The Annals of Applied Probability. 12, (2002), 607-636.

BINS AND BALLS: LARGE DEVIATIONS OF THE EMPIRICAL OCCUPANCY PROCESS

STÉPHANE BOUCHERON, FABRICE GAMBOA AND CHRISTIAN LÉONARD

ABSTRACT. In the random allocation model, balls are sequentially inserted at random into n exchangeable bins. The *occupancy score* of a bin denotes the number of balls inserted in this bin. The (random) distribution of occupancy scores defines the object of this paper: the *empirical occupancy measure* which is a probability measure over the integers. This measure-valued random variable packages many useful statistics. This paper characterizes the *Large Deviations* of the flow of empirical occupancy measures when n goes to infinity while the number of inserted balls remains proportional to n. The main result is a Sanov-like theorem for the empirical occupancy measure when the set of probability measures over the integers is endowed with metrics that are slightly stronger than the total variation distance. Thanks to a coupling argument, this result applies to the degree distribution of sparse random graphs.

1. INTRODUCTION

Consider the following classical model in random combinatorics. At each time $k = 1, 2, \dots,$ a ball is thrown into one bin among n. Let $\{1, \dots, n\}$ denote the set of bins. The set $\Omega^n = \{1, \dots, n\}^{\{1,2,\dots\}}$ of all sequences in $\{1, \dots, n\}$ is the natural space for the realizations of this experiment. For any $k \geq 1$, the canonical projection $B^n_k : \omega = (\omega_l)_{l\geq 1} \in \Omega^n \mapsto \omega_k \in \{1, \dots, n\}$ is the random variable: "name of the bin into which the k^{th} ball is thrown".

To make things easier, it is assumed that at time $k = 0$, all the bins are empty. The score of bin α at time $k \geq 0$ is defined by

$$S^n_k(\alpha) = \sum_{l=1}^{k} \mathbb{1}_{\{B^n_l=\alpha\}},$$

with the convention that $S^n_0 = 0$. Let us consider the time-scaling $k = \lfloor nt \rfloor$, $0 \leq t \leq T$ where $\lfloor s \rfloor$ is the integer part of s. We are interested in the time-rescaled evolution of the joint empirical distribution of the scores. This is described by the following empirical occupancy process from $[0, T]$ to the set $\mathcal{P}(\mathbb{N})$ of all probability measures on $\mathbb{N}$:

$$X^n_t = \frac{1}{n}\sum_{\alpha=1}^{n} \delta_{S^n_{\lfloor nt\rfloor}(\alpha)} = \sum_{i\geq 0} X^n_t(i)\delta_i \quad \in \mathcal{P}(\mathbb{N}), 0 \leq t \leq T$$

where δ stands for the Dirac measure and $X^n_t(i)$ is the proportion of bins with score i after $\lfloor nt \rfloor$ ball allocations. It should be clear that $(X^n_\cdot)$ satisfies a law of

1991 *Mathematics Subject Classification.* Primary:60F10 05A17 05C40 , Secondary: 05C80 05A99.

Key words and phrases. Random graphs, Large deviations, Occupancy, Orlicz spaces, Sanov theorem, Poisson approximation.

large numbers (see Proposition 2.3). We are interested in the *large deviations* of the measure-valued process $X^n_\cdot$ as n tends towards infinity.

The ambitions of this paper consist first in providing with results of wide applicability through the contraction principle and second in establishing a Sanov-like theorem in a dependent context. Moreover, the asymptotic behavior of the random allocation model is interesting because random allocations models arise naturally in many applications running from algorithm analysis [29] to statistics and statistical learning theory [6]. In this paper, we illustrate our main result by deriving a large deviation principle (LDP) for the degree distribution of sparse random graphs. This complements recent results by McKay and Wormald on moderate deviations [18].

1.1. The limitations of Poisson approximation.

The random allocations phenomenon is intimately connected with questions in Poisson approximation [1]. For any fixed t, X^n_t may be considered as the empirical measure of n identically distributed independent Poisson random variables conditioned on the fact that their sum is equal to $\lfloor nt \rfloor$. Let us denote by Y^n_t the empirical measure of n independent Poisson random variables with parameter t. By the Sanov theorem [9], Y^n_t satisfies a LDP in $\mathcal{P}(\mathbb{N})$ with good rate function $H(\nu \mid \mathbf{p}_t) \triangleq \sum_{i\in\mathbb{N}} \nu(i) \log \frac{\nu(i)}{\mathbf{p}_t(i)}$ (here $(\mathbf{p}_t(i))_{i\in\mathbb{N}}$ is the Poisson distribution with mean t). As the probability that the sum of n independent Poisson random variables with parameter t is equal to $\lfloor nt \rfloor$ is of order $1/\sqrt{n}$, we immediately get the following LDP upper bound for X^n_t:

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(X^n_t \in C) \leq - \inf_{\nu \in C} H(\nu \mid \mathbf{p}_t) \ ,$$

which by the way proves that, for fixed t, X^n_t satisfies a law of large numbers when all bins are initially empty. A very natural question is whether the lower bound also holds. Analyzing the asymptotic behavior of collections of dependent random variables by conditioning collections of independent random variables is a standard method in random combinatorics and statistical mechanics where it is related to the relation between micro and macro-canonical ensembles. This approach is all the more attractive as in the case of integer partitions profiles, the LDP can be recovered from the analysis of a system of independent random variables using a coupling argument [8]. In the language of statistical mechanics, random integer partitions correspond to Bose-Einstein statistics while the random allocations we are interested in, correspond to Maxwell-Boltzmann statistics [11].

In the case of random allocations, conditioning a collection of independent Poisson random variables has often been used to obtain Central Limit Theorems for X^n_t for fixed t (see [19, 22, 23, 24] and references therein). But the depoissonization arguments used in those papers which essentially go back to [2], required ingenuity. Facing technical difficulties in such a situation should not be surprising since at the Central Limit Theorem scale, the empty bin statistics $Y^n_t(0)$ and $X^n_t(0)$ behave differently, they have different variances (see for example [14]). Therefore in this paper, we do not use the depoissonization approach, we characterize the LDP for the flow of empirical occupancy measures by taking advantage of its simple dynamical behavior.

1.2. Related works.

Random allocations, urns models, or occupancy problems constitute basic objects in random combinatorics, statistics and theoretical computer science [13, 29]. Up to our knowledge, random allocations problems have mostly been investigated by two kinds of methods: combinatorial analysis and probabilistic techniques using characteristic functions. The combinatorial approach is illustrated in the Russian monograph [14]: in a first step appropriate generating functions are constructed and in a second step, asymptotic analysis is carried out using techniques from complex analysis. Such approaches provide results of unrivaled precision, but tend to be rather involved when dealing with infinite-dimensional random variables. Up to our knowledge, the probabilistic approach has heavily used the fact that multinomial random variables can be regarded as conditioned Poisson random variables (see for example [19, 22, 23, 24]), and few investigations have taken advantage of the simple structure of the allocation process and particularly of the underlying martingale structure [3].

1.3. **Outline of the paper.** The paper is organized as follows. Further definitions are first introduced in Section 2. The main result of the paper, Theorem 2.9, is stated at the end of this section. The LDP upper-bound is proved in Section 3 where a variational representation of the rate function is established. Thanks to Orlicz spaces techniques, the non-variational representation of the rate function is established in Section 4. The LDP lower bound is established in Section 5 thanks to the classical change of measure argument. The difficulty lies in the construction of a rich collection of absolutely continuous change of measures. In Section 6, the LDP for the flow of empirical occupancy measures is shown to hold with the same rate function when the topology is strengthened. In Section 7, a coupling argument allows to derive the LDP for the degree distribution of sparse random graphs from Theorem 2.9.

2. Main results

2.1. **The model.** Let us first refine the model description.

The kinetics of the process is as follows. If the k^{th} ball is allocated into a bin the score of which is $S^n_{k-1}(B^n_k) = i$, then:

$$\begin{aligned} X^n_{k/n}(i+1) &= X^n_{(k-1)/n}(i+1) + 1/n \\ X^n_{k/n}(i) &= X^n_{(k-1)/n}(i) - 1/n \\ X^n_{k/n}(j) &= X^n_{(k-1)/n}(j), \quad j \notin \{i, i+1\} \end{aligned}$$

and the value of the process remains constant on the time interval $[k/n, (k+1)/n)$. For any $i \geq 0$, each realization of $X^n(i)$ stands in the space $D([0,T],\mathbb{R})$ of right continuous left limited (càdlàg) paths from $[0,T]$ to $\mathbb{R}$. The sample path space of X^n is $D_{\mathcal{P}} \stackrel{\Delta}{=} D([0,T],\mathcal{P}(\mathbb{N}))$: the set of all $\nu : [0,T] \mapsto \mathcal{P}(\mathbb{N})$ such that $\nu(i) \in D([0,T],\mathbb{R})$ for all $i \geq 0$.
Let us endow $D_{\mathcal{P}}$ with its canonical filtration $(\mathcal{F}_t)_{0\leq t\leq T}$ where $\mathcal{F}_t = \sigma(\pi_s; 0 \leq s \leq t)$ is generated by the canonical projections $\pi_s : \nu \in D_{\mathcal{P}} \mapsto \nu_s \in \mathcal{P}(\mathbb{N})$ and the σ-field on $\mathcal{P}(\mathbb{N})$ is induced by the usual product σ-field on $\mathbb{R}^{\mathbb{N}}$. Similarly, Ω^n is endowed with the natural filtration $(\mathcal{A}^n_t)_{0\leq t\leq T}$ where $\mathcal{A}^n_t = \sigma(B^n_k; 1 \leq k \leq \lfloor nt \rfloor)$. The σ-fields on Ω^n and $D_{\mathcal{P}}$ are $\mathcal{A}^n_T$ and $\mathcal{F}_T$. Clearly, $X^n_t = X^n \circ \pi_t$, X^n is an $(\mathcal{A}^n_t)$-adapted process and the canonical process π is $(\mathcal{F}_t)$-adapted.

It is assumed that the bins are chosen uniformly and indepently at each time. This means that the probability measure $\mathbb{P}^n$ on Ω^n is the product of the uniform distribution on $\{1, \dots, n\}$:

$$\mathbb{P}^n(d\omega) = \bigotimes_{1 \le k \le \lfloor nT \rfloor} \left(\frac{1}{n} \sum_{1 \le \alpha \le n} \delta_\alpha(d\omega_k) \right)$$

2.2. **A larger class of models.** In this section, we describe slightly more genral allocation schemes, and state the corresponding *laws of large numbers.* This will be useful when deriving the LDP lower bound and when interpreting our large deviation results. In this larger class of models, the choice of the bin B^n_{k+1} at time $k+1$ depends on the whole empirical distribution $X^n_{k/n}$ at time k. Let us take a continuous function λ on $[0, T] \times \mathbb{N}$ such that:
i) λ has range included in $[a, \infty)$ for some $a > 0$ and
ii) there exists some integer M such that $\lambda(t, i) = 1$ for all $i \ge M$ and all t.
Conditionally on $\mathcal{A}^n_{k/n}$, the probability of choosing a bin with score i is

$$\mathbb{Q}^n(S^n_k(B^n_{k+1}) = i \mid \mathcal{A}^n_{k/n}) = \lambda(k/n, i) X^n_{k/n}(i) / \langle \lambda_{k/n}, X^n_{k/n} \rangle \tag{2.1}$$

where $\langle \lambda_{k/n}, X^n_{k/n} \rangle = \sum_{j \ge 0} \lambda(k/n, j) X^n_{k/n}(j)$. Let us remark that as $\inf \lambda \ge a > 0$, we have $\langle \lambda_{k/n}, X^n_{k/n} \rangle > 0$. The choice of the bin, among those of score i is uniform. Note that if $X^n_{k/n}(i; \omega) = 0$, then one cannot allocate the $(k+1)^{\text{th}}$ ball in a bin with score i :, the product form λX in (2.1) enforces this minimal consistency in the model. It is worth noting that under $\mathbb{Q}^n$, for any $d \ge M$ the $\mathbb{R}^d$-valued process formed by the projection of $X^n_\cdot$ on its first d coordinates is a Markov process. Indeed:

$$\langle \lambda_{k/n}, X^n_{k/n} \rangle = 1 + \sum_{i=0}^{M} (\lambda_{k/n}(i) - 1) X^n_{k/n}(i), \tag{2.2}$$

and hence for $d \ge M$, the law of the d-dimensional projection of X^n at time $(k+1)/n$ only depends on the value of the d-dimensional projection of X^n at time k/n. Under $\mathbb{Q}^n$, X^n is a projective limit of vector-valued Markov processes.
We shall see later (see Lemma 3.3) that the laws Q^n of X^n under $\mathbb{Q}^n$ are uniformly tight in $D_{\mathcal{P}}$ endowed with the topology of uniform convergence. Therefore we will only need to check that all converging subsequences have the same limit and this can be done by checking that converging subsequences of finite-dimensional distributions have the same limit. Because of the preceding remark it will even be enough to check this for the laws of d-dimensional projections of X^n with $d \ge M$. This point will be checked in Section 5.3.

Convention. Here and below, it is assumed that the value of all functions at score $i = -1$ is 0 : $X^n_t(-1), \nu_t(-1), \ell^\nu_t(-1), \dots = 0$.

Proposition 2.3 (Law of large numbers). *Let $\mathbb{Q}^n$ be specified by (2.1) and assume that $\inf_{t,i} \lambda(t, i) > 0$, that λ is continuous and that there exists an integer M such that $\lambda(t, i) = 1$ for $i \ge M$ and $t \ge 0$. Then the sequence (X^n) converges in law in $D_{\mathcal{P}}$ endowed with the topology of uniform convergence towards the process X with X_0 distributed according to δ_0 and $X_\cdot$ being the unique solution ν of*

$$\frac{d\nu_t}{dt}(i) = \ell^\nu_t(i-1)\nu_t(i-1) - \ell^\nu_t(i)\nu_t(i) \tag{2.4}$$

with initial condition $X_0 = \delta_0$ where $\ell^\nu_t(i) = \lambda(t, i) / \sum_{j \ge 0} \lambda(t, j) \nu_t(j)$.
In particular with $\mathbb{Q}^n = \mathbb{P}^n$ the limiting path is $\nu = \mathbf{p}$: the time-marginal flow of

the Poisson process with parameter 1, given by

$$\frac{d\mathbf{p}_t}{dt}(i) = \mathbf{p}_t(i-1) - \mathbf{p}_t(i) \tag{2.5}$$

Note that the special case of $\mathbb{P}^n$ is obtained with $\ell^\nu \equiv 1$, and that then $\sum_{i\geq 0} \ell_t^\nu(i)\nu_t(i) = 1$ for all t.

As mentioned in the introduction, we assume in the whole paper that for all n the initial distribution X_0^n is almost surely the Dirac mass (δ) at 0. This is only for the sake of simplicity. Actually, our main results still hold when the initial empirical occupancy measures satisfy a law of large number and an LDP and if the limiting initial occupancy distribution is sufficiently integrable.

2.3. **Large deviation principle.** Let $\mathcal{X}$ be a Hausdorff topological space endowed with some σ-field. A *rate function* on $\mathcal{X}$ is a function $I : \mathcal{X} \mapsto [0,\infty]$ which is lower semicontinuous. It is said to be a *good rate function* if in addition, its level sets $\{I \leq a\}$ are compact. A sequence $(X^n)_{n\geq 1}$ of random elements in $\mathcal{X}$ is said to satisfy the *large deviation principle* with rate function I if the sequence $(P^n)_{n\geq 1}$ of the corresponding laws on $\mathcal{X}$ satisfies the following

(1) *Upper bound*: For any measurable closed subset C of $\mathcal{X}$,

$$\limsup_{n\to\infty} \frac{1}{n}\log P^n(C) \leq -\inf_{x\in C} I(x)$$

(2) *Lower bound*: For any measurable open subset G of $\mathcal{X}$,

$$\liminf_{n\to\infty} \frac{1}{n}\log P^n(G) \geq -\inf_{x\in G} I(x)$$

2.4. **The topologies.** Considering $\mathcal{P}(\mathbb{N})$ as a subset of summable sequences $\ell_1(\mathbb{N})$, it is naturally endowed with the induced topologies of the pointwise convergence and of the norm topology. A metric for the pointwise topology is given for any $\pi,\rho \in \mathcal{P}(\mathbb{N})$, by $\sum_{i\geq 0} 2^{-i}|\pi(i)-\rho(i)|$, $\pi,\rho\in\mathcal{P}(\mathbb{N})$, while the norm on ℓ_1 induces the metric of the total variation

$$\|\pi-\rho\| = \sum_{i\geq 0} |\pi(i)-\rho(i)|,\ \ \pi,\rho\in\mathcal{P}(\mathbb{N}) \tag{2.6}$$

In fact, these topologies are equivalent. Clearly, the convergence in total variation implies the pointwise convergence. But the converse also holds, due to the dominated convergence theorem.
In this article, we shall not use the Skorokhod topology. The space $D_{\mathcal{P}}$ is endowed with the topology of uniform convergence associated with the norm

$$\|\nu-\mu\| = \sup_{0\leq t\leq T} \|\nu_t-\mu_t\|,\ \ \nu,\mu\in D_{\mathcal{P}}\ \ .$$

This makes $D_{\mathcal{P}}$ a nonseparable complete metric space. Note that $\mathcal{F}_T$ is smaller than the Borel σ-field of $D_{\mathcal{P}}$. This is the reason why only *measurable* closed and open sets are considered in the above statement of a LDP.

2.5. **The rate function.** It will be convenient to associate with each path $\nu \in D_{\mathcal{P}}$ the relaxed measure on $[0,T] \times \mathbb{N}$:

$$\bar{\nu}(dtdz) = \nu_t(dz)dt$$

A path $\nu \in D_{\mathcal{P}}$ is said to be *absolutely continuous* if for each $i \in \mathbb{N}$, there exists $\dot{\nu}(i)$ in $L_1([0,T],dt)$ such that $\nu_t(i) - \nu_0(i) = \int_{[0,t]} \dot{\nu}_s(i)\,ds$. For each absolutely continuous path ν, let us define v^ν, $\bar{\nu}$-almost everywhere by:

$$v_t^\nu(j) \triangleq -\sum_{i \le j} \dot{\nu}_t(i) \quad \text{for } j \ge 0 \ . \tag{2.7}$$

Let P be a probability measure and Q a non-negative measure on some measure space. The *relative entropy* of Q with respect to P is defined by:

$$H(Q|P) = \begin{cases} \mathbb{E}_Q \log \frac{dQ}{dP} & \text{if } Q \text{ is a probability measure and } Q \ll P \\ \infty & \text{otherwise.} \end{cases}$$

with the convention $0 \log 0 = 0$.

We are now in a position to define the rate function I, for any $\nu \in D_{\mathcal{P}}$:

$$I(\nu) \triangleq \begin{cases} \int_{[0,T]} H(v_t^\nu|\nu_t)dt & \text{if } \nu \text{ is absolutely continuous} \\ \infty & \text{otherwise.} \end{cases} \tag{2.8}$$

Note that for any ν satisfying $I(\nu) < \infty$, dt-almost everywhere v_t^ν is a *probability* measure on $\mathbb{N}$. Moreover, simple algebraic manipulations show that the time-derivative of the mean score is 1 dt-almost everywhere:

$$\sum_i i\,\dot{\nu}_t(i) = 1 \ .$$

The finiteness of the rate function warrants that balls are allocated with unit intensity.

2.6. **The main results.** The main result of the paper is the following theorem.

Theorem 2.9. *The sequence $(X^n)_{n\ge1}$ satisfies the* LDP *on $D_{\mathcal{P}}$ with the good rate function I.*

As the identities (2.7) are equivalent to:

$$\dot{\nu}_t(i) = v_t^\nu(i-1) - v_t^\nu(i) \quad \text{for} \quad i \ge 0 \ , \tag{2.10}$$

and as $I(\nu) = 0$ if and only if $v^\nu = \nu$, $\bar{\nu}$-almost everywhere, we obtain that $I(\nu) = 0$ if and only if for almost every t, $\dot{\nu}_t(i) = \nu_t(i-1) - \nu_t(i)$, for ν_t-almost all $i \ge 0$. Hence $\nu = \mathbf{p}$, which is in agreement with (2.5) in Proposition 2.3.
The rate function may actually be further interpreted:

Proposition 2.11. *The function I is a convex rate function.*
Let $\nu \in D_{\mathcal{P}}$, then $I(\nu) < \infty$ if and only if ν is absolutely continuous and there exists a measurable $\mathbb{R}^{\mathbb{N}}$-valued process ℓ^ν which is defined $\bar{\nu}$-almost everywhere, such that

(1) *the following master equation is satisfied $\bar{\nu}$-almost everywhere:*

$$\dot{\nu}_t(i) = \ell_t^\nu(i-1)\nu_t(i-1) - \ell_t^\nu(i)\nu_t(i), \ i \ge 0 \tag{2.12}$$

(2) $(\ell_t^\nu(i)\nu_t(i))_{i\ge0}$ *defines a probability on $\mathbb{N}$, for dt-almost every t*

(3) $\int_{[0,T]} [\sum_{i=0}^\infty \nu_t(i)\, \ell_t^\nu(i)\, \log \ell_t^\nu(i)]\, dt < \infty$.

Alternative expressions for $I(\nu)$ are

$$\begin{aligned} I(\nu) &= \int_{[0,T]\times\mathbb{N}} \ell^\nu \log \ell^\nu \, d\bar{\nu} = \int_{[0,T]} \left[\sum_{i=0}^{\infty} \nu_t(i)\, \ell^\nu_t(i) \, \log \ell^\nu_t(i) \right] dt \\ &= \int_{[0,T]} H(\ell^\nu_t \nu_t \mid \nu_t)\, dt \ , \end{aligned}$$

where ℓ^ν is any process satisfying the above properties 1, 2 and 3.

Proof. The convexity and the lower semicontinuity of I is a direct consequence of its variational representation obtained in Proposition 4.4.

If $I(\nu) < \infty$, then dt-almost everywhere v^ν_t is a probability measure on $\mathbb{N}$ which is absolutely continuous with respect to ν_t. Let $\ell^\nu_t = \frac{dv^\nu_t}{d\nu_t}$ be its Radon-Nykodym derivative. Clearly, property 2 holds. As

$$I(\nu) = \int_{[0,T]} H(v^\nu_t \mid \nu_t)\, dt = \int_{[0,T]} \langle \ell^\nu_t \log \ell^\nu_t, \nu_t \rangle \, dt \ , \tag{2.13}$$

property 3 is satisfied. Finally, property 1 is given by (2.10).
Conversely, let ℓ^ν_t satisfy conditions 1, 2 and 3. Let us set $v^\nu_t = \ell^\nu_t \nu_t$. Then, (2.12) is (2.10) which is equivalent to (2.7). Finally, conditions 2 and 3 with (2.13) imply that $I(\nu)$ is finite.
The alternative expressions for $I(\nu)$ follow from (2.13). □

Let ν be an absolutely continuous path. If $\nu_t(i) > 0$, (2.12) gives $\ell^\nu_t(i) = [-\sum_{j\le i} \dot{\nu}_t(j)]/\nu_t(i)$, so that $\ell^\nu_t(i)$ is uniquely defined up to $\bar{\nu}$-a.e. equality on $\{(t,i); \nu_t(i) > 0\}$. On the other hand, (2.12) and (2.13) are insensitive to the values of ℓ^ν on the complementary set $\{(t,i); \nu_t(i) = 0\}$. Therefore,

$$\ell^\nu_t(i) = \begin{cases} [-\sum_{j\le i} \dot{\nu}_t(j)]/\nu_t(i) & \text{if } \nu_t(i) > 0 \\ 1 & \text{if } \nu_t(i) = 0 \end{cases} \tag{2.14}$$

is a useful measurable inversion formula for ℓ^ν.

2.7. **Examples.** Let us take $T = 1$ to make things easier. By Proposition 2.3, we have the weak law of large numbers: $X^n \to \mathbf{p}$ where $\mathbf{p}$ is given by (2.5). By the upper bound of the large deviation principle in Theorem 2.9 and Borel-Cantelli's lemma, this convergence holds almost surely (considering an appropriate space $\Omega = \prod_{n\ge 1} \Omega^n$ and so on): we have a *strong law of large numbers.*
In particular, considering the final time $t = T = 1$, we obtain that almost surely, the limiting proportion of bins with a score at least equal to 4 is $\lim_{n\to\infty} X^n_1(\{4,5,\dots\})$ $= \mathbf{p}_1(\{4,5,\dots\}) = e^{-1}\sum_{i=4}^{\infty} 1/i! \approx 0.019$ and one may ask for the rate of convergence to zero of $\mathbb{P}^n(X^n_1(\{4,5,\dots\}) \ge 0.03)$ as n tends to infinity. The answer is given by Theorem 2.9, indeed by convexity:

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}^n(X^n_1(\{4,5,\dots\}) \ge 0.03) = -\inf\{I(\nu); \nu, \nu_1(\{4,5,\dots\}) \ge 0.03\}$$

since $\inf\{I(\nu); \nu, \nu_1(\{4,5,\dots\}) \ge 0.03\} = \inf\{I(\nu); \nu, \nu_1(\{4,5,\dots\}) > 0.03\}$ (this identity requires some work). As $\inf\{I(\nu); \nu, \nu_1(\{4,5,\dots\}) \ge 0.03\} < \infty$ (this also requires some work) and I has compact level sets, there is at least one ν^* in $D([0,1], \mathcal{P}(\mathbb{N}))$ such that $I(\nu^*) = \{I(\nu); \nu, \nu_1(\{4,5,\dots\}) \ge 0.03\} < \infty$.

More generally, let A be a measurable subset of $D_{\mathcal{P}}$ such that $I(\mathrm{int}(A)) = I(\mathrm{cl}(A)) < \infty$, where $I(B)$ stands for $\inf\{I(\nu); \nu \in B\}$. Then, we have

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}^n(X^n \in A) = -I(A).$$

In addition, suppose that there exists a unique ν^* in A such that $I(\nu^*) = I(A)$. Then,

$$\lim_{n\to\infty} \mathbb{P}^n(X^n \in \cdot \mid X^n \in A) = \delta_{\nu^*}(\cdot) \ . \tag{2.15}$$

Indeed, let U be any open neighbourhood of ν^*. By Theorem 2.9, we have

$$\begin{aligned} & \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}^n(X^n \in U^c \mid X^n \in A) \\ \leq \quad & \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}^n(X^n \in U^c \cap A) - \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}^n(X^n \in A) \\ \leq \quad & -[I(U^c \cap \mathrm{cl}(A)) - I(\mathrm{int}(A))] \ . \end{aligned}$$

As I has compact level sets, there exists some compact subset K such that $I(U^c \cap \mathrm{cl}(A)) = I(K \cap U^c \cap \mathrm{cl}(A))$ and there is some ν_U in the compact set $K \cap U^c \cap \mathrm{cl}(A)$ such that $I(K \cap U^c \cap \mathrm{cl}(A)) = I(\nu_U)$. As it is assumed that $I(\mathrm{int}(A)) = I(\nu^*) < I(\nu_U)$, the desired result is proved with an exponential rate of convergence.

3. The upper bound

3.1. **Statement of the upper bound.** As we will resort to duality arguments, let us first define a set of test functions. Let g be a real function on $\mathbb{N}$ (a sequence of real numbers), we set $Dg(j) \triangleq g(j+1) - g(j)$, for all $j \geq 0$. For any function $G : [0,T] \times \mathbb{N} \to \mathbb{R}$, let us denote for all $j \in \mathbb{N}$, $G(j) : t \in [0,T] \mapsto G_t(j)$ and for all $0 \leq t \leq T$, $G_t = (G_t(j))_{j\geq 0}$. The set of relevant test functions is

$$\mathcal{G} \triangleq \left\{ G : [0,T] \times \mathbb{N} \to \mathbb{R}; \sup_{t,j} |DG_t(j)| < \infty, G(j) \in \mathcal{C}, \forall j \in \mathbb{N} \right\}$$

where $\mathcal{C}$ is the space of all the functions $f : [0,T] \mapsto \mathbb{R}$ which are absolutely continuous and such that $f(T) = 0$. For any G in $\mathcal{G}$, we will denote by $\dot{G}_t$ the generalized derivative of G_t with respect to t , i.e.:

$$G_t(j) = -\int_{[t,T]} \dot{G}_s(j) ds, \quad t \in [0,T], j \in \mathbb{N} \ .$$

Let us also introduce the notation $\dot{\nu}(G)$. For all $G \in \mathcal{G}$ and $\nu \in D_{\mathcal{P}}$:

$$\dot{\nu}(G) \triangleq -\langle G_0, \nu_0 \rangle - \int_{[0,T]} \langle \dot{G}_t, \nu_t \rangle \, dt \tag{3.1}$$

The main result of the section is the following variational formulation of the large deviation upper bound.

Proposition 3.2. *For any closed measurable subset C of $D_{\mathcal{P}}$ we have*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}^n(X^n \in C) \leq -\inf_{\nu\in C} \sup_{G\in\mathcal{G}} \left\{ \dot{\nu}(G) - \int_{[0,T]} \log\langle \exp(DG_t), \nu_t \rangle \, dt \right\} \ .$$

Proof. In Lemma 3.7 the upper bound is proved for C measurable convex and compact. One removes the convexity restriction by means of a standard argument which is described in the proof of the upper bound of [28, Theorem 4.1]. Note that this argument still works in our context, since our space is locally convex and its topology is generated by measurable balls. In Lemma 3.3, it is shown that the laws of the X^n's are compactly supported. Combining these results, the upper bound holds for all measurable closed subsets. □

3.2. **Compactness considerations.** The following result shows that we should not be distracted by exponential tightness issues.

Lemma 3.3. *There exists a compact subset K of $D_{\mathcal{P}}$ such that for all $n \geq 1$ and all $\omega \in \Omega^n$, $X^n(\omega)$ belongs to K.*

Proof. We take advantage of the simple form of the sample paths of $X^n(\omega)$. Let x^n be any realization of X^n. As for any t and n,

$$\sup_{t \leq r,s < t+1/n} \|x^n_s - x^n_r\| = 2/n \ ,$$

we have for any $0 \leq t < t+\delta \leq T$, $\sup_{t \leq r,s<t+\delta} \|x^n_s - x^n_r\| \leq (1+n\delta)2/n = 2/n + 2\delta$.

On the other hand, for any t, the mean score per bin is $\sum_{i \geq 0} i x^n_t(i) \leq t$. Hence, for all $t \leq T$, x^n_t belongs to the relatively compact subset of $\mathcal{P}(\mathbb{N})$ consisting of the probability measures with their first moment bounded above by T. As $\mathcal{P}(\mathbb{N})$ is complete, this relatively compact subset is totally bounded.

As in the proof of Ascoli-Arzela's theorem, it follows from these considerations that for any $\epsilon > 0$, one can build a finite collection of open balls of $D_{\mathcal{P}}$ with radius ϵ which covers $\cup_{n \geq 1} \{X^n(\omega); \omega \in \Omega^n\}$. This means that it is a totally bounded set in the complete metric space $D_{\mathcal{P}}$. Therefore, its closure is compact.

In order to build this open covering, choose the centers of the ϵ-balls as piecewise constant paths (being constant on small enough intervals) with their values in a finite subset of $\mathcal{P}(\mathbb{N})$ (the set of the centers of small balls finitely covering the above totally bounded subset of $\mathcal{P}(\mathbb{N})$). □

3.3. **Exponential martingale.** We introduce a family of exponential martingales $Z^{G,n}$ which will allow us, by means of Lemma 3.4 below, to derive in Lemma 3.7 the upper bound for compact convex subsets.
For any G in $\mathcal{G}$ and $n \geq 1$, let us define the process $Z^{G,n}$ by

$$\begin{aligned} \frac{1}{n} \log Z^{G,n}_t \quad &\triangleq \quad \langle G_t, X^n_t \rangle - \langle G_0, X^n_0 \rangle - \int_{[0,t]} \langle \dot{G}_s, X^n_s \rangle \, ds \\ &\quad - \sum_{k=0}^{\lfloor nt \rfloor} \frac{1}{n} \log \sum_{j \geq 0} X^n_{k/n}(j) \exp\left(DG_{\frac{k+1}{n}}(j) \right) \end{aligned}$$

Note that since DG is bounded, there exists $c \geq 0$ such that $|G(i)| \leq c(1+i)$ for all $i \in \mathbb{N}$. As for all $0 \leq t \leq T$, $\sum_{i \geq 0} i X^n_t(i) \leq T$, it follows that all the terms in the definition of $Z^{G,n}$ are well defined and that it is a bounded process.

Lemma 3.4. *For any $G \in \mathcal{G}$ and $n \geq 1$, $(Z^{G,n}_t)_{0 \leq t \leq T}$ is a $\mathbb{P}^n$-martingale with respect to the filtration $(\mathcal{A}^n_t)_{0 \leq t \leq T}$. In particular, $\mathbb{E}_{\mathbb{P}^n} Z^{G,n}_T = 1$.*

Proof. It is enough to check that for any $0 \leq t \leq t+h \leq T$:

$$\mathbb{E}_{\mathbb{P}^n}\left(Z^{G,n}_{t+h}/Z^{G,n}_t \mid \mathcal{A}^n_t\right) = 1.$$

We have

$$\begin{aligned}\frac{1}{n}\log[Z^{G,n}_{t+h}/Z^{G,n}_t] &= \langle G_{t+h}, X^n_{t+h}\rangle - \langle G_t, X^n_t\rangle - \int_{[t,t+h]} \langle \dot{G}_s, X^n_s\rangle\, ds \\ &\quad - \sum_{k=\lfloor nt\rfloor+1}^{\lfloor n(t+h)\rfloor} \frac{1}{n}\log\sum_j X^n_{\frac{k}{n}}(j)\exp\left(DG_{\frac{k+1}{n}}(j)\right) \qquad (3.5)\end{aligned}$$

If $\lfloor n(t+h)\rfloor = \lfloor nt\rfloor$, the right-hand side vanishes and there is nothing to prove. Using cascade conditioning, all other cases reduce to $\lfloor n(t+h)\rfloor = \lfloor nt\rfloor + 1$. Furthermore, it is enough to consider the case $\lfloor nt\rfloor = nt$ and $1/n \leq h < 2/n$. Let $\frac{k}{n} \leq t < \frac{k+1}{n} \leq t+h$, then:

$$\langle G_{t+h}, X^n_{t+h}\rangle - \langle G_t, X^n_t\rangle = \int_{[t,t+h]} \langle \dot{G}_s, X^n_s\rangle\, ds + \langle G_{\frac{k+1}{n}}, X^n_{\frac{k+1}{n}}\rangle - \langle G_{\frac{k+1}{n}}, X^n_{\frac{k}{n}}\rangle\ .$$

Hence the right-hand side of (3.5) reduces to

$$\langle G_{\frac{k+1}{n}}, X^n_{\frac{k+1}{n}}\rangle - \langle G_{\frac{k+1}{n}}, X^n_{\frac{k}{n}}\rangle - \frac{1}{n}\log\sum_{j\geq 0} X^n_{\frac{k}{n}}(j)\ \exp\left(DG_{\frac{k+1}{n}}(j)\right)\ .$$

As

$$\mathbb{E}_{\mathbb{P}^n}\left[\exp\left(n[\langle G_{\frac{k+1}{n}}, X^n_{\frac{k+1}{n}}\rangle - \langle G_{\frac{k+1}{n}}, X^n_{\frac{k}{n}}\rangle]\right) \mid \mathcal{A}^n_t\right] = \sum_{j\geq 0} X^n_{\frac{k}{n}}(j)\ \exp DG_{\frac{k+1}{n}}(j)\ ,$$

the proof is completed. □

3.4. **Compact convex subsets.** Establishing the LDP upper bound for convex compact sets is now straightforward thanks to the following general min-max theorem due to Sion [27], see also [9, Exercice 2.2.38].

Theorem 3.6. *(Sion,1958) Let $\mathcal{K}(\theta, y)$ be convex and lower semicontinuous in y and concave and upper semicontinuous in θ. Let C be a compact convex set, then:*

$$\inf_{y\in C}\sup_{\theta}\mathcal{K}(\theta, y) = \sup_{\theta}\inf_{y\in C}\mathcal{K}(\theta, y).$$

This theorem may be applied in the following context. Let $C \subset D_{\mathcal{P}}$ be a convex and compact set, let the function $\mathcal{K}(G, \nu)$ for $G \in \mathcal{G}$ and $\nu \in D_{\mathcal{P}}$ be defined as

$$\mathcal{K}(G,\nu) \triangleq \dot{\nu}(G) - \int_{[0,T]} \log\langle \exp(DG_t), \nu_t\rangle dt.$$

Let us denote by $\mathcal{K}^n(G, \nu)$ the following discretized version of $\mathcal{K}$:

$$\mathcal{K}^n(G,\nu) \triangleq \dot{\nu}(G) - \sum_{k=0}^{\lfloor nT\rfloor} \frac{1}{n}\log\langle \exp(DG_{\frac{k+1}{n}}), \nu_{\frac{k}{n}}\rangle.$$

The functions $\mathcal{K}$ and $\mathcal{K}^n$ are convex with respect to ν thanks to the concavity of log, and concave with respect to G thanks to Hölder's inequality and the fact that log is increasing. The continuity (hence the lower semicontinuity) with respect to ν and the upper semicontinuity with respect to G follow from the definition of $\mathcal{G}$.

Lemma 3.7. *Let C be a measurable convex compact subset of $D_{\mathcal{P}}$, then*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}^n \left(X^n \in C \right) \leq - \inf_{\nu \in C} \sup_{G \in \mathcal{G}} \mathcal{K}(G,\nu) \ .$$

Proof. By an exponential Markov inequality, for any $G \in \mathcal{G}$:

$$\begin{aligned} \mathbb{P}^n \left(X^n \in C \right) &\leq \mathbb{P}^n \Big(e^{n\mathcal{K}^n(G,X^n)} \geq \inf_{\nu \in C} e^{n\mathcal{K}^n(G,\nu)} \Big) \\ &\leq \mathbb{E}_{\mathbb{P}^n} \Big(e^{n\mathcal{K}^n(G,X^n)} \Big) / \inf_{\nu \in C} e^{n\mathcal{K}^n(G,\nu)} \end{aligned}$$

As $\mathbb{E}_{\mathbb{P}^n} e^{n\mathcal{K}^n(G,X^n)} = \mathbb{E}_{\mathbb{P}^n} Z_T^{G,n}$, by Lemma 3.4 we have $\mathbb{E}_{\mathbb{P}^n} e^{n\mathcal{K}^n(G,X^n)} = 1$, so that

$$\mathbb{P}^n \left(X^n \in C \right) \leq \exp \big(-n \inf_{\nu \in C} \mathcal{K}^n(G,\nu) \big) \ .$$

We may now optimize with respect to $G \in \mathcal{G}$:

$$\mathbb{P}^n \left(X^n \in C \right) \leq \inf_G \exp \left(-n \inf_{\nu \in C} \mathcal{K}^n(G,\nu) \right) = \exp \left(-n \sup_{G \in \mathcal{G}} \inf_{\nu \in C} \mathcal{K}^n(G,\nu) \right) .$$

Letting n tend to infinity, one obtains

$$\limsup_n \frac{1}{n} \log \mathbb{P}^n \left(X^n \in C \right) \leq - \liminf_n \sup_G \inf_{\nu \in C} \mathcal{K}^n(G,\nu) \ .$$

It remains to prove that

$$\liminf_n \sup_{G \in \mathcal{G}} \inf_{\nu \in C} \mathcal{K}^n(G,\nu) \geq \sup_G \inf_{\nu \in C} \mathcal{K}(G,\nu) \ . \tag{3.8}$$

As DG is bounded and t-continuous and ν is right continuous, $\mathcal{K}^n$ converges pointwise towards $\mathcal{K}$. For any $G \in \mathcal{G}$, $\mathcal{K}^n(G,\cdot)$ is a sequence of pointwise converging continuous convex functions on a complete metric space. Hence, its converges uniformly towards $\mathcal{K}(G,\cdot)$ on the compact set C as n tends to infinity (see Theorem A.1 in the Appendix, for a proof of this result).
Let us take $\epsilon > 0$. For any n and G, let $\nu^{G,n} \in C$ be such that $\mathcal{K}^n(G,\nu^{G,n}) \leq \inf_{\nu \in C} \mathcal{K}^n(G,\nu) + \epsilon$. Because of the above uniform convergence, for any G, there exists $n^G \geq 1$ such that for all $n \geq n^G$: $\inf_{\nu \in C} \mathcal{K}^n(G,\nu) \geq \mathcal{K}^n(G,\nu^{G,n}) - \epsilon \geq \mathcal{K}(G,\nu^{G,n}) - 2\epsilon \geq \inf_{\nu \in C} \mathcal{K}(G,\nu) - 2\epsilon$. Hence,

$$\sup_{G \in \mathcal{G}} \liminf_{n\to\infty} \inf_{\nu \in C} \mathcal{K}^n(G,\nu) \geq \sup_{G \in \mathcal{G}} \inf_{\nu \in C} \mathcal{K}(G,\nu)$$

As, $\liminf_{n\to\infty} \sup_{G \in \mathcal{G}} \geq \sup_{G \in \mathcal{G}} \liminf_{n\to\infty}$, this proves (3.8).

Now applying Sion's minimax Theorem 3.6, the right hand side in (3.8) is identified with $\inf_{\nu \in C} \sup_{G \in \mathcal{G}} \mathcal{K}(G,\nu)$ and the proof is completed. □

4. The rate function

Our goal in this section is to prove that the rate function appearing in Proposition 3.2 is equal to the rate function I defined at (2.8).

This identification is stated in Proposition 4.4 below. It will be proved using the Riesz representation theorem in Orlicz spaces. Using Riesz representation theorem in L_2 would have been appropriate if we were facing a Gaussian situation, but the bins and balls model resort from the Poisson approximation problem. Orlicz spaces constitute a tailor-made framework to provide with non-variational representations of the rate function in such a case [15]. For the sake of completeness, let us first recall some basic facts about Orlicz spaces which are the extensions of the classical L_p spaces.

4.1. **Orlicz spaces.** A *Young function* θ is an even, convex, $[0,\infty]$-valued function satisfying $\theta(0)=0$, $\lim_{s\to+\infty}\theta(s)=+\infty$ and $\theta(s_0)<+\infty$ for some $s_0>0$. Let μ be a nonnegative bounded measure on the measurable space $(\Sigma,\mathcal{A})$. Consider the following vector spaces:

$$\begin{aligned} L_\theta &= \left\{f:\Sigma\to\mathbb{R},\ \exists a>0, \int_\Sigma \theta\Big(\frac{f}{a}\Big)\,d\mu<\infty\right\} \\ M_\theta &= \left\{f:\Sigma\to\mathbb{R},\ \forall a>0, \int_\Sigma \theta\Big(\frac{f}{a}\Big)\,d\mu<\infty\right\} \end{aligned}$$

where μ-almost everywhere equal functions are identified. Consider the following Luxemburg norm on L_θ:

$$\|f\|_\theta = \inf\left\{a>0, \quad \int_\Sigma \theta\Big(\frac{f}{a}\Big)\,d\mu\le 1\right\} \tag{4.1}$$

$(L_\theta,\|.\|_\theta)$ is a Banach space called the Orlicz space associated with θ. M_θ is a subspace of L_θ. If θ is a finite function, it is the closure of the space of step functions $\sum_{i=1}^n a_i \mathbb{1}_{A_i}$ under $\|\cdot\|_\theta$. For references, see [25]. Let θ^* be the convex conjugate of the Young function θ:

$$\theta^*(t)=\sup_{s\in\mathbb{R}}\{st-\theta(s)\}$$

As θ^* is also a Young function, one can consider the Orlicz space L_{θ^*}.
Hölder's inequality holds between L_θ and L_{θ^*}: For all $f\in L_\theta$ and $g\in L_{\theta^*}$,

$$fg\in L^1(\mu) \quad \text{and} \quad \int_\Sigma |fg|\,d\mu \le 2\|f\|_\theta\|g\|_{\theta^*} \tag{4.2}$$

By inequality (4.2), any g in L_{θ^*} defines a continuous linear form on L_θ for the duality bracket $\langle f,g\rangle=\int fg\,d\mu$. In the general case, the topological dual space of $(L_\theta,\|.\|_\theta)$ may be larger than L_{θ^*}. Nevertheless, we always have the following result:

Theorem 4.3. *Let θ be a finite Young function and θ^* its convex conjugate. The topological dual space of M_θ can be identified, using the previous duality bracket, with L_{θ^*}: $M'_\theta\simeq L_{\theta^*}$.*

For a proof of this result, see for instance [16, Section 4].

In the sequel, the relevant Young functions are τ and τ^* defined by:

$$\begin{aligned} \tau(x) &\overset{\Delta}{=} \exp(|x|)-|x|-1 \\ \tau^*(x) &= (|x|+1)\log(|x|+1)-|x| \ . \end{aligned}$$

4.2. **Variational representation of the rate function.**

Proposition 4.4. *For every $\nu\in D_{\mathcal{P}}$, we have*

$$I(\nu)=\sup_{G\in\mathcal{G}}\left\{\dot\nu(G)-\int_{[0,T]}\log\langle\exp(DG_t),\nu_t\rangle\,dt\right\} \ .$$

Proof. Let us take ν in $D_{\mathcal{P}}$. By (4.1), for any $G\in\mathcal{G}$, we have

$$\|DG\|_{\tau,\bar\nu}=\inf\left\{a>0; \int_{[0,T]\times\mathbb{N}} \tau(DG/a)\,d\bar\nu\le 1\right\} \ .$$

We define $K(\nu) \stackrel{\Delta}{=} \sup_{G\in\mathcal{G}} \mathcal{K}(G,\nu) = \sup_{G\in\mathcal{G}} \left\{ \dot{\nu}(G) - \int_{[0,T]} \log\langle \exp(DG_t), \nu_t\rangle \, dt \right\}$, so that for any $a > 0$ and $G \in \mathcal{G} : \dot{\nu}(G/a) \le K(\nu) + \int_{[0,T]} \log\langle \exp(DG_t/a), \nu_t\rangle \, dt$. Subtracting $\int_{[0,T]} \langle \nu_t, DG_t/a\rangle \, dt$ from both sides:

$$\begin{aligned} \dot{\nu}(G) - \int_{[0,T]} \langle \nu_t, DG_t/a\rangle \, dt &\le K(\nu) + \int_{[0,T]} [\log\langle \exp(DG_t/a), \nu_t\rangle - \langle DG_t/a, \nu_t\rangle] \, dt \\ &\stackrel{(a)}{\le} K(\nu) + \int_{[0,T]} \langle \exp(DG_t/a) - DG_t/a - 1, \nu_t\rangle \, dt \\ &\stackrel{(b)}{\le} K(\nu) + \int_{[0,T]} \langle \tau(DG_t/a), \nu_t\rangle \, dt \ , \end{aligned}$$

where (a) comes from $\log x \le x - 1$, and (b) from

$$e^x - x - 1 \le \exp|x| - |x| - 1 = \tau(x). \tag{4.5}$$

Choosing $a = \|DG\|_{\tau,\bar{\nu}}$, we obtain:

$$\dot{\nu}(G) - \int_{[0,T]\times\mathbb{N}} DG \, d\bar{\nu} \le [K(\nu) + 1] \|DG\|_{\tau,\bar{\nu}} \ .$$

As an analogue inequality can be proved while replacing G by $-G$, and as by Hölder's inequality (4.2): $\left| \int_{[0,T]\times\mathbb{N}} DG \, d\bar{\nu} \right| \le 2\|\mathbb{1}\|_{\tau^*,\bar{\nu}} \|DG\|_{\tau,\bar{\nu}} = 2T\tau^*(1)\|DG\|_{\tau,\bar{\nu}}$, we have:

$$|\dot{\nu}(G)| \le [K(\nu) + 1 + 2T\tau^*(1)] \|DG\|_{\tau,\bar{\nu}} \ .$$

Let us first prove that if $K(\nu) < \infty$, then $K(\nu) = I(\nu)$. Now, take ν such that $K(\nu) < \infty$. The above estimate implies that for all F, G in $\mathcal{G}$, if $DF = DG$ then $\dot{\nu}(F) = \dot{\nu}(G)$. It also implies that $\dot{\nu}$ is a continuous linear form on $M_\tau(\bar{\nu})$. By the Riesz representation theorem (Theorem 4.3), there exists ℓ^ν in $L_{\tau^*}(\bar{\nu})$ such that

$$\dot{\nu}(G) = \int_{[0,T]} \langle \ell_t^\nu \, DG_t, \nu_t\rangle \, dt, \quad \forall G \in \mathcal{G} \ . \tag{4.6}$$

It follows that

$$\begin{aligned} K(\nu) &= \sup_{G\in\mathcal{G}} \left\{ \dot{\nu}(G) - \int_{[0,T]} \log\langle \exp(DG_t), \nu_t\rangle \, dt \right\} \\ &\stackrel{(a)}{=} \sup_{G\in\mathcal{G}} \left\{ \int_{[0,T]} \Big(\langle \ell_t^\nu DG_t, \nu_t\rangle - \log\langle \exp(DG_t), \nu_t\rangle \Big) \, dt \right\} \\ &\stackrel{(b)}{=} \int_{[0,T]} \sup_{g\in\ell^\infty} \left\{ \langle g, \ell_t^\nu \nu_t\rangle - \log\langle e^g, \nu_t\rangle \right\} \, dt \\ &\stackrel{(c)}{=} \int_{[0,T]} H(\ell_t^\nu \nu_t \mid \nu_t) \, dt \ . \end{aligned}$$

Equation (a) follows from (4.6). In identity (b), assuming $g = DG_t \in \ell^\infty$, [26, Theorem 2] is used to exchange the supremum and the integral (we face a convex conjugate of a convex integral functional). Equation (c) is the variational representation of the relative entropy [9, Lemma 6.2.13]. Notice that if $K(\nu) < \infty$, then equation (c) implies that dt-almost everywhere $\ell_t^\nu \nu_t$ defines a probability measure on $\mathbb{N}$ which is absolutely continuous with respect to ν_t.

Because of Proposition 2.11, to complete the proof of the identity $K(\nu) = I(\nu)$ when $K(\nu) < \infty$, it remains to check that the master equation (2.12) is satisfied.

Choosing $G_t^{\varphi,i}(j) = \varphi_t \mathbb{1}_{\{i\}}(j)$ where φ is continuously differentiable and $\varphi_T = 0$, with (3.1) we have:

$$\begin{aligned}\dot{\nu}(G^{\varphi,i}) &= -\langle G_0^{\varphi,i}, \nu_0\rangle - \int_{[0,T]} \langle \dot{G}_t^{\varphi,i}, \nu_t\rangle \, dt \\ &= -\varphi_0 \nu_0(i) - \int_{[0,T]} \dot{\varphi}_t \nu_t(i) \, dt\end{aligned}$$

and equation (4.6) leads us to

$$\begin{aligned}\dot{\nu}(G^{\varphi,i}) &= \int_{[0,T]\times\mathbb{N}} \varphi D\mathbb{1}_{\{i\}} \ell^\nu \, d\bar{\nu} \\ &= \int_{[0,T]} \varphi_t [\ell_t^\nu(i-1)\nu_t(i-1) - \ell_t^\nu(i)\nu_t(i)] \, dt.\end{aligned}$$

Since these identities hold for all φ and i, the master equation (2.12) is satisfied and we have shown that $K(\nu) = I(\nu)$ whenever $K(\nu) < \infty$.

Finally if $I(\nu) < \infty$, by Proposition 2.11 we have (2.12) which is equivalent to (4.6) by the above-described computation. Now, according to the computation following (4.6), we obtain that $K(\nu) = I(\nu)$. This completes the proof of the proposition. □

A simple corollary of Proposition 2.11 is the following:

Corollary 4.7. *If ν satisfies $I(\nu) < \infty$, then for all $0 \le s \le t \le T$:*

$$\|\nu_t - \nu_s\| \le 2(t-s) \ .$$

As a matter of fact, the effective domain of the rate function I is included in the compact subset of $D_{\mathcal{P}}$ mentioned in Lemma 3.3.

5. The lower bound

In this section we prove the following lower bound.

Proposition 5.1. *For any open measurable subset U of $D_{\mathcal{P}}$ we have*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}^n(X^n \in U) \ge -\inf_{\nu\in U} I(\nu) \ .$$

Following a classical pattern, the proof is carried out using (exponential) changes of measure. Without loss of generality, one can assume that $I(U) < \infty$. The n^{th} change of measure associated with a path $\nu \in U$ satisfying $I(\nu) < \infty$, is the twisted probability measure $\mathbb{Q}^{\nu,n}$ defined by (2.1) with $\lambda = \ell^\nu$, provided that ℓ^ν is regular enough. The changes of measure are used through the following device. For any $\epsilon > 0$, we have

$$\begin{aligned}\frac{1}{n}\log \mathbb{P}^n\left(X^n \in U\right) &= \frac{1}{n}\log \mathbb{E}_{\mathbb{Q}^{\nu,n}}\left(\frac{d\mathbb{P}^n}{d\mathbb{Q}^{\nu,n}} \mathbb{1}_{\{X^n\in U\}}\right) \qquad (5.2)\\ &\ge \inf_{\xi\in V} \frac{1}{n}\log\frac{dP^n}{dQ^{\nu,n}}(\xi) + \frac{1}{n}\log \mathbb{Q}^{\nu,n}\left(X^n\in V\right) \\ &\ge \frac{1}{n}\log\frac{dP^n}{dQ^{\nu,n}}(\nu) - \epsilon - \epsilon\end{aligned}$$

for any small enough neighborhood V of ν with $V \subset U$ and any large enough n, provided that $\mathbb{Q}^{\nu,n}$ satisfies the three following properties:

- (α) For any open neighborhood V of ν, $\lim_n \frac{1}{n}\log \mathbb{Q}^{\nu,n}(X^n \in V) = 0$.
- (β) P^n is absolutely continuous with respect to $Q^{\nu,n}$.
- (γ) The map $\xi \in V \mapsto \frac{dP^n}{dQ^{\nu,n}}(\xi)$ is continuous at ν.

Here and in the sequel, P^n and $Q^{\nu,n}$ stand for the probability laws on $D_{\mathcal{P}}$ of X^n under $\mathbb{P}^n$ and $\mathbb{Q}^{\nu,n}$.

The sole finiteness of $I(\nu)$ does not ensure that these three properties are satisfied. Therefore we will focus on a subset of the effective domain of I : $\mathcal{D}_I$, that will be called the set of nice ν's.

5.1. **The nice ν's.** Property (α) will follow from the law of large numbers in Proposition 2.3. Property (β) will be easily checked if bounds are imposed on ν. Property (γ) will be checked if ν and ℓ^ν are sufficiently regular. We will assume that ν is nice according to the following definition.

Definition 5.3. *A path $\nu \in D_{\mathcal{P}}$ is said to be* nice *if*

(1) *ν belongs to $\mathcal{D}_I$: i.e. $I(\nu) < \infty$*
(2) *For all $t > 0$ and $i \in \mathbb{N}$, $\nu_t(i) > 0$.*
(3) *There exists $M \geq 0$ such that for all $i \geq M$ and all $0 \leq t \leq T$, $\ell^\nu_t(i) = 1$ and there exists $\beta > 0$ such that for all $t > 0$ and $i \in \mathbb{N}$, $\ell^\nu_t(i) \geq \beta$.*
(4) *For all $i \in \mathbb{N}$, $\ell^\nu(i)$ is C^2 with respect to t.*

Under condition 2, formula (2.14) allows to determine $\ell^\nu_t(i)$ as a function of ν for all i, while condition 4 allows its determination for all $t > 0$ and not only dt-almost everywhere.

Let ν be nice, the n^{th} *twisted probability measure* associated with it, is $\mathbb{Q}^{\nu,n}$ defined by (2.1) with $\lambda = \ell^\nu$. For all $n \geq 1$, under $\mathbb{Q}^{\nu,n}$, X^n is still a Markov process (as pointed out in Section 2.2) .
The main property of ν and $\mathbb{Q}^{\nu,n}$ when ν is nice, is stated in the following lemma.

Lemma 5.4 (Nice ν's are really nice.)**.** *For any nice ν,*

$$\sup_V \liminf_n Q^{\nu,n}\text{-}essinf_{\xi \in V}\left(\frac{1}{n}\log \frac{dP^n}{dQ^{\nu,n}}(\xi)\right) \geq -I(\nu),$$

where the supremum is taken over all open measurable neighborhoods V of ν.

This lemma together with (5.2) leads to the desired lower bound for any open measurable neighbourhood of any nice ν. In order to extend this result to the general case, the following density result will be needed.

Lemma 5.5 (Nice ν's are dense)**.** *For each ν in $\mathcal{D}_I$, there exists a sequence $(\nu_m)_{m\geq 1}$ of nice sample paths such that $\lim_{m\to\infty} \sup_{0\leq t\leq T} \|\nu - \nu_m\| = 0$ and $\lim_{m\to\infty} I(\nu_m) = I(\nu)$.*

The proofs of these lemmas are postponed after the proof of Proposition 5.1, at Sections 5.4 and 5.5.

5.2. **Proof of the lower bound.** With Proposition 2.3, Lemma 5.4 and Lemma 5.5 in hand, we can give a proof of lower bound.

Proof of Proposition 5.1. Let U be any open measurable subset of $D_{\mathcal{P}}$ with $I(U) < \infty$. For any $\epsilon > 0$, let $\nu_* \in U$ be such that $I(\nu_*) < I(U) + \epsilon$. By Lemma 5.5, there exists a sequence of nice sample paths $(\nu_m)_{m\geq 1}$ converging to ν_* in U such that $I(\nu_m)$ converges towards $I(\nu_*)$. Hence, there exists a nice ν in U such that

$I(\nu) < I(\nu_*) + \epsilon < I(U) + 2\epsilon$. Let $Q^{\nu,n}$ be the n^{th} twisted probability law of X^n associated with ν, then

$$\begin{aligned} & \liminf_n \frac{1}{n} \log \mathbb{P}^n(X^n \in U) \\ \overset{(a)}{\geq} \quad & \sup_{V\,:\,\nu \in V \subset U} \left[\liminf_n Q^{\nu,n}\text{-essinf}_{\xi \in V} \left(\frac{1}{n} \log \frac{dP^n}{dQ^{\nu,n}}(\xi) \right) + \liminf_n \frac{1}{n} \log Q^{\nu,n}(V) \right] \\ \overset{(b)}{\geq} \quad & \sup_{V : \nu \in V \subset U} \liminf_n \, Q^{\nu,n}\text{-essinf}_{\xi \in V} \left(\frac{1}{n} \log \frac{dP^n}{dQ^{\nu,n}}(\xi) \right) \\ \overset{(c)}{\geq} \quad & -I(\nu) \\ \geq \quad & -I(U) - 2\epsilon \ . \end{aligned}$$

where (a) follows from (5.2), (b) follows from Proposition 2.3 and (c) from Lemma 5.4. □

5.3. **Proof of Proposition 2.3.** The argument relies on the diffusion approximation techniques due to Kurtz (see [10] and references therein). As limiting distributions are degenerate, it is enough to rely on the following lemma which is an immediate consequence of Corollary 4.2 in [10, page 355].

Lemma 5.6. *Let Y^n be a sequence of $\mathbb{R}^d$-valued Markov chains with initial condition distributed according to μ, if the first-order differential equation $\frac{dy}{dt} = b(y,t)$ has a unique solution in $C^1([0,T], \mathbb{R}^d)$ for any initial condition y_0 in the support of μ, then if*

$$i) \quad \lim_n \sup_{\|y\|<r, t<T} n\mathbb{P}\{|Y^n_{k+1} - Y^n_k| > \epsilon \mid Y^n_k = y, k = \lfloor nt \rfloor\} = 0$$

and if the sequences

$$b_n(y,t) \triangleq n\mathbb{E}[Y^n_{k+1} - Y^n_k \mid Y^n_k = y, k = \lfloor tn \rfloor]$$

and

$$a_n(y,t) \triangleq n\mathbb{E}[(Y^n_{k+1} - Y^n_k)^\dagger (Y^n_{k+1} - Y^n_k) \mid Y^n_k = y, k = \lfloor tn \rfloor]$$

satisfy for each r and T:

$$ii) \quad \lim_n \sup_{\|y\| \leq r, t \leq T} |a_n(y,t)| = 0$$

and

$$iii) \quad \lim_n \sup_{\|y\| \leq r, t \leq T} |b_n(y,t) - b(y,t)| = 0$$

then the sequence of piecewise constant processes $t \mapsto Y^n_{\lfloor nt \rfloor}$ converges in distribution towards the process Y such that Y_0 is distributed according to μ and for each y_0 in the support set of μ, if $Y_0 = y_0$, $Y.$ is the unique solution of $\frac{dy}{dt} = b(y,t)$ with initial condition y_0.

Proof of Proposition 2.3. In Lemma 3.3, it is stated that the sequence $(Q^{\nu,n})_{n\geq 1}$ is tight in $D_{\mathcal{P}}$. Using observation 2.2 on page 4, it is thus enough to check the three conditions of Lemma 5.6 for the d-dimensional projections of X^n for all $d > M$. Condition (i) is trivially enforced.

The limiting ordinary differential equation is:

$$\frac{d\xi_t(i)}{dt} = \frac{1}{\langle \ell_t, \xi_t \rangle}[\ell_t^\nu(i-1)\,\xi_t(i-1) - \ell_t^\nu(i)\xi_t(i)], \quad i \in \mathbb{N} \ .$$

Because ν is assumed to be nice, $\ell_t^\nu(i)$ is bounded from below by β. Hence the limiting differential equation satisfies the local Lipschitz condition and has a unique solution.

Now we have to check conditions ii) and iii). For all $i \in \mathbb{N}$: the i^{th} coordinate of $b_n(x,t)$ equals:

$$n\mathbb{E}_{\mathbb{Q}^{\nu,n}}\left[X^n_{t+\frac{1}{n}}(i) - X^n_t(i)\big|X^n_t = x\right] = \frac{1}{\langle \ell_t, x\rangle}[\ell_t^\nu(i-1)\,x(i-1) - \ell_t^\nu(i)\,x(i)] \ .$$

Hence condition iii) is enforced.

The conditional covariation matrix is symmetric tridiagonal. The diagonal terms satisfy:

$$n\mathbb{E}_{\mathbb{Q}^{\nu,n}}\left[\text{Cov}(X^n_{t+\frac{1}{n}} - X^n_t)[i,i] \mid X^n_t = x\right] = \frac{\ell_t(i)x(i) + \ell_t(i-1)x(i-1)}{n\langle \ell_t, x\rangle}$$

The off-diagonal terms satisfy:

$$n\mathbb{E}_{\mathbb{Q}^{\nu,n}}\left[\text{Cov}(X^n_{t+\frac{1}{n}} - X^n_t)[i,i+1] \mid X^n_t = x\right] = -\frac{\ell_t(i)x(i)}{n\langle \ell_t, x\rangle} \ .$$

The sum of the absolute values of the coefficients of $a_n(x,t)$ is upper bounded by $4/n$, which warrants condition ii). □

5.4. **Proof of Lemma 5.4.** Let us first prove two preliminary results stated in Lemmas 5.7 and 5.8.

In this section, ℓ^ν is related to ν through equation (2.14), $\mathbb{Q}^{\nu,n}$ is defined by (2.1) and $Q^{\nu,n}$ is the corresponding law of X^n. Let us first derive an alternative form of the log-likelihood $\frac{dQ^{\nu,n}}{dP^n}$. For any $\xi \in D_{\mathcal{P}}$, let us denote the distribution function of ξ_t by:

$$F_\xi(t,i) \triangleq \sum_{j\le i} \xi_t(j)$$

and

$$\begin{aligned} I^{\nu,n}(\xi) \quad &\triangleq \quad \sum_{i\ge 0} F_\xi(0,i)\log \ell_0^\nu(i) - \sum_{i\ge 0} F_\xi(T,i)\log \ell_T^\nu(i) \\ &\quad + \sum_{k=1}^{\lfloor nT\rfloor -1} \frac{1}{n}\sum_{i\ge 0} F_\xi(\frac{k}{n},i)\frac{\log \ell^\nu_{\frac{k}{n}}(i) - \log \ell^\nu_{\frac{k-1}{n}}(i)}{1/n} - \sum_{k=0}^{\lfloor nT\rfloor}\frac{1}{n}\log\langle \ell^\nu_{\frac{k}{n}}, \xi_{\frac{k}{n}}\rangle \ .\end{aligned}$$

Lemma 5.7. *For any nice ν, $\frac{1}{n}\log\frac{dQ^{\nu,n}}{dP^n}(X^n) = I^{\nu,n}(X^n)$.*

Proof. The result follows from

$$\begin{aligned}
&\log \frac{dQ^{\nu,n}}{dP^n}(X^n) \\
= \quad & \sum_{k=0}^{\lfloor nT \rfloor -1} \sum_{i=0}^{\infty} 1\!\!1_{\{S_k^n(B_{k+1}^n)=i\}} \log\left(\frac{\ell^\nu_{\frac{k}{n}}(i)}{\langle \ell^\nu_{\frac{k}{n}}, X^n_{\frac{k}{n}}\rangle}\right) \\
\overset{(a)}{=} \quad & \sum_{k=0}^{\lfloor nT \rfloor -1} \sum_{i=0}^{\infty} \log \ell^\nu_{\frac{k}{n}}(i) \sum_{j\le i} n[X^n_{\frac{k}{n}}(j) - X^n_{\frac{k+1}{n}}(j)] - \sum_{k=0}^{\lfloor nT \rfloor -1} \log\langle \ell^\nu_{\frac{k}{n}}, X^n_{\frac{k}{n}}\rangle \\
\overset{(b)}{=} \quad & n\sum_{i\ge 0} \log \ell^\nu_0(i) \sum_{j\le i} X^n_0(j) - n\sum_{i\ge 0}\log \ell^\nu_T(i)\sum_{j\le i} X^n_T(j) \\
& + n\sum_{k=0}^{\lfloor nT \rfloor -1} \frac{1}{n}\sum_{i\ge 0} \frac{\log \ell^\nu_{\frac{k}{n}}(i) - \log \ell^\nu_{\frac{k-1}{n}}(i)}{1/n} \sum_{j\le i} X^n_{\frac{k}{n}}(j) - \sum_{k=0}^{\lfloor nT \rfloor} \log\langle \ell^\nu_{\frac{k}{n}}, X^n_{\frac{k}{n}}\rangle \ ,
\end{aligned}$$

where (a) comes from the identity

$$1\!\!1_{\{S_k^n(B_{k+1}^n)=i\}} = n\sum_{j\le i}\left(X^n_{\frac{k}{n}}(j) - X^n_{\frac{k+1}{n}}(j)\right) \ .$$

and (b) is Abel's transformation. □

Let us now define for all $\xi \in D_{\mathcal{P}}$,

$$\begin{aligned}
I^\nu(\xi) \quad \triangleq \quad & \sum_{i\ge 0} F_\xi(0,i)\log \ell^\nu_0(i) - \sum_{i\ge 0} F_\xi(T,i)\log \ell^\nu_T(i) \\
& + \int_{[0,T]} [\sum_{i\ge 0} F_\xi(t,i)\partial_t \log \ell^\nu_t(i)]\, dt - \int_{[0,T]} \log\langle \ell^\nu_t, \xi_t\rangle\, dt \ .
\end{aligned}$$

Lemma 5.8. *For any nice ν, we have: $I^\nu(\nu) = I(\nu)$.*

Proof. For any nice ν, we have

$$\begin{aligned}
I^\nu(\nu) \quad = \quad & \sum_{i\ge 0} F_\nu(0,i)\log \ell^\nu_0(i) - \sum_{i\ge 0} F_\nu(T,i)\log \ell^\nu_T(i) \\
& + \int_{[0,T]} [\sum_{i\ge 0} F_\nu(t,i)\partial_t \log \ell^\nu_t(i)]\, dt - \int_{[0,T]} \log\langle \ell^\nu_t, \nu_t\rangle\, dt \\
\overset{(a)}{=} \quad & \sum_{i\ge 0}\Big[F_\nu(0,i)\log \ell^\nu_0(i) - F_\nu(T,i)\log \ell^\nu_T(i) + \int_{[0,T]} F_\nu(t,i)\partial_t \log \ell^\nu_t(i)\, dt\Big] \\
\overset{(b)}{=} \quad & \sum_{i\ge 0}\int_{[0,T]} -\partial_t F_\nu(t,i)\log \ell^\nu_t(i)\, dt \\
\overset{(c)}{=} \quad & \sum_{i\ge 0}\int_{[0,T]} \ell^\nu_t(i)\nu_t(i)\log \ell^\nu_t(i)\, dt \\
\overset{(d)}{=} \quad & \int_{[0,T]} \langle \ell^\nu_t \log \ell^\nu_t, \nu_t\rangle\, dt \\
\overset{(e)}{=} \quad & I(\nu)
\end{aligned}$$

where (a) follows from Fubini's theorem and $\log\langle \ell^\nu_t, \nu_t\rangle = 0$ for all t, by 2) in Proposition 2.11, (b) follows from an integration by parts (ℓ^ν is t-differentiable), (c) is a direct consequence of the definition of ℓ^ν, see (2.14), (d) is Fubini's theorem again and (e) is stated in Proposition 2.11. □

We are now in a position to prove Lemma 5.4

Proof of Lemma 5.4, nice νs are really nice. Let us prove that for any $\epsilon > 0$, there exists an open neighborhood V of ν such that:

$$\liminf_n \inf\{-I^{\nu,n}(\xi); \xi \in V\} \geq \inf\{-I^\nu(\xi); \xi \in V\} - \epsilon. \tag{5.9}$$

Recall that:

$$\begin{aligned} I^{\nu,n}(\xi) &= \sum_{i\geq 0} F_\xi(0,i)\log \ell^\nu_0(i) - \sum_{i\geq 0} F_\xi(T,i)\log \ell^\nu_T(i) \\ &\quad + \sum_{k=1}^{\lfloor nT\rfloor -1} \frac{1}{n}\sum_{i\geq 0} F_\xi(\frac{k}{n},i)\frac{\log \ell^\nu_{\frac{k}{n}}(i) - \log \ell^\nu_{\frac{k-1}{n}}(i)}{1/n} - \sum_{k=0}^{\lfloor nT\rfloor} \frac{1}{n}\log\langle \ell^\nu_{\frac{k}{n}}, \xi_{\frac{k}{n}}\rangle \ . \end{aligned}$$

The third summand on the right-hand side may be decomposed as $A+B+C$ with :

$$\begin{aligned} A &= \sum_{k=1}^{\lfloor nT\rfloor -1} \frac{1}{n}\sum_i F_\xi(\frac{k}{n},i)\left[\frac{\log \ell^\nu_{\frac{k}{n}}(i) - \log \ell^\nu_{\frac{k-1}{n}}(i)}{1/n} - \partial_t \log \ell^\nu_{\frac{k-1}{n}}(i)\right] \\ B &= \sum_{k=1}^{\lfloor nT\rfloor -1} \frac{1}{n}\sum_i \left[F_\xi(\frac{k}{n},i) - F_\nu(\frac{k}{n},i)\right]\partial_t \log \ell^\nu_{\frac{k-1}{n}}(i) \\ C &= \sum_{k=1}^{\lfloor nT\rfloor -1} \frac{1}{n}\sum_i F_\nu(\frac{k}{n},i)\,\partial_t \log \ell^\nu_{\frac{k-1}{n}}(i). \end{aligned}$$

- Control of A. Thanks to 4. in Definition 5.3,

$$\left|\frac{\log \ell^\nu_{\frac{k}{n}}(i) - \log \ell^\nu_{\frac{k-1}{n}}(i)}{1/n} - \partial_t \log \ell^\nu_{\frac{k-1}{n}}(i)\right| \leq \frac{K}{n}$$

where K stands for any non-negative constant in this proof. As $|F_\xi| \leq 1$, $|A| \leq K/n$.
- Control of B. Thanks to 4. in Definition 5.3, $\sup_{t,i} |\partial_t \log \ell^\nu_t(i)| \leq K$, and it is possible to find an open neighborhood V of ν such that $\sup_{t,i} |F_\xi(t,i) - F_\nu(t,i)| \leq \epsilon$ for all ξ in V. Therefore, $|B| \leq K\epsilon$.
- Control of C. As a Riemann series (note that $\partial_t \log \ell^\nu_t$ is continuous thanks to 4. in Definition 5.3), $\lim_n C = \int_0^T \sum_i F_\nu(t,i)\partial_t \log \ell^\nu_t(i)\,dt$.

In order to control the fourth summand of the right-hand side of $I^{\nu,n}(\xi)$, note that for all $0 \leq t \leq T$, $|\langle \ell^\nu_t, \nu_t\rangle - \langle \ell^\nu_t, \xi_t\rangle| \leq \|\nu_t - \xi_t\|$. Indeed, thanks to 3. in

Definition 5.3:

$$\begin{aligned}\langle \ell_t^\nu, \xi_t \rangle &= \sum_{i<M} \ell_t^\nu(i)\xi_t(i) + \sum_{i\geq M} \ell_t^\nu(i)\,\xi_t(i) \\ &= \sum_{i<M} \ell_t^\nu(i)\nu_t(i) + \sum_{i\geq M} \xi_t(i) \\ &= 1 - \sum_{i\geq M} \ell_t^\nu(i)\,\nu_t(i) + \sum_{i\geq M} \xi_t(i) \\ &= 1 - \sum_{i\geq M} [\nu_t(i) - \xi_t(i)] \ .\end{aligned}$$

Therefore, for any $\epsilon > 0$, there exists an open neighborhood V of ν such that:

$$\sup_{\xi \in V} \left| \sum_{k=0}^{\lfloor nT \rfloor} \frac{1}{n} \left[\log\langle \ell^\nu_{\frac{k}{n}}, \xi_{\frac{k}{n}} \rangle - \log\langle \ell^\nu_{\frac{k}{n}}, \nu_{\frac{k}{n}} \rangle \right] \right| \leq K\epsilon \tag{5.10}$$

Combining the above arguments we have proved (5.9).

Note that (5.10) with the continuity of $\xi \mapsto F_\xi$ implies that I^ν is continuous at ν. Therefore:

$$\sup_V \inf\{-I^\nu(\xi); \xi \in V\} = -I^\nu(\nu) \tag{5.11}$$

Observing by Lemma 5.7 that $Q^{\nu,n}$ and P^n are mutually absolutely continuous measures and that :

$$Q^{\nu,n}\text{-essinf}_{\xi \in V} \left(\frac{1}{n} \log \frac{dP^n}{dQ^{\nu,n}}(\xi) \right) \geq \inf\{-I^{\nu,n}(\xi); \xi \in V\} \ ,$$

one completes the proof of the lemma combining this inequality, (5.9), (5.11) and Lemma 5.8. □

5.5. **Proof of Lemma 5.5, Nice ν's are dense.** To prove Lemma 5.5, we define two parametrized regularization procedures: see (5.12) and (5.19). Their relevant properties are stated in Lemmas 5.13 and 5.20. The proof of Lemma 5.5 which is a straightforward consequence of these preliminary results, is postponed after their proofs, at the end of this section.

The first regularization proceed by time-extension, mixing, and convolution by the following kernel

$$\zeta^\epsilon(s) \triangleq \begin{cases} \frac{2}{\epsilon}(1 - \frac{s}{\epsilon}) & \text{for } 0 \leq s \leq \epsilon \\ 0 & \text{otherwise} \end{cases}$$

Remember that $\mathbf{p}_t$ denotes the Poisson distribution with parameter t, it satisfies (2.5).

As the convolution by the regularizing kernel ζ^ϵ depends on sample paths up to time $T + \epsilon$, we first introduce a time-extension $\tilde{\nu}$ of ν. For any ν in $\mathcal{D}_I$, let $\tilde{\nu}$ be defined by:

$$\tilde{\nu} \triangleq \begin{cases} \tilde{\nu}_t = \nu_t & \text{for } t \in [0,T] \\ \frac{d\tilde{\nu}_t}{dt}(i) = \tilde{\nu}_t(i-1) - \tilde{\nu}_t(i) & \text{for } t > T \quad \text{and} \quad i \geq 0 \end{cases}$$

and let $\nu_t^{\alpha,\epsilon}$ be defined for all $t \geq 0$ by:

$$\nu^{\alpha,\epsilon} \triangleq \zeta^\epsilon * \nu^\alpha, \qquad \text{where} \qquad \nu_t^\alpha \triangleq (1-\alpha)\tilde{\nu}_t + \alpha \mathbf{p}_t \tag{5.12}$$

that is: $\nu_t^{\alpha,\epsilon} = \int_0^\infty \zeta^\epsilon(s)\nu_{t+s}^\alpha \, ds$.
The paths $\tilde{\nu}$, ν^α and $\nu^{\alpha,\epsilon}$ belong to $D([0,\infty),\mathcal{P}(\mathbb{N}))$. Notice that the restriction of $\nu^{\alpha,\epsilon}$ to $[0,T]$ satisfies condition 2 in the Definition 5.3 of nice ν's and that $\nu_\cdot^{\alpha,\epsilon}(i)$ and $\ell_\cdot^{\nu^{\alpha,\epsilon}}(i)$ are infinitely differentiable with respect to t for all i (see (2.14)). Let $\tilde{I}$ be defined by:

$$\tilde{I}(\tilde{\nu}) \triangleq \int_{[0,\infty)} H(v_t^{\tilde{\nu}} \mid \tilde{\nu}_t)\, dt \ ,$$

where $v_t^{\tilde{\nu}}$ is defined by (2.7). The following lemma summarizes the main properties of the regularized sample path $\nu^{\alpha,\epsilon}$.

Lemma 5.13. *If $I(\nu) < \infty$, the following statements hold:*

$$\tilde{I}(\nu^\alpha) \leq I(\nu), \tag{5.14}$$

$$\tilde{I}(\tilde{\nu}(\cdot + s)) \leq I(\nu), \quad \forall s \geq 0, \tag{5.15}$$

$$\tilde{I}(\nu^{\alpha,\epsilon}) \leq \tilde{I}(\nu^\alpha), \tag{5.16}$$

$$\lim_{\alpha\downarrow 0,\epsilon\downarrow 0} \sup_{t\leq T} \|\nu_t^{\alpha,\epsilon} - \nu_t\| = 0 \tag{5.17}$$

$$\lim_{\alpha\downarrow 0,\epsilon\downarrow 0} \tilde{I}(\nu^{\alpha,\epsilon}) = I(\nu) \ . \tag{5.18}$$

Proof. Note that as $\mathbf{p}_t(i) > 0$ for all $t > 0, i \in \mathbb{N}$, for all $\alpha > 0$ we have $\nu_t^\alpha(i) > 0$, $\nu_t^{\alpha,\epsilon}(i) > 0$, and thus $\ell_t^{\alpha,\epsilon}(i)$ is uniquely defined by (2.14).
As $\tilde{\nu}_t = \nu_t$ and $\ell_t^{\tilde{\nu}} = \ell_t^\nu$ for $t \leq T$ and $\ell_t^{\tilde{\nu}} = 1$ for $t > T$, we have:

$$\tilde{I}(\tilde{\nu}) = I(\nu) + \int_{[T,\infty)} \langle 1 \log 1, \tilde{\nu}_s\rangle \, ds = I(\nu).$$

The convexity of $\tilde{I}$ implies the inequalities (5.14) and (5.16).
The same remarks imply that time-shifting may only decrease the rate function, that is: (5.15).
The convergence (5.17) follows from

$$\begin{aligned}
& \sup_{t\leq T} \|\nu_t^{\alpha,\epsilon} - \nu_t\| \\
= \quad & \sup_{0\leq t\leq T} || \int_0^\infty ((1-\alpha)\tilde{\nu}_{t+s} + \alpha \mathbf{p}_{t+s})\, \zeta_s^\epsilon \, ds - \nu_t || \\
\leq \quad & \sup_{0\leq t\leq T} (1-\alpha) || \int_0^\infty (\tilde{\nu}_{t+s} - \tilde{\nu}_t)\zeta_s^\epsilon \, ds|| + \alpha || \int_0^\infty (\mathbf{p}_{t+s} - \tilde{\nu}_t)\zeta_s^\epsilon \, ds|| \\
\leq \quad & \sup_{0\leq t\leq T} (1-\alpha) || \int_0^\infty (\tilde{\nu}_{t+s} - \tilde{\nu}_t)\zeta_s^\epsilon \, ds|| + 2\alpha \quad (\text{as} \quad \|\tilde{\nu}_{t+s} - \tilde{\nu}_t\| \leq 2s) \\
\leq \quad & \int_0^\epsilon 2s\, \zeta_s^\epsilon ds + 2\alpha \\
\leq \quad & \epsilon + 2\alpha,
\end{aligned}$$

where the penultimate inequality is a consequence of Corollary 4.7.

Proof of (5.18). Now we identify $\nu^{\alpha,\epsilon}$ with its restriction to $[0,T]$. Inequalities (5.14) and (5.16) imply that $I(\nu^{\alpha,\epsilon}) \leq I(\nu)$ for all $\alpha, \epsilon > 0$. On the other hand, as I is lower semi-continuous, with (5.17), we obtain $\liminf_{\alpha,\epsilon} I(\nu^{\alpha,\epsilon}) \geq I(\nu)$. Hence, $\lim_{\alpha,\epsilon} I(\nu^{\alpha,\epsilon}) = I(\nu)$. □

The second regularization procedure operates directly on ℓ^ν. For any integer M, let us define Φ_M by $\Phi_M \ell^\nu_t(i) = \ell^\nu_t(i)$ for $i \leq M$, $t \geq 0$, and $\Phi_M \ell^\nu_t(i) = 1$ for all $i > M$ and $t \geq 0$. The associated sample path ν^M is defined by:

$$\dot{\nu}^M_t(i) = [\Phi_M \ell^\nu_t(i-1)]\nu^M_t(i-1) - [\Phi_M \ell^\nu_t(i)]\nu^M_t(i) \ . \tag{5.19}$$

Lemma 5.20. *If $I(\nu) < \infty$ and $\ell^\nu_\cdot(i)$ is t-continuous for all $i \geq 0$, the following statements hold:*

$$\lim_{M\to\infty} \sup_{t \leq T} \|\nu_t - \nu^M_t\| = 0 \tag{5.21}$$

$$\lim_{M\to\infty} I(\nu^M) = I(\nu) \ . \tag{5.22}$$

Proof. Thanks to the t-continuity of $\ell^\nu_\cdot(i)$ for all i, $\sup_{t,i} \Phi_M \ell^\nu_t(i) < \infty$. By construction, for any $i \leq M$, $\nu^M_t(i) = \nu_t(i)$ for all t. It follows that

$$\begin{aligned} & I(\nu) - I(\nu^M) \\ = & \int_{[0,T]} \left[\sum_{i \leq M} (\nu_t(i) - \nu^M_t(i))\ell^\nu(t,i) \log \ell^\nu(t,i) + \sum_{i>M} \nu_t(i)\ell^\nu_t(i) \log \ell^\nu_t(i) \right] dt \\ = & \int_{[0,T]} \left[\sum_{i>M} \nu_t(i)\ell^\nu_t(i) \log \ell^\nu_t(i) \right] dt. \end{aligned}$$

Letting M tend to infinity, by dominated convergence, the right-hand side vanishes and (5.22) is established.
Let us prove (5.21).

$$\begin{aligned} & \sum_{i\geq 0} |\nu^M_t(i) - \nu_t(i)| \\ = & \sum_i \left| \int_{[0,t]} (\dot{\nu}^M_s(i) - \dot{\nu}(i))\, ds \right| \\ = & \sum_{i>M} \left| \int_{[0,t]} (\nu^M_s(i-1) - \nu^M_s(i) - \ell^\nu_s(i-1)\nu_s(i-1) + \ell^\nu_s(i)\nu_s(i))\, ds \right| \\ \leq & \ 2\int_{[0,t]} \sum_{i>M} |\nu^M_s(i) - \nu_s(i))|\, ds + 2\sum_{i>M} | \int_{[0,t]} [\ell^\nu_s(i) - 1]\nu_s(i)|\, ds \\ \leq & \ 2\int_{[0,t]} \sum_{i\geq 0} |\nu^M_s(i) - \nu_s(i)|\, ds + h^M(t) \ , \end{aligned}$$

where $h^M(t) \triangleq 2\sum_{i>M} \int_{[0,t]} (\ell^\nu_s(i) + 1)\nu_s(i)\, ds$. Applying Gronwall's lemma:

$$\sum_{i\geq 0} |\nu^M_t(i) - \nu_t(i)| \leq h^M(t) + 2\int_0^t h^M(s) e^{2(t-s)}\, ds$$

Let us now remember that according to Dini's lemma, a sequence of continuous functions decreasing pointwise towards 0 on the compact interval $[0,T]$ is also uniformly convergent. Therefore to establish (5.21), it remains to note that h^M decreases pointwise to 0 as M tends to ∞. □

Proof of Lemma 5.5. Let ν be in $\mathcal{D}_I$. First apply the regularization (5.12). Then, apply the second regularization (5.19) to $\nu^{\alpha,\epsilon}$ for α, ϵ small enough. The resulting path is a nice path and the desired result follows from Lemmas 5.12 and 5.19. □

5.6. **What's in a large deviation.** The characterization of the LDP enables to investigate twisted allocation processes. Let us give a simple example. Assume that at time 0, all bins are empty and that at time t, $\ell_t(i) = \frac{\theta+i}{1+t}$, where $\theta > 0$. Let ν_t be the unique solution of the system:

$$\dot{\nu}_t(i) = \ell_t(i-1)\nu_t(i-1) - \ell_t(i)\nu_t(i) \ .$$

Little calculus establishes:

$$\nu_t(i) = \left(\frac{\theta}{t+\theta}\right)^\theta \frac{\Gamma(\theta+i)}{\Gamma(\theta)\Gamma(i+1)} \left(\frac{t}{t+\theta}\right)^i \ ,$$

which simplifies to

$$\nu_t(i) = \frac{1}{1+t}\left(\frac{t}{1+t}\right)^i,$$

when $\theta = 1$. Hence when $\theta = 1$, ν_t is a geometric distribution with expectation t., in the general case it is a the Pascal distribution of parameter $\theta, \frac{t}{1+t}$ (see for example [11] p. 166). The rate function is finite at ν,, for example for $\theta = 1$:

$$\begin{aligned} I(\nu) &= \int_0^T dt \left[\sum_i \frac{1+i}{(1+t)^2} \log\left(\frac{1+i}{1+t}\right)\left(\frac{t}{1+t}\right)^i\right] \\ &\le \int_0^T dt \left[\sum_i \frac{(1+i)^2}{(1+t)^3}\left(\frac{t}{1+t}\right)^i\right] \\ &\le \int_0^T dt \left[\frac{1+3t+2t^2}{(1+t)^2}\right] \\ &\le 2T - \ln(1+T). \end{aligned}$$

Hence, twisting the allocation probability may modify the integrability properties of the typical sample path. For example, if X is geometrically distributed then $X \log X$ does not have any more exponential moments.

6. Stronger topologies

One may wonder whether stronger topologies could be considered. Recall that X_t^n is very similar to the empirical measure of a Poisson random variable with parameter t. The latter satisfies the LDP with respect to the total variation distance with the good rate function $H(\cdot \mid \mathbf{p}_t)$. And it is not reasonable to think of test functions that could be larger than $i \log i$. As a matter of fact, if ν has finite relative entropy with respect to any Poisson distribution then $\langle i \log i, \nu\rangle < \infty$. But the distribution $\nu(i) \propto \frac{1}{i^2 \log^{2+\delta}(i)}$ with $\delta > 0$ has finite relative entropy with respect to any Poisson distribution although $\langle i \log^{1+\epsilon}(i), \nu\rangle = \infty$ as soon as $\epsilon \ge \delta$ (see [17] for approaches to extension of Sanov's theorem).

Let p be strictly larger than 1 and let $\mathcal{H}$ be the class of sequences defined by:

$$H = \{G = (G(i))_{i\in\mathbb{N}} : \ |G(0)| = 1 \ \text{ and for } i \ge 1, \ |DG(i]| \le \log^{\frac{1}{p}}(i)\}.$$

If $G \in \mathcal{H}$ then $|G(i)| \le i \log^{1/p}(i)$. In this section we consider the following metric on $D_{\mathcal{P}}$:

$$d_{\mathcal{H}}(\nu, \nu') \triangleq \sup_{s\in[0,T]} \sup_{G\in\mathcal{H}} \left(\langle G, \nu_s\rangle - \langle G, \nu'_s\rangle\right) \ . \tag{6.1}$$

Theorem 6.2. *The sequence $\big(X^n\big)$ satisfies the* LDP *with good rate function I on $D_{\mathcal{P}}$ equipped with the topology defined by the metric $d_{\mathcal{H}}$.*

The proof of Theorem 6.2 proceeds according to the following steps. The lower-semi-compactness of I under metric $d_{\mathcal{H}}$ and the exponential equivalence between (X^n) and the linearly interpolated process $(\widehat{X}^n)$ are established. Finally the exponential tightness of $(\widehat{X}^n)$ is established using an exponential Martingale argument. The theorem follows from the inverse contraction principle [9, theorem 4.2.4].

Lemma 6.3. *I is a good rate function under the topology induced by metric $d_{\mathcal{H}}$.*

Proof. The convexity and the lower-semi-continuity of I still hold. It is enough to prove that the finiteness of $I(\nu)$ (say $I(\nu) \le \alpha$ for some $\alpha > 0$), implies both an upper bound on the modulus of continuity under metric $d_{\mathcal{H}}$ and that there exists a compact set K_α of $\mathcal{P}(\mathbb{N})$ such that $\nu_t \in K_\alpha$.

Let $\nu \in D_{\mathcal{P}}$ be such that $I(\nu) < \infty$. For any $G \in \mathcal{H}$,

$$\begin{aligned}
\langle G, \nu_t - \nu_s\rangle &\overset{(a)}{=} \int_{[s,t]} \langle \ell_u^\nu DG, \nu_u\rangle\, du \\
&\overset{(b)}{=} \int_{[0,T]} \langle \mathbb{1}_{[s,t]} DG \ell_u^\nu, \nu_u\rangle\, du \\
&\overset{(c)}{\le} \left[\int_{[0,T]} \langle \mathbb{1}_{[s,t]}, \ell_u^\nu \nu_u\rangle du\right]^{1/q} \times \left[\int_{[0,T]} \langle |DG|^p, \ell_u^\nu \nu_u\rangle\, du\right]^{1/p} \\
&\overset{(d)}{\le} |t-s|^{1/q} \times \int_{[0,T]} \langle v, \ell_u^\nu \nu_u\rangle du
\end{aligned} \tag{6.4}$$

where we set $v(0) = 1$ and for $i \ge 1$, $v(i) = \log(i)$. Indeed, (a) follows from Proposition 2.11, (b) is immediate, (c) follows from Hölder's inequality and (d) from the definition of $\mathcal{H}$. Now

$$\begin{aligned}
\int_{[0,T]} \langle v, \ell_u^\nu \nu_u\rangle du &\overset{(a)}{\le} \int_{[0,T]} \langle (i-1)_{i\in\mathbb{N}}, \nu_u\rangle\, du + \int_{[0,T]} \langle \ell_u^\nu \log \ell_u^\nu, \nu_u\rangle\, du \\
&\overset{(b)}{\le} \langle (i), \nu_0\rangle\, T + \frac{T^2}{2} + I(\nu)\ \ ,
\end{aligned} \tag{6.5}$$

where (a) follows from the application of Young's inequality in the duality between τ and τ^*, i.e. $xy \le \tau(x) + \tau^*(y)$, and (b) from Proposition 2.11 again. Combining inequalities (6.4) and (6.5), we get:

$$\sup_{G\in\mathcal{H}} \langle G, \nu_t - \nu_s\rangle \le \big(\frac{T^2}{2} + \langle i, \nu_0\rangle\, T + I(\nu)\big)^{1/p} (t-s)^{1/q}\ \ . \tag{6.6}$$

To check the compact containment property under metric $d_{\mathcal{H}}$, it is enough to check that if $I(\nu) < \infty$ and $\langle \phi, \nu_0\rangle < \infty$ where $\phi(i) = (i \log i)$, then:

$$\langle (i\log i)_{i\in\mathbb{N}}, \nu_t\rangle \le \langle i \log i, \nu_0\rangle + I(\nu) + e(\frac{t^2}{2} + (1 + \langle i+1, \nu_0\rangle)t).$$

As $I(\nu) < \infty$, by Proposition 2.11:

$$\begin{aligned}\frac{d\langle\phi,\nu_t\rangle}{dt} &\overset{(a)}{=} \langle D\phi\,\ell^\nu_t,\nu_t\rangle\\ &\overset{(b)}{\leq} \sum \nu_t(i)\left(\ell^\nu_t(i)\log\ell^\nu_t(i) - (\ell^\nu_t(i)-1)\right)\\ &\quad+\sum\nu_t(i)\left(e^{D\phi(i)}-D\phi(i)-1\right)\\ &\quad+\sum\nu_t(i)\,D\phi(i)\\ &\overset{(c)}{\leq} \langle\phi(\ell^\nu_t),\nu_t\rangle + e\left(t+1+\langle i+1,\nu_0\rangle\right),\end{aligned}$$

where inequality (b) follows from Young's inequality in the duality between τ and τ^*, and inequality (c) follows from $D\phi(i) \leq 1+\log(i+1)$, $\langle\ell^\nu,\nu\rangle = 1$ and $\langle i+1,\nu_t\rangle = \langle i+1,\nu_0\rangle$.

Integration with respect to t finishes the proof. □

In the sequel, $\widehat{X}^n$ denotes the linearly interpolated version of X^n:

$$\widehat{X}^n_t \triangleq X^n_{\frac{\lfloor nt\rfloor}{n}} + \big(t-\frac{\lfloor nt\rfloor}{n}\big)\big(X^n_{\frac{\lceil nt\rceil}{n}} - X^n_{\frac{\lfloor nt\rfloor}{n}}\big) \ .$$

By Theorem 4.2.13 in [9], the exponential equivalence between X^n and $\widehat{X}^n$ warrants that $\widehat{X}^n$ satisfies the LDP with good rate function I under the topology induced by the total variation distance.

Lemma 6.7. *X^n and $\widehat{X}^n$ are exponentially equivalent under the topology defined by metric $d_{\mathcal{H}}$.*

Let us denote by $L^n(k) \triangleq S^n_{k-1}(B^n_k)$ the occupancy score at time $\frac{k-1}{n}$ of the bins where the k^{th} allocation takes place at time $\frac{k}{n}$.

Proof. As:

$$\sum_{i\geq 0} G(i)[X^n_{k/n}(i) - X^n_{(k+1)/n}(i)] = \frac{1}{n}DG(L^n(k) \leq \frac{1}{n}\log^{1/p}(L^n_k) \ ,$$

we have

$$\begin{aligned}\mathbb{P}\left[\sup_{t\leq T}\sup_{G\in\mathcal{H}}\langle G, X^n_t - \widehat{X}^n_t\rangle \geq \eta\right] &\leq \mathbb{P}\left[\sup_{k\leq nT}\frac{1}{n}\log L^n(k)\geq\eta\right]\\ &\leq nT\mathbb{P}\left[L^n_{\lfloor nT\rfloor}\geq e^{n\eta}\right] \ .\end{aligned}$$

But at time $T = k/n$, ess-sup $\left(S^n_{k-1}(B^n_k)\right)$ is smaller than nT. Hence:

$$\forall\eta,\,\forall n, e^{n\eta} > nT \Rightarrow \sup_{t\leq T}\sup_{G\in\mathcal{H}}\langle G, X^n_t - \widehat{X}^n_t\rangle < \eta.$$

So for all $\eta > 0$:

$$\lim\frac{1}{n}\log\mathbb{P}\left[\sup_{t\leq T}\sup_{G\in\mathcal{H}}\langle G, X^n_t-\widehat{X}^n_t\rangle\geq\eta\right] = -\infty \ .$$

□

It remains to show that $(\widehat{X}^n)$ also satisfies the LDP with the good rate function I with respect to the topology defined by (6.1). By the inverse contraction principle [9], it is enough to check the exponential tightness.

Lemma 6.8. *The sequence $(\widehat{X}^n)$ is exponentially tight under the topology induced by metric $d_{\mathcal{H}}$.*

In the proof, we will use the following inequality obviously obtained from Cauchy-Schwartz one. Let $q(i)$ be a probability on $\mathbb{N}$ having finite expectation. Then for some universal constant C we have,

$$\sum_{i\geq 0} q(i)^{3/4} \leq C\big(\sum_{i\geq 0} i\, q(i)\big)^{1/2} < \infty \ . \tag{6.9}$$

Proof. Indeed, in view of Lemma 6.3, it is enough to check that for any $\alpha > 0$,

$$\limsup_n \frac{1}{n} \log \mathbb{P}\left\{I(\widehat{X}^n) > \alpha\right\} < 0 \ .$$

Note first that:

$$I(\widehat{X}^n) = \sum_{k=0}^{\lfloor nT \rfloor} \frac{1}{n} \log \frac{1}{X_{k-1}^n(L_k^n)}$$

Denote by Z_m the following quantity:

$$Z_m \triangleq \prod_{k=0}^{m} \frac{[\widehat{X}_{k-1}^n(S_{k-1}^n(B_k^n))]^{-1/4}}{\sum_{i\geq 0}[\widehat{X}_{k-1}^n(i)]^{3/4}} \ .$$

One may check that Z_m is an $\mathcal{A}_m$-martingale. This entails:

$$\mathbb{E}_{\mathbb{P}^n}\left[\exp\left(\frac{n}{4} I(\widehat{X}^n) - \sum_{k=1}^{\lfloor nt \rfloor} \log[\sum_{i\geq 0}(X^n_{\frac{k-1}{n}}(i))^{3/4}]\right)\right] = \mathbb{E}_{\mathbb{P}^n}[Z_1] \ . \tag{6.10}$$

Now

$$Z_1 = \frac{[X_0^n(S_0^n(B_1^n))]^{-1/4}}{\sum_i [X_0^n(i)]^{3/4}} \leq [X_0^n(S_0^n(B_1^n))]^{-1/4} = [\nu_0(S_0^n(B_1^n))]^{-1/4} \ ,$$

thus using the initial remark:

$$\mathbb{E}_{\mathbb{P}^n}[Z_1] \leq \sum_i [\nu_0(i)]^{3/4} < \infty.$$

On the other hand, by (6.9) and recalling $\langle (i)_{i\in\mathbb{N}}, \widehat{X}_T^n\rangle \leq \langle (i)_{i\in\mathbb{N}}, \widehat{X}_0^n\rangle + T$

$$\sum_{i\geq 0} [\widehat{X}_T^n(i)]^{3/4} \leq C(\langle i, \widehat{X}_0^n\rangle + T)^{1/2} \triangleq K,$$

thus:

$$Z_{\lfloor nT \rfloor} \geq K^{-\lfloor nT \rfloor} \prod_{k=1}^{\lfloor nT \rfloor} [X_{k-1}^n(S_{k-1}^n(B_k^n))]^{-1/4} \ . \tag{6.11}$$

Finally:

$$\begin{aligned}
\mathbb{P}\left\{I(\widehat{X}^n) \geq \alpha\right\} &= \mathbb{P}\left\{K^{-\lfloor nT \rfloor} \exp(\frac{n}{4} I(\widehat{X}^n)) \geq K^{-\lfloor nT \rfloor} \exp(\frac{n\alpha}{4})\right\} \\
&\leq \mathbb{P}\left\{Z_{\lfloor nT \rfloor} \geq K^{-\lfloor nT \rfloor} \exp(\frac{n\alpha}{4})\right\} \\
&\leq \mathbb{E}_{\mathbb{P}^n}[Z_1]\, K^{\lfloor nT \rfloor} \exp(-\frac{n\alpha}{4})
\end{aligned}$$

As $\mathbb{E}_{\mathbb{P}^n} Z_1 < \infty$,

$$\limsup_n \frac{1}{n} \log \mathbb{P}\left\{ I(\widehat{X}^n) > \alpha \right\} \leq - \left[-T \log K + \frac{\alpha}{4} \right]$$

which is negative for sufficiently large α. □

7. An application to random graphs

The charm of probabilistic approaches pertains partly to their robustness: the validity of a limit theorem often extends to models that are variants of the original model. This is illustrated in this section: Theorem 2.9 is used to characterize the Large Deviations of the degree sequence of sparse random graphs.

In the Erdös-Rényi $\mathcal{G}(n, \lfloor tn \rfloor)$ model for random graphs, $\lfloor tn \rfloor$ edges are inserted at random among n vertices. When t remains fixed while n tends to infinity, the model deals with sparse random graphs (with average degree $2t$). The degree of vertex i after $k = \lfloor nt \rfloor$ edge insertions is denoted $U_i^n(k)$. Any (random) graph defines an empirical probability measure $V^n(k)$ on $\mathbb{N}$:

$$V_t^n \triangleq \frac{1}{n} \sum_{i=1}^{n} \delta_{U_i^n(\lfloor tn \rfloor)} \ .$$

This empirical measure is called the *degree distribution of the graph*, it is one of the fundamental objects of study in random graph theory [4].

If vertices are identified with bins and edge extremities with balls, the degree distribution may be regarded as a conditioned empirical occupancy measure. Several papers [21, 5, 7] actually describe the random graph generation process as a conditioned allocation process: balls are allocated by pairs, and two balls belonging to the same pair may not be allocated in the same bin (to prevent the formation of self-loops), and the pair of balls corresponding to edge k may not be allocated into a pair of bins i and i' which has already received a pair a balls previously.

The conditioning approach is fruitful when establishing upper bounds, but it runs into difficulties when trying to prove LDP. Hence, we will resort on coupling and exponential approximation arguments to derive an LDP for the degree distribution of sparse random graphs. We indeed establish the following LDP for the degree distribution:

Theorem 7.1. *In the $\mathcal{G}(n, \lfloor tn \rfloor)$ random graph model, the empirical degree distribution satisfies a* LDP *with the good rate function:*

$$I'(\mu) \triangleq \inf \left\{ I(\nu) : \quad \nu \in D_{\mathcal{P}}([0, 2t], \mathcal{P}(\mathbb{N})), \, \nu_{2t} = \mu \right\}$$

The theorem is a consequence of the following coupling lemma and Theorem in [9, theorem 4.2.13.].

Lemma 7.2. *There exists a sequence of probability spaces over which one may define a random variable $\left(Y_t^n\right)_{t \leq T}$ distributed like $\left(X_t^n\right)$ and another random variable $\left(W_t^n\right)_{t \leq T}$ distributed like $\left(V_t^n\right)_{t \leq T}$ and such that for any $\epsilon > 0$:*

$$\lim_n \frac{1}{n} \log \mathbb{P}\left\{ \sup_t \|Y_{2t}^n - W_t^n\| > \epsilon \right\} = -\infty \ .$$

Proof. The coupling space is defined as follows. After step k, $2k$ balls have been inserted into the n bins and k edges have been inserted among the n vertices. At step $k+1$, a couple of indices (i,i') is picked uniformly at random, a ball is inserted into bin i and another ball is inserted into bin i' (both bins may be identical). If $i \neq i'$ and if the edge $\{i,i'\}$ had not been inserted previously then the edge $\{i,i'\}$ is inserted, otherwise a new couple of indices is picked at random until the couple defines a new edge, then this edge is inserted into the random graph under construction.

Notice that the probability that the pair of bins that receive the two balls at step k differs from the pair of vertices adjacent to the k^{th} edge is equal to $\frac{1}{n}+(1-\frac{1}{n})\frac{2k}{n(n-1)} \leq \frac{1}{n}(1+2T)$. Let Δ^n denote the total number of steps with index less than nT at which the insertion in the random allocation process and the insertion in the graph construction process differ. It is worth noting that

$$\sup_{t\leq T} \|Y^n_{2t} - W^n_t\| \leq \frac{8\Delta^n}{n} \ . \tag{7.3}$$

As S^n is a sum of independent Bernoulli random variables with success probability $(\frac{1}{n}+(1-\frac{1}{n})\frac{2k}{n(n-1)})_{k\leq nT}$, $\mathbb{E}\,\Delta^n \leq T(1+T)$ and $\mathrm{Var}(\Delta^n) \leq T(1+2T)$. Now applying Bernstein's inequality, we get:

$$\mathbb{P}\left\{\Delta^n \geq T(1+T)+s\right\} \leq \exp -\left[\frac{s^2}{2(T(1+2T)+s/3)} \ .\right] \ . \tag{7.4}$$

The lemma follows by combining inequalities (7.4) and (7.3). □

Note that the full LDP does not follow immediately if the graph process is considered as a conditioned allocation process as in [21, 5, 7]: the conditioning event is not measurable with respect to the flow of empirical occupancy measure. For sparse random graphs, Theorem 7.1 complements the results reported in [18] where events with polynomially small probability are characterized.

Appendix A

While proving Lemma 3.7, the following convergence result has been used. It is probably a standard result, but we did not find references for it.

Theorem A.1. *Let X be a complete metric vector space and $(f_n)_{n\geq 1}$ be a sequence of numerical convex lower semicontinuous functions on X such that for all $x \in X$, $\sup_{n\geq 1} f_n(x) < \infty$.*
If $(f_n)_{n\geq 1}$ converges pointwise to f: $\lim_{n\to\infty} f_n(x) = f(x) \in \mathbb{R}$ for all $x \in X$, then it also converges uniformly on any compact subset.

Proof. The argument essentially relies on Baire's and Dini's theorems.

Let us denote $g_n = \sup_{k\geq n} f_k$. Under our assumptions, (g_n) is a sequence of convex finite lower semicontinuous functions on X. Since X is a complete metric vector space, a consequence of Baire's theorem is that any convex finite lower semicontinuous function on X is continuous, see [12, Thm. 9, p. 112]. Therefore, (g_n) is a nonincreasing sequence of continuous functions converging pointwise to $\limsup_n f_n = f$. By Dini's theorem, (g_n) converges uniformly on any compact subset of X. In particular, f is continuous in restriction to any compact subset.

Let us denote h_n the convex lower semicontinuous envelope of $\inf_{k\geq n} f_n$. It defines a nondecreasing sequence of finite convex lower semicontinuous functions on

X which converges pointwise to the convex lower semicontinuous envelope $\bar{f}$ of f. By Dini's theorem again, (h_n) converges uniformly on any compact subset to $\bar{f}$. But, we have already seen that in restriction to any compact subset, f is continuous. Hence, we have: $\bar{f} = f$ on any compact and (h_n) converges uniformly on any compact subset to f.

As $h_n \leq f_n \leq g_n$, we have proved that (f_n) converges uniformly on any compact subset. □

References

[1] A. Barbour, L. Holst and S. Janson. *Poisson Approximation,*, Clarendon Press, (1992).

[2] M.S. Bartlett. The characteristic function of a conditional statistic. *J. London Math. Soc.* **13**, 62-67 (1938).

[3] D. Barraez and S. Boucheron. A functional central limit theorem for empirical occupancy measures. Computer Science, Université Paris-Sud, Preprint (1998), `http://www.lri.fr/ bouchero/PUB/dara97.ps.gz`.

[4] Bollobàs, B. *Random Graphs*. Academic Press. (1985).

[5] Bollobàs, B. and Frieze, A. M. On matching and Hamiltonian cycles in random graphs. in *Random Graphs 83. Annals of Discrete Mathematics*, **28**, 23–46, (1985).

[6] S. Boucheron and D. Gardy. An urn model from learning theory. *Random Structures and Algorithms,* **1**, 43–67, (1997).

[7] V. Chvatal. Almost all graphs with 1.44 edges are 3 colourable. *Random Structures and Algorithms,* **2**, 11–28, (1991).

[8] A. Dembo, A. Vershik and O. Zeitouni. Large deviations for integer partitions, preprint, (1998).

[9] A. Dembo and O. Zeitouni. *Large Deviation Techniques and Applications*, Springer-Verlag, (1999).

[10] S. Ethier and T.G. Kurtz. *Markov processes, characterization and convergence*, J. Wiley, (1986).

[11] W. Feller. *Probability theory, I*, J. Wiley and sons (1968).

[12] J.R. Giles. Convex analysis with application in the differentiation of convex functions. In *Research Notes in Math.* **58**. Pitman, London, (1992).

[13] N.L. Johnson and S. Kotz. *Urn models and their application. An approach to modern discrete probability theory.* John Wiley & Sons (1977).

[14] V.F. Kolcin, B.A. Sevast'janov and V.P. Cistjakov. *Random allocations.* John Wiley & Sons. XI, (1978).

[15] C. Léonard. Large deviations for particle systems associated with spatially homogeneous Boltzmann type equations. *Probability Theory and Related Fields,* **101**, 1-44, (1985).

[16] C. Léonard. Minimizers of energy functionals under not very integrable constraints. *preprint,* (2000).

[17] C. Léonard and J. Najim. An extension of Sanov's theorem. Application to the Gibbs conditioning principle. *preprint, MODAL'X, Université Paris-X, number 6,* (2000).

[18] B.D. McKay and N.C. Wormald. The degree sequence of a random graph. *Random Structures and Algorithms,* **11**, 97–117, (1998).

[19] C. Morris. Central Limit Theorems for multinomial sums. *Annals of Statistics,* **3**, 165–168, (1975).

[20] Palka, Z. On the number of vertices of given degree in a random graph. *Journal of Graph Theory* **8,** 267–270, (1984).

[21] Pittel, B., Spencer J. and Wormald N. Sudden emergence of a giant k-core in a random graph. *Journal of Combinatorial Theory, Series B,* **67**, 111–151, (1996).

[22] M.P. Quine. A functional Central Limit Theorem for generalized occupancy numbers. *Stochastic Processes and Applications,* **9**, 109-115, (1979).

[23] M.P. Quine and J. Robinson. A Berry-Esseen bound for an occupancy problem. *Annals of Probability,* **10**, 663–671, (1982).

[24] M.P. Quine and J. Robinson. Normal approximations to sums of scores based on occupancy numbers. *Annals of Probability,* **12**, 794–804, (1984).

[25] R. Rao and R. Ren. *Theory of Orlicz spaces,* Marcel Dekker Inc., (1991).

[26] Rockafellar, R.T. (1968) Integrals which are convex functionals. *Pacific Journal of Mathematics,* **24**, 525–538.
[27] M. Sion. On general Minimax theorems. *Pacific Journal of Mathematics,* , , 171–176, (1958).
[28] S.R.S Varadhan. *Large deviations and applications,* SIAM, (1984).
[29] J.S. Vitter and P. Flajolet. Average-case analysis of algorithms and data structures. Van Leeuwen, Jan (ed.), *Handbook of theoretical computer science.* Vol. A., Elsevier Science Publishers. 431-524 (1990).

Contents

Laboratoire de Recherche en Informatique, UMR CNRS 8623, Bât. 490, Université Paris-Sud, 91405 Orsay cedex, France

Laboratoire de Statistiques et de Probabilités, UMR C5583, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse cedex 4

Modal-X, Université Paris 10, Bât. G., 200, avenue de la République. 92001 Nanterre Cedex France, and Ecole Polytechnique, CMAP, 91128 Palaiseau Cedex, France

E-mail address: Stephane.Boucheron@lri.fr
E-mail address: gamboa@cict.fr
E-mail address: Christian.Leonard@modalx.u-paris10.fr