j), \quad S_4 := S(Y')\Delta S((Y')^k).
\end{aligned}$$

Further, let $S_0 = S(Y)\cap S(Y^i)\cap S(Y^j)$ and $S_0' = S(Y')\cap S((Y')^j)\cap S((Y')^k)$. By Proposition 4.4(ii), it follows that conditioned on Ω_n,

$$\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0} \le \sum_{s_1\in S_1, s_2\in S_2} \mathbf{1}_{d_{S_0}(s_1,\partial K)\le 2n^{-1/d}\rho_n}\mathbf{1}_{d_{S_0}(s_2,\partial K)\le 2n^{-1/d}\rho_n}\mathbf{1}_{d_{S_0}(s_1,s_2)\le 2n^{-1/d}\rho_n}.$$

Conditioned on E_3, the sets S_1, $S_2\cup S_3$, and S_4 are pairwise disjoint:

$$\begin{aligned}
&\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h(Z')|^2\mathbf{1}_E\mathbf{1}_{\Omega_n}\mathbf{1}_{E_3}\Big]\\
&\le C\mathbf{1}_{E_3}T^4 n^{-4-2\alpha/d}\rho_n^{4d}\mathbb{E}\big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_E\mathbf{1}_{\Omega_n}\big].
\end{aligned}$$

By conditioning on the realization of all hidden chains H, we obtain

$$\begin{aligned}
&\mathbb{E}\big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_E\mathbf{1}_{\Omega_n}\big]\\
&= \mathbb{E}\big[\mathbb{E}\big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}\mathbf{1}_E\mathbf{1}_{\Omega_n}|H\big]\big]\\
&\le \mathbb{E}\left[\mathbb{E}\left[\sum_{\substack{s_1\in S_1, s_2\in S_2\\ s_1'\in S_3, s_2'\in S_4}} \mathbf{1}_{d_{S_0}(s_1',\partial K)\le 2n^{-1/d}\rho_n}\mathbf{1}_{d_{S_0}(s_1,s_2)\le 2n^{-1/d}\rho_n}\mathbf{1}_{d_{S_0'}(s_1',s_2')\le 2n^{-1/d}\rho_n}\mathbf{1}_E\mathbf{1}_{\Omega_n}|H\right]\right]\\
&\le \mathbb{E}\mathbb{E}\Bigg[\sum_{s_2\in S_2, s_1'\in S_1} \mathbf{1}_{d_{S_0}(s_1',\partial K)\le 2n^{-1/d}\rho_n}\mathbf{1}_E\mathbf{1}_{\Omega_n}\\
&\qquad \mathbb{E}\Bigg[\sum_{s_1\in S_1, s_2'\in S_4} \mathbf{1}_{d_{S_0}(s_1,s_2)\le 2n^{-1/d}\rho_n}\mathbf{1}_{d_{S_0'}(s_1',s_2')\le 2n^{-1/d}\rho_n}\Big|s_1', s_2\Bigg]\Big|H\Bigg].
\end{aligned}$$

Now, conditioned on H, s_1', and s_2, we have independence in the innermost expectation. Therefore, the above is bounded by

$$\mathbb{E}\Bigg[\sum_{s_2\in S_2, s_1'\in S_1} \mathbf{1}_{d_{S_0}(s_1',\partial K)\le 2n^{-1/d}\rho_n}\mathbf{1}_E\mathbf{1}_{\Omega_n}4T^2 2^d n^{-2}\rho_n^{2d}\Bigg] \le CT^4 n^{-2}\rho_n^{2d}n^{-\alpha/d}\rho_n^{\alpha}.$$

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

Then,

$$\begin{aligned}&\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{\Omega_n}\mathbf{1}_{E_3}\Big]\\ &\leq \quad C\mathbf{1}_{E_3}T^8 n^{-6-3\alpha/d}\rho_n^{6d+\alpha}.\end{aligned} \tag{61}$$

Finally, for the event E_2, assuming that $|i-j|\geq |\mathcal{S}|K$, the other case being identical, we have

$$\begin{aligned}&\mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}\mathbf{1}_{\Delta_{j,k}h(Y')\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{\Omega_n}\mathbf{1}_{E_2}\Big]\\ &\leq \mathbb{E}\Big[\mathbf{1}_{\Delta_{i,j}h(Y)\neq 0}|\Delta_j h(Z)|^2|\Delta_k h\big(Z'\big)|^2\mathbf{1}_E\mathbf{1}_{\Omega_n}\mathbf{1}_{E_2}\Big]\\ &\leq C\mathbf{1}_{E_2}T^6 n^{-5-3\alpha/d}\rho_n^{5d+\alpha}.\end{aligned} \tag{62}$$

Using (59), (60), (62), and (61) leads to

$$\begin{aligned}B_{|R|}(h)\leq C\Big((1-\epsilon)^T+\mathbb{P}\big(\Omega_n^c\big)+\mathbf{1}_{E_1}T^4 n^{-4-2\alpha/d}\rho_n^{4d}\\ +\mathbf{1}_{E_2}T^6 n^{-5-3\alpha/d}\rho_n^{5d+\alpha}+\mathbf{1}_{E_3}T^8 n^{-6-3\alpha/d}\rho_n^{6d+\alpha}\Big).\end{aligned}$$

Then,

$$\begin{aligned}&\sum_{i=0}^{|R|-1}\sum_{j,k\notin A}\mathbf{1}_{i\neq j\neq k}B_{|R|}(h)\\ &\leq C\Big(n^3(1-\epsilon)^T+n^3\mathbb{P}\big(\Omega_n^c\big)+T^6 n^{-3-2\alpha/d}\rho_n^{4d}+T^7 n^{-3-3\alpha/d}\rho_n^{5d+\alpha}+T^8 n^{-3-3\alpha/d}\rho_n^{6d+\alpha}\Big)\\ &\leq C\big(n^{-3-2\alpha/d}(\ln n)^{10+4\epsilon'}\big),\end{aligned}$$

where we have chosen $K=c\ln n$, for a suitable $c>0$, independent of n, using also (58) and the definition of ρ_n.

Funding information

The first author's research is supported in part by grant no. 524678 from the Simons Foundation. The second author was partially supported by a TRIAD NSF grant (award 1740776) and the FNR grant APOGee at Luxembourg University (R-AGR-3585-10-C).

Competing interests

There were no competing interests to declare which arose during the preparation or publication process for this article.

References

[1] BOUSQUET, O. AND HOUDRÉ, C. (2019). Iterated jackknives and two-sided variance inequalities. In *High Dimensional Probability VIII: The Oaxaca Volume* (Progress in Probability **74**), Birkhäuser, Cham, pp. 33–40.
[2] CHATTERJEE, S. (2008). A new method for normal approximation. *Ann. Prob.* **36**, 1584–1610.
[3] CHATTERJEE, S. (2014). A short survey of Stein's method. In *Proceedings of the International Congress of Mathematicians: Seoul 2014*, Vol. IV, Kyung Moon Sa, Seoul, pp. 1–24.
[4] CHEN, L. H. Y., GOLDSTEIN, L. AND SHAO, Q.-M. (2014). *Normal Approximation by Stein's Method*. Springer, Berlin, Heidelberg.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press

[5] CHEN, P., SHAO, Q.-M. AND XU, L. (2020). *A universal probability approximation method: Markov process approach*. Preprint. Available at https://arxiv.org/abs/2011.10985.
[6] CHU, D., SHAO, Q.-M. AND ZHANG, Z. (2019). Berry–Esseen bounds for functionals of independent random variables. Presented at the Symposium in Memory of Charles Stein (1920–2016). Available at https://projecteuclid.org/journals/annals-of-probability/volume-47/issue-1/BerryEsseen-bounds-of-normal-and-nonnormal-approximation-for-unbounded-exchangeable/10.1214/18-AOP1255.full.
[7] DURBIN, R., EDDY, S., KROGH, A. AND MITCHISON, G. (1998). *Biological Sequence Analysis*. Cambridge University Press.
[8] ENGLUND, G. (1981). A remainder term estimate for the normal approximation in classical occupancy. *Ann. Prob.* **9**, 684–692.
[9] GORODEZKY, I. AND PAK, I. (2012). Generalized loop-erased random walks and approximate reachability. *Random Structures Algorithms* **44**, 201–223.
[10] GRABCHAK, M., KELBERT, M. AND PARIS, Q. (2020). On the occupancy problem for a regime-switching model. *J. Appl. Prob.* **57**, 53–77.
[11] HOUDRÉ, C. AND KERCHEV, G. (2019). On the rate of convergence for the length of the longest common subsequences in hidden Markov models. *J. Appl. Prob.* **56**, 558–573.
[12] HOUDRÉ, C. AND MA, J. (2016). On the order of the central moments of the length of the longest common subsequences in random words. In *High Dimensional Probability VII: The Cargèse Volume* (Progress in Probability **71**), Birkhäuser, Cham, pp. 105–136.
[13] KENDALL, W. S. AND MOLCHANOV, I. (eds) (2010). *New Perspectives in Stochastic Geometry*. Oxford University Press.
[14] LACHIÈZE-REY, R. AND PECCATI, G. (2017). New Berry–Esseen bounds for functionals of binomial point process. *Ann. Appl. Prob.* **27**, 1992–2031.
[15] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Prob.* **20**, 32 pp.
[16] RHEE, W. AND TALAGRAND, M. (1986). Martingale inequalities and the jackknife estimate of the variance. *Statist. Prob. Lett.* **4**, 5–6.
[17] WILSON, D. B. (1996). Generating random spanning trees more quickly than the cover time. In *STOC '96: Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, Association for Computing Machinery, New York, pp. 296–303.

https://doi.org/10.1017/apr.2021.40 Published online by Cambridge University Press