Large Deviation Principle for General Occupancy Models

Jim X. Zhang*
Division of Applied Math
Brown University
Providence, R.I.02912
USA

Paul Dupuis†
Division of Applied Math
Brown University
Providence, R.I.02912
USA

December 1, 2004

Abstract

We use process level large deviation analysis to obtain the rate function for a general family of occupancy problems. Our interest is the asymptotics of the empirical distributions of various quantities (such as the fraction of urns that contain a given number of balls). In the general setting, balls are allowed to land in a given urn depending on the urn's contents prior to the throw. We discuss a parametric family of statistical models which includes *Maxwell-Boltzmann*, *Bose-Einstein* and *Fermi-Dirac* statistics as special cases. A process level large deviation analysis is conducted and the rate function for the original problem is then characterized, via the contraction principle, by the solution to a calculus of variations problem. We conjecture that the solution to the variational problem coincides with that of a finite dimensional minimization problem.

1 Introduction

Occupancy problems center on the distribution of r balls that have been thrown into n urns. In the simplest scenario each ball is equally likely to land in any of the urns, i.e., each ball is independently assigned to a given urn

*Research supported in part by the National Science Foundation (NSF-DMS-0306070).
†Research supported in part by the National Science Foundation (NSF-DMS-0306070 and NSF-DMS-0404806) and the Army Research Office (DAAD19-02-1-0425).

with probability $1/n$. In this case, we say that the urn model uses *Maxwell-Boltzmann* (MB) statistics. This model has been studied for decades and applied in diverse fields such as computer science, biology, and statistics. See [2, 6, 7] and the references therein. However, balls may also enter the urns in a nonuniform way. An important generalization is to allow the likelihood that the ball lands in a given urn to depend on its contents prior to the throw, as in *Bose-Einstein* (BE) and *Fermi-Dirac* (FD) statistics. See [7, 4, 9] and the references therein.

For MB statistics, many results have been obtained using "exact" methods. For example, combinatorial methods are used in [5] and methods that use generating functions are discussed in [7]. Although they do not directly involve approximations, the implementation of these methods can be difficult. For example, in combinatorial methods one has to deal with the difference of events using the inclusion-exclusion formula and the resulting computations can involve large errors. In the moment generating function approach in [7] similar difficulties occur.

Large deviations approximations give an attractive alternative to both of these approaches. One reason is that they offer good approximations with just modest computation. A second, perhaps more important reason, is that qualitative insights can be obtained. In [8] the LDP for the MB model is obtained, and the rate function exhibited in more-or-less explicit form.

In the present paper, we discuss a parametric family of statistical models, of which the previously mentioned MB, BE and FD statistics are all special cases. We assume there are n urns and that $\lfloor Tn \rfloor$ balls are thrown into them (where $\lfloor s \rfloor$ denotes the integer part of s), and analyze the asymptotic properties as n goes to ∞. A typical problem of interest is to characterize the large deviation asymptotics of the empirical distribution after all the balls are thrown. For example, one can wish to estimate the probability that at most half of the urns are empty after all the balls are thrown. A direct analysis of this problem is hard, and instead we lift the problem to the process level and analyze the large deviation asymptotics at this process level. Once the process level large deviation analysis is done, one can apply the Contraction Mapping Theorem to answer the original question. We conjecture that the variational problem that results from the contractions principle can in fact be solved explicitly (as was done in [8] for MB), and the formula is stated in Section 6.

Although process level large deviations are by now quite standard, there are several interesting features, both qualitative and technical, which distinguish occupancy models and place them outside the range of existing theory. The most significant of these as far as the proof is concerned are the singular transition rates that occur in the (Markovian) process level description of the model. We will use a weak convergence approach that is naturally suited to these problems and results in a nicely compact and self contained proof, and one that can easily accommodate further generalization of the model. A second very interesting feature is the previously mentioned possibility for explicit solutions to the variational problems that arise in the process level approximations.

The outline of the paper is as follows. In Section 2 the parametric family of occupancy problem is described in detail. A dynamical system characterization of the random occupancy process is given, and a representation for certain exponential integrals is given in terms of a "controlled" occupancy process. From this representation formula one can identify the large deviation rate function immediately. In Section 3 we prove the lower bound for the Laplace principle, which corresponds to the large deviation upper bound. In Section 4, the rate function I is studied more closely so as to deal the technical difficulty of the singular transition rates. In Section 5, we prove the upper bound for the Laplace principle which corresponds to the large deviation lower bound. Finally, in Section 6 we conjecture a simplified formula of the rate function for the process at a given fixed time.

2 Preliminaries and Main Result

In this section, we formulate the problem of interest and state the LDP. The proof is given in sections that follow. As described in the introduction, we focus on the asymptotic behavior of the general occupancy problem.

The general occupancy problem has the same structure as the Maxwell-Boltzmann occupancy problem, except that in the general problem urns are distinguished according to the number of balls contained therein. The full collection of models will be indexed by a parameter a. This parameter takes values in the set $(0, \infty] \cup \{-1, -2, \ldots\}$, and its interpretation is as follows. Suppose that a ball is about to be thrown into a fixed set of urns, and that any two urns (labeled say A and B) are selected. An urn is said to be of *category* i if it contains i balls. Suppose that urn A is of category i, while B is of category j. Then the probability that the ball is thrown into urn A, conditioned on the state of all the urns and that the ball is thrown into either urn A or B, is

$$\frac{a+i}{(a+i)+(a+j)}.$$

When $a = \infty$ we interpret this to mean that the two urns are equally likely.

Also, when $a < 0$ we use this ratio to define the probabilities only when $0 \leq i \vee j \leq -a$ and $i < -a$ or $j < -a$, so the formula gives a well defined probability. The probability that a ball is placed in an urn of category $-a$ is 0. Thus under this model, urns can only be of category $0, 1, \ldots - a$, and we only throw balls into categories $0, 1, \ldots - a - 1$.

In this setup, certain special cases are distinguished. The cases $a = 1$, $a = \infty$, $a = -I$ correspond to what are called *Bose-Einstein* statistics, *Maxwell-Boltzmann* statistics, and *Fermi-Dirac* statistics, respectively.

Suppose that before we throw a ball there are already tn balls in all the urns, and further suppose that the occupancy state is $(x_0, x_1, \ldots x_{I+})$. Here $x_i, i = 0, \ldots, I$ denotes the fraction of urns that contain i balls, and x_{I+} denotes the fraction containing more than I balls. Then the "un-normalized" or "relative" probability of throwing into a category i urn with $i \leq I$ is simply $(a + i)x_i$. Let us temporarily abuse notation, and let $x_{I+1}, x_{I+2}, \ldots$ denote the exact fraction in each category i with $i > I$. Since there are tn balls in the urns before we throw, $\sum_{i=0}^{\infty} ix_i = t$. Thus the (normalized and true) probability that the ball is placed in an urn that contains exactly i balls, $i = 0, \ldots I$, is $\frac{(a+i)x_i}{a+t}$, and the probability that the ball is placed in an urn that has more than I balls is $1 - \sum_{j=0}^{I} \frac{a+j}{a+t} x_j$.

An explicit construction of this process is as follows. To simplify, we assume the empty initial condition, i.e., all urns are empty. One can consider other initial conditions, with only simple notational changes in the results to be stated below. We introduce a time variable t that ranges from 0 to T. At a time t that is of the form l/n, with $0 \leq l \leq \lfloor nT \rfloor$ an integer, l balls have been thrown. Let $X^n(t) = \left\{ X_0^n(t), X_1^n(t), \ldots X_I^n(t), X_{I+}^n(t) \right\}$ be the *occupancy state* at that time. As noted previously, $X_i^n(t)$ denotes the fraction of urns that contain i balls at time t, $i = 0, \ldots I$, and $X_{I+}^n(t)$ the fraction of urns that contain more than I balls. The definition of X^n is extended to all $t \in [0, T]$ not of the form l/n by piecewise linear interpolation. Note that $X^n(t)$ is indeed a probability vector in $\mathbb{R}^{I+2}$. If

$$\mathcal{S}_I \doteq \left\{ x \in \mathbb{R}^{I+2} : x_i \geq 0, 0 \leq i \leq I + 1 \text{ and } \sum_{i=0}^{I+1} x_i = 1 \right\},$$

then for any $t \in [0, T]$, $X^n(t) \in \mathcal{S}_I$. Thus X^n takes values in $\mathcal{U} \doteq C\left([0, T], \mathcal{S}_I\right)$. We equip $\mathcal{U}$ with the usual supremum norm and on $\mathcal{S}_I$ we take the usual L_1 norm.

It will be convenient to work with the following "dynamical system" representation. For $x \in \mathbb{R}^{I+2}$ and $t \in [0, -a1_{\{a<0\}} + \infty 1_{\{a>0\}})$ define the vector $\rho(t,x) \in \mathbb{R}^{I+2}$ by

$$\rho_k(t,x) = \frac{a+k}{a+t} x_k, \qquad \text{for } k = 0, \ldots I, \tag{2.1}$$

and

$$\rho_{I+1}(t,x) = 1 - \sum_{k=0}^{I} \frac{a+k}{a+t} x_k.$$

A direct calculation shows that if

$$x \in \mathcal{S}_I \quad \text{ and } \quad \sum_{k=0}^{I+1} k x_k \leq t, \tag{2.2}$$

then $\rho(t,x)$ is indeed a probability vector in $\mathbb{R}^{I+2}$, i.e., $\rho(t,x) \in \mathcal{S}_I$. We can then define a family of independent random vector fields

$$\left\{ y^{i,n}(\cdot) : i = 0, 1, \ldots \lfloor nT \rfloor - 1, \lfloor nT \rfloor \right\}$$

that take values in

$$\Lambda \doteq \{ e_{j+1} - e_j, 0 \leq j \leq I \} \cup \{0\}$$

and with distributions

$$P\left\{ y^{i,n}(x) = v \right\} = \begin{cases} \rho_k\left(\frac{i}{n}, x\right) & \text{if } \ v = e_{k+1} - e_k \quad 0 \leq k \leq I \\ \rho_{I+1}\left(\frac{i}{n}, x\right) & \text{if } \ v = 0 \end{cases}.$$

Finally, we define $X^n(l/n)$ recursively by

$$X^n\left((i+1)/n\right) = X^n\left(i/n\right) + \frac{1}{n} y^{i,n}\left(X^n\left(i/n\right)\right),$$

and the initial condition $X^n(0) = (1, 0, \ldots 0)$. Observe that the increments $\left\{ y^{i,n}\left(X^n\left(i/n\right)\right)\right\}$ are conditionally distributed according to $\rho\left(\frac{i}{n}, X^n(i/n)\right)$, and thus the process X^n is obviously Markovian and will have the same distribution as the occupancy process described previously.

Often one is interested in the large deviations of the empirical occupancy measure at the terminal time T, namely $X^n(T)$. We study this by analyzing the large deviation properties of the whole process X^n and then using the Contraction Mapping Theorem. The Laplace formulation will be used to perform the process level analysis. Let F be any bounded and continuous function on $\mathcal{U}$. The processes X^n are said to satisfy the Laplace principle with rate function I if the following two conditions hold:

1. For each $K < \infty$, the set $\{\varphi \in \mathcal{U} : \ I(\varphi) \leq K\}$ is compact in $\mathcal{U}$.

2.
$$\lim_{n\to\infty} -\frac{1}{n} \log E \exp\left[-nF(X^n)\right] = \inf_{\varphi\in\mathcal{U}} \left[I(\varphi) + F(\varphi)\right].$$

Since X^n takes values in a Polish space, the notions of Laplace principle and large deviation principle are equivalent [3, Theorem 1.2.1] .

Define the $I+2$ by $I+2$ matrix

$$M = \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

Let $\varphi \in \mathcal{U}$ be given with $\varphi_0(0) = 1$. Suppose there is a Borel measurable function $\theta : [0,T] \longmapsto \mathcal{S}_I$ such that for any $t \in [0,T]$

$$\varphi(t) = \varphi(0) + \int_0^t M\theta(s)ds. \tag{2.3}$$

We interpret $\theta_i(s)$ as the rate at which balls are thrown into urns that contain i balls at time s. Moreover $\theta(s)$ is unique in the sense that if another $\tilde{\theta} : [0,T] \longmapsto \mathcal{S}_I$ satisfies (2.3) then $\tilde{\theta} = \theta$ a.e. on $[0,T]$. We call φ a *valid* occupancy state process if there exists $\theta : [0,T] \longmapsto \mathcal{S}_I$ satisfying (2.3). In this case θ is called the occupancy rate process associated with φ. It is easy to observe that if φ is valid then $\varphi(s)$ satisfies (2.2) for all $s \in [0,T]$. This shows that $\rho\left(s, \varphi(s)\right) \in \mathcal{S}_I$.

The relative entropy function will be used throughout the paper and we define it now. For two probability measures α and β on a Polish space $\mathcal{A}$, the relative entropy of α with respect to β is defined by

$$R\left(\alpha||\beta\right) \doteq \int_{\mathcal{A}} \left(\log \frac{d\alpha}{d\beta}\right) d\alpha$$

whenever α is absolutely continuous with respect to β (and with the convention that $0 \log 0 = 0$). In all other cases we set $R\left(\alpha||\beta\right) = \infty$. When two probability vectors ρ and $\nu \in \mathcal{S}_I$ appear in the relative entropy function, we interpret them as probability measures on the simplex $\{0, 1, \ldots, I, I+1\}$, and thus

$$R\left(\rho||\nu\right) \doteq \sum_{i=0}^{I+1} \rho_i \log \frac{\rho_i}{\nu_i}.$$

As observed before, when $\varphi(s)$ is *valid,* $\rho(s,\varphi(s)) \in \mathcal{S}_I$, which makes $R(\theta(s)||\rho(s,\varphi(s)))$ well defined. For such φ define

$$I(\varphi) = \int_0^T R(\theta(s)||\rho(s,\varphi(s)))ds. \tag{2.4}$$

If φ is not valid then define $I(\varphi) = \infty$. In the next three sections we will prove the urn models constructed in this section satisfy the Laplace principle with rate function I. In particular, in Section 3 we will prove

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\exp\left[-nF(X^n)\right] \geq \inf_{\varphi\in\mathcal{U}}\left[I(\varphi)+F(\varphi)\right],$$

and in Section 5 we will prove

$$\limsup_{n\to\infty} -\frac{1}{n}\log E\exp\left[-nF(X^n)\right] \leq \inf_{\varphi\in\mathcal{U}}\left[I(\varphi)+F(\varphi)\right].$$

These bounds are equivalent to the large deviation upper and lower bound [3]. In Section 4, we will prove several properties of the rate function I, and in particular show that I has compact level sets.

It will turn out that certain representation formulas for exponential integrals simplify proving the Laplace principle. Consider a controlled process $\bar{X}^n(t)$ constructed as follows. The process dynamics are of the same general structure as those of X^n, save that $y^{i,n}(X^n(i/n))$ is replaced by a sequence of controlled random vectors $\bar{y}^{i,n}$. Let $(\mathcal{V},\mathcal{A})$ be a measurable space and $\mathcal{Y}$ a Polish space and let $\tau(dy|x)$ be a family of probability measures on $\mathcal{Y}$ parameterized by $x \in \mathcal{V}$. We call $\tau(dy|x)$ a *stochastic kernel* on $\mathcal{Y}$ given $\mathcal{V}$ if for every Borel subset E of $\mathcal{Y}$ the function mapping $x \in \mathcal{V} \mapsto \tau(E|x) \in [0,1]$ is measurable. The conditional distributions of the controlled random vectors will be specified by a sequence $\left\{\nu^{i,n} : i = 0,1,\ldots,\lfloor nT\rfloor\right\}$, where each quantity $\nu^{i,n} = \nu^{i,n}(x_0,x_1,x_2,\ldots x_i)$ is interpreted as a stochastic kernel on Λ given $(\mathcal{S}_I)^{i+1}$. We call such a sequence $\left\{\nu^{i,n} : i = 0,1,\ldots,\lfloor nT\rfloor\right\}$ an *admissible* control sequence. Each control $\nu^{i,n}$ will give rise to a corresponding relative entropy term in the representation formula.

The controlled process is determined by

$$\begin{aligned} \bar{X}^n((i+1)/n) &= \bar{X}^n(i/n) + \frac{1}{n}\bar{y}^{i,n} \qquad \text{for } i = 0,1,\ldots\lfloor nT\rfloor \\ \bar{X}^n(0) &= (1,0,\ldots 0), \end{aligned}$$

where $\bar{y}^{i,n}$ has the conditional distribution $\nu^{i,n}\left(\bar{X}^n(0),\ldots,\bar{X}^n(i/n)\right)$. The random vectors $\bar{X}^n(i/n)$ and $\bar{y}^{i,n}$ are defined recursively in the following order:

$$\bar{X}^n(0), \bar{y}^{0,n}, \bar{X}^n(1/n), \bar{y}^{1,n}, \bar{X}^n(2/n), \ldots, \bar{X}^n(\lfloor nT \rfloor /n), \bar{y}^{\lfloor nT \rfloor ,n}.$$

For all $n \in \mathbb{N}$ the controlled random vectors $\bar{X}^n(i/n)$ and $\bar{y}^{i,n}$ are defined on a common probability space $\left(\bar{\Omega}, \bar{\mathcal{F}}, \bar{P}\right)$, and expectation on this space is denoted by $\bar{E}$. Define

$$\rho^{i,n} \doteq \rho\left(i/n, \bar{X}^n(i/n)\right),$$

where $\rho(t,x)$ was defined previously in (2.1) . Then by [3, Proposition 1.4.2] and the chain rule for relative entropy [3, Theorem C.3.1]

$$-\frac{1}{n}\log E\exp\left[-nF(X^n)\right] = \inf_{\{\nu^{i,n}\}} \bar{E}\left[F\left(\bar{X}^n\right) + \frac{1}{n}\sum_{i=0}^{\lfloor nT \rfloor} R\left(\nu^{i,n} || \rho^{i,n}\right)\right], \tag{2.5}$$

where the infimum is over all the admissible control sequences $\left\{\nu^{i,n}\right\}$.

3 The Large Deviation Upper Bound

In this section, we prove

$$\liminf_{n\to\infty} -\frac{1}{n}\log E\exp\left[-nF(X^n)\right] \geq \inf_{\varphi\in\mathcal{U}}\left[I(\varphi) + F(\varphi)\right],$$

which corresponds to the large deviation upper bound. By (2.5) it is enough to show that

$$\liminf_{n\to\infty}\inf_{\{\nu^{i,n}\}} \bar{E}\left[F\left(\bar{X}^n\right) + \frac{1}{n}\sum_{i=0}^{\lfloor nT \rfloor} R\left(\nu^{i,n} || \rho^{i,n}\right)\right] \geq \inf_{\varphi\in\mathcal{U}}\left[I(\varphi) + F(\varphi)\right].$$

For $0 \leq l \leq \lfloor nT \rfloor$ and $t \in [l/n, l/n + 1/n)$, define

$$\hat{X}^n(t) = \bar{X}^n(l/n).$$

Thus $\hat{X}^n$ is the piecewise constant interpolation of the occupancy process. Note that for all ω

$$\sup_{t\in[0,T]}\left|\hat{X}^n(t) - \bar{X}^n(t)\right| \leq \frac{1}{n}.$$

Therefore if $\bar{X}^n$ converges weakly to $\bar{X}$, then also $\hat{X}^n$ converges weakly to $\bar{X}$.

We do the same thing for the controlled measures, and for $t \in [l/n, l/n+1/n)$ set

$$\hat{\nu}^n(t) = \nu^{l,n} \quad \text{and} \quad \hat{\rho}^n(t) = \rho\left(\frac{l}{n}, \hat{X}^n(t)\right).$$

Note that because relative entropy is nonnegative and $(\lfloor nT \rfloor + 1)/n \geq T$,

$$\frac{1}{n}\sum_{i=0}^{\lfloor nT \rfloor} R\left(\nu^{i,n} || \rho^{i,n}\right) \geq \int_0^T R\left(\hat{\nu}^n(t) || \hat{\rho}^n(t)\right) dt. \tag{3.1}$$

For an $\mathcal{S}_I-$valued process $\{y(t)\}$ we define $\langle y \rangle$ as its indefinite integral, i.e.,

$$\langle y \rangle(t) \doteq \int_0^t y(s) ds, \quad \text{for } 0 \leq t \leq T.$$

Then $\langle y \rangle$ can be viewed as a vector of sub-cumulative distribution functions taking values in the space

$$\mathcal{Q} \doteq \left\{\langle y \rangle, y : [0,T] \mapsto \mathcal{S}_I \text{ is measurable}\right\}.$$

We consider $\mathcal{Q}$ as a subset of $C([0,T], \mathbb{R}^{I+2})$ with the inherited topology. Since each component of $\langle y \rangle$ is Lipschitz continuous (with Lipschitz constant 1), the Arzela-Ascoli Theorem implies that $\mathcal{Q}$ is precompact. Choose a convergent sequence $\langle y_n \rangle$, with limit z. Since each component of z inherits the Lipschitz continuity and monotonicity of the $\langle y_n \rangle$, a vector of non-negative derivatives y exist, and for any $0 \leq s \leq t \leq T$ these derivatives satisfy

$$\int_s^t \sum_{i=1}^{I+1} y_i(u) du = \sum_{i=1}^{I+1} [z_i(t) - z_i(s)] = t - s.$$

This implies that $y \in \mathcal{S}_I$ for almost every $t \in [0,T]$, and hence $z \in \mathcal{Q}$. We conclude that $\mathcal{Q}$ is compact.

Let $m \otimes y$ be the vector of sub-probability measures generated by $\langle y \rangle$, i.e., for each $0 \leq i \leq I+1$ and $0 \leq a \leq T$

$$(m \otimes y)_i((-\infty, a]) \doteq \frac{\langle y \rangle_i(a)}{T}.$$

Each component of $m \otimes y$ can be viewed as taking values in the space of sub-probability measures with the topology of weak convergence, and then $m \otimes y$ as taking values in the product space with corresponding product topology. However, we can also consider $m \otimes y$ as a probability measure on $[0,T] \times \{0,1,\ldots I, I+1\}$, with the topology of weak convergence on this space. These two topologies are clearly equivalent, and since uniform convergence of sub-cumulative distribution functions implies the weak convergence of the corresponding sub-probability measures, the mapping $\langle y \rangle \mapsto m \otimes y$ is continuous. The Continuous Mapping Theorem then implies the following result.

Lemma 3.1. *Let Y_n and Y be $\mathcal{S}_I$-valued random processes. If $\langle Y_n \rangle$ converges weakly to $\langle Y \rangle$ then $m \otimes Y_n$ converges weakly to $m \otimes Y$.*

We will also need conditions under which $\langle Y_n \rangle$ will converge weakly to $\langle Y \rangle$. Define $\mathcal{V} \doteq \mathcal{D}([0,T] : \mathcal{S}_I)$ to be the space of functions that map $[0,T]$ into $\mathcal{S}_I$, are right continuous, and have left-hand limits. Note that $\mathcal{U} \subset \mathcal{V}$. We equip $\mathcal{V}$ with the standard *Skorohod metric* $s(\cdot,\cdot)$ so that $(\mathcal{V}, s)$ is a Polish space (cf. [1]). If $y_n \in \mathcal{V}$, $s(y_n, y) \to 0$, and if $y \in \mathcal{U}$, then in fact $y_n(t) \to y(t)$ uniformly in $t \in [0,T]$, and hence $\langle y_n \rangle \to \langle y \rangle$. Another application of the Continuous Mapping Theorem gives the following.

Lemma 3.2. *Suppose a sequence of $\mathcal{V}$-valued processes Y_n converges weakly to the $\mathcal{U}$-valued process Y. Then $\langle Y_n \rangle$ converges weakly to $\langle Y \rangle$, and hence $m \otimes Y_n$ converges weakly to $m \otimes Y$.*

We will also need the following formula, which can be verified directly from the given definitions:

$$\int_0^T R\left(\hat{\nu}^n(t) || \rho^n(t)\right) dt = T R\left(m \otimes \hat{\nu}^n || m \otimes \rho^n\right). \tag{3.2}$$

Next we will prove the key weak convergence theorem used in the process level analysis.

Theorem 3.3. *Define a sequence of controlled processes and controls $(\bar{X}^n(t), \hat{\nu}^n(t))$ as above. Then*

$$\left\{ \left(\bar{X}^n, \langle \hat{\nu}^n \rangle \right), n \in \mathbb{N} \right\}$$

is tight. For any sequence from $\left\{ \left(\bar{X}^n, \langle \hat{\nu}^n \rangle \right), n \in \mathbb{N} \right\}$ consider a further subsequence that converges in distribution to (X, η). Then the limit processes have the following properties:

1. *There exists an S_I−valued process θ so that $\eta = \langle \theta \rangle$ w.p.1.*
2. *The process X is a valid occupancy process and the process θ is the occupancy rate process associated with X, i.e., w.p.1.*

$$X(t) = X(0) + \int_0^t M\theta(s) ds \quad \textit{for all } t \in [0,T].$$

Proof. Both $\bar{X}^n$ and $\langle \hat{\nu}^n \rangle$ are uniformly (in n and ω) Lipschitz continuous. Hence by the Arzela-Ascoli Theorem, $\left\{ \left(\bar{X}^n, \langle \hat{\nu}^n \rangle \right), n \in \mathbb{N} \right\}$ is tight. Let (X, η) denote the weak limit of a convergent subsequence. Since the second component takes values in $\mathcal{Q}$ and this space is compact, there exists a measurable $\mathcal{S}_I$-valued process $\theta(t)$ so that $\eta = \langle \theta \rangle$.

Notice that for each $0 \le i \le I+1$ and $0 \le l \le \lfloor nT \rfloor$, and with the notational conventions $e_{I+2} = e_{I+1}$ and $\hat{\nu}_{-1}(t) = 0$,

$$\begin{aligned} &\bar{X}_i^n \left(\frac{l+1}{n} \right) - \bar{X}_i^n \left(\frac{l}{n} \right) \\ &= \frac{1}{n} 1_{\left\{ \bar{y}^{l,n} = e_i - e_{i-1} \right\}} - \frac{1}{n} 1_{\left\{ \bar{y}^{l,n} = e_{i+1} - e_i \right\}} \\ &= \frac{1}{n} \hat{\nu}_{i-1}^n \left(\frac{l}{n} \right) - \frac{1}{n} \hat{\nu}_i^n \left(\frac{l}{n} \right) + \frac{1}{n} \hat{Y}_i^n \left(\frac{l}{n} \right), \end{aligned} \tag{3.3}$$

with $\hat{Y}_i^n \left(\frac{l}{n} \right)$ implicitly defined by

$$\hat{Y}_i^n \left(\frac{l}{n} \right) = \left[1_{\left\{ \bar{y}^{l,n} = e_i - e_{i-1} \right\}} - \hat{\nu}_{i-1}^n \left(\frac{l}{n} \right) \right] - \left[1_{\left\{ \bar{y}^{l,n} = e_{i+1} - e_i \right\}} - \hat{\nu}_i^n \left(\frac{l}{n} \right) \right].$$

In the same way that we defined $\hat{\nu}^n, \hat{X}^n$ on the whole $[0,T]$ by piecewise constant interpolation, we can also define $\hat{Y}^n(t)$ on $[0,T]$. Let $\mathcal{F}_l^n$ be the natural filtration, i.e., the $\sigma-$algebra generated by $\left\{ \bar{X}^n(0), \bar{X}^n(1/n), \ldots, \bar{X}^n(l/n) \right\}$. Then

$$E \left[1_{\left\{ \bar{y}^{l,n} = e_i - e_{i-1} \right\}} | \mathcal{F}_l^n \right] = \hat{\nu}_{i-1}^n \left(\frac{l}{n} \right),$$

which shows that $\left\{ \hat{Y}_i^n \left(\frac{l}{n} \right) \right\}$, $0 \le l \le \lfloor nT \rfloor$ is a martingale difference with respect to $\mathcal{F}_l^n$.

We have observed that $\left\{ \sum_{l=0}^k \hat{Y}_i^n \left(\frac{l}{n} \right) : 0 \le k \le \lfloor nT \rfloor \right\}$ is a martingale with respect to $\mathcal{F}_k^n$. It is also easy to see that $E \left[\hat{Y}_i^n \left(\frac{l}{n} \right) \right]^2 = O(1)$. Summing (3.3) shows that for any $0 \le l \le \lfloor nT \rfloor$

$$\bar{X}_i^n (l/n) - \bar{X}_i^n (0) = \langle \hat{\nu}_{i-1}^n \rangle (l/n) - \langle \hat{\nu}_i^n \rangle (l/n) + \langle \hat{Y}_i^n \rangle (l/n).$$

Owing to the fact that the jumps in the discrete time processes are uniformly bounded, if $t \in \left[\frac{l}{n}, \frac{l+1}{n} \right)$ for some $0 \le l \le \lfloor nT \rfloor$ then $\langle \hat{\nu}^n \rangle (t) = \langle \hat{\nu}^n \rangle (l/n) + O(1/n)$ and $\langle \hat{Y}^n \rangle (t) = \langle \hat{Y}^n \rangle (l/n) + O(1/n)$, where the $O(1/n)$ term does not depend on ω. Since the Lipschitz continuity of $\bar{X}^n$ implies $\left| \bar{X}^n(t) - \bar{X}^n(s) \right| \le |t - s|$, for any $s, t \in [0,T]$

$$\bar{X}_i^n (t) - \bar{X}_i^n (0) = \langle \hat{\nu}_{i-1}^n \rangle (t) - \langle \hat{\nu}_i^n \rangle (t) + \langle \hat{Y}_i^n \rangle (t) + g_i^n(t).$$

where $g_i^n(t)$ converges to 0 uniformly in t and in ω. Recalling that $\left\{\hat{Y}_i^n\left(\frac{l}{n}\right)\right\}$, $0 \leq l \leq \lfloor nT \rfloor$ is a martingale difference and $E\left(\langle \hat{Y}_i^n \rangle^2 \left(\frac{l}{n}\right)\right) = O(1/n)$, by a standard martingale inequality

$$\langle \hat{Y}_i^n \rangle (t) \to 0 \text{ uniformly for } t \in [0,T], w.p.1.$$

Since $\langle \hat{Y}_i^n \rangle (t) + g_i^n(t)$ converges to 0 w.p.1 and $\left(\bar{X}_i^n(t) - \bar{X}_i^n(0), \langle \hat{\nu}_{i-1}^n \rangle (t) - \langle \hat{\nu}_i^n \rangle (t)\right)$ converges weakly to $(X_i(t) - X_i(0), \eta_{i-1}(t) - \eta_i(t))$,

$$X_i(t) - X_i(0) = \eta_{i-1}(t) - \eta_i(t) \qquad w.p.1.$$

Recall that we have proved the existence of a process θ so that $\eta(t) = \int_0^t \theta(s)ds$. Thus the last display can be rewritten

$$X_i(t) - X_i(0) = \langle \theta \rangle_{i-1}(t) - \langle \theta \rangle_i(t) \qquad w.p.1,$$

which is indeed

$$X(t) = X(0) + \int_0^t M\theta(s)ds \qquad w.p.1.$$

□

Theorem 3.4. *Define I by (2.4) for any of the occupancy models discussed in Section 2. If $F : \mathcal{U} \mapsto \mathbb{R}$ is bounded and continuous, then*

$$\liminf_{n\to\infty} -\frac{1}{n} \log E \exp\left[-nF(X^n)\right] \geq \inf_{\varphi\in\mathcal{U}} \left[I(\varphi) + F(\varphi)\right].$$

Proof. Owning to the representation formula (2.5) it is enough to show that

$$\liminf_{n\to\infty} \inf_{\{\nu^{i,n}\}} \bar{E}\left[F\left(\bar{X}^n\right) + \frac{1}{n}\sum_{i=0}^{\lfloor nT \rfloor} R\left(\nu^{i,n} || \rho^{i,n}\right)\right] \geq \inf_{\varphi\in\mathcal{U}} \left[I(\varphi) + F(\varphi)\right].$$

Consider any admissible sequence $\left\{\nu^{i,n}\right\}$. Then (3.1) and (3.2) imply

$$\begin{aligned} &\bar{E}\left[F\left(\bar{X}^n\right) + \frac{1}{n}\sum_{i=0}^{\lfloor nT \rfloor} R\left(\nu^{i,n} || \rho^{i,n}\right)\right] \\ &\geq \bar{E}\left[F\left(\bar{X}^n\right) + \int_0^T R\left(\hat{\nu}^n(t) || \hat{\rho}^n(t)\right) dt\right] \\ &= \bar{E}\left[F\left(\bar{X}^n\right) + TR\left(m \otimes \hat{\nu}^n || m \otimes \hat{\rho}^n\right)\right]. \end{aligned}$$

By Theorem 3.3 we know for any subsequence of $\mathbb{N}$, there exists a subsubsequence such that $\left\{\left(\hat{X}^n, \langle\hat{\nu}^n\rangle\right)\right\}$ converges in distribution to a limit $(X, \langle\theta\rangle)$.

Let

$$\mathcal{W} \doteq \left\{x \in \mathcal{V} : \sum_{i=1}^{I+1} ix(t) \leq t, \text{ for } 0 \leq t \leq T\right\}.$$

Due to our construction of the controlled process $\hat{X}^n$, we know that for each ω, $\hat{X}^n(\omega) \in \mathcal{W}$. For $a \in (0,\infty) \cup \{-1,-2,\ldots\}$ and $x \in \mathcal{W}$ define $g(x)$ by

$$(g(x))_i(t) \doteq \frac{a+i}{a+t} x_i(t), \qquad \text{for } 0 \leq i \leq I$$

and

$$(g(x))_{I+1}(t) \doteq 1 - \sum_{i=0}^{I} \frac{a+i}{a+t} x_i(t).$$

Then g maps $\mathcal{W}$ to $\mathcal{V}$. The case $a = \infty$ is defined as the obvious limit. When $a \in (0,\infty]$ g is clearly bounded and continuous. When $a < 0$ the boundedness of g is not as trivial but still elementary. We know that when $a < 0$, balls are only thrown among the categories $0, 1, \ldots, -a-1$. Thus if there are n urns there can at most be $-an$ balls thrown, and therefore $T \leq -a$. When $T = -a$ all the urns have exactly $-a$ balls, which is not an interesting case to study. We therefore assume $T < -a$. Also, because of the same restriction on the possible categories we can (without loss) Therefore

$$||g|| \leq \frac{-i-a}{-t-a} \leq \frac{-a}{-a-T} = \frac{a}{a+T},$$

which shows that g is bounded. The argument to show continuity is similar and omitted.

With these definitions we have $g\left(\hat{X}^n\right)(t) = \hat{\rho}^n(t)$ and $g(X)(t) = \rho(t, X(t))$. Since $\hat{X}^n, X \in \mathcal{W}$, we have $\hat{\rho}^n$ and $\rho(\cdot, X(\cdot)) \in \mathcal{V}$. By the Continuous Mapping Theorem and the definition of $\hat{\rho}^n$, weak convergence of $\hat{X}^n$ to X implies weak convergence of $\hat{\rho}^n$ to $\rho(t, X(t))$ in $\mathcal{V}$. Applying Lemma 3.2, we have that $m \otimes \hat{\rho}^n$ converges weakly to $m \otimes \rho$. Similarly, by Lemma 3.1 the weak convergence of $\langle\hat{\nu}^n\rangle$ to $\langle\theta\rangle$ implies the weak convergence of $m \otimes \hat{\nu}^n$ to $m \otimes \theta$.

Now applying Fatou's Lemma (for weak convergence) and using the lower semicontinuity of relative entropy,

$$\begin{aligned}
&\liminf_{n\to\infty} \bar{E}\left[F\left(\bar{X}^n\right)+\frac{1}{n}\sum_{i=0}^{\lfloor nT\rfloor} R\left(\nu^{i,n}||\rho^{i,n}\right)\right] \\
&\geq \liminf_{n\to\infty} \bar{E}\left[F\left(\bar{X}^n\right)+TR\left(m\otimes\hat{\nu}^n||m\otimes\hat{\rho}^n\right)\right] \\
&\geq \bar{E}\left[F\left(X\right)+TR\left(m\otimes\theta||m\otimes\rho\right)\right] \qquad (3.4) \\
&= \bar{E}\left[F\left(X\right)+\int_0^T R\left(\theta(t)||\rho(t,X(t))\right)dt\right].
\end{aligned}$$

As proved in Theorem 3.3, $X(t) = X(0) + \int_0^t M\theta(s)ds$, therefore by the definition (2.4) of the rate function $I(\varphi)$,

$$\int_0^T R\left(\theta(t)||\rho(t,X(t))\right)dt = I(X).$$

Thus (3.4) yields

$$\liminf_{n\to\infty} \inf_{\{\nu_i^n\}} \bar{E}\left[F\left(\bar{X}^n\right)+\frac{1}{n}\sum_{i=0}^{\lfloor nT\rfloor} R\left(\nu^{i,n}||\rho^{i,n}\right)\right] \geq \inf_{\varphi\in\mathcal{U}}\left[I(\varphi)+F(\varphi)\right].$$

Hence we complete the proof of the large deviation upper bound. □

4 Properties of the Rate Function.

In this section, we will prove some important properties of the rate function, some of which will be used later on to prove the large deviation lower bound.

Theorem 4.1. *Let I be defined as in (2.4). Then for any $K \in [0,\infty)$ the level set $\{\varphi \in \mathcal{U} : I(\varphi) \leq K\}$ is compact.*

Proof. As is always the case in the weak convergence approach, the proof of compactness of level sets is essentially a deterministic analogue of the proof of the large deviation upper bound, and hence omitted. See [3, Proposition 6.2.4] for the proof in an analogous situation. □

Theorem 4.2 (Zero Cost Trajectory). *For $t \in [0,T]$ let $f(t) = \left(1+\frac{t}{a}\right)^{-a}$ ($f(t) = e^{-t}$ in the case $a = \infty$),*

$$\phi_i(t) = \frac{(-t)^i}{i!} f^{(i)}(t) \qquad \text{for } 0 \leq i \leq I,$$

and let $\phi_{I+1}(t) = 1 - \sum_{i=0}^{I} \phi_i(t)$. Then $I(\phi) = 0$.

Proof. We first assume $a \neq \infty$. It is easy to see that for any $0 \leq i < \infty$,

$$\frac{(-t)^i}{i!} f^{(i)}(t) \geq 0 \quad \text{and} \quad \sum_{i=0}^{\infty} \frac{(-t)^i}{i!} f^{(i)}(t) = 1. \tag{4.1}$$

Thus ϕ as defined in the statement of the theorem is indeed a probability vector. It is also clearly a continuously differentiable function. We will show that

$$\dot{\phi}(t) = M\rho(t, \phi(t)). \tag{4.2}$$

If so, then the occupancy rate process θ associated to ϕ is indeed $\rho(t, \phi(t))$, and thus by the definition of rate function

$$I(\phi) = \int_0^T R\left(\theta(t) || \rho(t, \phi(t))\right) dt = 0.$$

To show (4.2) we calculate $\phi_i(t) = \frac{(-t)^i}{i!} f^{(i)}(t)$ for $0 \leq i \leq I$ explicitly:

$$\phi_i(t) = \frac{t^i}{i!} \frac{\prod_{j=0}^{i-1}(a+j)}{a^i} \left(1 + \frac{t}{a}\right)^{-a-i}.$$

Hence the derivative satisfies

$$\begin{aligned} \dot{\phi}_i(t) &= \frac{a+i-1}{a+t} \phi_{i-1}(t) - \frac{a+i}{a+t} \phi_i(t) \\ &= \rho_{i-1}(t, \phi(t)) - \rho_i(t, \phi(t)) \\ &= (M\rho(t, \phi(t))_i, \end{aligned}$$

where the second equality is due to the definition of $\rho(t, \phi(t))$. The case of $\phi_{I+1}(t)$ is also a straightforward calculation and hence omitted.

Next we consider the case when $a = \infty$. In this case $f(t) = e^{-t}$, and the validity of (4.2) can be directly verified. ☐

Lemma 4.3. *For every choice of the parameter a there exist $\delta > 0$ and $0 < K < \infty$ so that the zero cost trajectory $\phi(t)$ is away from the boundary of $\mathcal{S}_I$, i.e.,*

$$\phi_i(t) \geq \delta t^K \tag{4.3}$$

for any $0 \leq i \leq I + 1$.

Proof. Note that when $a > 0$, $0 \leq i \leq I$ and $0 \leq t \leq T$,

$$\begin{aligned} \phi_i(t) &= \frac{t^i}{i!} \frac{\prod_{j=0}^{i-1}(a+j)}{a^i} \left(1 + \frac{t}{a}\right)^{-a-i} \\ &\geq \frac{t^i}{I!} \left(1 + \frac{T}{a}\right)^{-a-I}, \end{aligned} \tag{4.4}$$

and because of (4.1) we have

$$\begin{aligned}\phi_{I+1}(t) &= 1 - \sum_{i=0}^{I} \phi_i(t) &(4.5)\\ &\geq \frac{(-t)^{I+1}}{(I+1)!} f^{(I+1)}(t)\\ &\geq \frac{t^{I+1}}{(I+1)!}\left(1+\frac{T}{a}\right)^{-a-I-1}. &(4.6)\end{aligned}$$

For the case $a < 0$ we have $T < -a$ and $a \leq -I-1$. Recall that for $0 \leq i \leq I$

$$\phi_i(t) = \frac{t^i}{i!}\frac{\prod_{j=0}^{i-1}(a+j)}{a^i}\left(1+\frac{t}{a}\right)^{-a-i}.$$

Since $a + I \leq -1$, $a + j \leq -1$ for each $0 \leq j \leq I$, and thus

$$\phi_i(t) \geq \frac{t^i}{I!}\frac{1}{(-a)^i}\left(1+\frac{t}{a}\right)^{-a-i}.$$

Moreover since $a < 0$ and $-a-I > 0$, for each fixed a and $i \leq I$, $\left(1+\frac{t}{a}\right)^{-a-i}$ is monotone decreasing in t. Therefore

$$\phi_i(t) \geq \frac{t^i}{I!}\left(-\frac{1}{a}\right)^i\left(1+\frac{T}{a}\right)^{-a-i}.$$

Lastly, since $T < -a$ and $a < 0$, $0 < 1 + T/a < 1$. Thus $\left(1+\frac{T}{a}\right)^{-a-i}$ is monotone increasing in i, and therefore

$$\phi_i(t) \geq \frac{t^i}{I!}\left(-\frac{1}{a}\right)^i\left(1+\frac{T}{a}\right)^{-a}.$$

For $\phi_{I+1}(t)$ we have

$$\begin{aligned}\phi_{I+1}(t) &= 1 - \sum_{i=0}^{I} \phi_i(t)\\ &\geq \frac{(-t)^{I+1}}{(I+1)!} f^{(I+1)}(t)\\ &= \frac{t^{I+1}}{(I+1)!}\frac{\prod_{j=0}^{I}(a+j)}{a^{I+1}}\left(1+\frac{t}{a}\right)^{-a-I}\\ &\geq \frac{t^{I+1}}{(I+1)!}\left(-\frac{1}{a}\right)^{I+1}\left(1+\frac{T}{a}\right)^{-a}.\end{aligned}$$

The last inequality follows exactly the same reasoning as for $0 \le i \le I$.

Finally, for the case $a = \infty$ we just take limits on (4.4) and (4.5), and use that

$$\lim_{a\to\infty}\left(1+\frac{T}{a}\right)^{-a-I} = e^{-T}.$$

If follows that for every choice of the parameter a there exist $\delta > 0$ and $0 < K < \infty$ so that the zero cost trajectory $\phi(t)$ satisfies (4.3) for any $0 \le i \le I+1$. □

Lemma 4.4. *For a given value of a let the parameters δ and K be as in (4.3). Let $\varphi \in \mathcal{U}$ satisfy $I(\varphi) < \infty$. Then for any $\varepsilon > 0$ there exists $\varphi^\varepsilon \in \mathcal{U}$ such that*

1. $I(\varphi^\varepsilon) \le I(\varphi)$,
2. $d(\varphi, \varphi^\varepsilon) \le \varepsilon$,
3. $\varphi_i^\varepsilon(t) \ge \varepsilon\delta t^K$ *for all* $t \in [0,T]$ *and* $i = 0, 1, \ldots I, I+1$.

Proof. For any $\varepsilon > 0$ and $\varphi \in \mathcal{U}$, let

$$\varphi^\varepsilon = (1-\varepsilon)\,\varphi + \varepsilon\bar{\varphi},$$

where $\bar{\varphi}$ is the zero cost trajectory. Then $\varphi^\varepsilon \in \mathcal{U}$. From the definition of $\rho(t,x)$ in (2.1) it follows that $\rho(t,x)$ is linear in x. Also, recalling the definition of $I(\varphi)$ in (2.4) and the joint convexity of relative entropy, we find that $I(\varphi)$ is convex in φ. Therefore

$$\begin{aligned} I(\varphi^\varepsilon) &\le (1-\varepsilon)\,I(\varphi) + \varepsilon I(\bar{\varphi}) \\ &= (1-\varepsilon)\,I(\varphi) \\ &\le I(\varphi). \end{aligned}$$

Since $d\,(\varphi, \bar{\varphi}) \le 1$

$$d(\varphi, \varphi^\varepsilon) \le \varepsilon d\,(\varphi, \bar{\varphi}) \le \varepsilon,$$

and also $\varphi^\varepsilon \ge \varepsilon\bar{\varphi} \ge \varepsilon\delta t^K$. □

The final theorem of this section is essential in proving the large deviation lower bound.

Definition 4.5 (Good Path). *We call an occupancy process $\varphi \in \mathcal{U}$ a "good path" if there exist constants $0 < \delta', K' < \infty$ so that $\varphi_i(t) \ge \delta' t^{K'}$ for $t \in [0,T]$ and $0 \le i \le I+1$.*

Definition 4.6 (Good Control). *We call an occupancy rate process θ a "good control" if the process θ is piecewise constant on $[0,T]$, with a finite number of intervals of constancy. In other words, there exist a finite number of intervals $[r_i, s_i], 1 \leq i \leq m$ so that $[0,T] = \cup_{i=1}^m [r_i, s_i]$, and $\theta(t)$ is a constant vector on each (r_i, s_i). In addition, we assume there exists $0 < \sigma < T$ so that θ is "pure" on $[0,\sigma)$, in the sense that for any interval of constancy $(r,s) \subset [0,\sigma)$, there exists $i, 0 \leq i \leq I+1$ such that $\theta_i(t) = 1$ for $t \in (r,s)$.*

Theorem 4.7. *For a good path $\varphi \in \mathcal{U}$ assume $I(\varphi) < \infty$. Let δ', K' be the associated constants in the definition of a good path. For any $\varepsilon > 0$ there exists a good control θ^* and associated $\sigma > 0$ so that if φ^* is the occupancy path associated to θ^*, then*

1. $I(\varphi^*) \leq I(\varphi) + \varepsilon$,
2. $d(\varphi^*, \varphi) \leq \varepsilon$,
3. *if* $t < \sigma$ *and* $\theta_i^*(t) = 1$ *then* $\varphi_i^*(t) > \delta' \sigma^{K'}$.

Proof. For a $\sigma > 0$ that will be specified later on, we construct a pure control $\theta^*(t), t \in [0,\sigma)$ as follows. For $0 \leq i \leq I$ let $\theta_i^*(t) = 1$ if

$$\sum_{j=0}^{i} j\varphi_j(\sigma) + \sum_{k=i+1}^{I+1} i\varphi_k(\sigma) \leq t < \sum_{j=0}^{i} j\varphi_j(\sigma) + \sum_{k=i+1}^{I+1} (i+1)\varphi_k(\sigma),$$

and let $\theta_{I+1}^*(t) = 1$ if

$$\sum_{j=0}^{I} j\varphi_j(\sigma) + (I+1)\varphi_{I+1}(\sigma) \leq t < \sigma.$$

Observe that the component φ_i will increase only during the interval when $\theta_{i-1}^*(t) = 1$, and that it decreases to its final value while $\theta_i^*(t) = 1$. Observe also that $\varphi^*(\sigma) = \varphi(\sigma)$. Hence for $t < \sigma$, if $\theta_i^*(t) = 1$ then $\varphi_i^*(t) > \varphi_i^*(\sigma) \geq \delta' \sigma^{K'}$.

Now assume that $0 < a < \infty$. For such i and t,

$$\begin{aligned} \rho_i(t, \varphi^*(t)) &= \frac{a+i}{a+t}\varphi_i^*(t) \\ &\geq \frac{a}{a+T}\delta'\sigma^{K'} \\ &= \delta''\sigma^{K'}, \end{aligned} \tag{4.7}$$

where δ'' is defined as $\frac{a}{a+T}\delta'$. Note that the above bound is true for all $0 \leq i \leq I+1$.

Recall that when $a < 0$ we can assume without loss that $a + 1 + I \leq 0$, and that no balls are placed in urns that currently contain more than I balls. It follows that

$$\sum_{j=0}^{I+1} j\varphi_j(\sigma) = \sigma,$$

and that $\theta_{I+1}(t)$ is always zero. Hence the same will be true of $\theta^*_{I+1}(t)$, i.e., $\theta^*_{I+1}(t) = 0$ for all $t \in [0, \sigma]$. For $0 \leq i \leq I$, we have

$$\begin{aligned}
\rho_i(t, \varphi^*(t)) &\geq \frac{a+i}{a+t}\delta'\sigma^{K'} \\
&\geq \frac{a+I}{a+t}\delta'\sigma^{K'} \\
&\geq \frac{a+I}{a}\delta'\sigma^{K'} \\
&\geq -\frac{1}{a}\delta'\sigma^{K'}.
\end{aligned}$$

Thus for such i there exists a constant $\delta'' > 0$ so that $\rho_i(t, \varphi^*(t)) \geq \delta''\sigma^{K'}$ when $\theta^*_i(t) = 1$.

Finally, when $a = \infty$ we can choose $\delta'' = \delta'$ and (4.7) will hold.

This completes the construction of θ^* and φ^* on $[0, \sigma)$. The lower bounds on the ρ_i and the fact that θ^* is pure on $[0, \sigma]$ imply

$$\int_0^\sigma R\left(\theta^*(t) || \rho(t, \varphi^*(t))\right) dt \leq -\sigma \log\left(\delta''\sigma^{K'}\right).$$

Now let us choose σ small enough so that $\int_0^\sigma R\left(\theta^*(t) || \rho(t, \varphi^*(t))\right) dt \leq \varepsilon/2$ and $\sup_{t\in[0,\sigma)} |\varphi^*(t) - \varphi(t)| < \varepsilon$. Also, recall that under the construction $\varphi^*(\sigma) = \varphi(\sigma)$.

The construction of controls on $[\sigma, T]$ is easier. Let $\theta(t)$ be the rate process associated with $\varphi(t)$ by (2.3). For $M \in \mathbb{N}$ we partition $[\sigma, T]$ into M subintervals of length $c_M = (T - \sigma)/M$. For each s that $\sigma + lc_M \leq s \leq \sigma + (l+1)c_M$ where $0 \leq l \leq (M-1)$, let

$$\theta^{(M)}(s) = \frac{\int_{\sigma+lc_M}^{\sigma+(l+1)c_M} \theta(t)dt}{c_M}.$$

Let $\varphi^{(M)}(t)$ be the occupancy path associated with $\theta^{(M)}(t)$. Then it is easy to check that $\varphi^{(M)}(t)$ coincides with $\varphi(t)$ on the "partition points" in $[\sigma, T]$, i.e., those points of the form $\{\sigma + lc_M : 0 \le l \le (M-1)\}$. Thus for M large enough (e.g., $M > (T-\sigma)/\varepsilon$), $\sup_{t\in[\sigma,T]} \left|\varphi^{(M)}(t) - \varphi(t)\right| < \varepsilon$.

Because $\varphi(t)$ is good, when $t > \sigma$, we have $\varphi(t) > \delta' t^{K'} \ge \delta' \sigma^{K'} > 0$. Therefore $\varphi(t)$ is uniformly bounded away from the boundary after time σ. As $M \to \infty, \theta^{(M)}(t)$ converges to $\theta(t)$ and $\varphi^{(M)}(t)$ converges to $\varphi(t)$ a.e., and thus by the Lebesgue Dominated Convergence Theorem

$$\lim_{M\to\infty} \int_\sigma^T R\left(\theta^M(t)||\rho(t|\varphi^M(t))\right) dt = \int_\sigma^T R\left(\theta(t)||\rho(t|\varphi(t))\right) dt.$$

Now choose $M < \infty$ large enough so that $\int_\sigma^T R\left(\theta^M(t)||\rho(t|\varphi^M(t))\right) dt < \int_\sigma^T R\left(\theta(t)||\rho(t|\varphi(t))\right) dt + \varepsilon/2$. Let θ^* be defined as it was previously on $[0,\sigma]$, and set it equal to θ^M on $[\sigma, T]$. We have

$$\begin{aligned} I(\varphi^*) &= \int_\sigma^T R\left(\theta^M(t)||\rho(t|\varphi^M(t))\right) dt + \int_0^\sigma R\left(\theta^*(t)||\rho(t|\varphi^*(t))\right) dt \\ &< \int_\sigma^T R\left(\theta(t)||\rho(t|\varphi(t))\right) dt + \varepsilon/2 + \varepsilon/2 \\ &\le I(\varphi) + \varepsilon. \end{aligned}$$

Thus we complete the proof. □

5 The Large Deviation Lower Bound

Theorem 5.1. *Define I by (2.4) for any of the occupancy models discussed in Section 2. If $F : \mathcal{U} \mapsto \mathbb{R}$ is bounded and continuous, then*

$$\limsup_{n\to\infty} -\frac{1}{n} \log E \exp\left[-nF(X^n)\right] \le \inf_{\varphi\in\mathcal{U}} \left[I(\varphi) + F(\varphi)\right].$$

Proof. According to (2.5), the theorem follows if

$$\limsup_{n\to\infty} \inf_{\{\nu^{i,n}\}} \bar{E}\left[F\left(\bar{X}^n\right) + \frac{1}{n}\sum_{i=0}^{\lfloor nT\rfloor} R\left(\nu^{i,n}||\rho^{i,n}\right)\right] \le \inf_{\varphi\in\mathcal{U}} \left[I(\varphi) + F(\varphi)\right].$$

For any $\varphi \in \mathcal{U}$ such that $I(\varphi) < \infty$, Lemma 4.4 and Theorem 4.7 imply that for any $\varepsilon > 0$ there exists (φ^*, θ^*) with the properties described in Theorem 4.7. Since F is continuous in $\mathcal{U}$, we only need to show that there exists a sequence of admissible controls $\{\nu^n\}$ so that

$$\limsup_{n\to\infty} \bar{E}\left[F\left(\bar{X}^n\right) + \frac{1}{n}\sum_{i=0}^{\lfloor nT\rfloor} R\left(\nu^{i,n}||\rho^{i,n}\right)\right] \le I(\varphi^*) + F(\varphi^*).$$

The latter inequality will follow if we can find a sequence of admissible $\{\nu^n\}$ such that

$$\limsup_{n\to\infty} \bar{E}\left[\sum_{i=0}^{\lfloor nT\rfloor} R\left(\nu^{i,n}||\rho^{i,n}\right)\right] \leq I(\varphi^*), \tag{5.1}$$

and such that if $\bar{X}^n$ is the occupancy process constructed under $\{\nu^n\}$ then for any small $b > 0$

$$\limsup_{n\to\infty} \bar{P}\left\{d\left(\bar{X}^n, \varphi^*\right) > b\right\} = 0. \tag{5.2}$$

In other words, $\bar{X}^n$ converges to φ^* in probability.

To prove the desired inequalities (5.1) and (5.2) we need to construct the proper $\{\nu^n\}$. Recall that $\{\nu^n\}$ can depend in any measurable way on the "past," and so we could, in principle, use such information in constructing the controls. However, it turns out that we can construct the controls without reference to the controlled process (so-called "open loop" controls). Let θ^* be the good control as described in Theorem 4.7. We know that θ^* is piecewise constant and pure up to time $\sigma > 0$. We also know that before time σ, if $\theta_i^*(t) = 1$ then both $\rho_i(t, \varphi^*(t))$ and $\varphi_i^*(t)$ are greater than a fixed value $\zeta > 0$. We can also assume for the same value of ζ that both $\rho_i(t, \varphi^*(t))$ and $\varphi_i^*(t)$ are greater than ζ for all $i \in [0, 1, \ldots I, I+1]$ and all $t \in [\sigma, T]$.

Although the limit trajectory stays away from the boundary after time σ, there is no guarantee that the random process $\bar{X}^n$ is uniformly bounded away. In order to handle this possibility, we use a stopping time argument similar to one used in [8].

Let (l_n/n) be the minimum of the first time such that for some i, $\bar{X}_i^n(l_n/n) \leq \zeta/2$ and $\theta_i^*(l_n/n) > 0$, and the fixed deterministic time T. This is the first time the random process is close to the boundary, and hence there is the possibility of a large contribution to the total cost [note that when $\theta_i^*(l_n/n) = 0$ there is no contribution to the cost regardless of the value of $\bar{X}_i^n(l_n/n)$]. The control $\{\nu^n\}$ is then defined by

$$\nu^{i,n} = \begin{cases} \theta^*(i/n) & \text{if } i \leq l_n \\ \rho\left(i/n, \bar{X}^n(i/n)\right) & \text{if } i > l_n. \end{cases}$$

Prior to the stopping time, we use exactly what θ^* suggests, and after the stopping time we follow the law of large number trajectory (and therefore incur no additional cost).

Now we apply Theorem 3.3. Thus given any subsequence we have convergence along a further subsequence as indicated in the theorem, with limit

$\left(\bar{X}, [\bar{\theta}]\right)$. Using a standard argument by contradiction, it will be enough to prove (5.1) and (5.2) for this convergent processes. Let $\tau^n = (l_n/n) \leq T$. Note that because the applied controls are pure, the process $\bar{X}^n(t)$ is deterministic prior to σ, and also that prior to this time, the time derivatives of $\bar{X}^n(t)$ and $\varphi^*(t)$ are piecewise constant. In fact, the two derivatives are identical except possibly on a bounded number of intervals each of length less than $1/n$ (the points where they may disagree are all located with within distance $1/n$ of the endpoints of the intervals of constancy of $\dot{\varphi}^*(t)$). Thus for large n we cannot have $\tau^n < \sigma$. Since the range of τ^n is a bounded set in $\mathbb{R}$, we can also assume τ^n converges in distribution to a limit τ, and without loss we assume the convergence is along the same subsequence. Since $\tau^n \geq \sigma$ for large n we have $\tau \geq \sigma$ w.p.1. It is easy to check that the limit control processes w.p.1 satisfies

$$\bar{\theta}(t) = \begin{cases} \theta^*(t) & \text{if } t \leq \tau \\ \rho\left(t, \bar{X}(t)\right) & \text{if } t > \tau \end{cases} .$$

Owing to the definition of τ^n, if $\tau < T$ then $\bar{X}_i(\tau) \leq \zeta/2$ for some $i \in [0, 1, \ldots I, I+1]$ (although $\varphi_i^*(t) \geq \zeta$ when $t \in [\sigma, T]$).

We use that $\bar{\theta}(t) = \theta^*(t)$ when $t \leq \tau$ and that $\theta^*(t)$ is deterministic. As shown in Theorem 3.3, $\left(\bar{X}, \theta^*\right)$ satisfies (2.3) for $t \in [0, \tau]$. Thus for $t \in [0, \tau]$.

$$\bar{X}(t) = \varphi^*(t) \qquad \text{w.p.1.}$$

This forms a contradiction since

$$\bar{X}_i(\tau) \leq \zeta/2 < \zeta \leq \varphi_i^*(\tau).$$

Therefore $\tau = T$, and thus for all $t \in [0, T]$

$$\bar{X}(t) = \varphi^*(t) \qquad \text{w.p.1.}$$

This also indicates that the weak limit of the random processes is indeed limit (φ^*, θ^*), which implies (5.2). To prove (5.1), we use the weak convergence and the Dominated Convergence Theorem:

$$\begin{aligned} & \limsup_{n \to \infty} \bar{E}\left[\frac{1}{n} \sum_{i=0}^{\lfloor nT \rfloor} R\left(\nu^{i,n} || \rho^{i,n}\right)\right] \\ = & \int_0^T R\left(\theta^*(t) || \rho(t, \varphi^*(t))\right) \\ = & \; I(\varphi^*). \end{aligned}$$

This completes the proof. □

6 Explicit Formula for the LDP of the Process at a Given Time

In the previous sections we have identified the large deviation rate function (2.4) for a class of occupancy problems. The large deviation principle for the process at a given fixed time can then be expressed in terms of the solution to a calculus of variations problem. In this section we state a conjecture on the solution to this problem. The explicit formula is analogous to one obtained in [8] for the case of MB statistics, and it is possible that the techniques developed there could be used here as well. At the present time, however, we prefer to simply state the result as a conjecture in order to pursue a potentially more general approach that would include such generalizations as urn models with balls of different types.

Since the Maxwell-Boltzmann case is rigorously analyzed in [8], we also assume $a < \infty$ (the formal statement can of course be obtained as a limit). By the Contraction Mapping Theorem, the large deviation rate function for an ending point ω is given by

$$\mathcal{J}(\omega) \doteq \inf_{\varphi \in C([0:T]:\mathcal{S}_I), \varphi(T)=\omega.} I(\varphi).$$

Define

$$(a)_i \doteq \prod_{j=0}^{i-1}(a+j)$$

for all $i \in \mathbb{N}$, and also

$$Q_i^a(x) \doteq \left(\frac{x}{a}\right)^i \frac{(a)_i}{i!} \left(1 + \frac{x}{a}\right)^{-a-i}$$

for all $i \in \mathbb{N}$ and $x \in [0, T]$.

Denote $\pi^k = \{\pi_0^k, \pi_1^k, \ldots, \pi_\infty^k\} \in \mathbb{R}_\infty$ for all $0 \leq k \leq I+1$, where π_i^k represents the probability of throwing i additional balls into the kth category. Denote $\pi = (\pi^0, \pi^1, \ldots, \pi^I, \pi^{I+1})$, so that $\pi \in \mathbb{R}_\infty^{I+2}$. For any given $\alpha \in \mathcal{S}_I$, we say $\pi = (\pi^0, \pi^1, \ldots, \pi^I, \pi^{I+1}) \in \mathcal{F}(\alpha, \omega, T)$ if

$$\sum_{j=0}^{\infty} \pi_j^k = 1 \quad 0 \leq k \leq I+1, \quad \sum_{k=0}^{I+1} \alpha_k \sum_{j=0}^{\infty} j \pi_j^k = T,$$

and

$$\omega_i = \sum_{k=1}^{i} \alpha_k \pi_{i-k}^k \quad 0 \leq i \leq I.$$

With the above notation, we conjecture that in the case of empty initial conditions

$$\mathcal{J}(\omega) = \min_{\pi^0} R\left(\pi^0 || Q^a(T)\right),$$

where π^0 satisfies

$$\pi_i^0 = \omega_i \quad i = 0, \ldots, I, \ \sum_{i=0}^{\infty} \pi_i^0 = 1, \quad \text{and} \ \sum_{i=0}^{\infty} i\pi_i^0 = T.$$

Moreover for a general initial condition α,

$$\mathcal{J}(\alpha, \omega) = \min_{\pi \in \mathcal{F}(\alpha,\omega,T)} \sum_{k=0}^{I+1} \alpha_k R\left(\pi^k \left\| Q^{a+k}\left(\frac{a+k}{a} T\right)\right.\right).$$

References

[1] P. Billingsley. *Convergence of Probability Measures.* Wiley, New York, 1968.

[2] A. Charalambides. A unified derivation of occupancy and sequential occupancy distributions. In *Advances in Combinatorial Methods and Applications to Probability and Statistics, N. Balakrishnan Ed.*, pages 259–273, 1997.

[3] P. Dupuis and R. Ellis. *A Weak Convergence Approach to Large Deviations.* Wiley Interscience, New York, 1997.

[4] W. Feller. *An Introduction to Probability Theory and Its Applications Vol. 1.* John Wiley, New York, 1968.

[5] W. Feller. *An Introduction to Probability Theory and Its Applications Vol.2.* John Wiley, New York, 1971.

[6] L. Holst. On the coupon collectors and other urn problems. *International Statistical Review*, 54(1):15–27, 1986.

[7] N.L.Johnson and S. Kotz. *Urn Models and their Applications.* John Wiley, 1977.

[8] P. Dupuis, C. Nuzman and P. Whiting. Large deviations asymptotics for occupancy problems. *The Annals of Probability*, 32:2765–2818, 2004.

[9] S. Ross. *Introduction to Probability Models (4th Edn.).* Wiley, 198.