Electron. Commun. Probab. **29** (2024), article no. 2, 1–13.
https://doi.org/10.1214/23-ECP569
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Cutoff in the Bernoulli-Laplace urn model with swaps of order n

Joseph S. Alameda* Caroline Bang† Zachary Brennan‡
David P. Herzog§ Jürgen Kritschgau¶ Elizabeth Sprangel‖

Abstract

This paper considers the (n,k)-Bernoulli–Laplace urn model in the case when there are two urns containing n balls each, with two different colors of balls (*red* and *white*). In our setting, the total number of red and white balls is the same. Our focus is on the large-time behavior of the corresponding Markov chain tracking the number of red balls in a given urn assuming that the number of selections k at each step obeys $\alpha \leq k/n \leq \beta$, where α, β are constants satisfying $0 < \alpha < \beta < \frac{1}{2}$. Under this assumption, cutoff in the total variation distance is established and a cutoff window is provided. The results in this paper solve an open problem posed by Eskenazis and Nestoridi in [8].

Keywords: Markov chain; mixing times; cutoff; Bernoulli-Laplace model.
MSC2020 subject classifications: 60J10; 37A25.
Submitted to ECP on August 15, 2022, final version accepted on December 15, 2023.

1 Introduction

In this paper, we study the (n,k)-*Bernoulli-Laplace model*. In the model, there are two urns, a *left* urn and a *right* urn, each of which contains exactly n balls. Of the total $2n$ balls contained in both urns, n are colored red and n are colored white. Starting from a coloration of n balls in each urn, at each step k balls are selected uniformly at random without replacement from each urn, with the selections from each urn independent from one another. The selected balls are then swapped and placed in the opposite urn. The process then repeats itself. Letting X_t^x denote the number of red balls in the left urn after t swaps with $X_0^x = x$ red balls initially in the urn, the process X^x is Markov. Our main goal is to understand how long it takes for the chain to be within $\epsilon > 0$ of its stationary distribution π in total variation. Our focus in this paper is on the case when k satisfies $\alpha \leq k/n \leq \beta$, where the constants α, β are such that $0 < \alpha < \beta < \frac{1}{2}$. The main result of this paper resolves an open question posed by Eskenazis and Nestoridi in [8].

Our interest in the (n,k)-Bernoulli-Laplace model comes from shuffling large decks of cards. Mapping the above model to this setting, the deck of cards has size $2n \gg 1$

*United States Naval Academy, United States of America. E-mail: alameda@usna.edu
†Iowa State University, United States of America. E-mail: cbang@iastate.edu
‡Iowa State University, United States of America. E-mail: brennanz@iastate.edu
§Iowa State University, United States of America. E-mail: dherzog@iastate.edu
¶Carnegie Mellon University, United States of America. E-mail: jkritsch@andrew.cmu.edu
‖Iowa State University, United States of America. E-mail: sprangel@iastate.edu

and at each step of the shuffle we cut the deck into two equal piles of n cards, shuffle each deck independently and perfectly, reassemble the deck and then move the top k cards to the bottom. This process repeats itself until sufficient mixing is achieved. From this description, it follows that the (n,k)-Bernoulli-Laplace model describes this card shuffling algorithm *without* the separate step of shuffling each of the smaller decks independently and perfectly at each step. See [14] for further details.

1.1 Preliminaries

Before discussing existing results in the literature and the results of this paper, we first fix some notation and terminology.

Throughout, $\mathcal{X} = \{0,1,2,\dots,n\}$ denotes the state space of the Bernoulli–Laplace chain. By comparing the resulting Bernoulli–Laplace chains that swap k balls at each step and $n-k$ balls at each step, we may assume without loss of generality that $k \le n/2$. For $x \in \mathcal{X}$, we use the notation X^x to denote the Bernoulli–Laplace chain with $X_0^x = x$, as described formally above. In order to define the chain rigorously, let $\text{Hyp}(j,\ell,m)$ denote the hypergeometric distribution of m objects selected without replacement from a total of j objects, $\ell \le j$ of which are type 1 and $j-\ell$ are type 2. That is, if $H \sim \text{Hyp}(j,\ell,m)$, then

$$\mathbf{P}(H = z) = \frac{\binom{\ell}{z}\binom{j-\ell}{m-z}}{\binom{j}{m}}, \quad z \in [0 \vee (m-(j-\ell)), \ell \wedge m] \cap \mathbf{N}.$$

Letting $\{H_-^x, H_+^x\}_{x\in\mathcal{X}}$ denote a collection of independent random variables with $H_-^x \sim \text{Hyp}(n,x,k)$ and $H_+^x \sim \text{Hyp}(n,n-x,k)$, the one-step transition of the Bernoulli–Laplace chain is defined by

$$\mathbf{P}(X_1^x = y) = \mathbf{P}(H_+^x - H_-^x = y-x), \quad x,y \in \mathcal{X}. \tag{1.1}$$

An explicit formula for the transition can be readily computed from (1.1) (see, for example, [8]). However, it will not be particularly useful in our analysis.

It is known that the Bernoulli–Laplace chain has a unique stationary distribution π which is $\text{Hyp}(2n,n,n)$ (see [18]). In this paper, we analyze the distance between the Bernoulli–Laplace chain and π in total variation as $n \to \infty$ when $k = k(n)$ satisfies further assumptions. In particular, we note that all of the quantities above depend on the parameter n. Throughout, unless we must emphasize it (as in the paragraph below), this dependence is suppressed.

Let

$$d^{(n)}(t) = \max_{x\in\mathcal{X}} \|\mathbf{P}(X_t^x \in \cdot\,) - \pi(\,\cdot\,)\|_{TV} = \frac{1}{2}\max_{x\in\mathcal{X}} \sum_{y\in\mathcal{X}} |\mathbf{P}(X_t^x = y) - \pi(y)|, \tag{1.2}$$

and define for fixed $\epsilon > 0$ the *mixing time* $t_{\text{mix}}^{(n)}(\epsilon)$ by $t_{\text{mix}}^{(n)}(\epsilon) = \min\{t \in \mathbf{N} : d^{(n)}(t) \le \epsilon\}$. We say that the Bernoulli–Laplace model exhibits *cutoff* if

$$\lim_{n\to\infty} \frac{t_{\text{mix}}^{(n)}(\epsilon)}{t_{\text{mix}}^{(n)}(1-\epsilon)} = 1 \quad \text{for all fixed} \quad \epsilon \in (0,1). \tag{1.3}$$

If the Bernoulli–Laplace model exhibits cutoff and if for every $\epsilon \in (0,1)$ fixed there exists a constant $c(\epsilon)$ and a sequence w_n satisfying

$$w_n = o(t_{\text{mix}}^{(n)}(1/2)) \qquad \text{and} \qquad t_{\text{mix}}^{(n)}(\epsilon) - t_{\text{mix}}^{(n)}(1-\epsilon) \le c(\epsilon) w_n \quad \text{for all} \quad n, \tag{1.4}$$

we say that the Bernoulli–Laplace model has *cutoff window* w_n. For other preliminaries concerning mixing times of Markov chains, see [13].

1.2 Previous results and statement of the main result

Existing results on mixing times for the (n, k)-Bernoulli–Laplace model largely focus on the case where the number of selections k is *much smaller* than the number of balls n in each urn. The earliest works of Diaconis and Shahshahani [4, 5] and Donnely, Floyd and Sudbury [6] treat the case when $k = 1$ and establish cutoff in total variation and in the separation distance, respectively. Diaconis and Shahshahani proved their results by analyzing random walks on Cayley graphs of the symmetric group where edges correspond to transpostions. These results were extended to random walks on distance regular graphs in [1]. See also [16] for the Bernoulli-Laplace model with multiple urns in the case when $k = 1$. We refer to [17] which contains a signed generalization of the model.

The case when $k > 1$ in the two-urn Bernoulli–Laplace model was first studied by Nestoridi and White [14], where a number of estimates were deduced for the mixing time in different regimes of the parameter k as it depends on n. These estimates were made sharp in the case when k satisfies $k = o(n)$ as $n \to \infty$. In particular, when $k = o(n)$ as $n \to \infty$ it was shown in [8] that

$$\frac{n}{4k}\log n - \frac{c(\epsilon)n}{k} \leq t_{\text{mix}}(\epsilon) \leq \frac{n}{4k}\log n + \frac{3n}{k}\log(\log n \vee 2) + O\Big(\frac{n}{\epsilon^4 k}\Big). \tag{1.5}$$

Thus, the Bernoulli–Laplace model with $k = o(n)$ exhibits cutoff at $\frac{n}{4k}\log n$ with cutoff window $\frac{n}{k}\log(\log n \vee 2)$. Estimates deduced in the case when k/n satisfies $k/n \in (0, 1/2)$ and $k/n \to \lambda$ as $n \to \infty$ for some $\lambda \in (0, 1/2)$ were not optimal. That is, in [14], in this case it was shown that

$$\frac{\log n}{2|\log(1 - 2\frac{k}{n})|} - c(\epsilon) \leq t_{\text{mix}}(\epsilon) \leq \frac{\log(n/\epsilon)}{2\lambda(1-\lambda)}. \tag{1.6}$$

It was conjectured in [8, Question 1] that the lower bound in (1.6) is sharp, and it was left there as an open problem to determine the mixing time of the (n, k)-Bernoulli–Laplace model when k is of order n as $n \to \infty$. In this paper, we solve this problem.

In what follows, for $k/n \in (0, 1/2)$ we define a sequence of times t_n by

$$t_n = \frac{\log n}{2|\log(1 - 2\frac{k}{n})|}. \tag{1.7}$$

Remark 1.1. Note that when $k = o(n)$, the bound (1.5) implies that the Bernoulli–Lapace model exhibits cutoff at time t_n with cutoff window $\frac{n}{k}\log(\log n \vee 2) \vee \log n$.

Remark 1.2. When $k/n \in (0, 1/2)$ satisfies $k/n \to \lambda \in (0, 1/2)$ as $n \to \infty$ it is claimed in [8] that

$$t_{\text{mix}}(\epsilon) \geq \frac{\log n}{2|\log(1 - 2\lambda)|} - c(\epsilon).$$

It appears that the argument in [14], which is referenced to in [8], shows the lower bound in (1.6). Note that the lower bound (1.6) implies the claimed bound in [8] in the case when $k/n \to \lambda$ sufficiently fast as $n \to \infty$.

Remark 1.3. The case when $k/n \in (0, 1/2]$ satisfies $k/n \to 1/2$ as $n \to \infty$ was also studied in [14]. If $k = \frac{n}{2} - c$ with $2 \leq c < 1 + \log_6 n$, then it was shown in [14] that

$$\|\mathbf{P}(X_t^0 \in \cdot\,) - \pi(\,\cdot\,)\|_{TV} \leq 12\pi^2 n^2 \Big(\frac{6^{c-1}}{n}\Big)^{2t-2}. \tag{1.8}$$

For example, if $c \geq 2$ is bounded in n, then the righthand side in (1.8) decays to zero as $n \to \infty$ if we set $t = 2 + 1/\log\log n$. However, the above estimate becomes worse as c gets closer to $1 + \log_6 n$.

Throughout this paper, we employ the following assumption on n, k.

Assumption 1.4. *There exist constants α, β with $0 < \alpha < \beta < \frac{1}{2}$ such that $\alpha \le \frac{k}{n} \le \beta$.*

Our main result is the following:

Theorem 1.5. *Suppose that in the (n,k)-Bernoulli-Laplace chain, Assumption 1.4 is satisfied with parameters α, β. For any $\epsilon \in (0,1)$, there exist positive constants $c(\epsilon, \alpha, \beta)$ and $C(\epsilon, \alpha, \beta)$ depending only on ϵ, α, β such that*

$$t_n - c(\epsilon, \alpha, \beta) \le t_{mix}(\epsilon) \le t_n + C(\epsilon, \alpha, \beta). \tag{1.9}$$

In particular, under Assumption 1.4, the (n,k)-Bernoulli-Laplace model exhibits cutoff at time t_n with cutoff window 1.

The proof of the lower bound in (1.9) can be established under Assumption 1.4 following the calculations in [14]. We provide the details for completeness in Section 2. These calculations use standard methods developed in [19, 20] by Wilson. Our main contribution in this work is the proof of the upper bound in (1.9).

The proof of the upper bound (1.9) is split into two parts. In the first part, we show that two *path coupled* copies of the chain, defined below in Section 2, are *sufficiently close* with high probability after $t_n + \kappa$ steps where κ is a large parameter to be determined later. The *high probability* here depends on both n and κ being large. The second part of the proof shows that the total variation distance of two copies of the chain started at these sufficiently close initial conditions is order $1/\kappa^2$ after one additional step. Picking a large $\kappa = \kappa(\epsilon, \alpha, \beta) \ge 1$ accordingly, we will then arrive at the upper bound in (1.9) using the strong Markov property.

Aside from the sometimes tedious differences in asymptotics between the cases when $k = o(n)$ and when (n,k) satisfies Assumption 1.4, both cases make use of the first and second eigenvalues of the chain [8]. Here, this is done in the first part of the proof. However, the second part of the proof requires more work. In this part, as in [8] we use the representations

$$X_1^x \stackrel{d}{=} x + H_+^x - H_-^x \qquad \text{and} \qquad X_1^y \stackrel{d}{=} y + H_+^y - H_-^y. \tag{1.10}$$

Critical to the proof of the main results in the case when $k = o(n)$ in [8] is the total variation approximation of each of the hypergeometrics above using the binomial distribution. This follows the work of Diaconis and Freedman [3]. When Assumption 1.4 is satisfied, this approximation is false [7, Theorem 2]. To get around this issue, we derive a local limit theorem for the hypergeometric distributions (1.10) which is strong enough to ensure that a discrete normal is a good approximation in total variation. We then compare each of the approximating discrete normals to see that they are close in total variation.

This paper is organized as follows: In Section 2, Theorem 1.5 is proven assuming two results, which are established later in Section 3 and Section 4.

Further notation Throughout, we make use of a large parameter $\kappa \ge 1$ (see, for example, (2.3) below) to be determined later. For real-valued sequences (a_n) and (b_n), we use the notation $a_n = O_\kappa(b_n)$ to mean that there exists a constant $C > 0$ possibly depending $\kappa \ge 1$ such that $|a_n| \le C|b_n|$ for all $n \ge n_0$, for some $n_0 \in \mathbf{N}$. We use $a_n \lesssim b_n$ to denote the existence of a constant $C > 0$ independent of $\kappa \ge 1$ such that $a_n \le Cb_n$ for all $n \ge n_0$, for some $n_0 \in \mathbf{N}$. If $H \sim \text{Hyp}(j, \ell, m)$, we assume that H is distributed on $\{0, 1, \dots, m\}$ by extending the probability mass function to be zero outside of its support.

2 Key results and proof of Theorem 1.5

In this section, we state the main results needed to prove Theorem 1.5, namely Proposition 2.2 and Theorem 2.3 below. These results respectively correspond to *part one* and *part two* of the proof of the upper bound in (1.9), as discussed in the introduction. We then conclude this section by proving Theorem 1.5 assuming Proposition 2.2 and Theorem 2.3 hold. Later, we establish Theorem 2.3 and Proposition 2.2 in, respectively, Section 3 and Section 4.

Before stating these results, we first note some facts about the eigenvalues and eigenfunctions of the Bernoulli–Laplace chain. Although all eigenvalues and eigenfunctions of the chain are known [5, 11, 14], their expressions beyond the first few are complicated. Thus similar to [8], we only employ the first and second eigenvalues and eigenfunctions.

For $x \in \mathcal{X}$, let

$$f_1(x) = 1 - \frac{2x}{n} \quad \text{and} \quad f_2(x) = 1 - \frac{2(2n-1)x}{n^2} + \frac{2(2n-1)x(x-1)}{n^2(n-1)}.$$

The functions f_1 and f_2 are the first and second eigenfunctions of the Bernoulli–Laplace chain, and their respective eigenvalues are $f_1(k)$ and $f_2(k)$ [8]. Note that $|f_1| \le 1$ on $\mathcal{X}$ and, after a short exercise optimizing $f_2(x)$ on $\mathcal{X}$, it follows that

$$-\frac{1}{2n-2} \le f_2(x) \le 1 \quad \text{for all} \quad x \in \mathcal{X},\, n \ge 2. \tag{2.1}$$

In addition to these facts, we also need the next lemma, which follows after a short exercise using the definition of eigenvalue and eigenfunction.

Lemma 2.1. *For all $t \ge 0$ and all $x \in \mathcal{X}$, we have the identities*

$$\mathbf{E} f_1(X_t^x) = f_1(k)^t f_1(x) \quad \textit{and} \quad \mathbf{E} f_1(X_t^x)^2 = \frac{1}{2n-1} + \frac{2n-2}{2n-1} f_2(k)^t f_2(x). \tag{2.2}$$

In order to state Proposition 2.2, we make use of the following grand coupling $Y^0, Y^1, \ldots, Y^n$ of the Markov chains $X^0, X^1, \ldots, X^n$.

Grand (path) coupling *First, label the balls in all left urns at time t separately from 1 to n so that each red ball has a smaller label than each white ball. Next, label all balls in all right urns from $n+1$ to $2n$ so that each red ball has a smaller label than each white ball. Uniformly and independently select subsets $A_{\mathit{left}} \subseteq \{1, \ldots, n\}$ and $A_{\mathit{right}} \subseteq \{n+1, \ldots, 2n\}$ with $|A_{\mathit{left}}| = |A_{\mathit{right}}| = k$. To obtain the state of each Markov chain at time $t+1$, swap the balls indexed by elements of A_{left} in each left urn with the balls in the corresponding right urn with index belonging to A_{right}.*

Let $\kappa \ge 1$ be large enough so that

$$\kappa^6 \big(1 - 2\alpha + 2\alpha^2\big)^\kappa \le 1. \tag{2.3}$$

Note that such a choice is possible since $\alpha \in (0, 1/2)$ by Assumption 1.4. The precise choice of κ will be made later in this section as it depends also on $\epsilon > 0$. Define sets

$$I(\kappa) = \Big\{x \in \mathcal{X} \,:\, |x - \frac{n}{2}| \le \kappa\sqrt{n}\Big\} \text{ and } F(\kappa) = \Big\{(x,y) \in I(\kappa)^2 \,:\, |x - y| \le \frac{\sqrt{n}}{\kappa^3}\Big\} \tag{2.4}$$

and for $x, y \in \mathcal{X}$, let

$$\tau_{x,y}(\kappa) = \min\big\{t \,:\, (Y_t^x, Y_t^y) \in F(\kappa)\big\}. \tag{2.5}$$

Proposition 2.2. *Suppose that Assumption 1.4 is satisfied. Then*

$$\max_{x,y \in \mathcal{X}} \mathbf{P}\big\{\tau_{x,y}(\kappa) > t_n + \kappa\big\} \lesssim \frac{1}{\kappa^2}. \tag{2.6}$$

Proposition 2.2 will be proved later in Section 4 using the first and second eigenvalues of the chain as well as the structure of the grand path coupling. This result, in particular, corresponds to *part one* in the proof of the upper bound in (1.9), as discussed in the introduction. The other main result, which corresponds to *part two* of the proof of the upper bound in (1.9), is as follows:

Theorem 2.3. *Suppose that Assumption 1.4 is satisfied. Then*

$$\max_{(x,y)\in F(\kappa)} \|\mathbf{P}(X_1^x \in \,\cdot\,) - \mathbf{P}(X_1^y \in \,\cdot\,)\|_{\mathrm{TV}} \lesssim \frac{1}{\kappa^2}. \tag{2.7}$$

The proof of Theorem 2.3 will be given in Section 3, and it relies heavily on a total variation approximation of the hypergeometric random variables in (1.10) when $x, y \in F(\kappa)$.

Assuming Proposition 2.2 and Theorem 2.3, we now conclude Theorem 1.5. This argument also uses the facts concerning the first and second eigenvalues/eigenfunctions of the chain, as detailed above in this section. The proof below uses well-established methods and is very similar to the combination of [8, Proof of Theorem 1] and [14, Section 4.3]. We provide the details for completeness.

Proof of Theorem 1.5. Let $\epsilon \in (0,1)$. We first prove the upper bound in (1.9). Let $x, y \in \mathcal{X}$ be arbitrary, $A \subset \mathcal{X}$, and $t = t_n + \kappa + 1$ where $\kappa \geq 1$ satisfies (2.3). Using the grand coupling introduced below Lemma 2.1, we see that Proposition 2.2 and the strong Markov property at time $\tau_{x,y}(\kappa)$ imply

$$\begin{aligned}
&|\mathbf{P}(X_t^x \in A) - \mathbf{P}(X_t^y \in A)| \\
&= |\mathbf{P}(Y_t^x \in A) - \mathbf{P}(Y_t^y \in A)| \\
&\leq 2\mathbf{P}(\tau_{x,y}(\kappa) > t-1) + |\mathbf{P}(Y_t^x \in A,\, \tau_{x,y}(\kappa) \leq t-1) - \mathbf{P}(Y_t^y \in A,\, \tau_{x,y}(\kappa) \leq t-1)| \\
&\lesssim \frac{1}{\kappa^2} + \max_{\substack{z,w\in F(\kappa)\\ s=1,2,\ldots,t}} |\mathbf{P}(Y_s^z \in A) - \mathbf{P}(Y_s^w \in A)|.
\end{aligned} \tag{2.8}$$

For any $s \geq 1$ and $z, w \in \mathcal{X}$ we have

$$\begin{aligned}
|\mathbf{P}(Y_s^z \in A) - \mathbf{P}(Y_s^w \in A)| &\leq \|\mathbf{P}(Y_s^z \in \,\cdot\,) - \mathbf{P}(Y_s^w \in \,\cdot\,)\|_{TV} \\
&\leq \|\mathbf{P}(Y_1^z \in \,\cdot\,) - \mathbf{P}(Y_1^w \in \,\cdot\,)\|_{TV} &(2.9)\\
&= \|\mathbf{P}(X_1^z \in \,\cdot\,) - \mathbf{P}(X_1^w \in \,\cdot\,)\|_{TV}. &(2.10)
\end{aligned}$$

Combining (2.8) and (2.10) with Theorem 2.3 and using the fact that $A \subset \mathcal{X}$ was arbitrary we obtain

$$\max_{x\in\mathcal{X}} \|\mathbf{P}(X_t^x \in \,\cdot\,) - \pi(\,\cdot\,)\|_{TV} \leq \max_{x,y\in\mathcal{X}} \|\mathbf{P}(X_t^x \in \,\cdot\,) - \mathbf{P}(X_t^y \in \,\cdot\,)\|_{TV} \lesssim \frac{1}{\kappa^2}.$$

Picking $\kappa = \kappa(\epsilon, \alpha, \beta) > 0$ large enough finishes the proof of the upper bound in (1.9).

Next, for the claimed lower bound in (1.9), let $s_n = t_n - c_0$ for some constant $c_0 = c_0(\epsilon, \alpha, \beta) > 0$ to be determined. Recalling that π has $\mathrm{Hyp}(2n, n, n)$ distribution, let H be any random variable with $H \sim \mathrm{Hyp}(2n, n, n)$. By Lemma 2.1, $\mathbf{E} f_1(H) = 0$ and $\mathbf{E} f_1(H)^2 = 1/(2n-1)$. By Assumption 1.4, let $n_0 = n_0(\alpha) \geq 2$ be large enough so that $f_2(k) > 0$ for $n \geq n_0$. For $n \geq n_0$, let $\Delta_n = \log f_2(k) - 2\log f_1(k)$ and note that Assumption 1.4 implies

$$|f_2(k) - f_1(k)^2| \lesssim n^{-1}, \qquad |\Delta_n| \lesssim n^{-1}, \qquad \text{and} \tag{2.11}$$

$$f_2(k)^{s_n} = f_1(k)^{2s_n} e^{s_n \Delta_n} = \frac{e^{s_n\Delta_n}}{n f_1(k)^{2c_0}}, \quad n \geq n_0. \tag{2.12}$$

Combining (2.11)-(2.12) with Lemma 2.1 and (1.7) we find that for $n \geq n_0$

$$\begin{aligned}\Big|\mathrm{Var}(f_1(X^0_{s_n})) - \frac{1}{2n-1}\Big| &= \Big|\frac{2n-2}{2n-1} f_2(k)^{s_n} - f_1(k)^{2s_n}\Big| \\ &= f_1(k)^{2s_n}\Big|e^{s_n\Delta_n} - 1 + e^{s_n\Delta_n}(\frac{2n-2}{2n-1} - 1)\Big| \leq K\frac{\log n}{n^2}\end{aligned}$$

for some constant $K = K(c_0, \alpha, \beta) > 0$. Hence, for $\delta > 0$, letting

$$B = \big\{x \in \mathcal{X} \,:\, \sqrt{\delta(2n-1)}|f_1(x)| \leq 1\big\},$$

we note that Chebychev's inequality implies

$$\pi(B) = 1 - \pi(B^c) \geq 1 - \mathbf{E} f_1(H)^2 \delta(2n-1) = 1 - \delta.$$

On the other hand, letting

$$\tilde{B} = \{x \in \mathcal{X} \,:\, \sqrt{\delta(2n-1)}|f_1(x) - f_1(k)^{s_n}| \leq 1\},$$

we note by picking $c_0 = c_0(\delta, \alpha) > 0$ large enough, $B \cap \tilde{B} = \emptyset$. Hence by Chebychev's inequality, $n \geq n_0$ implies

$$\mathbf{P}(X^0_{s_n} \in B) = 1 - \mathbf{P}(X^0_{s_n} \in B^c) \leq 1 - \mathbf{P}(X^0_{s_n} \in \tilde{B}) = \mathbf{P}(X^0_{s_n} \in \tilde{B}^c) \leq \delta + 2K\delta\frac{\log n}{n}.$$

This then implies

$$\max_{x \in \mathcal{X}} \|\pi(\,\cdot\,) - \mathbf{P}(X^x_{s_n} \in \,\cdot\,)\|_{TV} \geq |\pi(B) - \mathbf{P}(X^0_{s_n} \in B)| \geq 1 - 2\delta - 2K\delta\frac{\log n}{n}.$$

Pick $\delta = (1 - \frac{\epsilon}{2})/2$ and note that the claimed lower bound in (1.9) now follows. □

3 Proof of Theorem 2.3

Considering (1.10), the proof of Theorem 2.3 relies critically on a local limit theorem for the hypergeometric distribution in the relevant parameter ranges. This is the content of Proposition 3.1 below and one of the main results of this paper. Once we establish Proposition 3.1, we then use it to conclude Theorem 2.3 at the end of the section.

To setup the statement of Proposition 3.1, we let $\mathcal{X}_k = \{0, 1, 2, \ldots, k\}$ and $\phi(x) = \exp(-\frac{x^2}{2})/\sqrt{2\pi}$, $x \in \mathbf{R}$, denote the probability density function of the standard normal distribution on $\mathbf{R}$. We say that a random variable Z has *discrete normal distribution* on $\mathcal{X}_k$ with parameters $\mathfrak{m} \in \mathbf{R}, \mathfrak{s} > 0$, denoted by $Z \sim \mathrm{dN}_k(\mathfrak{m}, \mathfrak{s})$, if

$$\mathbf{P}(Z = j) = \frac{1}{\mathfrak{s}\mathcal{N}_{\mathfrak{m},\mathfrak{s}}}\phi\Big(\frac{j - \mathfrak{m}}{\mathfrak{s}}\Big),\ j \in \mathcal{X}_k, \quad \text{where} \quad \mathcal{N}_{\mathfrak{m},\mathfrak{s}} = \sum_{j \in \mathcal{X}_k} \frac{\phi\big(\frac{j-\mathfrak{m}}{\mathfrak{s}}\big)}{\mathfrak{s}}. \tag{3.1}$$

For any $x \in \mathcal{X}$, introduce parameters

$$p_x = \frac{x}{n},\ \ q_x = 1 - p_x, \quad \sigma_x = 1 \vee \sqrt{kp_xq_x(1 - k/n)}, \quad \text{and} \quad \mathcal{N}_x = \mathcal{N}_{kp_x,\sigma_x}. \tag{3.2}$$

For any random variable W, we let μ_W denote its distribution. We have the following result.

Proposition 3.1 (Local limit theorem)**.** *Suppose that Assumption 1.4 is satisfied and let $x \in \mathcal{X}$ be such that $x = n/2 + O_\kappa(\sqrt{n})$. If $H \sim \mathrm{Hyp}(n, x, k)$ and $Z \sim \mathrm{dN}_k(kp_x, \sigma_x)$, then for any $\zeta \in (0, 1/3)$ we have*

$$\|\mu_H - \mu_Z\|_{\mathrm{TV}} = O_\kappa\big(n^{-\frac{1-3\zeta}{2}}\big). \tag{3.3}$$

Remark 3.2. Any hypergeometric random variable $H \sim \mathrm{Hyp}(n,x,k)$ can be written as $H = \sum_{i=1}^{k} \xi_i$ where the ξ_i are independent, non-identically distributed Bernoulli random variables [7, 10]. Although one can use this representation to deduce asymptotic normality of H using Barry-Esseen type bounds [2, Theorem 1.1 and Theorem 1.2], we found it challenging to obtain as sharp estimates on the "bulk" of the distributions to have (3.3). We refer the reader to [15, Theorem 5] for conditions on the sum $\sum_{i=1}^{k} \xi_i$ to have a local limit theorem in the usual sense. Our proof relies on estimates specific to the hypergeometric distribution deduced in [12].

Remark 3.3. The righthand side of (3.3) can be improved to $O_\kappa(n^{-1/2})$ using the results in [12]. However, to keep the presentation simpler and more self-contained, we deduce the slightly weaker estimate above.

Before proving Proposition 3.1, we first establish some preliminary results. First, we need a technical lemma concerning the asymptotic behavior of the normalization constant $\mathcal{N}_x$ in (3.2) when $x = n/2 + O_\kappa(\sqrt{n})$.

Lemma 3.4. *Suppose Assumption 1.4 is satisfied and that $x \in \mathcal{X}$ satisfies $x = n/2 + O_\kappa(\sqrt{n})$. Then $\mathcal{N}_x = 1 + O_\kappa(n^{-1/2})$.*

Proof. The proof is done using integral comparison, and we only prove the needed lower bound as the upper bound is done similarly. For simplicity in the proof, below we suppress the dependence of the parameters in (3.2) on x. To obtain the lower bound, note that if $\mathcal{Z}$ is a standard normal random variable on $\mathbf{R}$ we have for all n large enough

$$\begin{aligned}\mathcal{N} &\geq \sum_{j=0}^{\lfloor kp \rfloor} \frac{e^{-\frac{(j-kp)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} + \sum_{j=\lfloor kp \rfloor}^{k} \frac{e^{-\frac{(j-kp)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} - \frac{1}{\sqrt{2\pi\sigma^2}} \\ &\geq \int_0^k \frac{e^{-\frac{(u-kp)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\, du - \frac{2}{\sqrt{2\pi\sigma^2}} = 1 - \mathbf{P}(\mathcal{Z} \leq -kp/\sigma) - \mathbf{P}(\mathcal{Z} \geq kq/\sigma) - \frac{2}{\sqrt{2\pi\sigma^2}}\end{aligned}$$

where the $-2/\sqrt{2\pi\sigma^2}$ resulted from shifting Riemann rectangles in the integral comparison and we used the substitution $u' = (u-kp)/\sigma$ to get the last equality. Employing Assumption 1.4 and recalling that $x = \frac{n}{2} + O_\kappa(\sqrt{n})$, we find that the parameters in (3.2) satisfy

$$p = \frac{1}{2} + O_\kappa(n^{-1/2}), \;\; q = \frac{1}{2} + O_\kappa(n^{-1/2}), \;\; \sigma = O_\kappa(\sqrt{n}). \tag{3.4}$$

Applying Chebychev's inequality and using (3.4), we obtain for n large enough

$$\mathbf{P}(\mathcal{Z} \leq -kp/\sigma) \leq \frac{\sigma^2}{k^2p^2} = O_\kappa(n^{-1}).$$

Similarly, $\mathbf{P}(\mathcal{Z} \geq kq/\sigma) = O_\kappa(n^{-1})$ and $\frac{2}{\sqrt{2\pi\sigma^2}} = O_\kappa(n^{-1/2})$. We thus conclude the result. □

In addition to the above result, we will also employ the following lemma deduced in [12, Lemma 1] as well as Hoeffding's inequality [9]. We note that Lemma 3.5 is a reformulation of [12, Lemma 1] suited for our purposes.

To setup the statement of the results, for $\zeta \in (0, 1/3)$ and $x \in \mathcal{X}$ define

$$L_x = \inf\left\{j \in \mathcal{X}_k \,:\, \frac{j-kp_x}{\sigma_x} \geq -\sigma_x^\zeta\right\} \;\text{ and }\; R_x = \sup\left\{j \in \mathcal{X}_k \,:\, \frac{j-kp_x}{\sigma_x} \leq \sigma_x^\zeta\right\}. \tag{3.5}$$

Lemma 3.5. *Consider the parameters introduced in* (3.2) *and* (3.5)*. Suppose that Assumption 1.4 is satisfied and $x \in \mathcal{X}$ satisfies $x = \frac{n}{2} + O_\kappa(\sqrt{n})$. If $H \sim Hyp(n,x,k)$, for* *any $j \in \mathcal{X}_k$ let $r(j,x) \in \mathbf{R}$ be defined by $\mathbf{P}(H=j) = \frac{1}{\sigma_x}\phi((j-kp_x)/\sigma_x)\exp(r(j,x))$ when $\mathbf{P}(H=j) \neq 0$ and $r(j,x)=0$ otherwise. Then*

$$\max_{j\in\{L_x,L_x+1,\ldots,R_x-1,R_x\}} |r(j,x)| = O_\kappa(n^{-\frac{1-3\zeta}{2}}).$$

Theorem 3.6 (Hoeffding's inequality)**.** *Suppose that $\xi_1,\xi_2,\ldots,\xi_j$ are independent Bernoulli random variables and $S_j = \xi_1+\cdots+\xi_j$. Then for any $t \geq 0$ we have*

$$\mathbf{P}\{|S_j - \mathbf{E}S_j| \geq t\} \leq 2e^{-\frac{2t^2}{j}}.$$

We are now prepared to prove Proposition 3.1. For notational simplicity in the following arguments, we will suppress the dependence on x in the parameters in (3.2) and (3.5), and we will use the shorthand notation

$$\varphi(j) = \mathbf{P}[Z=j] = \frac{\phi((j-kp)/\sigma)}{\mathcal{N}\sigma} \quad \text{and} \quad u_j = \frac{j-kp}{\sigma}. \tag{3.6}$$

Proof of Proposition 3.1. Using the definitions of L, R, as well as Assumption 1.4 and $x = \frac{n}{2} + O_\kappa(\sqrt{n})$, applying the triangle inequality and then optimizing the normal density produces

$$\begin{aligned}
2\|\mu_H - \mu_Z\|_{TV} &\leq \sum_{j=R+1}^{k}\varphi(j) + \sum_{j=0}^{L-1}\varphi(j) + \mathbf{P}(H<L) + \mathbf{P}(H>R) + \sum_{j=L}^{R}|\mathbf{P}(H=j)-\varphi(j)| \\
&\leq \frac{(k+1)}{\sqrt{2\pi}\mathcal{N}\sigma}e^{-\frac{\sigma^{2\zeta}}{2}} + \mathbf{P}(H<L) + \mathbf{P}(H>R) + \sum_{j=L}^{R}|\mathbf{P}(H=j)-\varphi(j)| \\
&=: T_1 + T_2 + T_3 + T_4.
\end{aligned}$$

Note first that our assumptions and Lemma 3.4 give $T_1 = O_\kappa(n^{-1/2})$. For T_2+T_3, recall from Remark 3.2 that $H = \sum_{i=1}^k \xi_i$ where the ξ_i are independent Bernoulli random variables and that $\mathbf{E}H = kp$. Using the definitions of L and R, Assumption 1.4 and $x = \frac{n}{2} + O_\kappa(\sqrt{n})$, Theorem 3.6 implies

$$T_2 + T_3 = \mathbf{P}(H-kp < L-kp) + \mathbf{P}(H-kp > R-kp) = O_\kappa\Big(e^{-2\frac{\sigma^{2+2\zeta}}{k}}\Big) = O_\kappa(n^{-1/2}).$$

Finally for T_4, first using the triangle inequality and then applying Lemma 3.4 and Lemma 3.5, we obtain for n large enough

$$T_4 \leq \sum_{j=L}^{R}\frac{\phi(u_j)}{\sigma}|1-\frac{1}{\mathcal{N}}| + \sum_{j=L}^{R}\Big|\mathbf{P}(H=j) - \frac{\phi(u_j)}{\sigma}\Big| \leq |\mathcal{N}-1| + \sum_{j=L}^{R}\frac{\phi(u_j)}{\sigma}|e^{r(j,x)}-1| \tag{3.7}$$

$$\begin{aligned}
&\leq O_\kappa(n^{-1/2}) + \sum_{j=L}^{R}\frac{\phi(u_j)}{\sigma}|r(j,x)|e^{|r(j,x)|} \\
&\leq O_\kappa(n^{-1/2}) + \mathcal{N}O_\kappa\big(n^{-\frac{1-3\zeta}{2}}\big) \\
&= O_\kappa\big(n^{-\frac{1-3\zeta}{2}}\big).
\end{aligned} \tag{3.8}$$

Combining the estimates finishes the proof. □

We now employ Proposition 3.1 in order to see what we have left to estimate to conclude Theorem 2.3. Let $x, y \in \mathcal{X}$ and assume without loss of generality that $x \neq y$. Let $Z_-^x \sim \mathrm{dN}_k(kp_x,\sigma_x)$, $Z_+^x \sim \mathrm{dN}_k(kq_x,\sigma_x)$, $Z_-^y \sim \mathrm{dN}_k(kp_y,\sigma_y)$ and $Z_+^y \sim \mathrm{dN}_k(kq_y,\sigma_y)$

be independent and independent of the hypergeometric random variables in (1.10). Recalling that μ_W denotes the distribution of the random variable W, first observe that

$$\begin{aligned}
&\mu_{H_+^x - H_-^x + x} - \mu_{H_+^y - H_-^y + y} \qquad (3.9)\\
&= (\mu_{H_+^x - H_-^x + x} - \mu_{H_+^x - Z_-^x + x} + \mu_{H_+^x - Z_-^x + x} - \mu_{Z_+^x - Z_-^x + x} + \mu_{Z_+^x - Z_-^x + x} - \mu_{Z_+^x - Z_-^y + x})\\
&\quad - (\mu_{H_+^y - H_-^y + y} - \mu_{H_+^y - Z_-^y + y} + \mu_{H_+^y - Z_-^y + y} - \mu_{Z_+^y - Z_-^y + y}) + (\mu_{Z_+^x - Z_-^y + x} - \mu_{Z_+^y - Z_-^y + y}).
\end{aligned}$$

Also, recall that for any random variables X, Y, W distributed on $\mathbf{Z}$ with W independent of X and Y, we have $\|\mu_{X+W} - \mu_{Y+W}\|_{TV} \le \|\mu_X - \mu_Y\|_{TV}$. This fact, (1.10), independence and the triangle inequality give

$$\begin{aligned}
\|\mathbf{P}(X_1^x \in \cdot\,) - \mathbf{P}(X_1^y \in \cdot\,)\|_{TV} \le &\sum_{\substack{z \in \{x,y\}\\ \mathrm{sg} \in \{+,-\}}} \|\mu_{H_{\mathrm{sg}}^z} - \mu_{Z_{\mathrm{sg}}^z}\|_{TV}\\
&+ \|\mu_{\eta + Z_+^x} - \mu_{Z_+^y}\|_{TV} + \|\mu_{Z_-^x} - \mu_{Z_-^y}\|_{TV}
\end{aligned}$$

where $\eta = x - y$. Under Assumption 1.4 and $x, y \in I(\kappa)$, Proposition 3.1 implies $\sum_{z,\mathrm{sg}} \|\mu_{H_{\mathrm{sg}}^z} - \mu_{Z_{\mathrm{sg}}^z}\|_{TV} = O_\kappa(n^{-1/4})$. Thus, to conclude Theorem 2.3, we are left to show:

Lemma 3.7. *Suppose that Assumption 1.4 is satisfied and let $x, y \in \mathcal{X}$ be such that $(x, y) \in F(\kappa)$ with $x \ne y$. Consider the random variables $Z_-^x \sim d\mathrm{N}_k(kp_x, \sigma_x)$, $Z_+^x \sim d\mathrm{N}_k(kq_x, \sigma_x)$, $Z_-^y \sim d\mathrm{N}_k(kp_y, \sigma_y)$ and $Z_+^y \sim d\mathrm{N}_k(kq_y, \sigma_y)$, and let $\eta = x - y$. Then*

$$\|\mu_{\eta + Z_+^x} - \mu_{Z_+^y}\|_{TV} + \|\mu_{Z_-^x} - \mu_{Z_-^y}\|_{TV} \lesssim \frac{1}{\kappa^2}. \tag{3.10}$$

Proof. For simplicity, let $Z_1 = Z_+^x$ and $Z_2 = Z_+^y$ and note that it suffices to show that $\|\mu_{\eta + Z_1} - \mu_{Z_2}\|_{TV} \lesssim \kappa^{-2}$ assuming $\eta \ge 0$. Let $\mathcal{Y}_\eta = \mathcal{X}_k \cap (\mathcal{X}_k + \eta)$ and $\mathcal{N}_1 = \mathcal{N}_{kq_x, \sigma_x}$ and $\mathcal{N}_2 = \mathcal{N}_{kq_y, \sigma_y}$ denote the normalization constants for Z_1 and Z_2, respectively. Set $J = [\frac{k}{2} - 4\kappa\sqrt{n}, \frac{k}{2} + 4\kappa\sqrt{n}]$. Note that by the triangle inequality

$$\begin{aligned}
2\|\mu_{\eta + Z_1} - \mu_{Z_2}\|_{TV} \le &\sum_{j \in \mathcal{Y}_\eta \cap J} |\mathbf{P}(Z_1 = j - \eta) - \mathbf{P}(Z_2 = j)|\\
&+ \sum_{j \in \mathcal{Y}_\eta \cap J^c} \mathbf{P}(Z_1 = j - \eta) + \sum_{j \in \mathcal{Y}_\eta \cap J^c} \mathbf{P}(Z_2 = j)\\
&+ \sum_{j \in \mathcal{X}_k \setminus (\mathcal{X}_k + \eta)} \mathbf{P}(Z_2 = j) + \sum_{j \in (\mathcal{X}_k + \eta) \setminus \mathcal{X}_k} \mathbf{P}(Z_1 = j - \eta)\\
=: &\,\mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4 + \mathcal{T}_5.
\end{aligned}$$

We next show how to estimate $\mathcal{T}_4$. The term $\mathcal{T}_5$ can be estimated analogously. Observe that since $\eta = x - y \ge 0$ and $j \in \mathcal{X}_k \setminus (\mathcal{X}_k + \eta)$, then $j \le \eta - 1$. Since $(x, y) \in F(\kappa)$, we have $\eta = |x - y| \le \sqrt{n}/\kappa^3$. Applying Assumption 1.4 and using $(x, y) \in F(\kappa)$, we find that $1 \vee (\frac{1}{2} - \kappa n^{-1/2})\sqrt{n\alpha(1-\beta)} \le \sigma_y \le \sqrt{n}$ and $kq_y \ge \alpha(\frac{n}{2} - \kappa\sqrt{n})$. Applying Lemma 3.4, for all n large enough we obtain

$$\mathcal{T}_4 \le \sum_{j=0}^{\eta - 1} \mathbf{P}(Z_2 = j) \le \eta \mathbf{P}(Z_2 = \eta - 1) = \frac{\eta}{\mathcal{N}_2 \sigma_y} \phi\Big(\frac{\eta - 1 - kq_y}{\sigma_y}\Big) \le C_0 e^{-\epsilon_0 n}$$

for some constants $C_0, \epsilon_0 > 0$ independent of n.

We next estimate $\mathcal{T}_2$. Term $\mathcal{T}_3$ can be estimated analogously. Let $\mathcal{Z}$ be a standard normal random variable. Using $\kappa \ge 1$ and the assumptions that $(x, y) \in F(\kappa)$, $0 < k/n < 1/2$, we obtain the bounds

$$\frac{k}{2} + 4\kappa\sqrt{n} \ge \kappa\sqrt{n} + \eta + kq_x + 1 \quad \text{and} \quad \frac{k}{2} - 4\kappa\sqrt{n} \le -\kappa\sqrt{n} + \eta - 1 + kq_x. \tag{3.11}$$

Consequently, using integral comparison (with rectangles of width one), monotonicity of the standard normal density away from the mean, and Lemma 3.4 we have

$$\mathcal{T}_2 \le \frac{2}{\mathcal{N}_1\sigma_x} + \int_{J^c} \frac{\phi\big(\frac{u-\eta-kq_x}{\sigma_x}\big)}{\mathcal{N}_1\sigma_x}\, du \lesssim \frac{2}{\mathcal{N}_1}\mathbf{P}\Big\{\mathcal{Z} \ge \frac{\kappa\sqrt{n}}{\sigma_x}\Big\} \le \frac{2\sigma_x^2}{n\mathcal{N}_1}\frac{\mathbf{E}\mathcal{Z}^2}{\kappa^2} \lesssim \frac{1}{\kappa^2}$$

where we first made the integral substitution $u' = (u-\eta-kq_x)/\sigma_x$ and then used (3.11) to obtain the second asymptotic inequality.

Turning finally to the remaining term $\mathcal{T}_1$, we first do a calculation to show that that $\sigma_x^{-1} - \sigma_y^{-1} = O_\kappa(n^{-1})$ for $(x,y) \in F(\kappa)$. Let $f(z) = (1-z)z$ and note by the Mean Value Theorem, for $z_1, z_2 \in [0,1]$ we have $|f(z_1) - f(z_2)| \le |z_1 - z_2|$. We thus find that there exists a constant $C_1 > 0$ independent of n such that for all n large enough

$$\begin{aligned}\left|\frac{1}{\sigma_x} - \frac{1}{\sigma_y}\right| = \left|\frac{\sigma_y^2 - \sigma_x^2}{\sigma_x\sigma_y(\sigma_x+\sigma_y)}\right| &= \frac{1}{\sqrt{k(1-k/n)}}\left|\frac{p_yq_y - p_xq_x}{\sqrt{p_yq_yp_xq_x}(\sqrt{p_yq_y}+\sqrt{p_xq_x})}\right| \\ &\le C_1\frac{|f(y/n) - f(x/n)|}{\sqrt{n}} \le C_1\frac{|x-y|}{n^{3/2}} = O_\kappa(n^{-1}) \qquad (3.12)\end{aligned}$$

as $(x,y) \in F(\kappa)$.

Next, by first using calculations similar to (3.7) and then appealing to (3.12), we get

$$\begin{aligned}\mathcal{T}_1 &= \sum_{j\in\mathcal{Y}_\eta\cap J}\left|\frac{\phi\big(\frac{j-\eta-kq_x}{\sigma_x}\big)}{\sigma_x} - \frac{\phi\big(\frac{j-kq_y}{\sigma_y}\big)}{\sigma_y}\right| + O_\kappa(n^{-1/2}) \\ &\le \sum_{j\in\mathcal{Y}_\eta\cap J}\left|\frac{\phi\big(\frac{j-\eta-kq_x}{\sigma_x}\big)}{\sigma_x} - \frac{\phi\big(\frac{j-kq_y}{\sigma_y}\big)}{\sigma_x}\right| + \sum_{j\in\mathcal{Y}_\eta\cap J}\phi\Big(\frac{j-kq_y}{\sigma_y}\Big)\left|\frac{1}{\sigma_x} - \frac{1}{\sigma_y}\right| + O_\kappa(n^{-1/2}) \\ &\le \sum_{j\in\mathcal{Y}_\eta\cap J}\left|\frac{\phi\big(\frac{j-\eta-kq_x}{\sigma_x}\big)}{\sigma_x} - \frac{\phi\big(\frac{j-kq_y}{\sigma_y}\big)}{\sigma_x}\right| + 9\kappa\sqrt{n}\left|\frac{1}{\sigma_x} - \frac{1}{\sigma_y}\right| + O_\kappa(n^{-1/2}) =: \mathcal{T}_1' + O_\kappa(n^{-1/2}),\end{aligned}$$

where on the penultimate line above we used $\phi \le 1$ and the number of integers in the interval J is bounded above by $9\kappa\sqrt{n}$. To estimate $\mathcal{T}_1'$, observe that there exists a universal constant $C_2 > 0$ such that $|\phi(x) - \phi(y)| \le C_2|x-y|$ for all $x, y \in \mathbf{R}$. Hence using (3.12) again, the definition of J and $(x,y) \in F(\kappa)$ we obtain

$$\begin{aligned}\mathcal{T}_1' &\le \frac{C_2}{\sigma_x}\sum_{j\in\mathcal{Y}_\eta\cap J}\left|\frac{j-kq_y}{\sigma_y} - \frac{j-\eta-kq_x}{\sigma_x}\right| \\ &= \frac{C_2}{\sigma_x}\sum_{j\in\mathcal{Y}_\eta\cap J}\left|\frac{j-kq_y}{\sigma_x} - \frac{j-\eta-kq_x}{\sigma_x}\right| + \frac{C_2}{\sigma_x}\left|\frac{1}{\sigma_x} - \frac{1}{\sigma_y}\right|\sum_{j\in\mathcal{Y}_\eta\cap J}|j-kq_y| \\ &\le \frac{C_2}{\sigma_x}\frac{18\kappa\sqrt{n}\eta}{\sigma_x} + \frac{C_2}{\sigma_x}\left|\frac{1}{\sigma_x} - \frac{1}{\sigma_y}\right|45\kappa^2 n = \frac{18C_2 n}{\kappa^2\sigma_x^2} + O(n^{-1/2}) \lesssim \frac{1}{\kappa^2}\end{aligned}$$

where we used $|j - kq_y| \le 5\kappa\sqrt{n}$ for $j \in J$ and $y \in I(\kappa)$. Also, on the last line above, we used the fact that $(x,y) \in F(\kappa)$ implies $\eta = |x-y| \le \sqrt{n}/\kappa^3$ and Assumption 1.4 with $x \in I(\kappa)$ implies $\sigma_x \ge 1 \vee (\frac{1}{2} - \kappa n^{-1/2})\sqrt{n\alpha(1-\beta)}$. □

4 Proof of Proposition 2.2

The goal of this section is to prove Proposition 2.2. First, below in Lemma 4.1, we use the grand coupling introduced in Section 2 to estimate $\mathbf{E}|Y_t^x - Y_t^y|$ for $x, y \in \mathcal{X}$. We then use this estimate along with the facts established in Section 2 concerning the first and second eigenvalues/eigenfunctions of the chain to conclude Proposition 2.2.

Lemma 4.1. *Let $x, y \in \mathcal{X}$. Then for all $t \geq 0$*

$$\mathbf{E}|Y_t^x - Y_t^y| \leq \left(1 - \frac{2k(n-k)}{n^2}\right)^t |x - y|.$$

Proof. Let $x, y \in \mathcal{X}$ with $x > y$. By the Markov property, it suffices to show the claimed bound for $t = 1$. Notice that if $x - y = 1$, then by the definition of the coupling, direct calculation gives (see also [14, Equation (15)])

$$\mathbf{E}[|Y_1^x - Y_1^y|] = \left(1 - \frac{2k(n-k)}{n^2}\right).$$

Hence, for general $x > y$ it follows from the triangle inequality that

$$\mathbf{E}[|Y_1^x - Y_1^y|] \leq \sum_{i=0}^{x-y-1} \mathbf{E}[|Y_1^{y+i} - Y_1^{y+i+1}|] = (x-y)\left(1 - \frac{2k(n-k)}{n^2}\right).$$

This finishes the proof. $\square$

Proof of Proposition 2.2. Using a union bound we obtain

$$\begin{aligned}\mathbf{P}\big\{\tau_{x,y}(\kappa) > t_n + \kappa\big\} \leq \mathbf{P}\Big\{Y_{t_n}^x \notin I(\kappa)\Big\} &+ \mathbf{P}\Big\{Y_{t_n}^y \notin I(\kappa)\Big\} \\ &+ \mathbf{P}\Big\{Y_{t_n+\kappa}^x \notin I(\kappa)\Big\} + \mathbf{P}\Big\{Y_{t_n+\kappa}^y \notin I(\kappa)\Big\} \\ &+ \mathbf{P}\Big\{Y_{t_n}^x, Y_{t_n}^y \in I(\kappa), |Y_{t_n+\kappa}^x - Y_{t_n+\kappa}^y| > \frac{\sqrt{n}}{\kappa^3}\Big\}.\end{aligned} \tag{4.1}$$

We first work to control the first four terms on the righthand side of (4.1). Observe that Lemma 2.1 implies that for any $z \in \mathcal{X}$ and $t \geq 0$

$$\begin{aligned}\mathbf{P}\{Y_t^z \notin I(\kappa)\} &= \mathbf{P}\{f_1(Y_t^z)^2 > 4\kappa^2/n\} \\ &\leq \frac{n}{4\kappa^2}\mathbf{E}[f_1(Y_t^z)^2] = \frac{n}{4\kappa^2}\left(\frac{1}{2n-1} + \left(\frac{2n-2}{2n-1}\right) f_2(k)^t f_2(z)\right).\end{aligned} \tag{4.2}$$

Hence using (2.1), (2.11) and the definition of t_n in (1.7) we find that for all n large enough (so that $f_2(k) > 0$)

$$f_2(k)^{t_n+\kappa} f_2(z) \leq f_2(k)^{t_n} = e^{t_n \log f_2(k)} = e^{2t_n \log f_1(k)} e^{t_n \Delta_n} \lesssim n^{-1}. \tag{4.3}$$

From this and (4.2) it follows that

$$\mathbf{P}\Big\{Y_{t_n}^x \notin I(\kappa)\Big\} + \mathbf{P}\Big\{Y_{t_n}^y \notin I(\kappa)\Big\} + \mathbf{P}\Big\{Y_{t_n+\kappa}^x \notin I(\kappa)\Big\} + \mathbf{P}\Big\{Y_{t_n+\kappa}^y \notin I(\kappa)\Big\} \lesssim \frac{1}{\kappa^2}.$$

Lastly, we turn our attention to bounding the remaining term on the righthand side of (4.1). Using the Markov property, Lemma 4.1 and the choice of κ in (2.3) we have

$$\begin{aligned}\mathbf{P}\Big\{Y_{t_n}^x, Y_{t_n}^y \in I(\kappa), |Y_{t_n+\kappa}^x - Y_{t_n+\kappa}^y| > \frac{\sqrt{n}}{\kappa^3}\Big\} &\leq \max_{x,y \in I(\kappa)} \mathbf{P}\{|Y_\kappa^x - Y_\kappa^y| \geq \frac{\sqrt{n}}{\kappa^3}\} \\ &\leq \max_{x,y \in I(\kappa)} \frac{\kappa^3}{\sqrt{n}} \mathbf{E}|Y_\kappa^x - Y_\kappa^y| \\ &\leq 2\kappa^4\left(1 - \frac{2k(n-k)}{n^2}\right)^\kappa \lesssim \frac{1}{\kappa^2}.\end{aligned}$$

Combining these estimates completes the proof. $\square$

References

[1] E. D. Belsley. Rates of convergence of random walk on distance regular graphs. *Probab. Theory Related Fields*, 112(4):493–533, 1998. MR1664702

[2] P. Deheuvels, M. L. Puri, and S. S. Ralescu. Asymptotic expansions for sums of nonidentically distributed bernoulli random variables. *Journal of Multivariate Analysis*, 28(2):282–303, 1989. MR0991952

[3] P. Diaconis and D. Freedman. De Finetti's theorem for Markov chains. *Ann. Probab.*, 8(1):115–130, 1980. MR0556418

[4] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(2):159–179, 1981. MR0626813

[5] P. Diaconis and M. Shahshahani. Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.*, 18(1):208–218, 1987. MR0871832

[6] P. Donnelly, P. Lloyd, and A. Sudbury. Approach to stationarity of the Bernoulli-Laplace diffusion model. *Adv. in Appl. Probab.*, 26(3):715–727, 1994. MR1285456

[7] W. Ehm. Binomial approximation to the poisson binomial distribution. *Statistics & Probability Letters*, 11(1):7–16, 1991. MR1093412

[8] A. Eskenazis and E. Nestoridi. Cutoff for the Bernoulli-Laplace urn model with $o(n)$ swaps. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(4):2621–2639, 2020. MR4164850

[9] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963. MR0144363

[10] S. Hui and C.J. Park. The representation of hypergeometric random variables using independent Bernoulli random variables. *Communications in Statistics-Theory and Methods*, 43(19):4103–4108, 2014. MR3264931

[11] K. Khare and H. Zhou. Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. *Ann. Appl. Probab.*, 19(2):737–777, 2009. MR2521887

[12] S. N. Lahiri, A. Chatterjee, and T. Maiti. Normal approximation to the hypergeometric distribution in nonstandard cases and a sub-Gaussian Berry-Esseen theorem. *J. Statist. Plann. Inference*, 137(11):3570–3590, 2007. MR2363278

[13] D. A. Levin and Y. Peres. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, 2017. MR3726904

[14] E. Nestoridi and G. White. Shuffling large decks of cards and the Bernoulli-Laplace urn model. *J. Theoret. Probab.*, 32(1):417–446, 2019. MR3908920

[15] V. V. Petrov. Sums of independent random variables. In *Sums of Independent Random Variables*. De Gruyter, 2022.

[16] F. Scarabotti. Time to reach stationarity in the Bernoulli-Laplace diffusion model with many urns. *Adv. in Appl. Math.*, 18(3):351–371, 1997. MR1436486

[17] C. H. Schoolfield, Jr. A signed generalization of the Bernoulli-Laplace diffusion model. *J. Theoret. Probab.*, 15(1):97–127, 2002. MR1883200

[18] D. Taïbi. Une généralisation du modèle de diffusion de Bernoulli-Laplace. *J. Appl. Probab.*, 33(3):688–697, 1996. MR1401466

[19] D. B. Wilson. Mixing time of the rudvalis shuffle. *Electronic Communications in Probability*, 8:77–85, 2003. MR1987096

[20] D. B. Wilson. Mixing times of lozenge tiling and card shuffling markov chains. *The Annals of Applied Probability*, 14(1):274–325, 2004. MR2023023

Acknowledgments. We graciously acknowledge support from the National Science Foundation through grants DMS-1855504 (D.P.H.) and DMS-1839918 (C.B., Z.B., J.K., E.S.). We also acknowledge fruitful conversations with Evita Nestoridi and Julia Komjathy about the topics of this paper. We thank the anonymous referees for helpful comments and suggestions, and for helping improve the strength of our main result from an earlier version of the paper.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4], ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe `http://www.lockss.org/`
[2]EJMS: Electronic Journal Management System: `https://vtex.lt/services/ejms-peer-review/`
[3]IMS: Institute of Mathematical Statistics `http://www.imstat.org/`
[4]BS: Bernoulli Society `http://www.bernoulli-society.org/`
[5]Project Euclid: `https://projecteuclid.org/`
[6]IMS Open Access Fund: `https://imstat.org/shop/donation/`