

# A martingale minimax exponential inequality for Markov chains

Raphaël Cerf<sup>\*†</sup>

October 28, 2022

## Abstract

We prove a new inequality controlling the large deviations of the empirical measure of a Markov chain. This inequality is based on the martingale used by Donsker and Varadhan and the minimax theorem. It holds for convex sets and it requires to take an infimum over the starting point. In the case of a compact space, this inequality is a partial improvement of the large deviations estimates of Donsker and Varadhan. In the case of a non compact space, we condition on the event that the process visits  $n$  times a compact subset of the space and we still obtain a control on the exponential scale.

There is an abundant literature devoted to obtaining exponential inequalities for Markov chains. A typical goal is to generalize to the Markovian setting the nice inequalities existing for i.i.d. random variables, like the Bernstein, Hoeffding or Chernoff inequalities. Of course, some hypothesis are required on the Markov chains in order to derive an exponential inequality. Without pretending to make an extensive review, let us mention some important and interesting works on this subject. León and Perron [15] have derived Hoeffding bounds for discrete reversible Markov chains which have an optimal exponential rate. Cattiaux and Guillin [5] have obtained deviations bounds with the help of general functional inequalities. Kargin [14] derives a Bernstein–Hoeffding inequality from a spectral gap condition. His result has been recently improved by Naor, Rao and Regev [17]. Chazottes and Redig [7] have introduced some assumptions on the moments of the coupling time. Gao, Guillin and Wu [13] use a method of transportation–information inequality to establish a type of Bernstein inequality. Dedecker and Gouëzel [8] prove an Hoeffding inequality for geometrically ergodic Markov chains. Bertail and Cléménçon [3], Adamczak and Bednorz [1] rely on regeneration techniques. Moulos and Anantharam [16] have obtained optimal Chernoff and Hoeffding bounds for finite state Markov chains, with a constant prefactor.

We shall follow here a different strategy, based on a refinement of the martingale technique employed by Donsker and Varadhan [11]. This note is essentially a little complement to the classical paper [11]. We prove a new exponential inequality controlling the large deviations of the empirical measure of a Markov chain. This inequality is based on two ingredients: the martingale used by Donsker and Varadhan and the minimax theorem (notice that the minimax theorem has also been used in the derivation of Chernoff’s type inequalities for independent random variables, see [6, 9]). This inequality holds for convex sets and it requires to take an infimum over the starting point. In the case of a compact state space, this inequality is a partial improvement of the large deviations estimates of Donsker and Varadhan. In the case of a non compact space, we condition on the event

---

<sup>\*</sup>DMA, Ecole Normale Supérieure, CNRS, PSL Research University, 75005 Paris.

<sup>†</sup>Laboratoire de Mathématiques d’Orsay, Université Paris-Sud, CNRS, Université Paris–Saclay, 91405 Orsay.

that the process visits  $n$  times a compact subset of the space and we still obtain a control on the exponential scale.

**The model.** Let  $X$  be a separable metric space. Let  $(X_n)_{n \in \mathbb{N}}$  be a stationary Markov chain whose state space is  $X$ . We denote by  $P_x$  and  $E_x$  the probability and expectation for the Markov chain  $(X_n)_{n \in \mathbb{N}}$  starting from  $X_0 = x$ . For  $u$  a continuous bounded function from  $X$  to  $\mathbb{R}$ , we set

$$\pi u(x) = E_x(u(X_1)).$$

This notation is imported from [11], where  $\pi$  stands for the transition kernel of the Markov chain. We suppose that for any real-valued function  $u$  which is bounded and continuous on  $X$ , the map  $x \mapsto \pi u(x)$  is continuous on  $X$ .

We define the local time  $L_n$  at time  $n \geq 1$  by setting, for  $A$  a Borel subset of  $X$ ,

$$L_n(A) = \sum_{k=0}^{n-1} 1_{\{X_k \in A\}}. \quad (1)$$

The local time  $L_n$  is a random measure on  $X$ . Let  $Y$  be a compact subset of  $X$ . For  $n \geq 1$ , we define

$$\tau(Y, n) = \inf \{ k \geq 1 : L_k(Y) = n \}.$$

In case the Markov chain visits  $A$  strictly less than  $n$  times, the random time  $\tau(Y, n)$  is infinite. We set, for  $A$  a Borel subset of  $Y$ ,

$$L_n^Y(A) = \frac{1}{n} L_{\tau(Y, n)}(A) = \frac{1}{n} \sum_{k=0}^{\tau(Y, n)-1} 1_{\{X_k \in A\}},$$

where the sum runs over all the non-negative integers in the case that  $\tau(Y, n) = \infty$ . We denote by  $\mathcal{M}(Y)$  the set of the Borel probability measures on  $Y$ , endowed with the topology of weak convergence. The space  $Y$  being compact, the space  $\mathcal{M}(Y)$  is also compact (see chapter 1, section 5 in [4]). Whenever  $\tau(Y, n)$  is finite, the random measure  $L_n^Y$  is a random element of  $\mathcal{M}(Y)$ . We denote by  $\mathcal{S}^*(Y)$  the collection of the real-valued continuous functions  $u$  defined on  $X$  which satisfy

$$\exists c_1, c_2 \in \mathbb{R} \quad \forall y \in X \quad 0 < c_1 \leq u(y) \leq c_2, \quad (2)$$

$$\forall y \in X \setminus Y \quad u(y) \geq \pi u(y). \quad (3)$$

The condition (3) means that  $u$  is  $\pi$ -superharmonic outside of  $Y$ . We define finally

$$\forall \mu \in \mathcal{M}(Y) \quad I(\mu) = - \inf_{u \in \mathcal{S}^*(Y)} \int_Y \ln \left( \frac{\pi u}{u} \right)(x) \mu(dx).$$

Let  $\tilde{Y}$  be a Borel set containing  $Y$  and all the possible exit points from  $Y$ , i.e., such that  $Y \subset \tilde{Y}$  and

$$\forall y \in Y \quad P_y(X_1 \in \tilde{Y}) = 1.$$

We state next our general exponential inequality.

**Theorem.** *For any closed convex subset  $C$  of  $\mathcal{M}(Y)$ , we have*

$$\forall n \geq 1 \quad \inf_{x \in \tilde{Y}} P_x(L_n^Y \in C) \leq \exp \left( -n \inf_{\mu \in C} I(\mu) \right).$$

An interesting feature of this theorem is that it does not require that the space  $X$  itself is compact. Without further assumptions, it might very well happen that the process starting from  $x \in Y$  does not return  $n$  times to  $Y$ . Yet on the event that  $\{L_n^Y \in C\}$ , we have necessarily that  $\tau(Y, n) < \infty$  (because  $C \subset \mathcal{M}(Y)$ ). An unpleasant feature of the inequality is the infimum over the starting point. It could well be that, for some set  $Y$ , we have to choose  $\tilde{Y} = X$ , and that for some Markov chains, the infimum over  $X$  is 0. However there are certainly many cases where the inequality provides an interesting control. For instance, if the diameter of the jumps of the Markov chain are bounded, then the set  $\tilde{Y}$  can be taken to be an enlargement of  $Y$ , and if in addition the space is finite-dimensional, then  $\tilde{Y}$  would also remain compact. A variant of the inequality was obtained for the symmetric random walk on the lattice  $\mathbb{Z}^d$  in [2] and it was a crucial tool to study the random walk penalized by its range. In this situation, we can take for  $Y$  any finite subset of  $\mathbb{Z}^d$  and for  $\tilde{Y}$  the points of  $\mathbb{Z}^d$  at distance at most 1 from  $Y$ . In the case that the space  $X$  is already compact, we can apply the theorem with the choice  $Y = X$ . In this case,  $\tau(X, n) = n$ ,  $nL_n^Y = L_n$  and we obtain the following corollary.

**Corollary.** *Suppose that the space  $X$  is compact. Let  $\mathcal{U}_1$  be the collection of the continuous positive functions  $u$  defined on  $X$ . We define*

$$\forall \mu \in \mathcal{M}(X) \quad I(\mu) = - \inf_{u \in \mathcal{U}_1} \int_X \ln \left( \frac{\pi u}{u} \right) (x) \mu(dx). \quad (4)$$

For any closed convex subset  $C$  of  $\mathcal{M}(X)$ , we have

$$\forall n \geq 1 \quad \inf_{x \in X} P_x \left( \frac{1}{n} L_n \in C \right) \leq \exp \left( - n \inf_{\mu \in C} I(\mu) \right). \quad (5)$$

This corollary is a partial strengthening of the large deviations upper bound of Donsker and Varadhan (see [11], theorem 1), which we restate next: for any closed subset  $C$  of  $\mathcal{M}(X)$ , we have

$$\forall x \in X \quad \limsup_{n \rightarrow \infty} \frac{1}{n} \ln P_x \left( \frac{1}{n} L_n \in C \right) \leq - \inf_{\mu \in C} I(\mu), \quad (6)$$

where the function  $I$  is given by formula (4). This is the action functional of the large deviations principle proved by Donsker and Varadhan under additional hypothesis [11]. Their work contains also several results on  $I$ , as well as alternative formulas for  $I$ . In their work [11], Donsker and Varadhan make several hypothesis on the transition kernel  $\pi$  of the Markov chain, however these hypothesis are important for the proof of the large deviations lower bound, they are not needed for the proof of the large deviations upper bound. The upper bound (6) of Donsker and Varadhan is better than inequality (5) in the sense that it holds for any starting point  $x$  and any closed subset  $C$  of  $\mathcal{M}(X)$ , not necessarily convex. The inequality (5) is better in the sense that it is a genuine exponential inequality, which holds for any value  $n \geq 1$ . Its weak side is that we have to take the infimum over all starting points and that it holds only for convex sets.

Let us explain how we could derive the corollary starting from the results of Donsker and Varadhan [11]. It is a classical fact that the quantity of the lefthand side of (5) is supermultiplicative, i.e., that

$$\forall m, n \geq 1 \quad \inf_{x \in X} P_x \left( \frac{1}{m+n} L_{m+n} \in C \right) \geq \inf_{x \in X} P_x \left( \frac{1}{n} L_n \in C \right) \times \inf_{x \in X} P_x \left( \frac{1}{m} L_m \in C \right),$$

see [10], chapter IV on uniform large deviations. Equivalently, the sequence

$$-\frac{1}{n} \ln \left( \inf_{x \in X} P_x \left( \frac{1}{n} L_n \in C \right) \right)$$

is subadditive. Unfortunately, we cannot apply directly the classical subadditive lemma, because the sequence might become infinite. So we rely instead of the following variant.

**Lemma.** *Let  $f : \mathbb{N} \setminus \{0\} \rightarrow [0, \infty]$  be a subadditive map, i.e.,*

$$\forall m, n \in \mathbb{N} \setminus \{0\} \quad f(m+n) \leq f(m) + f(n).$$

*We have*

$$\liminf_{n \rightarrow \infty} \frac{f(n)}{n} = \inf_{n \geq 1} \frac{f(n)}{n}.$$

The proof is very simple. By the subadditive property,

$$\forall d, n \in \mathbb{N} \setminus \{0\} \quad \frac{f(dn)}{dn} \leq \frac{f(n)}{n}$$

and we conclude by sending first  $d$  to  $\infty$  and then taking the infimum over  $n$  in  $\mathbb{N} \setminus \{0\}$ . That's it! Applying the lemma, we conclude that

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \ln \left( \inf_{x \in X} P_x \left( \frac{1}{n} L_n \in C \right) \right) = \inf_{n \geq 1} -\frac{1}{n} \ln \left( \inf_{x \in X} P_x \left( \frac{1}{n} L_n \in C \right) \right),$$

or equivalently

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \ln \left( \inf_{x \in X} P_x \left( \frac{1}{n} L_n \in C \right) \right) = \sup_{n \geq 1} \frac{1}{n} \ln \left( \inf_{x \in X} P_x \left( \frac{1}{n} L_n \in C \right) \right). \quad (7)$$

Rewriting (7) as an inequality and combining it with the inequality (6), we obtain the inequality (5). It should be noted, though, that the proof of the large deviations upper bound (6) involves a delicate exchange between an infimum and a supremum, that is a sort of minimax theorem (see formula (2.5) and thereafter in [11]). So the previous argument looks more tortuous than the proof of the theorem, that we start next.

The crucial ingredient to derive the relevant probabilistic estimates on the Markov chain is the family of martingales used by Donsker and Varadhan, which we define thereafter. Let  $u$  be a function belonging to  $\mathcal{S}^*(Y)$ . For  $n \geq 0$ , we set

$$M_n = \left( \prod_{k=0}^{n-1} \frac{u(X_k)}{\pi u(X_k)} \right) u(X_n).$$

We claim that, under  $P_x$ , the process  $(M_n)_{n \in \mathbb{N}}$  is a martingale. We denote by  $E_x$  the expectation for the Markov chain  $(X_n)_{n \in \mathbb{N}}$  starting from  $X_0 = x$ . Let us fix  $n \geq 1$ . We have

$$E_x(M_n | X_0, \dots, X_{n-1}) = \left( \prod_{k=0}^{n-1} \frac{u(X_k)}{\pi u(X_k)} \right) E_x(u(X_n) | X_{n-1})$$

$$= \left( \prod_{k=0}^{n-1} \frac{u(X_k)}{\pi u(X_k)} \right) \pi u(X_{n-1}) = M_{n-1}.$$

The random time  $T = \tau(Y, n) - 1$  is a stopping time for  $(M_n)_{n \in \mathbb{N}}$ . Indeed, for  $t \geq 1$ , the event  $\{T = t\}$  is measurable with respect to  $L_{t+1}$ , hence also to  $X_1, \dots, X_t$ . We apply next the optional stopping theorem:

$$\forall t \geq 0 \quad E_x(M_{t \wedge T}) = E_x(M_0) = u(x). \quad (8)$$

We are interested in controlling  $L_{\tau(Y, n)} = L_{T+1}$ , so we express  $M_n$  with the help of the local time  $L_{n+1}$  as follows:

$$\begin{aligned} M_n &= \exp \left( \sum_{0 \leq k \leq n} \ln \frac{u(X_k)}{\pi u(X_k)} \right) \pi u(X_n) \\ &= \exp \left( \int_{y \in X} \ln \frac{u(y)}{\pi u(y)} dL_{n+1}(y) \right) \pi u(X_n). \end{aligned} \quad (9)$$

We bound from below the lefthand member of (8) and we use the identity (9) to get (recall that  $u$  is non-negative, and so is  $M_n$ )

$$\begin{aligned} E_x(M_{t \wedge T}) &\geq E_x \left( M_{t \wedge T} \mathbf{1}_{\{T < \infty\}} \right) \\ &= E_x \left( \exp \left( \int_{y \in X} \ln \frac{u(y)}{\pi u(y)} dL_{t \wedge T+1}(y) \right) \pi u(X_{t \wedge T}) \mathbf{1}_{\{T < \infty\}} \right). \end{aligned} \quad (10)$$

Recall that the function  $u$  belongs to  $\mathcal{S}^*(Y)$ . The  $\pi$ -superharmonicity of  $u$  on  $X \setminus Y$  implies that

$$\int_{y \in X} \ln \frac{u(y)}{\pi u(y)} dL_{t \wedge T+1}(y) \geq \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} dL_{t \wedge T+1}(y). \quad (11)$$

Reporting inequality (11) in inequalities (10) and (8), we get

$$u(x) \geq E_x \left( \exp \left( \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} dL_{t \wedge T+1}(y) \right) \pi u(X_{t \wedge T}) \mathbf{1}_{\{T < \infty\}} \right). \quad (12)$$

On the event  $\{T < \infty\}$ , we have

$$\begin{aligned} \forall y \in Y \quad \lim_{t \rightarrow \infty} L_{t \wedge T+1}(y) &= L_{T+1}(y) = nL_n^Y(y), \\ \lim_{t \rightarrow \infty} X_{t \wedge T} &= X_T. \end{aligned}$$

Applying Fatou's lemma, we deduce from (12) that

$$u(x) \geq E_x \left( \exp \left( n \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} dL_n^Y(y) \right) \pi u(X_T) \mathbf{1}_{\{T < \infty\}} \right).$$

Yet  $X_T$  belongs to  $Y$ , thus

$$\pi u(X_T) = E_{X_T}(u(X_1)) \geq \inf_{y \in Y} E_y(u(X_1)) \geq \inf_{y \in \tilde{Y}} u(y) \quad (13)$$

and

$$u(x) \geq \left( \inf_{y \in \tilde{Y}} u(y) \right) E_x \left( \exp \left( n \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} dL_n^Y(y) \right) 1_{\{T < \infty\}} \right). \quad (14)$$

Since  $u$  belongs to  $\mathcal{S}^*(Y)$ , then the infimum on the right is positive. Taking now the infimum over  $x \in \tilde{Y}$ , we obtain

$$\inf_{x \in \tilde{Y}} E_x \left( \exp \left( n \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} dL_n^Y(y) \right) 1_{\{T < \infty\}} \right) \leq 1. \quad (15)$$

We proceed by bounding from below the lefthand member of (15) as follows: for any  $x \in \tilde{Y}$ ,

$$\begin{aligned} E_x \left( \exp \left( n \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} dL_n^Y(y) \right) 1_{\{T < \infty\}} \right) \\ \geq E_x \left( \exp \left( n \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} dL_n^Y(y) \right) 1_{\{L_n^Y \in C\}} \right) \\ \geq \exp \left( n \inf_{\mu \in C} \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} d\mu(y) \right) P_x(L_n^Y \in C). \end{aligned} \quad (16)$$

Recall that  $C \subset \mathcal{M}(Y)$ , so that  $L_n^Y \in C$  implies automatically that  $T < \infty$ . Let us define

$$\phi_n(C) = \inf_{x \in \tilde{Y}} P_x(L_n^Y \in C). \quad (17)$$

The previous inequalities (15) and (16) yield that

$$\phi_n(C) \leq \exp \left( -n \inf_{\mu \in C} \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} d\mu(y) \right). \quad (18)$$

This inequality holds for any function  $u$  in  $\mathcal{S}^*(Y)$ . In order to get a functional which is convex, we perform a change of functions and we set  $\phi = \ln u$ . We denote by  $\mathcal{T}^*$  the image of  $\mathcal{S}^*(Y)$  under this change of functions, i.e.,

$$\mathcal{T}^* = \{ \ln u : u \in \mathcal{S}^*(Y) \}.$$

We rewrite inequality (18) as follows: for any  $\phi \in \mathcal{T}^*$ ,

$$\phi_n(C) \leq \exp \left( -n \inf_{\mu \in C} \int_{y \in Y} -\ln(e^{-\phi(y)} \pi e^{\phi(y)}) d\mu(y) \right). \quad (19)$$

We define a map  $\Phi$  on  $\mathcal{T}^* \times \mathcal{M}(Y)$  by

$$\Phi(\phi, \mu) = \int_{y \in Y} -\ln(e^{-\phi(y)} \pi e^{\phi(y)}) d\mu(y).$$

Optimizing the previous inequality (19) over the function  $\phi$  in  $\mathcal{T}^*$ , we get

$$\phi_n(C) \leq \exp \left( -n \sup_{\phi \in \mathcal{T}^*} \inf_{\mu \in C} \Phi(\phi, \mu) \right). \quad (20)$$

The map  $\Phi$  is linear in  $\mu$ . We shall next prove that it is convex in  $\phi$ . In fact, the convexity in  $\phi$  is a consequence of the convexity of the functions  $t \in \mathbb{R} \mapsto \exp(t)$  and  $t \in \mathbb{R}^+ \mapsto -\ln(t)$ . The delicate point is to check that the domain  $\mathcal{T}^*$  of definition of  $\Phi$  is convex. Let  $\phi, \psi$  belong to  $\mathcal{T}^*$  and let  $\alpha, \beta \in ]0, 1[$  be such that  $\alpha + \beta = 1$ . There exist  $u, v \in \mathcal{S}^*(Y)$  such that  $\phi = \ln u$ ,  $\psi = \ln v$ , whence

$$\alpha\phi + \beta\psi = \ln(u^\alpha v^\beta),$$

and we have to check that  $u^\alpha v^\beta$  is in  $\mathcal{S}^*(Y)$ . The continuity and condition (2) are straightforward, the only delicate point is the  $\pi$ -superharmonicity (3). Let  $y$  be a fixed point in  $X \setminus Y$ . We apply the Hölder inequality to the random variables  $u^\alpha(X_1)$ ,  $v^\beta(X_1)$ , with respect to the measure  $P_y$  and with the exponents  $p = 1/\alpha$ ,  $q = 1/\beta$ . We obtain

$$E_y(u^\alpha(X_1)v^\beta(X_1)) \leq E_y(u(X_1))^\alpha E_y(v(X_1))^\beta = (\pi u(y))^\alpha (\pi v(y))^\beta.$$

We finally use the fact that  $u, v$  are  $\pi$ -superharmonic in  $X \setminus Y$  to conclude that

$$\forall y \in X \setminus Y \quad u^\alpha(y)v^\beta(y) \geq \pi u^\alpha v^\beta(y).$$

Moreover the set  $C$  is compact and convex. For any fixed  $\phi \in \mathcal{T}^*$ , the map  $\mu \in \mathcal{M}(Y) \mapsto \Phi(\phi, \mu)$  is linear and continuous for the weak convergence. For any fixed  $\mu \in \mathcal{M}(Y)$ , the map  $\phi \in \mathcal{T}^* \mapsto \Phi(\phi, \mu)$  is convex. We are in position to apply the famous minimax theorem (see [12]):

$$\sup_{\phi \in \mathcal{T}^*} \inf_{\mu \in C} \Phi(\phi, \mu) = \inf_{\mu \in C} \sup_{\phi \in \mathcal{T}^*} \Phi(\phi, \mu). \quad (21)$$

Reporting in (20), we conclude that

$$\phi_n(C) \leq \exp\left(-n \inf_{\mu \in C} I(\mu)\right), \quad (22)$$

where  $I(\mu)$  is defined by

$$I(\mu) = \sup_{\phi \in \mathcal{T}^*} \Phi(\phi, \mu). \quad (23)$$

Coming back to the set of functions  $\mathcal{S}^*(Y)$ , we have

$$I(\mu) = \sup_{u \in \mathcal{S}^*(Y)} \int_{y \in Y} \ln \frac{u(y)}{\pi u(y)} d\mu(y) = - \inf_{u \in \mathcal{S}^*(Y)} \int_{y \in Y} \ln \left(\frac{\pi u}{u}\right)(y) d\mu(y), \quad (24)$$

and so we get the inequality stated in the theorem.

## References

- [1] Radoslaw Adamczak and Witold Bednorz. Exponential concentration inequalities for additive functionals of Markov chains. *ESAIM Probab. Stat.*, 19:440–481, 2015.
- [2] Nathanaël Berestycki and Raphaël Cerf. The random walk penalised by its range in dimensions  $d \geq 3$ . *Ann. H. Lebesgue*, 4:1–79, 2021.
- [3] Patrice Bertail and Gabriela Cioł ek. New Bernstein and Hoeffding type inequalities for regenerative Markov chains. *ALEA Lat. Am. J. Probab. Math. Stat.*, 16(1):259–277, 2019.

- [4] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [5] Patrick Cattiaux and Arnaud Guillin. Deviation bounds for additive functionals of Markov processes. *ESAIM Probab. Stat.*, 12:12–29, 2008.
- [6] Raphaël Cerf. *On Cramér’s theory in infinite dimensions*, volume 23 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2007.
- [7] Jean-René Chazottes and Frank Redig. Concentration inequalities for Markov processes via coupling. *Electron. J. Probab.*, 14:no. 40, 1162–1180, 2009.
- [8] Jérôme Dedecker and Sébastien Gouëzel. Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electron. Commun. Probab.*, 20:no. 64, 12, 2015.
- [9] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*. Jones and Bartlett Publishers, Boston, MA, 1993.
- [10] J.-D. Deuschel and D. W. Stroock. *Large deviations*. Academic Press, 1989.
- [11] M. D. Donsker and S. R. S. Varadhan. Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.*, 28:1–47; *ibid.* 28 (1975), 279–301, 1975.
- [12] K. Fan. Minimax theorems. *Proc. Nat. Acad. Sci.*, 39:42–47, 1953.
- [13] F. Gao, A. Guillin, and L. Wu. Bernstein-type concentration inequalities for symmetric Markov processes. *Theory Probab. Appl.*, 58(3):358–382, 2014.
- [14] Vladislav Kargin. A large deviation inequality for vector functions on finite reversible Markov Chains. *The Annals of Applied Probability*, 17(4):1202 – 1221, 2007.
- [15] Carlos A. León and François Perron. Optimal Hoeffding bounds for discrete reversible Markov chains. *The Annals of Applied Probability*, 14(2):958 – 970, 2004.
- [16] Vrettos Moulos and Venkat Anantharam. Optimal chernoff and hoeffding bounds for finite state markov chains, 2019.
- [17] Assaf Naor, Shravas Rao, and Oded Regev. Concentration of Markov chains with bounded moments. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):2270 – 2280, 2020.