

Chapter 7.3. Selected Topics: Quantitative Limit Theorems

Suppose we observe a simple random walk on \mathbb{Z} , and we keep track of the number of times the walker crosses a (directed) edge (e.g. adding one card to a pile at each crossing). Over time we will observe a random height profile of the pile of cards emerging at large time scales.

For instance, in the symmetric case, [we have seen](#) that for this specific Markov chain, the number of times the edge $(n, n + 1)$ is crossed (during an excursion from 0) is a Galton-Watson process (indexed by n). So, as we follow the chain over a long time, the curve will be distributed as a sum of independent Galton-Watson processes. Moreover, in this example, from the position of the random walk at the final time, we can exactly infer the difference between the number of crossings of the edge $(n, n + 1)$ and the number of crossings of $(n + 1, n)$ (i.e. the difference between the blue bars and orange bars in the video). For instance, if the final position is $X_T = 5$, then the walker has crossed the edges $(n, n + 1)$ and $(n + 1, n)$ with $n < 0$ or $n \geq 5$, the same number of times; while, for $0 \leq n \leq 4$, the edge $(n, n + 1)$ has been crossed once more than $(n + 1, n)$.

A more interesting example is a random walk on \mathbb{Z}_N .

In this case, the difference between the blue bars and the orange bars (referring to the video) will give us the number of times the walker wound around the loop, information that cannot be retrieved solely by knowing the final position.

As we look at Markov chains on large graph embedded in higher dimensional manifolds, things become more interesting.

Preliminaries for Markov Currents

In general, for a Markov chain on a finite or countable state space S , we may want to consider a random observable that counts how many times the chain passes through a given edge.

Notation for probabilities on edges

Let us denote by E the space of edges in the graph induced by the transition probabilities $(p_{x,y})$, that is $E := \{(x, y) : p_{x,y} > 0\}$. As usual, we assume that $p_{x,y}$ is fixed from the start and denote $e \in E$ a generic edge. As for probabilities on the vertices, a probability on the space of edges is identified with a collection $(j_e)_{e \in E}$ with $j_e \geq 0$ and $\sum_e j_e = 1$. The space $\mathcal{P}(E)$ of probability measures on E is [equipped with its own Lévy distance](#); if $|E|$ is finite, this is the usual topology of the simplex $\{j_e \geq 0, \sum_e j_e = 1\}$ regarded as a subset of $\mathbb{R}^{|E|}$, or more generally (if E is countable), the Lévy distance is topologically equivalent (once we enumerate the edges $\{e_0, e_1, \dots\}$) to

$$d(j, j') = \sum_k |j_{e_k} - j'_{e_k}| 2^{-k} \quad (1)$$

That is, a sequence j^n converges to j iff all the values j_e^n converge.

For $\mu \in \mathcal{P}(S)$ and $f: S \rightarrow \mathbb{R}$ we usually denote the integral $\mu(f) = \sum_x \mu_x f(x)$. Similarly for $j \in \mathcal{P}(E)$ and $\xi: E \rightarrow \mathbb{R}$, we denote $j(\xi) := \sum_{e \in E} j_e \xi(e)$.

If j is a probability on edges, then we denote $\mu^j \in \mathcal{P}(S)$ the probability on vertices given by

$$\mu_x^j = \sum_{y: (x,y) \in E} J_{x,y} \quad (2)$$

and p^j the transition probabilities

$$p_{x,y}^j = J_{x,y} / \mu_x^j \quad (3)$$

Currents as probabilities on edges

The example above motivates us to consider the measure on the edges

$$J^T := \frac{1}{T} \sum_{t=0}^{T-1} \delta_{(X_t, X_{t+1})}$$

$$J_{x,y}^T \equiv \frac{|\{t \leq T-1 : X_t = x, X_{t+1} = y\}|}{T}$$

In other words J_e^T counts the number of times the chain passed through the (directed) edge e up to time T , normalized by the total number of jumps (that is, T). J^T is a probability measure on the space of edges E of S . We can regard J^T as a map on the **Skorohod space of paths**

$$J^T: D(S) \rightarrow \mathcal{P}(E)$$

associating to each path a probability on the edges, according to the number of times the path \mathbf{X} crossed each edge up to time T .

Remark. J^T is a richer observable than the occupation measure $\pi^T := \sum_{t=0}^{T-1} \delta_{X_t}$, which encodes the (normalized) number of times the chain visited each vertex. Indeed

$$\mu_x^{J^T} := \sum_y J_{x,y}^T = \pi_x^T \quad (4)$$

so we can recover π^T from J^T . Regarding J^T as a probability on $S \times S$, this tells us that π^T is just its projection on the first component. For similarities with 1-currents on manifolds, sometimes J^T is referred to as a **current**.

Problems and estimates of interest

We already know that π^T converges to the invariant measure m (for irreducible, positive-recurrent chains), as $T \rightarrow \infty$. Each time the chain visits a vertex x , there is a probability $p_{x,y}$ to cross the edge (x, y) ; and as $T \rightarrow \infty$, the (irreducible, positive-recurrent) Markov chain \mathbf{X} will visit x a fraction m_x of the times. Thus we expect that

$$\lim_T J_{x,y}^T = m_x p_{x,y} \quad (5)$$

which is indeed compatible with Equation 4.

In this chapter we want to understand a more quantitative convergence result for J^T , which in turn will imply Equation 5, and via Equation 4 similar bounds for π^T . This is a highly motivated problem, let us consider an example which may give us an idea¹.

¹Fluctuations of this type of mathematical objects (e.g. J^T) have been shown to have a rather deep meaning, providing a universal dynamical notion of the concept of entropy in complex systems.

Example 0.1. In the Transmission Control Protocol for internet communication, a server x will choose the next directly-connected server y to send a packet of data it received, depending on the final destination of the packet (this is inferred via subnet mask, so a packet indeed bounces across many servers, and does not follow an optimal geodesic path). It is a fair assumption to think that the packets perform a Markov chain with some transition probabilities $p_{x,y}$. A pair (x, y) of directly-connected servers will have a maximum connection capacity $C_{x,y}$ defined over a small amount of time, and the transmission will fail if this capacity is exceeded. If we observe the traffic over a period of time, we can estimate $p_{x,y}$ very precisely, and may want to know how to assign the capacities $C_{x,y}$ to minimize the probability of failures. That is, we need to know, for an arbitrary assignment $C_{x,y}$

$$\mathbb{P}(J_{x,y}^T \leq C_{x,y}, \forall x, y \in S) \approx ?$$

Main results: Statements

Definition 0.1. Recall Equation 2, Equation 3 and define the function $I: \mathcal{P}(E) \rightarrow [0, \infty]$

$$I(j) := \begin{cases} \sum_{x,y} \mu_x^j p_{x,y}^j \log(p_{x,y}^j / p_{x,y}) & \text{if } \sum_y J_{x,y} = \sum_y J_{y,x} \\ +\infty & \text{otherwise} \end{cases} \quad (6)$$

Definition 0.2 (Donsker-Varadhan functional). Define the function² $I_{DV}: \mathcal{P}(E) \rightarrow [0, \infty]$

$$I_{DV}(\mu) := \inf\{I(j), j \in \mathcal{P}(E) : \mu^j = \mu\} \quad (7)$$

In Exercise 0.3, Exercise 0.4 and Exercise 0.5 a more explicit expression and some properties of I_{DV} are shown.

Remark. Considering the definition of the relative entropy $H(\cdot|\cdot)$, $I(\cdot)$ can also be written as

$$I(j) := \begin{cases} \sum_x \mu_x^j H(p_{x,\cdot}^j | p_{x,\cdot}) & \text{if } \sum_y J_{x,y} = \sum_y J_{y,x} \\ +\infty & \text{otherwise} \end{cases} \quad (8)$$

Our main result is the following

Theorem 0.1. Consider an irreducible Markov chain on a finite state space. Then for any closed set $\mathcal{C} \subset \mathcal{P}(E)$ it holds

$$\mathbb{P}(J^T \in \mathcal{C}) \leq e^{-T \inf_{j \in \mathcal{C}} I(j) + o(T)} \quad (9)$$

This in turn implies an exponential version of Equation 5 (in probability) and, via Equation 4, a similar result for π^T .

Corollary 0.1. Define the typical current as $\bar{j}_{x,y} := m_x p_{x,y}$. Then, under the same conditions as Theorem 0.1, $I(j) = 0$ iff $j = \bar{j}$. In particular for each $\varepsilon > 0$ there exists a constant $c_\varepsilon > 0$ such that

$$\mathbb{P}\left(\sum_{e \in E} |J_e^T - \bar{j}_e| \geq \varepsilon\right) \leq e^{-T(c_\varepsilon + o(1))}$$

Corollary 0.2. Under the same hypotheses as Theorem 0.1, for any closed set $\mathcal{C} \subset \mathcal{P}(S)$ it holds

$$\mathbb{P}(\pi^T \in \mathcal{C}) \leq e^{-T \inf_{\mu \in \mathcal{C}} I_{DV}(\mu) + o(T)} \quad (10)$$

²This is called the *Donsker-Varadhan functional*, named after a famous series of papers by the NYU mathematicians M.R.Donsker and S.R.S.Varadhan in the 70s. The papers are mostly about continuous-time Markov processes (so can be rather technical), but the discrete-time adaptation is straightforward.

One may prove (beyond the scope of this note) that I is optimal in the following sense: for each measurable $\mathcal{A} \subset \mathcal{P}(E)$

$$e^{-T \inf_{j \in \text{Interior}(\mathcal{A})} I(j) + o(T)} \leq \mathbb{P}(J^T \in \mathcal{A}) \leq e^{-T \inf_{j \in \text{Closure}(\mathcal{A})} I(j) + o(T)}$$

(and a similar statement holds for I_{DV}).

Remark. The condition $\sum_y J_{x,y} = \sum_y J_{y,x}$ implies that the measure μ^j is invariant for the induced transition probabilities p^j . Thus, Theorem 0.1 can be interpreted as follows: *the probability that the observed current J^T fluctuates near a given j decays exponentially as $\exp(-TI(j))$. The rate function $I(j)$ represents the relative entropy of the modified dynamics p^j with respect to the original dynamics p , averaged over the p^j -invariant measure μ^j .*

Similarly, Corollary 0.2 quantifies the rarity of observing an empirical measure π^T close to μ (where μ is not necessarily the invariant measure). The cost is given by $I_{DV}(\mu)$, which quantitatively expresses the “lack of P -invariance” of μ , see e.g. the bound Exercise 0.4.

Main results: Proofs

This section is dedicated to the proof of Theorem 0.1, Corollary 0.1, Corollary 0.2.

Definition 0.3. A function $F: \mathcal{P}(E) \rightarrow [0, \infty]$ is **lower semicontinuous** (or **lsc** for short) if for each $c \geq 0$ the set $\{j \in \mathcal{P}(E) : F(j) \leq c\}$ is closed.

The following lemma states that, if we prove an inequality of the type Equation 9, replacing I by some function I_α depending on a parameter, we can get the same inequality optimizing over α . This is not just a consequence of taking the inf over α on both sides of the inequality, since we have to exchange the \sup_α (the sup appearing because of the minus sign in the exponential), with the \inf_j .

Lemma 0.1. For \mathcal{U} an arbitrary index set, and $\alpha \in \mathcal{U}$, let $I_\alpha: \mathcal{P}(E) \rightarrow [0, \infty]$ be a lsc function. Suppose that, for each $\alpha \in \mathcal{U}$, Equation 9 holds with I replaced by I_α , that is

$$\overline{\lim}_T \frac{1}{T} \log \mathbb{P}(J^T \in \mathcal{C}) \leq - \inf_{j \in \mathcal{C}} I_\alpha(j), \quad \text{for all } \mathcal{C} \text{ closed} \quad (11)$$

Then Equation 9 holds with I replaced by

$$\hat{I}(j) := \sup_{\alpha \in \mathcal{U}} I_\alpha(j)$$

Proof. By general principles of Large Deviations Theory, there exists an optimal lower semicontinuous function \bar{I} such that Equation 11 holds (with \bar{I} in place of I_α) on compact sets. Since we assumed E finite, $\mathcal{P}(E)$ is compact, and thus each closed set is compact. Therefore Equation 11 reads as

$$\bar{I} \geq I_\alpha \quad \text{for all } \alpha \in \mathcal{U}$$

from which $\bar{I} \geq \sup_{\alpha \in \mathcal{U}} I_\alpha$. □

Lemma 0.2. For $\xi \in C(E)$, define $\phi_\xi \in C(S)$ as

$$\phi_\xi(x) := \log \left(\sum_y e^{\xi(x,y)} p_{x,y} \right)$$

Then

$$\mathbb{E} \left[\exp \left(\sum_{t=0}^{T-1} \xi(X_t, X_{t+1}) - \phi_\xi(X_t) \right) \right] = 1 \quad (12)$$

Proof. Define $q_{x,y} := e^{\xi(x,y) - \phi_\xi(x)} p_{x,y}$. Then $q_{x,y} \geq 0$ and $\sum_y q_{x,y} = 1$. Therefore $(q_{x,y})$ are Markov transition probabilities, and we can construct a measure \mathbb{Q} on $D(S)$. We have that

$$\mathbb{Q}(X_0 = x_0, \dots, X_T = x_T) = \exp\left(\sum_{t=0}^{T-1} \xi(x_t, x_{t+1}) - \phi_\xi(x_t)\right) \mathbb{P}(X_0 = x_0, \dots, X_T = x_T)$$

Thus Equation 12 is just the expected value of the constant function $\mathbf{1}$ under \mathbb{Q} . \square

Lemma 0.3. For ξ, ϕ_ξ as in Lemma 0.2, and $\mathcal{A} \subset \mathcal{P}(E)$

$$\mathbb{P}(J^T \in \mathcal{A}) \leq \exp\left(-\inf_{j \in \mathcal{A}} j(\xi) - \mu^j(\phi_\xi)\right)$$

Proof. The exponent in Equation 12 can be written as $J^T(\xi) - \pi^T(\phi_\xi)$. Recalling Equation 4, and simply just multiplying and dividing by the same factor

$$\begin{aligned} \mathbb{P}(J^T \in \mathcal{A}) &= \mathbb{E}\left[e^{-T(J^T(\xi) - \mu^{J^T}(\phi_\xi))} e^{\sum_{t=0}^{T-1} \xi(X_t, X_{t+1}) - \phi_\xi(X_t)} \mathbf{1}_{\mathcal{A}}(J^T)\right] \\ &\leq \sup_{j \in \mathcal{A}} e^{-T(j(\xi) - \mu^j(\phi_\xi))} \mathbb{E}\left[e^{\sum_{t=0}^{T-1} \xi(X_t, X_{t+1}) - \phi_\xi(X_t)} \mathbf{1}_{\mathcal{A}}(J^T)\right] \\ &\leq \exp(-T \inf_{j \in \mathcal{A}} (j(\xi) - \mu^j(\phi_\xi))) \mathbb{E}\left[e^{\sum_{t=0}^{T-1} \xi(X_t, X_{t+1}) - \phi_\xi(X_t)}\right] \end{aligned}$$

which coincides with the statement in view of Lemma 0.2. \square

For a function $f: S \rightarrow \mathbb{R}$, let $df(x, y) := f(y) - f(x)$. Define

$$\mathcal{P}_*(E) := \{j \in \mathcal{P}(E) : j(df) = 0, \text{ for all } f \in C(S)\} = \left\{j \in \mathcal{P}(E) : \sum_y j_{x,y} = \sum_z j_{z,x}, \text{ for all } x \in S\right\}$$

Lemma 0.4. For any compact set $\mathcal{K} \subset \mathcal{P}(E)$ such that $\mathcal{K} \cap \mathcal{P}_*(E) = \emptyset$, it holds

$$\mathbb{P}(J^T \in \mathcal{K}) = 0, \quad \text{for all } T \text{ large enough.}$$

Proof. Given $T \geq 1$, let $y_0 = X_T, \dots, y_N = X_0$ be a path in $D(S)$ connecting X_T to X_0 in a minimal number of steps (thus at most $|E|$ steps). Define the current

$$\tilde{j}^T = \frac{1}{T+N} \sum_{t=0}^{T+N-1} \delta_{(X_t, X_{t+1})}$$

Telescoping the sum

$$\begin{aligned} J^T(df) &= \frac{1}{T} \sum_{t=0}^{T-1} f(X_{t+1}) - f(X_t) = (f(X_T) - f(X_0))/T \\ \tilde{j}^T(df) &= 0 \end{aligned}$$

On the other hand, by Equation 1, $d(J^T, \tilde{j}^T) \leq |E|/T$. In particular J^T is at distance no more than $|E|/T$ from $\mathcal{P}_*(E)$ with probability 1. Since \mathcal{K} is compact and $\mathcal{P}_*(E)$ is closed (indeed compact as we assume S finite), if these two sets do not intersect they are at a strictly positive distance. Therefore $J^T \notin \mathcal{K}$ with probability 1, for $|E|/T$ small enough. \square

Theorem 0.1. Lemma 0.3 can be interpreted as: Equation 9 holds with I replaced by

$$I_\xi(j) := j(\xi) - \mu^j(\phi_\xi)$$

Lemma 0.4 can be interpreted as: Equation 9 holds with I replaced by

$$I_*(j) := \begin{cases} 0 & \text{if } j \in \mathcal{P}_*(E). \\ +\infty & \text{otherwise.} \end{cases}$$

I_ξ is continuous (in particular lsc), I_* is lsc on $\mathcal{P}(E)$. Therefore by Lemma 0.1 the inequality Equation 9 holds with rate

$$\hat{I}(j) = \begin{cases} \sup_\xi I_\xi(j) & \text{if } j \in \mathcal{P}_*(E). \\ +\infty & \text{otherwise.} \end{cases}$$

Now take $\xi(x, y) = \log(j_{x,y}/p_{x,y})$ (the definition of $\xi(x, y)$ if $j_{x,y} = 0$ is irrelevant). To get

$$\begin{aligned} I_\xi(j) &= \sum_{x,y} j_{x,y} \log(j_{x,y}/p_{x,y}) - \sum_x \mu_x^j \log\left(\sum_z j_{x,z}\right) \\ &= \sum_{x,y} \mu_x^j p_{x,y}^j \log(\mu_x^j p_{x,y}^j / p_{x,y}) - \sum_x \mu_x^j \log \mu_x^j \\ &= \sum_{x,y} \mu_x^j p_{x,y}^j \log(p_{x,y}^j / p_{x,y}) = I(j) \end{aligned}$$

Therefore, for $j \in \mathcal{P}_*(E)$, $\hat{I}(j) = \sup_\xi I_\xi(j) \geq I(j)$. Since the inequality holds trivially ($+\infty = +\infty$) outside $\mathcal{P}_*(E)$, we have that Equation 9 holds. \square

Corollary 0.1. Given two probabilities $\mu, \nu \in \mathcal{P}(S)$, the function $\nu \mapsto \sum_y \nu_y \log \nu_y / \mu_y$ is minimized at $\nu = \mu$ (where it vanishes). Therefore, in Equation 6, each term in the sum over x is non-negative.

If $I(j) = 0$, it must hold $p_{x,y}^j = p_{x,y}$ for every y and every x such that $\mu_x^j > 0$ (p^j is not defined if $\mu_x^j = 0$ so we can arbitrarily set it to $p_{x,y}^j = p_{x,y}$ everywhere). But then the condition $j \in \mathcal{P}_*(E)$, implies that μ^j is invariant for $p_{x,y}^j \equiv p_{x,y}$. By irreducibility m is the only invariant probability for $(p_{x,y})$, thus $\mu^j = m$. Thus $j_{x,y} = \mu_x^j p_{x,y}^j = m_x p_{x,y} = \bar{j}_{x,y}$.

Now, the set $\mathcal{K}_\varepsilon := \{j : \sum_e |j_e - \bar{j}_e| \geq \varepsilon\}$ is a compact set at positive distance from \bar{j} . Thus $c_\varepsilon := \inf_{j \in \mathcal{K}_\varepsilon} I(j) > 0$, since I is lsc. \square

Proof. For $\mathcal{C} \subset \mathcal{P}(S)$ closed, define $\mathcal{C}' := \{j : \mu^j \in \mathcal{C}\}$ (so just the preimage of \mathcal{C} w.r.t. the map $j \mapsto \mu^j$). \mathcal{C}' is closed in $\mathcal{P}(E)$, and thus since $\pi^T = \mu^{J^T}$

$$\mathbb{P}(\pi^T \in \mathcal{C}) = \mathbb{P}(J^T \in \mathcal{C}') \leq e^{-T \inf_{j \in \mathcal{C}'} I(j) + o(T)} = e^{-T \inf_{\mu \in \mathcal{C}} I_{\text{DV}}(\mu) + o(T)} \quad (13)$$

\square

Additional Exercises

Exercise 0.1. Find an explicit bound (similar to Equation 9) for h^T the number of windings of a nearest neighbors random walk on \mathbb{Z}_N . *Hint: write the number of windings h^T in terms of J^T .*

Exercise 0.2. Let $F: \mathcal{P}(E) \rightarrow \mathbb{R}$ be continuous (and bounded, which is a consequence of E being finite). Prove that

$$\mathbb{E} \left[e^{TF(J^T)} \right] \leq e^{T \sup_j (F(j) - I(j)) + o(T)}$$

Exercise 0.3. Prove the formula

$$I_{DV}(\mu) = \sup_{u \in C(S), u > 0} \sum_x \mu_x \log \left(\frac{u(x)}{(Pu)(x)} \right) \quad (14)$$

Solution

In principle this can be solved with Lagrange multipliers and extensive but elementary computations. A slightly more abstract application of the technique however allows a quicker approach.

The infimum in the definition Equation 7 may be taken over $j \in \mathcal{P}_*(E)$, since $I(j) = +\infty$ otherwise. Therefore one can write $J_{x,y} = \mu_x q_{x,y}$, where the Markov transition probabilities $Q = (q_{x,y})_{x,y}$ satisfy $\mu Q = \mu$ (from $j \in \mathcal{P}_*(E)$).

We can use an arbitrary function $f \in C(S)$ as *Lagrange multiplier* to impose the constraint $\mu Q = \mu$. Indeed, for $Q = (q_{x,y})_{x,y}$

$$I_{DV}(\mu) = \inf_Q \sup_{f \in C(S)} \sum_{x,y} \mu_x q_{x,y} \log \left(\frac{q_{x,y}}{p_{x,y}} \right) + \mu((I - Q)f)$$

where now the infimum is taken over an *arbitrary* transition probabilities $Q = (q_{x,y})_{x,y}$, since the supremum over f equals $+\infty$ unless $\mu = \mu Q$ is satisfied.

The key point now is the following. The expression in the r.h.s. of which we are taking the $\inf_Q \sup_f$ is

- convex (thus continuous) in Q , for Q in the convex compact set $\sum_y q_{x,y} = 1, q_{x,y} \geq 0$.
- affine (thus continuous) in f , for $f \in C(S)$.

We can therefore use Sion's minmax principle, and exchange the order of the inf and sup to get

$$I_{DV}(\mu) = \sup_{f \in C(S)} \inf_Q \sum_{x,y} \mu_x q_{x,y} \log \left(\frac{q_{x,y}}{p_{x,y}} \right) + \mu((I - Q)f) = \sup_{f \in C(S)} \mu(f) + \inf_Q \sum_{x,y} \mu_x q_{x,y} \log \left(\frac{q_{x,y}}{p_{x,y} e^{f(y)}} \right)$$

Now the infimum over $q_{x,y}$ is easily computed (since we do not need to impose the "difficult" condition $\mu = Q\mu$, just $\sum_y q_{x,y} = 1$), and we get

$$I_{DV}(\mu) = \sup_{f \in C(S)} \sum_x \mu_x f(x) - \mu_x \log \left(\sum_y p_{x,y} e^{f(y)} \right)$$

which gives the wanted expression once we set $u = e^f$.

This proof ultimately suggests that, if we repeat the proof of Theorem 0.1 for π^T instead of J^T , it would have been enough to optimize over $\xi(x, y) = f(y)$ (depending only on y instead of a generic function of both x and y), consistently with the fact that the bound for π^T is a weaker result, and indeed can be derived from the exponential bound for J^T .

Exercise 0.4. Prove the explicit bounds

$$\mathcal{E}(\sqrt{\varrho}, \sqrt{\varrho}) \leq I_{DV}(\mu) \leq \sum_{x,y} \mu_x \mu_y \log \left(\frac{\mu_y}{p_{x,y}} \right) \quad (15)$$

where \mathcal{E} is the [Dirichlet form](#) and $\varrho_x := \mu_x/m_x$ is density of μ w.r.t. the invariant measure m .

Solution

Following the beginning of [Exercise 0.3](#), we can write I_{DV} as an infimum over Markov transition probabilities $Q = (q_{x,y})$ that leave μ invariant: $\mu = \mu Q$. The choice $q_{x,y} = \mu_y$ leaves μ invariant, as easily verified, and yields the stated upper bound for I_{DV} .

As for the lower bound, in [Equation 14](#) take $u = \sqrt{\varrho}$ (the case where ϱ vanishes at some point is easily dealt with, since these points are irrelevant and one can set $u(x) = 1$ if $\varrho(x) = 0$). The expression in [Equation 14](#) for this particular choice of u is not larger than the supremum over u , thus

$$I_{DV}(\mu) \geq \sum_x \mu_x \log \left(\frac{\sqrt{\varrho}(x)}{(P\sqrt{\varrho})(x)} \right) = - \sum_x m_x \log \left(1 - \frac{((I-P)\sqrt{\varrho})(x)}{\sqrt{\varrho}(x)} \right)$$

Apply $-\log(1-z) \geq z$ for each term of the sum to get

$$I_{DV}(\mu) \geq \sum_x m_x \varrho(x) \frac{((I-P)\sqrt{\varrho})(x)}{\sqrt{\varrho}(x)}$$

which simplifies to the wanted inequality.

Exercise 0.5. Prove that $I_{DV}(\mu) = 0$ iff $\mu = m$ is the invariant measure (recall that we assumed the chain to be irreducible). Deduce, as in [Corollary 0.1](#) that for some $c_\varepsilon > 0$

$$\mathbb{P} \left(\sum_{x \in S} |\pi_x^T - m_x| \geq \varepsilon \right) \leq e^{-T(c_\varepsilon + o(1))}$$

where $o(1)$ vanishes as $T \rightarrow \infty$.

Solution

Since $m = \mu^{\bar{j}}$ and $I(\bar{j}) = 0$

$$0 \leq I_{DV}(\mu) \leq I(\bar{j}) = 0$$

Conversely, if $I_{DV}(\mu) = 0$, then $\mathcal{E}(\sqrt{\varrho}, \sqrt{\varrho}) = 0$. Since we assumed the chain to be irreducible, this implies that ϱ is constant, [from point c. of the Proposition on Dirichlet forms](#) with $f = g$. This is equivalent to $\mu = m$ since they are both probabilities.

The exponential bound follows the same argument as [Corollary 0.1](#), since it only uses that $I(\cdot)$ has a unique root.