*Bernoulli* **18**(4), 2012, 1341–1360
DOI: 10.3150/11-BEJ376

# Conditional large and moderate deviations for sums of discrete random variables. Combinatoric applications

FABRICE GAMBOA[1,*], THIERRY KLEIN[1,**] and CLÉMENTINE PRIEUR[2]

[1]*Institut de Mathématiques de Toulouse, Université de Toulouse, F-31062 Toulouse, France.*
*E-mail:* [*]*fabrice.gamboa@math.univ-toulouse.fr;* [**]*thierry.klein@math.univ-toulouse.fr*
[2]*Université Joseph Fourier, Laboratoire Jean Kunzmann, Equipe Projet INRIA MOISE, Grenoble, France.*
*E-mail: clementine.prieur@imag.fr*

We prove large and moderate deviation principles for the distribution of an empirical mean conditioned by the value of the sum of discrete i.i.d. random variables. Some applications for combinatoric problems are discussed.

*Keywords:* combinatoric problems; conditional distribution; large and moderate deviation principles

## 1. Introduction

In many random combinatorial problems (see Janson [16]), the distribution of the interesting statistic is the law of an empirical mean built on an independent and identically distributed (i.i.d.) sample conditioned by some exogenous integer random variable (r.v.). In general, this exogenous r.v. is also itself a sample mean built on integer r.v.s. In the whole paper $\mathbb{N}^*$ will denote the set $\{1, 2, \ldots\}$ of positive integers, $\mathbb{N} = \mathbb{N}^* \cup 0$, and $\mathbb{Z}$ will be the set of all integers. Hence, a general frame for this kind of problem may be formalized as follows. Let $(p_n)$ be a sequence of integers and $(q_n)$ be a sequence of positive integers. Further, let $\mathbf{X} = (X_j^{(n)})_{n\in\mathbb{N}^*, j=1,\ldots,q_n}$ and $\mathbf{Y} = (Y_j^{(n)})_{n\in\mathbb{N}^*, j=1,\ldots,q_n}$ be two triangular arrays of random variables. Both arrays are such that on their lines the r.v.s are i.i.d. Moreover, it is assumed that the elements of the array $\mathbf{X}$ are integers. The interesting distribution is then the law of $(q_n)^{-1}T_n := (q_n)^{-1}\sum_{j=1}^{q_n} Y_j^{(n)}$ conditioned on a specific value of $S_n := \sum_{j=1}^{q_n} X_j^{(n)}$. That is the conditional distribution

$$\mathcal{L}_n := \mathcal{L}\big((q_n)^{-1}T_n | S_n = p_n\big),$$

where $(p_n)$ is some given integer sequence. When the distribution of $(X_j^{(n)}, Y_j^{(n)})$ does not depend on $n$, the Gibbs conditioning principle [6,7,25] states that $\mathcal{L}_n$ converges weakly to the degenerated distribution concentrated on a point $\chi$ depending on the conditioning value (see Corollary 2.2). Around the Gibbs conditioning principle, general limit theorems yielding the asymptotic behavior of the conditioned sum are given in [14,19,24]. Asymptotic expansions for the distribution of the conditioned sum are proved in [12,22]. In this paper, our aim is to prove a large deviation principle for $\mathcal{L}_n$. Roughly speaking, this means that we will give an exponential equivalent for this conditional distribution. On a finer scale, we prove a large deviation principle for $\tilde{\mathcal{L}}_n := \mathcal{L}(\sqrt{\frac{a_n}{q_n}}(T_n - b_n)|S_n = p_n)$, where $b_n$ is a centering factor specified in Theorem 2.2 of Section 2.2.3 and $a_n$ is a decreasing positive sequence of real numbers with $a_n \to 0$, $a_n q_n \to +\infty$. We then say that $\tilde{\mathcal{L}}_n$ satisfies a moderate deviation principle [7], Section 3.7. Our work follows the nice ones of Janson [15,16]. In these last papers, a central limit theorem with moment convergence is proved. The starting point in the proof is a simple representation of the conditional characteristic function as an inverse Fourier transform. This representation was first given by Bartlett [4], Equation (16). To establish large and moderate deviation principles, we will make use of Gärtner–Ellis theorem in which an asymptotic evaluation of the Laplace transform is needed. For this purpose, we first transcribe the Bartlett formula to get a simple integral representation for the conditional Laplace transform (see Lemma 3.1). The main result of [16] is quite general as it only requires assumptions on the three first moments of $(\mathbf{X}, \mathbf{Y})$. Here, we need further assumptions. However, contrarily to [16], Section 2, we do not restrict to the central case (conditioning on $S_n = \mathbb{E}(S_n)$) nor on the "pseudo" central case (conditioning on $S_n = \mathbb{E}(S_n) + \mathcal{O}(\sqrt{q_n}\sigma_{X_i^{(n)}})$, with $\sigma_X^2 = \operatorname{var}(X)$). In [22], the authors study general saddle point approximations for multidimensional discrete empirical means and obtain an approximation formula for conditional probabilities. We focus here on the exponential part of this formula, stating a full large deviation principle (see Theorem 2.1). Using some classical tools of convex analysis we give an explicit natural and elegant form for the rate function. Furthermore, we complement our study by stating a moderate deviation principle for the conditional law (see Theorem 2.2). As usual, the rate function is quadratic and the scaling factor is the asymptotic variance, which can be interpreted here as a residual variance in some linear regression model, generalizing the factor found in [16] for the central case. The paper is organized as follows. In the next section, to be self contained, we first recall some classical results on large deviation principles. Then we state our main results: a large deviation principle and a moderate deviation principle for conditioned sums. Section 3 is devoted to the proofs. In Section 4, we will show a large deviation principle for the conditional empirical measure. It is easily derived from Theorem 2.1. In Section 5, we discuss the application of our main results to some combinatorial examples. In particular, we obtain the rate function for the moderate deviations in the context of random forests. We also discuss possible extensions to more general models.

# 2. Main results

## 2.1. Large and moderate deviations

### 2.1.1. *Some generalities*

Let us first recall what is a large deviation principle (LDP) (see, e.g., [7,13]). In the whole paper, $(a_n)$ is a decreasing positive sequence of real numbers with $\lim_{n\to\infty} a_n = 0$.

We say that a sequence $(R_n)$ of probability measures on a measurable Hausdorff space $(U, \mathcal{B}(U))$ satisfies a LDP with rate function $I$ and speed $(a_n)$ if:

(i) $I$ is lower semi continuous (l.s.c.), with values in $\mathbb{R}^+ \cup \{+\infty\}$.

(ii) For any measurable set $A$ of $U$:

$$-I(\operatorname{int} A) \le \liminf_{n\to\infty} a_n \log R_n(A) \le \limsup_{n\to\infty} a_n \log R_n(A) \le -I(\operatorname{clo} A),$$

where $I(A) = \inf_{\xi\in A} I(\xi)$ and $\operatorname{int} A$ (resp., $\operatorname{clo} A$) is the interior (resp., the closure) of $A$.

We say that the rate function $I$ is good if its level set $\{x \in U\colon\ I(x) \le a\}$ is compact for any $a \ge 0$. More generally, a sequence of $U$-valued random variables is said to satisfy a LDP if their distributions satisfy a LDP.

To be self-contained, we also recall some definitions and results which will be used in the sequel (we refer to [7,13] for more on large deviations).

*Laplace and Fenchel–Legendre transforms.* To begin with, let $Z$ be an integer random variable and define the span of $Z$ by $m_Z := \sup\{m \in \mathbb{N}, \exists b \in \mathbb{N}, \operatorname{Supp}(Z) \subset m\mathbb{Z} + b\}$. Let $\varphi_Z$ denote the characteristic function of $Z$. When $Z$ is square integrable, $\sigma_Z^2$ denotes its variance. For $\tau$ lying in $\operatorname{dom}\psi_Z := \{\tau \in \mathbb{R} : \mathbb{E}[\exp(\tau Z)] < +\infty\}$, we define $\psi_Z(\tau) := \ln \mathbb{E}[\exp(\tau Z)]$ as the cumulant generating function of $Z$. Obviously, $\psi_Z$ is analytic in the interior of $\operatorname{dom}\psi_Z$. We denote by $R_Z$ the interior of the range of $\psi'_Z$. It is well known that $R_Z$ is a subset of the interior of the convex hull of the support of $Z$. These two subsets of $\mathbb{R}$ coincide whenever $\psi_Z$ is essentially smooth (see definition below). Further, let $\psi_Z^*$ denote the Fenchel–Legendre transform of $\psi_Z$ [7], Definition 2.2.2, page 26. For any $\tau^* \in R_Z$, there exists a unique $\tau_{\tau^*} \in \operatorname{dom}\psi_Z$ such that $\psi'_Z(\tau_{\tau^*}) = \tau^*$ and we may define $Z^{*,\tau^*}$ as a r.v. on $\mathbb{N}$ having the following distribution

$$\mathbb{P}(Z^{*,\tau^*} = k) = \exp[k\tau_{\tau^*} - \psi_Z(\tau_{\tau^*})]\mathbb{P}(Z = k) \qquad (k \in \mathbb{N}). \tag{2.1}$$

It is well known that $\mathbb{E}(Z^{*,\tau^*}) = \tau^*$. For more details on the relationships between $\psi_Z$, $\psi_Z^*$, $Z^{*,\tau^*}$, we refer to the book [3].

Let now $(Z, W)$ be a random vector of $\mathbb{R}^2$. We naturally extend some of the previous notations to $(Z, W)$. For example, $\psi_{Z,W}$ is the cumulant generating function built on $(Z, W)$ defined on $\operatorname{dom}\psi_{Z,W} \subset \mathbb{R}^2$ and $\psi_{Z,W}^*$ denotes the Fenchel–Legendre transform of $(Z, W)$.

*Convex functions.* Let $f$ be a proper convex function on $\mathbb{R}^k$. That is $f$ is convex and valued in $\mathbb{R} \cup \{+\infty\}$. We say that $f$ is essentially smooth whenever it is differentiable on the nonempty interior of $\operatorname{dom} f$ and it is steep. That is, for any vector $c$ lying on the boundary of $\operatorname{dom} f$

$$\lim_{x\to c, x\in \operatorname{int}\operatorname{dom} f} \|\nabla f(x)\| = +\infty,$$

where $\nabla f(x)$ denotes the gradient of $f$ at point $x$.

*Gärtner–Ellis corollary.*

**Corollary 2.1 (Gärtner–Ellis [7], Theorem 2.3.6c, page 44).** *Let $(Z_n)$ be a sequence of random variables valued in $\mathbb{R}$, $(a_n)_n$ a decreasing positive sequence of real numbers with $\lim_{n\to\infty} a_n = 0$. Define $\Lambda_n(\theta) = \ln \mathbb{E}e^{\theta Z_n}$. Assume that*

1. *for all* $\theta \in \mathbb{R}$, $a_n \Lambda_n(\theta/a_n) \to \Lambda(\theta) \in ]-\infty, +\infty]$,
2. $0$ *lies in the interior of* $\operatorname{dom}(\Lambda(\theta))$ *and* $\Lambda(\theta)$ *is essentially smooth and lower semi continuous*.

*Then* $(Z_n)$ *satisfies a LDP with good rate function* $\Lambda^*$ *and speed* $a_n$.

## 2.2. Main results

### 2.2.1. *The model*

For $n \in \mathbb{N}^*$, let $(X^{(n)}, Y^{(n)})$ be a random vector with $X^{(n)} \in \mathbb{Z}$. We assume that $(X^{(n)}, Y^{(n)})$ converges in law to $(X, Y)$ where $X$ is a non-essentially constant integer valued r.v. Notice that it implies that $\psi_X$ is strictly convex, that $X^{(n)}$ is not essentially constant and that $\psi_{X^{(n)}}$ is strictly convex for $n$ large enough. Further let $((X_i^{(n)}, Y_i^{(n)}))_{1\le i\le n}$ be an i.i.d. sample having the same distribution as $(X^{(n)}, Y^{(n)})$.

Let, for $n \in \mathbb{N}^*$ and $q_n \in \mathbb{N}^*$, $S_n = X_1^{(n)} + \cdots + X_{q_n}^{(n)}$ and $T_n = Y_1^{(n)} + \cdots + Y_{q_n}^{(n)}$. In the whole paper, $p_n$ will be a sequence of integers such that $\mathbb{P}(S_n = p_n) > 0$.

### 2.2.2. *Large deviations*

**Theorem 2.1.** *Let* $p_n \in \mathbb{Z}$, $q_n \in \mathbb{N}^*$, $\lambda \in \mathbb{R}$ *such that* $p_n/q_n \in R_{X^{(n)}} \to \lambda \in R_X$. *Assume that*

1. *the functions* $\psi_{X,Y}$ *and* $\psi_{X_n,Y_n}$ *are essentially smooth*, *and let* $\tau$ *be the unique real such that* $\psi_X'(\tau) = \lambda$,
2. $\operatorname{dom}\psi_Y = \operatorname{dom}\psi_{Y^{(n)}} = \mathbb{R}$, *and* $\forall u \in \mathbb{R}$ $|\mathbb{E}[\mathrm{e}^{uY^{(n)}} - \mathrm{e}^{uY}]| \to 0$,
3. *there exists* $r > 0$ *such that* $I_\tau := [\tau - r, \tau + r] \subset (\operatorname{dom}\psi_X) \cap (\bigcap_{n\ge 1} \operatorname{dom}\psi_{X^{(n)}})$ *and*

$$\forall u \in \mathbb{R}, \forall s \in I_\tau, \qquad \sup_{t\in\mathbb{R}} \bigl|\mathbb{E}\bigl[\mathrm{e}^{(\mathrm{i}t+s)X^{(n)}+uY^{(n)}} - \mathrm{e}^{(\mathrm{i}t+s)X+uY}\bigr]\bigr| \to 0. \tag{2.2}$$

*Then the distribution of* $(T_n/q_n)$ *conditioned by the event* $\{S_n = p_n\}$ *satisfies a LDP with good rate function* $\psi_{X,Y}^*(\lambda, \cdot) - \psi_X^*(\lambda)$ *and speed* $(q_n)^{-1}$.

### 2.2.3. *Moderate deviations*

Let $\xi$ lying in the interior of $\operatorname{dom}\psi_X$ and consider the random vector $(\check{X}_\xi, \check{Y}_\xi)$ whose distribution is given, for any $k \in \mathbb{N}$ and real Borel set $A$, by

$$\mathbb{P}(\check{X}_\xi = k, \check{Y}_\xi \in A) = \exp[-\psi_X(\xi) + k\xi]\mathbb{P}(X = k, Y \in A). \tag{2.3}$$

We define in the same way the random vector $(\check{X}_\xi^{(n)}, \check{Y}_\xi^{(n)})$. Obviously, $\check{X}_\xi$ has the same distribution as $X^{*,\xi^*}$ with $\xi^* = \psi_X'(\xi)$. Further, let $\alpha_\xi^2$ be the variance of the residual $\check{\varepsilon}_\xi$ for the linear regression of $\check{Y}_\xi$ on $\check{X}_\xi$:

$$\check{\varepsilon}_\xi := \bigl(\check{Y}_\xi - \mathbb{E}(\check{Y}_\xi)\bigr) - \frac{\operatorname{cov}(\check{X}_\xi, \check{Y}_\xi)}{\operatorname{var}(\check{X}_\xi)}\bigl(\check{X}_\xi - \mathbb{E}(\check{X}_\xi)\bigr). \tag{2.4}$$

This variance is the asymptotic variance in the central limit theorem for such conditioned variable as it can be seen in Theorem 2.1 of [16]. The explicit form of $\alpha_\xi^2$ is then

$$\alpha_\xi^2 = \operatorname{var}(\check{Y}_\xi) - \operatorname{cov}(\check{X}_\xi \check{Y}_\xi)^2 / \operatorname{var}(\check{X}_\xi).$$

Then we get the following result.

**Theorem 2.2.** *Let $p_n \in \mathbb{Z}, q_n \in \mathbb{N}^*, \lambda \in \mathbb{R}$ such that $\lambda$ (resp., $p_n/q_n$) lies in $R_X$ (resp., $R_{X^{(n)}}$) and $p_n/q_n \to \lambda$. Assume that*

1. *the functions $\psi_{X,Y}$ and $\psi_{X_n,Y_n}$ are essentially smooth,*
2. *there exists $r_0 > 0$ such that*

$$B_0 := ]-r_0, r_0[ \subset (\operatorname{dom} \psi_Y) \cap \Bigl(\bigcap_{n\geq 1} \operatorname{dom} \psi_{Y^{(n)}}\Bigr),$$

*and let $\tau$ (resp., $\tau_n$) be the unique real such that $\psi'_X(\tau) = \lambda$ (resp., $\psi'_{X^{(n)}}(\tau_n) = p_n/q_n$),*

3. *there exists $r > 0$ such that $I_\tau := [\tau - r, \tau + r] \subset (\operatorname{dom} \psi_X) \cap (\bigcap_{n\geq 1} \operatorname{dom} \psi_{X^{(n)}})$ and*

$$\forall s \in I_\tau, \qquad \sup_{t\in\mathbb{R}} \bigl|\mathbb{E}\bigl[\mathrm{e}^{(\mathrm{i}t+s)X^{(n)}} - \mathrm{e}^{(\mathrm{i}t+s)X}\bigr]\bigr| \to 0, \tag{2.5}$$

$$\sup_n \sup_{(s,v)\in I_\tau\times B_0} \mathbb{E}\Bigl(\mathrm{e}^{sX^{(n)}+v(Y^{(n)}-\mathbb{E}(\check{Y}^{(n)}_{\tau_n}))}\Bigr) < \infty, \tag{2.6}$$

4. *$(a_n)$ satisfies $a_n q_n \to +\infty$.*

*Then the distribution of $(\sqrt{\frac{a_n}{q_n}}(T_n - q_n\mathbb{E}(\check{Y}^{(n)}_{\tau_n})))$ conditioned by the event $\{S_n = p_n\}$ satisfies a LDP with good rate function $J(\cdot) = \frac{(\cdot)^2}{2\alpha_\tau^2}$ and speed $a_n$.*

***Remark 2.1.*** As $(q_n)^{-1} = \mathrm{o}(a_n)$, we say that the distribution of $(\sqrt{\frac{a_n}{q_n}}(T_n - q_n\mathbb{E}(\check{Y}^{(n)}_{\tau_n})))$ conditioned by the event $\{S_n = p_n\}$ satisfies a moderate deviation principle (MDP).

**Corollary 2.2.** *If the set of assumptions of either Theorem 2.1 or Theorem 2.2 is satisfied, then $\mathcal{L}_n$ converges in distribution towards the degenerate distribution concentrated on $E(\check{Y}_\tau)$.*

# 3. Proofs

For $p_n \in \mathbb{Z}$, such that $\mathbb{P}(S_n = p_n) \neq 0$, let for $u \in \mathbb{R}$

$$f_n(u) := \frac{1}{q_n} \log \mathbb{E}[\exp(uT_n)|S_n = p_n] \in \mathbb{R} \cup \{+\infty\}. \tag{3.1}$$

In order to apply Gärtner–Ellis corollary, we have to prove that $f_n(u)$ converges when $n \to \infty$. The next two subsections yield a simple representation of $f_n(u)$ using the Fourier transform.

## 3.1. A simple representation using Fourier transform

Recall that we set $\varphi_Z(t) := \mathbb{E}(e^{itZ})$. An obvious but useful lemma follows.

**Lemma 3.1 (Bartlett's formula, see Equation (16) in [4]).** *Let $Z$ be an integer r.v. and $W$ be an integrable r.v. Then, for any nonnegative integer $k$ lying in the support of $Z$,*

$$\mathbb{E}[W|Z=k] = \frac{\int_{-\pi}^{\pi} \mathbb{E}[W \exp(\mathrm{i}tZ)] \exp(-\mathrm{i}kt)\,\mathrm{d}t}{\int_{-\pi}^{\pi} \varphi_Z(t) \exp(-\mathrm{i}kt)\,\mathrm{d}t}.$$

## 3.2. Some changes of probability

One of the main tool to prove large deviation results is the use of changes of probability. In this section, we review the different changes of probability used in this paper.

(a) Let $\lambda \in R_Z$. We define $\tau \in \operatorname{dom}\psi_Z$ by $\psi_Z'(\tau) = \lambda$. We then introduce $Z^{*,\lambda}$ as a random variable valued on $\mathbb{N}$:

$$\mathbb{P}(Z^{*,\lambda} = k) = \exp[k\tau - \psi_Z(\tau)]\mathbb{P}(Z = k) \qquad (k \in \mathbb{N}). \tag{3.2}$$

We have $\mathbb{E}(Z^{*,\lambda}) = \lambda$. This change of probability is quite classical in large deviation theory. In order to prove Lemma 3.3, we also define $Z^{(n),*,p_n/q_n}$, replacing $Z$ by $Z^{(n)}$, $\lambda$ by $p_n/q_n$ and $\tau$ by $\tau_n$. Then $\mathbb{E}(Z^{(n),*,p_n/q_n}) = p_n/q_n$, as needed to apply Lemma 3.2.

(b) For $u$ in $\operatorname{dom}\psi_Y$, define $\widehat{X}_u$ by

$$\mathbb{P}(\widehat{X_u} = k) = \exp[-\psi_Y(u)]\mathbb{E}\bigl[\exp(uY)\mathbb{1}_{\{X=k\}}\bigr]. \tag{3.3}$$

Similarly, replacing $(X, Y)$ by $(X^{(n)}, Y^{(n)})$, we define $\widehat{X}_u^{(n)}$. The r.v. $\widehat{X_u}$ and $\widehat{X}_u^{(n)}$ appear naturally when applying the inversion of Fourier transform in the proof of Theorem 2.1.

(c) For the moderate deviations, the asymptotic is different (see Theorem 2.2). Therefore the r.v.s $Y^{(n)}$ have to be centered. The centering factor and the rate function are closely related to the following change of probability. Let $\xi$ lying in the interior of $\operatorname{dom}\psi_X$ and consider the random vector $(\check{X}_\xi, \check{Y}_\xi)$ whose distribution is given, for any $k \in \mathbb{N}$ and real Borel set $A$, by

$$\mathbb{P}(\check{X}_\xi = k, \check{Y}_\xi \in A) = \exp[-\psi_X(\xi) + k\xi]\mathbb{P}(X = k, Y \in A). \tag{3.4}$$

We define in the same way the random vector $(\check{X}_\xi^{(n)}, \check{Y}_\xi^{(n)})$. Obviously, $\check{X}_\xi$ has the same distribution as $X^{*,\xi^*}$ with $\xi^* = \psi_X'(\xi)$. Further, let $\alpha_\xi^2$ be the variance of the residual $\check{\varepsilon}_\xi$ for the linear regression of $\check{Y}_\xi$ on $\check{X}_\xi$:

$$\check{\varepsilon}_\xi := \bigl(\check{Y}_\xi - \mathbb{E}(\check{Y}_\xi)\bigr) - \frac{\operatorname{cov}(\check{X}_\xi, \check{Y}_\xi)}{\operatorname{var}(\check{X}_\xi)}\bigl(\check{X}_\xi - \mathbb{E}(\check{X}_\xi)\bigr). \tag{3.5}$$

Note that $J(y) = \frac{y^2}{\alpha_\tau^2}$ is the rate function in Theorem 2.2. Moreover, the centering factor is $\mathbb{E}(\check{Y}^{(n)}_{\tau_n})$. Hence, the change of probability used in the proof of Theorem 2.1 (see change of probability (3.3) above) has to be modified, according to this centering factor. This leads to the change of probability (3.6) below.

(d) Let $\tau$ (resp., $\tau_n$) be such that $\psi'_X(\tau) = \lambda$ (resp., $\psi'_{X^{(n)}}(\tau_n) = p_n/q_n$). Define the random variable $\widehat{X^{(n)}_u}$ distributed on $\mathbb{N}$ by:

$$\mathbb{P}\big(\widehat{X^{(n)}_u} = k\big) = \mathrm{e}^{-\psi_{\widetilde{Y}^{(n)}}(u_n)}\mathbb{E}\big[\mathrm{e}^{u\widetilde{Y}^{(n)}/\sqrt{a_n q_n}}\mathbb{1}_{\{X^{(n)}=k\}}\big], \tag{3.6}$$

where $\widetilde{Y}^{(n)} = Y^{(n)} - \mathbb{E}(\check{Y}^{(n)}_{\tau_n})$, $u_n = u/\sqrt{a_n q_n}$.

## 3.3. Laplace lemmas

We begin this section with a variation on a lemma first due to Laplace, see [11]. To be self contained, we give also the sketch of its proof.

**Lemma 3.2.** *Let $p_n \in \mathbb{Z}$, $q_n \in \mathbb{N}^*$, $\lambda \in \mathbb{R}$. Let $Z$ be a non-essentially constant square integrable integer r.v. with span $m_Z \in \mathbb{N}^*$. Let $(Z^{(n)})$ be a sequence of nonnegative i.i.d. integer random variables also having span $m_n$. Let $Z^{(n)}_1, \dots, Z^{(n)}_n$ be an i.i.d. sample distributed as $Z^{(n)}$. Assume that*

1. $\|\varphi_{Z^{(n)}} - \varphi_Z\|_\infty \underset{n\to+\infty}{\longrightarrow} 0$,
2. *the means of $Z$ and $Z^{(n)}$ are equal respectively to $\lambda$ and $p_n/q_n$ with $p_n/q_n \underset{n\to+\infty}{\longrightarrow} \lambda$,*
3. $\sigma^2_{Z^{(n)}} \underset{n\to+\infty}{\longrightarrow} \sigma^2_Z$,
4. $\mathbb{E}(|Z^{(n)} - p_n/q_n|^3)$ *is uniformly bounded.*

*Then, when $n$ tends to infinity*

$$\mathbb{P}\left(\sum_{j=1}^{q_n} Z^{(n)}_j = p_n\right) = \frac{m}{\sqrt{2\pi q_n}\sigma_Z}\mathrm{e}^{-b^2/(2m^2)}\big(c_{n,m_n,b_n} + \mathrm{o}(1)\big) \tag{3.7}$$

*with* $\sup Z^{(n)} \subset m_n\mathbb{Z} + b_n$ *and $c_{n,m_n,b_n}$ is bounded.*

***Remark 3.1.*** Notice that Assumption 1 implies that $m_n = m_Z$ for $n$ large enough.

**Proof of Lemma 3.2.** The inversion of the Fourier transform yields

$$\mathbb{P}\left(\sum_{j=1}^{q_n} Z^{(n)}_j = p_n\right) = \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}p_n t}\varphi^{q_n}_{Z^{(n)}}(t)\frac{\mathrm{d}t}{2\pi} = \int_{-\pi}^{\pi} [\mathrm{e}^{-\mathrm{i}p_n/q_n t}\varphi_{Z^{(n)}}(t)]^{q_n}\frac{\mathrm{d}t}{2\pi}. \tag{3.8}$$

Recall that $\sup Z^{(n)} = m_n\mathbb{Z} + b_n$ and that $m_n = m_Z$ for $n$ large enough. Hence, without lost of generality we assume $m_n = m_Z = m$ and $b_n = b$ for all $n$. In the interval $[0, 2\pi[$, $\varphi_{Z^{(n)}}(t) = \varphi_Z(t) = 1$ if and only if $t = \frac{2l\pi}{m}$ for $l = 0, \dots, m-1$.

$$\mathbb{P}\left(\sum_{j=1}^{q_n} Z_j^{(n)} = p_n\right) = \int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i}p_n t}\varphi_{Z^{(n)}}^{q_n}(t)\frac{\mathrm{d}t}{2\pi} = \int_{-\pi/m}^{-\pi/m+2\pi} [\mathrm{e}^{-\mathrm{i}p_n/q_n t}\varphi_{Z^{(n)}}(t)]^{q_n}\frac{\mathrm{d}t}{2\pi}. \tag{3.9}$$

Set $f_n(t) := [\mathrm{e}^{-\mathrm{i}p_n/q_n t}\varphi_{Z^{(n)}}(t)]^{q_n}$.

$$\begin{aligned}\int_{-\pi/m}^{-\pi/m+2\pi} f_n(t) &= I_1 + \sum_{l=0}^{m-1} I_{2,l} + \sum_{l=0}^{m-2} I_{3,l} + I_4\\ &:= \int_{-\pi/m}^{-\delta} f_n(t) + \sum_{l=0}^{m-1}\int_{2l\pi/m-\delta}^{2l\pi/m+\delta} f_n(t) + \sum_{l=0}^{m-2}\int_{2l\pi/m+\delta}^{2(l+1)\pi/m-\delta} f_n(t) \\ &\quad + \int_{2(m-1)\pi/m+\delta}^{-\pi/m+2\pi} f_n(t).\end{aligned} \tag{3.10}$$

There exist nonnegative integer valued random variables $Y_n$ such that $Z^{(n)} = mY_n + b$. Set $u = t - 2\pi l/m$.

Then

$$\begin{aligned}I_{2,l} &= \frac{1}{2\pi}\int_{-\delta}^{\delta} \mathrm{e}^{\mathrm{i}(u+2l\pi/m)(bq_n-p_n)}\varphi_{Y_n}^{q_n}(mu)\,\mathrm{d}u\\ &= \frac{1}{2\pi}\mathrm{e}^{\mathrm{i}2l(\pi/m)(bq_n-p_n)}\int_{-\delta}^{\delta} \mathrm{e}^{\mathrm{i}u(bq_n-p_n)}\varphi_{Y_n}^{q_n}(mu)\,\mathrm{d}u.\end{aligned}$$

On one hand, using a Taylor expansion of order 2 of $\mathrm{e}^{\mathrm{i}tb}\mathrm{e}^{-p_n/q_n t}\varphi_{Z^{(n)}}(t) = \mathrm{e}^{\mathrm{i}tb}\varphi_{Z^{(n)}-p_n/q_n}(t)$, we get

$$\varphi_{Z^{(n)}-p_n/q_n}(t) = 1 - \frac{t^2}{2}\sigma_n^2 - t^3\mathbb{E}\left(\mathrm{i}\big(Z^{(n)} - p_n/q_n\big)^3\int_0^1 \frac{(1-y)^2}{2}\mathrm{e}^{\mathrm{i}yt(Z_n-p_n/q_n)}\,\mathrm{d}y\right),$$

where $\sigma_n^2$ states for $\sigma_{Z^{(n)}}^2$. Now as $\mathbb{E}(|Z^{(n)} - p_n/q_n|^3)$ is bounded and as $\sigma_n^2 \to \sigma_Z^2$ we can find a positive number (independent of $n$) $\delta < \pi$ such that for $|t| < \delta$ and for $n$ large enough

$$|\mathrm{e}^{-\mathrm{i}p_n/q_n t}\varphi_{Z^{(n)}}(t)| = |\varphi_{Z^{(n)}-p_n/q_n}(t)| \le 1 - \frac{\sigma_Z^2 t^2}{4}. \tag{3.11}$$

On the other hand, one has both

$$\xi_l := \sup_{2l\pi/m+\delta \le t \le 2(l+1)\pi/m-\delta} |\varphi_Z(t)| < 1$$

and

$$\xi_{n,l} := \sup_{2l\pi/m+\delta \le t \le 2(l+1)\pi/m-\delta} |\varphi_{Z^{(n)}}(t)| < 1.$$

Further, as $\|\varphi_{Z^{(n)}} - \varphi_Z\|_\infty \to 0$, we get $\xi_{n,l} \to \xi_l$. Let $\epsilon > 0$ be such that $\xi_l + \epsilon < 1$. For $n$ large enough $\xi_{n,l} \le \xi_l + \epsilon$. $|I_{3,l}|, l = 0, \dots, m-1$, are bounded by $(\xi_l + \epsilon)^{q_n}/(2\pi)$, hence are exponentially small. In a similar way, $|I_1|$ and $|I_4|$ are exponentially small. To deal with $I_1$, one performs the variable change $mu = \sqrt{q_n}\sigma_Z t$, and use both (3.11) and inequality $\log(1-\theta) \le -\theta$, ($\theta \in [0, 1[$) to conclude by using both central limit and Lebesgue theorems. It concludes the proof with $c_{n,m,b} = \sum_{l=0}^{m-1} \mathrm{e}^{\mathrm{i}2l(\pi/m)(bq_n - p_n)}$. □

We now give an extension of the previous lemma involving not only the probability for the sum to be equal to the mean of $Z^{(n)}$ but equal to any *good* rational number.

**Lemma 3.3.** *Let $Z$ be a nondegenerated integer r.v. with span $m \in \mathbb{N}^*$. Let $(Z_j^{(n)})_j$ be a sequence of i.i.d. integer random variables. Let $p_n \in \mathbb{Z}$, $q_n \in \mathbb{N}^*$, $\lambda \in \mathbb{R}$ such that $p_n/q_n \in R_{Z^{(n)}} \to \lambda \in R_Z$. Let $\tau$ (resp., $\tau_n$) be the unique real such that $\psi'_Z(\tau) = \lambda$ (resp., $\psi'_{Z^{(n)}}(\tau_n) = p_n/q_n$). We make the following assumptions.*

1. *There exists $r > 0$ such that $I_\tau := [\tau - r, \tau + r] \subset (\bigcap_{n\ge1} \operatorname{dom} \psi_{Z^{(n)}}) \cap (\operatorname{dom} \psi_Z)$.*
2. $$\forall s \in I_\tau, \qquad \lim_n \sup_{t\in\mathbb{R}} \left|\mathbb{E}\left[\left(\mathrm{e}^{(\mathrm{i}t+s)Z^{(n)}} - \mathrm{e}^{(\mathrm{i}t+s)Z}\right)\right]\right| = 0. \tag{3.12}$$

*Then, when $n$ goes to infinity*

$$\mathbb{P}\left(\sum_{j=1}^{q_n} Z_j^{(n)} = p_n\right) = \mathrm{e}^{-q_n \psi^*_{Z^{(n)}}(p_n/q_n)} \frac{m}{\sqrt{2\pi q_n}\sigma_{Z^{*,\lambda}}} \mathrm{e}^{-b^2/(2m^2)}\left(c_{n,m,b} + \mathrm{o}(1)\right), \tag{3.13}$$

*where $\sigma^2_{Z^{*,\lambda}}$ is the variance of $Z^{*,\lambda}$ defined in (2.1) and* $\sup Z \subset m\mathbb{Z} + b$.

**Proof.** Using the multinomial formula, we may write

$$\mathbb{P}\left(\sum_{j=1}^{q_n} Z_j^{(n),*,p_n/q_n} = p_n\right) = \mathrm{e}^{q_n((p_n/q_n)\tau_n - \psi_{Z^{(n)}}(\tau_n))}\mathbb{P}\left(\sum_{j=1}^{q_n} Z_j^{(n)} = p_n\right),$$

where $\tau_n$ is the unique real such that $\psi'_{Z^{(n)}}(\tau_n) = p_n/q_n$. Hence,

$$\mathbb{P}\left(\sum_{j=1}^{q_n} Z_j^{(n)} = p_n\right) = \mathrm{e}^{-q_n \psi^*_{Z^{(n)}}(p_n/q_n)}\mathbb{P}\left(\sum_{j=1}^{q_n} Z_j^{(n),*,p_n/q_n} = p_n\right),$$

where $Z_1^{(n),*,p_n/q_n}, \dots, Z_n^{(n),*,\lambda}$ are i.i.d. r.v.s having the distribution defined by

$$\mathbb{P}\left(Z^{(n),*,p_n/q_n} = k\right) = \exp\left(k\tau_n - \psi_{Z_1^{(n)}}(\tau_n)\right)\mathbb{P}\left(Z_1^{(n)} = k\right).$$

Further, the expectation of this last r.v. is $p_n/q_n$. Let us now check the assumptions of Lemma 3.2.

- Assumption 2 of Lemma 3.2 is satisfied by construction of $Z^{(n),*,p_n/q_n}$.
- Let us prove that $\mathbb{E}|Z^{(n),*,p_n/q_n} - p_n/q_n|^3$ is bounded. Using Hölder inequality, we get that

$$\begin{aligned}
&\mathbb{E}\big|Z^{(n),*,p_n/q_n} - p_n/q_n\big|^3 \\
&\quad = \sum_{k=0}^{\infty}\left|k - \frac{p_n}{q_n}\right|^3 \mathrm{e}^{k\tau_n - \psi_{Z^{(n)}}(\tau_n)}\mathbb{P}\big(Z_1^{(n)} = k\big) \\
&\quad = \mathrm{e}^{\psi^*_{Z^{(n)}}(p_n/q_n)}\mathbb{E}\left(\left|Z^{(n)} - \frac{p_n}{q_n}\right|^3 \mathrm{e}^{\tau_n(Z^{(n)} - p_n/q_n)}\right) \\
&\quad \le \mathrm{e}^{\psi^*_{Z^{(n)}}(p_n/q_n)}\left(\mathbb{E}\left(\left(Z^{(n)} - \frac{p_n}{q_n}\right)^4 \mathrm{e}^{\tau_n(Z^{(n)} - p_n/q_n)}\right)\right)^{3/4}\big(\mathbb{E}\big(\mathrm{e}^{\tau_n(Z^{(n)} - p_n/q_n)}\big)\big)^{1/4}.
\end{aligned}$$

  Using classical arguments on convex functions [23], we get that $\tau_n \underset{n\to+\infty}{\longrightarrow} \tau$. Hence, by Assumptions 1 and 2 of Lemma 3.3 we get that $\mathbb{E}|Z^{(n),*,\lambda} - \lambda|^3$ is bounded.
- Similar arguments yield that

$$\|\varphi_{Z_1^{(n),*,p_n/q_n}} - \varphi_{Z^{*,\lambda}}\| \to 0 \tag{3.14}$$

  and that $\sigma^2_{Z_n^*} \to \sigma^2_{Z^*}$.

Hence, all the assumptions of Lemma 3.2 are satisfied and we may conclude using Lemma 3.2. □

## 3.4. Proof of Theorem 2.1

Let, for $t \in \mathbb{R}$ and $u \in \mathbb{R}$,

$$\Phi_{X^{(n)},Y^{(n)}}(t,u) := \mathbb{E}\big(\exp\big[\mathrm{i}tX^{(n)} + uY^{(n)}\big]\big).$$

On one hand, using Lemma 3.1, we may write, for $u \in \mathbb{R}$ and $n$ large enough,

$$f_n(u) = \frac{1}{q_n}\log\frac{\int_{-\pi}^{\pi}\mathrm{e}^{-\mathrm{i}p_n t}\Phi^{q_n}_{X^{(n)},Y^{(n)}}(t,u)\,\mathrm{d}t}{\int_{-\pi}^{\pi}\mathrm{e}^{-\mathrm{i}p_n t}\Phi^{q_n}_{X^{(n)},Y^{(n)}}(t,0)\,\mathrm{d}t}. \tag{3.15}$$

Using twice equation (3.8), we may rewrite (3.15) as

$$f_n(u) = \frac{1}{q_n}\left[\log\mathbb{P}\left(\sum_{j=1}^{q_n}\widehat{X}^{(n)}_{u,j} = p_n\right) - \log\mathbb{P}(S_n = p_n)\right] + \psi_{Y^{(n)}}(u), \tag{3.16}$$

where $\widehat{X}_{u,1}^{(n)}, \dots, \widehat{X}_{u,n}^{(n)}$ are independent copies of $\widehat{X}_u^{(n)}$ defined in Section 3.2 by equation (3.3). In order to apply Lemma 3.3 to $\widehat{X}_{u,i}^{(n)}$, let us prove that

$$\forall s \in I_\tau, \qquad \lim_n \sup_{t\in\mathbb{R}} \bigl|\mathbb{E}\bigl[\bigl(\mathrm{e}^{(\mathrm{i}t+s)\widehat{X}_u^{(n)}} - \mathrm{e}^{(\mathrm{i}t+s)\widehat{X}_u}\bigr)\bigr]\bigr| = 0. \tag{3.17}$$

We have, for all $s \in I_\tau$,

$$\begin{aligned}
&\bigl|\mathbb{E}\bigl[\bigl(\mathrm{e}^{(\mathrm{i}t+s)\widehat{X}_u^{(n)}} - \mathrm{e}^{(\mathrm{i}t+s)\widehat{X}_u}\bigr)\bigr]\bigr| \\
&\quad = \bigl|\mathbb{E}\bigl[\bigl(\mathrm{e}^{(\mathrm{i}t+s)X^{(n)}}\mathrm{e}^{uY^{(n)}}\mathrm{e}^{-\psi_Y^{(n)}(u)} - \mathrm{e}^{(\mathrm{i}t+s)X}\mathrm{e}^{uY}\mathrm{e}^{-\psi_Y(u)}\bigr)\bigr]\bigr|.
\end{aligned}$$

Now by Assumption 2 of Theorem 2.1, we have $\mathrm{e}^{-\psi_Y^{(n)}(u)} = \mathrm{e}^{-\psi_Y(u)} + \epsilon_n$, with $\epsilon_n \to 0$. Hence, we have

$$\begin{aligned}
&\bigl|\mathbb{E}\bigl[\bigl(\mathrm{e}^{(\mathrm{i}t+s)\widehat{X}_u^{(n)}} - \mathrm{e}^{(\mathrm{i}t+s)\widehat{X}_u}\bigr)\bigr]\bigr| \\
&\quad \le \mathrm{e}^{-\psi_Y(u)}\bigl|\mathbb{E}\bigl[\mathrm{e}^{uY^{(n)}}\mathrm{e}^{(s+\mathrm{i}t)X^{(n)}} - \mathrm{e}^{uY}\mathrm{e}^{(s+\mathrm{i}t)X}\bigr]\bigr| + \epsilon_n\bigl|\mathbb{E}\bigl[\bigl(\mathrm{e}^{(\mathrm{i}t+s)X^{(n)}}\mathrm{e}^{uY^{(n)}}\bigr)\bigr]\bigr|.
\end{aligned}$$

The right-hand side of this last inequality goes to 0 by Assumption 3 of Theorem 2.1. It remains to prove that $p_n/q_n$ (resp., $\lambda$) belongs to $R_{\hat{X}_u^{(n)}}$ (resp., $R_{\hat{X}_u}$). Since $\psi_{X,Y}$ is essentially smooth, using Proposition 9.7 in [2] $R_X$ (resp., $R_{\hat{X}_u}$) is the interior of the convex hull of the support of $X$ (resp., $\hat{X}_u$) since $X$ and $X_u$ have the same support we have $R_{\hat{X}_u} = R_X$. In a similar way, we can prove that $R_{\hat{X}_u^{(n)}} = R_{\hat{X}^{(n)}}$. Hence, we have $p_n/q_n \in R_{\hat{X}_u^{(n)}}$. Applying Lemma 3.3 we obtain, for $u \in \mathbb{R}$

$$f(u) := \lim_{n\to\infty} \frac{1}{q_n} \log \mathbb{E}[\exp(uT_n)|S_n = p_n] = -[\psi^*_{\widehat{X}_u}(\lambda) - \psi_Y(u) - \psi^*_X(\lambda)]. \tag{3.18}$$

The convex dual function $f^*$ of $f$ is given by

$$\begin{aligned}
f^*(y) &:= \sup_{u\in\mathbb{R}}[uy - f(u)] \\
&= \sup_{u\in\mathbb{R}}\bigl(uy + [\psi^*_{\widehat{X}_u}(\lambda) - \psi_Y(u)]\bigr) - \psi^*_X(\lambda) \\
&= \sup_{(\xi,u)\in\operatorname{dom}\psi_{X,Y}}\left[uy + \xi\frac{p}{q} - \psi_{X,Y}(\xi,u)\right] - \psi^*_X(\lambda) \\
&= \psi^*_{X,Y}\left(\frac{p}{q}, y\right) - \psi^*_X(\lambda).
\end{aligned} \tag{3.19}$$

As $\psi_{X,Y}$ is essentially smooth, using Theorem 26.3 in [23], we deduce that $\psi^*_{X,Y}$ is essentially strictly convex. Hence, using once more Theorem 26.3 in [23], we may deduce that $f$ is essentially smooth. Therefore, we can apply Gärtner–Ellis Corollary 2.1 (see Theorem 2.3.6(c) in [7]) and conclude.

## 3.5. Proof of Theorem 2.2

Let $\tilde{T}_n = T_n - q_n \mathbb{E}(\check{Y}_{\tau_n}^{(n)})$ and

$$g_n(u) = a_n \log\bigl(\mathbb{E}(\mathrm{e}^{\tilde{T}_n u/\sqrt{a_n q_n}} | S_n = p_n)\bigr).$$

Proceeding as in the proof of Theorem 2.1, we have

$$\begin{aligned} g_n(u) &= a_n \log \frac{\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i} p_n t} \Phi^{q_n}_{X^{(n)}, Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}(t, u/\sqrt{a_n q_n})}{\int_{-\pi}^{\pi} \mathrm{e}^{-\mathrm{i} p_n t} \Phi^{q_n}_{X^{(n)}, Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}(t, 0)} \\ &= a_n \left( \log \mathbb{P}\left( \sum_{j=1}^{q_n} \widehat{\widehat{X_{u,j}^{(n)}}} = p_n \right) - \log \mathbb{P}(S_n = p_n) \right) + a_n q_n \psi_{Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}\bigl(u/\sqrt{a_n q_n}\bigr), \end{aligned}$$

where $\widehat{\widehat{X_{u,j}^{(n)}}}$ are i.i.d. r.v. on $\mathbb{Z}$ with distribution defined in Section 3.2 by Equation (3.6).

In order to use Lemma 3.3, we first have to prove that

$$\forall s \in I_\tau, \qquad \lim_n \sup_{t \in \mathbb{R}} \bigl| \mathbb{E}\bigl[\mathrm{e}^{(\mathrm{i}t+s)\widehat{\widehat{X_u^{(n)}}}} - \mathrm{e}^{(\mathrm{i}t+s)\widehat{\widehat{X_u}}}\bigr]\bigr| = 0. \tag{3.20}$$

We have

$$\bigl|\mathbb{E}\bigl[\mathrm{e}^{(\mathrm{i}t+s)\widehat{\widehat{X_u^{(n)}}}} - \mathrm{e}^{(\mathrm{i}t+s)\widehat{\widehat{X_u}}}\bigr]\bigr| \le C \mathrm{e}^{-\psi_{Y - \mathbb{E}(\check{Y}_{\xi_0})}(0)} \bigl|\mathbb{E}\bigl[\mathrm{e}^{(s+\mathrm{i}t)X^{(n)}} - \mathrm{e}^{(s+\mathrm{i}t)X}\bigr]\bigr|,$$

which goes to zero by assumption (2.5). As in the proof of Theorem 2.1 it is easy to prove that $R_{\widehat{\widehat{X_u}}} = R_X$ and $R_{\widehat{\widehat{X_u^{(n)}}}} = R_{X^{(n)}}$.

Using Lemma 3.3 we obtain, for $u \in \mathbb{R}$

$$g_n(u) \overset{n \to +\infty}{\sim} -q_n a_n \bigl[\psi^*_{\widehat{\widehat{X_u^{(n)}}}}(p_n/q_n) - \psi_{Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}\bigl(u/\sqrt{a_n q_n}\bigr) - \psi^*_{X^{(n)}}(p_n/q_n)\bigr].$$

Define

$$H_n(h) = \sup_{\xi \in \mathbb{R}} \left[ \xi \frac{p_n}{q_n} - \psi_{X^{(n)}, Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}(\xi, h) \right].$$

As

$$\psi^*_{\widehat{\widehat{X_u^{(n)}}}}(p_n/q_n) = \sup_x \left( \frac{p_n}{q_n} x - \psi_{\widehat{X}_u^{(n)}}(x) \right),$$

and

$$\psi_{\widehat{X}_u^{(n)}}(x) = \psi_{X^{(n)}, Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}\bigl(x, u/\sqrt{a_n q_n}\bigr) - \psi_{Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}\bigl(u/\sqrt{a_n q_n}\bigr),$$

we get

$$g_n(u) \overset{n\to+\infty}{\sim} -q_n n a_n\Big(H_n\big(u/\sqrt{a_n q_n}\big) - H_n(0)\Big). \tag{3.21}$$

We claim that if $\lim_n h_n = 0$, then

$$\lim_n \frac{H_n(h_n) - H_n(0)}{h_n^2} = -\frac{\alpha_{\tau_n}^2}{2} + \mathrm{O}(1). \tag{3.22}$$

Assuming that (3.22) is true, and as $\alpha_{\tau_n}^2 \to \alpha_\tau^2$, we get that

$$\lim_n g_n(u) = -u^2 \frac{\alpha_\tau^2}{2}.$$

We easily conclude, since

$$g^*(y) = \sup_u \Big\{ uy + \lim_n g_n(u) \Big\} = \frac{y^2}{2\alpha_\tau^2}.$$

It remains to prove that (3.22) is true. Recall that

$$H_n(h) = \sup_{\xi\in\mathbb{R}} \bigg[ \xi \frac{p_n}{q_n} - \psi_{X^{(n)}, Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}(\xi, h) \bigg].$$

In the sequel, $\psi_x'$ (resp., $\psi_y'$) will denote the partial derivative of $\psi_{X^{(n)}, Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}(\xi, h)$ with respect to the first (resp., second) variable. On one hand, by assumption (2.6), we can define on $I_\tau \times B_0$ the function $F_n$ by:

$$F_n(\xi, h) = \psi_x'(\xi, h) - p_n/q_n.$$

We then deduce from the implicit function theorem that there exists a neighborhood of $(\tau_n, 0)$ on which:

$$H_n(h) = \xi_n(h) \frac{p_n}{q_n} - \psi_{X^{(n)}, Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)})}(\xi_n(h), h),$$

with

$$\xi_n'(h) = -\frac{\psi_{xy}''(\xi_n(h), h)}{\psi_{xx}''(\xi_n(h), h)}.$$

We can then compute the derivatives of $H_n$ (in the sequel we omit the argument $(\xi_n(h), h)$ in the derivatives). We have (with obvious notations)

$$H_n'(h) = -\psi_y',$$

$$H_n''(h) = \frac{(\psi_{x,y}'')^2}{\psi_{x,x}''} - \psi_{y,y}'',$$

$$H_n^{(3)}(h) = \left(\frac{\psi_{x,y}''}{\psi_{x,x}''}\right)^3 \psi_{x,x,x}^{(3)} - 3\left(\frac{\psi_{x,y}''}{\psi_{x,x}''}\right)^2 \psi_{x,x,y}^{(3)} + 3\frac{\psi_{x,y}''}{\psi_{x,x}''} \psi_{x,y,y}^{(3)} - \psi_{y,y,y}^{(3)}.$$

Replacing the partial derivative of $\psi$ by its expression, we get

$$H_n'(0) = 0 \quad \text{and} \quad H_n''(0) = -\alpha_{\tau_n}^2.$$

On the other hand, using a Taylor expansion, we get

$$H_n(h_n) - H_n(0) = h_n H_n'(0) + \frac{h_n^2}{2} H_n''(0) + \frac{h_n^3}{6} H_n^{(3)}(z_n), \qquad z_n \in [0, h_n]. \tag{3.23}$$

Hence, (3.23) becomes

$$H_n(h_n) - H_n(0) = -\frac{h_n^2 \alpha_{\tau_n}^2}{2} + \frac{h_n^3}{6} H_n^{(3)}(z_n), \qquad z_n \in [0, h_n]. \tag{3.24}$$

Now the expression of $H_n^{(3)}$ is a rational fraction of some partial derivatives of $\mathbb{E}(\mathrm{e}^{\xi X^{(n)} + h(Y^{(n)} - \mathbb{E}(\check{Y}_{\tau_n}^{(n)}))})$. The denominator of this rational fraction is bounded away from 0 as it converges to a variance and numerator is bounded by (2.6). Hence, $H_n^{(3)}$ is bounded and the claim is proved.

# 4. LDP for the conditional empirical measure

Let $L_n := \frac{1}{q_n} \sum_{j=1}^{q_n} \delta_{Y_j^{(n)}}$. In this section, we will give a LDP for the law of $L_n$ conditioned on the event $\{S_n = p_n\}$.

**Lemma 4.1 (Exponential tightness).** *Under the Assumptions of Theorem 2.1, the law of $L_n$ conditioned on the event $\{S_n = p_n\}$ is exponentially tight.*

**Proof.** We set $|T|_n = \sum_{i=1}^{q_n} |Y_i^{(n)}|$. For $\delta > 0$ set $K_\delta = \{\nu \in M_1(\mathbb{R}), \int |x| \nu(\mathrm{d}x) \le \delta\}$. Now for any $u > 0$

$$\mathbb{P}(L_n \in K_\delta^c | S_n = p_n) = \mathbb{P}\left(\sum_{i=1}^{q_n} \left|Y_i^{(n)}\right| > \delta q_n \middle| S_n = p_n\right) \le \mathrm{e}^{-q_n(\delta u - g_n(u))},$$

where

$$g_n(u) := \frac{1}{q_n} \log \mathbb{E}[\exp(u|T|_n) | S_n = p_n] \in \mathbb{R} \cup \{+\infty\}.$$

The last inequality is obtained using Markov inequality (note that the Assumptions of Theorem 2.1 imply that $\mathbb{E}(\mathrm{e}^{uY^{(n)}}) < \infty, \forall u$).

Now we apply Theorem 2.1 to $|Y_i^{(n)}|$ and get that $\lim_n g_{n(u)} := g(u)$ where $g(u)$ is the analogue of $f(u)$ see (3.18) when $Y_i^{(n)}$ is replaced by $|Y_i^{(n)}|$. This concludes the proof of Lemma 4.1. □

**Theorem 4.1.** *Under the Assumptions of Theorem 2.1 the law of $L_n$ conditioned on the event $\{S_n = p_n\}$ satisfies a LDP with good rate function*

$$\Lambda^*(\nu) = \sup_{\varphi \in C_b(\mathbb{R})} \{\langle \varphi, \nu \rangle - \Lambda(\varphi)\}, \qquad \nu \in M_1(\mathbb{R}),$$

*where, for $\varphi \in C_b(\mathbb{R})$,*

$$\Lambda(\varphi) := \log \mathbb{E}\big(\mathrm{e}^{\langle \varphi, \delta_{Y_1^n} \rangle} | S_n = p_n\big).$$

**Proof.** We proceed exactly as for the classical proof of Sanov theorem (see, e.g., Theorem 6.2.10 in [7]). The finite dimensional LDP is a straightforward application of Theorem 2.1, and the exponential tightness has been proved in Lemma 4.1. □

# 5. Examples

In this section, we give two examples of applications and one counter example. These examples are borrowed from [16].

## 5.1. Occupancy problem

In the classical occupancy problem (see [16] and the references therein for more details), $m$ balls are distributed at random into $N$ urns. The resulting numbers of balls $Z_1, \dots, Z_N$ have a multinomial distribution, and it is well known that this equals the distribution of $(X_1, \dots, X_N)$ conditioned on $\sum_{i=1}^N X_i = m$, where $X_1, \dots, X_N$ are i.i.d. with $X_i \sim \mathcal{P}(\lambda^*)$, for an arbitrary $\lambda^* > 0$. The classical occupancy problem studies the number $W$ of empty urns; this is thus $\sum_{i=1}^N \mathbb{1}_{\{X_i=0\}}$ conditioned on $\sum_{i=1}^N X_i = m$.

Now suppose that $m = p_n \to \infty$ and $N = q_n \to \infty$ with $\frac{p_n}{q_n} \to \lambda$. Take $X_i^{(n)} \sim \mathcal{P}(\lambda_n^*)$. Note that we do not assume that $\lambda_n^* = p_n/q_n$ and $\lambda^* = \lambda$ which is the case in Janson's work. It is easy to see that Assumption 3 of Theorem 2.1 is fulfilled and that $\psi_{X,Y}$ is essentially smooth. Moreover, for $(x, y) \in \mathbb{R}^2$ and $\lambda > 0$, we have

$$\psi_X(x) = -\lambda^* + \lambda^* \mathrm{e}^x,$$

$$\psi_X^*(\lambda) = \lambda \log\left(\frac{\lambda}{\lambda^*}\right) + \lambda^* - \lambda,$$

$$\psi_{X,Y}(x, y) = -\lambda^* + \log\big(\mathrm{e}^{\lambda^* \exp(x)} - 1 + \mathrm{e}^y\big).$$

Hence, we can apply Theorem 2.1. Here the function $\psi_{X,Y}^*$ does not have any explicit form. Assumptions of Theorem 2.2 are obviously fulfilled. We have

$$\mathbb{E}\big(\check{Y}_{\tau_n}^{(n)}\big) = \mathrm{e}^{-\lambda_n^* \exp(\tau_n)},$$

$$\mathbb{P}(\check{X}_\tau = k) = \mathrm{e}^{-\lambda^* \exp(\tau)} (\mathrm{e}^\tau \lambda^*)^k / k!.$$

Hence, $\check{X}_\tau$ is Poisson with parameter $\lambda e^\tau$. An easy calculation gives

$$\begin{aligned}\operatorname{cov}(\check{X}_\tau, \check{Y}_\tau) &= -\lambda^* e^\tau e^{-\lambda^* \exp(\tau)},\\ \operatorname{var}(\check{Y}_\tau) &= e^{-\lambda^* \exp(\tau)}\bigl(1 - e^{-\lambda^* \exp(\tau)}\bigr).\end{aligned}$$

Hence, $\alpha_\tau^2 = e^{-\lambda^* \exp(\tau)}(1 - e^{-\lambda^* \exp(\tau)} + \lambda^* e^\tau e^{-\lambda^* \exp(\tau)})$. Now, as $\tau = \log(\frac{\lambda}{\lambda^*})$, we get $J(\cdot) = \frac{(\cdot)^2 e^{\lambda^*}}{1 - e^{-\lambda^*} + \lambda^* e^{-\lambda^*}}$ in the particular case where $\lambda^* = \lambda$. Notice that a functional LDP is given in [5].

***Remark 5.1.*** Theorem 2.1 allows us to deal with other statistics than $\sum_{i=1}^n \mathbb{1}_{\{X_i=0\}}$. For example, statistics of the form $\sum_{i=1}^n f(X_i, Z_i)$ where $Z_1, \dots, Z_n$ are i.i.d. and independent from $X_1, \dots, X_n$. Let us describe the particular case of bootstrap see [8]. Let $Z_1, \dots, Z_n$ be i.i.d. real valued random variables. We choose from $(Z_1, \dots, Z_n)$ at random with replacement a $m$ sample $Z_1^*, \dots, Z_m^*$. Then $\sum_{i=1}^m f(Z_i^*)$ is distributed as $\sum_{i=1}^n X_i f(Z_i)$ conditioned on $\sum_{i=1}^n X_i = m$, where $X_i \sim \mathcal{P}(\lambda), i = 1, \dots, n$, for any $\lambda > 0$. Hence, we get the same kind of conditioning as for the occupancy problem.

## 5.2. Branching processes

Consider a Galton–Watson process, beginning with one individual, where the number of children of an individual is given by a random variable $X$ having finite moments. Assume further that $\mathbb{E}(X) = 1$. We number the individuals as they appear. Let $X_i$ be the number of children of the $i$th individual. It is well known (see Example 3.4 in [16] and the references therein) that the total progeny is $n \geq 1$ if and only if

$$S_k := \sum_{i=1}^k X_i \geq k \qquad \text{for } 0 \leq k < n \qquad \text{but} \qquad S_n = n - 1. \tag{5.1}$$

This type of conditioning is different from the one studied in the present paper, but Janson proves [16], Example 3.4, that if we ignore the order of $X_1, \dots, X_n$, conditioning on (5.1) is equivalent to conditioning on $S_n = n - 1$. Hence, we can study variables of the kind $Y_i = f(X_i)$. Considering the case where $Y_i = \mathbb{1}_{\{X_i=3\}}$, the $\sum_{i=1}^n Y_i$ is the number of families with three children. Now choosing $X_i \sim \mathcal{P}(\lambda)$, we compute the rate function as in Example 5.1.

## 5.3. Random forests

*The model*

Consider a uniformly distributed random labelled rooted forest with $m$ vertices and $N < m$ roots. Without loss of generality, we may assume that the vertices are $1, \dots, m$ and, by symmetry, that the roots are the $N$ first vertices (i.e., $1, \dots, N$). Following [16], this model can be realized as follows:

1. The sizes of the $N$ trees in the forest are distributed as $X_1, \dots, X_N$ conditioned on $\sum_{i=1}^N X_i = m$, where $X_i$ are i.i.d. with the Borel distribution for some arbitrary $\mu \in )0, 1/\mathrm{e}]$. Recall that $X$ has the Borel distribution with parameter $\mu \in ]0, 1/e]$ if

$$\mathbb{P}(X = l) = \frac{1}{T(\mu)} \frac{l^{l-1}}{l!} \mu^l, \qquad l \in \mathbb{N}^*, z \in ]0, 1/\mathrm{e}], \tag{5.2}$$

   where $T(\mu) = \sum_{l=1}^{\infty} \frac{l^{l-1}}{l!} \mu^l$ is the well-known tree function. In the following, we use classical properties of $T$ (see, e.g., [10]).
2. Further the tree number $i$ is drawn uniformly among the trees of size $X_i$.

A classical quantity of interest is the number of trees of size $K$ in the forest (see, e.g., [18,20, 21]). It means we choose $Y_i = \mathbb{1}_{\{X_i = K\}}$. Let us now assume that we conditione on $\sum_{i=1}^N X_i = m$ with $m = p_n \to +\infty$, $N = q_n \to +\infty$ with $p_n/q_n \to \lambda$. Take $X_i^{(n)} \sim \mathrm{Bo}(\mu_n)$ with $\mu_n \to \mu \in ]0, 1/\mathrm{e}]$. It is easy to see that the assumptions of Theorem 2.1 are satisfied. We do not have any simple expression for $\psi_{X,Y}^*$ and hence for the rate function. Concerning moderate deviations, it is easier to see that the assumptions of Theorem 2.2 are satisfied. We have

$$\mathbb{E}\big(\check{Y}_{\tau_n}^{(n)}\big) = \mathrm{e}^{-\psi_X^{(n)}(\tau_n)} \mathrm{e}^{K\tau_n} \mathbb{P}\big(X^{(n)} = K\big),$$
$$\mathbb{P}(\check{X}_\tau = k) = \mathrm{e}^{-\psi_X(\tau)} \mathrm{e}^{k\tau} \mathbb{P}(X = k).$$

A calculation gives

$$\mathrm{cov}(\check{X}_\tau, \check{Y}_\tau) = \mathrm{e}^{-\psi_X(\tau)} \mathrm{e}^{K\tau} \mathbb{P}(X = K) \left( K - \sum_{j=1}^{+\infty} j \mathrm{e}^{-\psi_X(\tau)} \mathrm{e}^{j\tau} \mathbb{P}(X = j) \right),$$
$$\mathrm{var}(\check{Y}_\tau) = \sum_{j=1}^{+\infty} j^2 \mathrm{e}^{-\psi_X(\tau)} \mathrm{e}^{j\tau} \mathbb{P}(X = j) - \left( \sum_{j=1}^{+\infty} j \mathrm{e}^{-\psi_X(\tau)} \mathrm{e}^{j\tau} \mathbb{P}(X = j) \right)^2.$$

To give an explicit formula for the rate function let us consider the particular case where $\mu = \lambda$. In that case, we have $\alpha_\tau^2 = \mathrm{e}^{-1} \mathrm{e}^{1/\mu^2} (1 - \mathrm{e}^{-1} \mathrm{e}^{1/\mu^2} + \frac{1}{\mu^3(\mu+1)})$. Now applying Theorem 2.2, we get $J(\cdot) = \frac{(\cdot)^2}{2\alpha_\tau^2}$.

## 5.4. Hashing

*The model*

Hashing with linear probing can be regarded as throwing $n$ balls sequentially into $m$ urns at random; the urns are arranged in a circle and a ball that lands in an occupied urn is moved to the next empty urn, always moving in a fixed direction. The length of the move is called the displacement of the ball, and we are interested in the sum of all displacements which is a random variable noted $d_{m,n}$. We assume $n < m$.

After throwing all balls, there are $N = m - n$ empty urns. These divide the occupied urns into blocks of consecutive urns. For convenience, we consider the empty urn following a block as belonging to this block. Janson [15] proved that the length of the blocks (counting the empty urn) and the sum of displacements inside each block are distributed as $(X_1, Y_1), \dots, (X_N, Y_N)$ $(N = m - n)$ conditioned on $\sum_{i=1}^N X_i = m$, where $(X_i, Y_i)$ are i.i.d. copies of a pair $(X, Y)$ of random variables. $X$ has the Borel distribution

$$\mathbb{P}(X = l) = \frac{1}{T(\lambda)} \frac{l^{l-1}}{l!} \lambda^l, \qquad l \in \mathbb{N}^*, z \in [0, \mathrm{e}^{-1}[, \tag{5.3}$$

where $T(\lambda) = \sum_{l=1}^{\infty} \frac{l^{l-1}}{l!} \lambda^l$ is the well-known tree function and $\lambda$ is an arbitrary number with $0 < \lambda \le \mathrm{e}^{-1}$. The conditional distribution of $Y$ given $X = l$ is the same as the distribution of $d_{l,l-1}$.

Unfortunately, using Janson's results [10,15,16], we can prove that the joint Laplace transform of $(X_1, Y_1)$ is only defined on $(-\infty, a) \times (-\infty, 0)$ for some positive $a$. Hence, our results can not be applied. Nevertheless, in a forthcoming work, we will study conditioned LDP for self-normalized sums in the spirit of [17]. In that case, the Laplace will be defined.

## 5.5. Bose–Einstein statistics

This example is borrowed from [14]. Consider $N$ urns. Put $n$ indistinguishable balls in the urns in such a way that each distinguishable outcome has the same probability that is,

$$1 \Big/ \binom{n + N - 1}{n},$$

see, for example, [9]. Let $Z_k$ be the number of balls in the $k$th urn. It is well known that $(Z_1, \dots, Z_N)$ is distributed as $(X_1, \dots, X_N)$ conditioned on $\sum_{i=1}^N X_i = n$, where $X_1, \dots, X_N$ are i.i.d. with a geometric distribution. As for Example 5.1, we can get a L.D.P. for variables of the form $\sum_i h(X_i)$ if $\operatorname{dom} \psi_{h(X_i)} = \mathbb{R}$.

## 5.6. Possible extensions

Among possible extensions, let us mention the case where the variables $Y_i$ are independent but do not have the same distribution. This case occurs in [14], Examples 2 and 3, where the quantity of interest is the law of $\sum_{i=1}^N h_i(X_i)$ conditioned on the event $\sum_{i=1}^N X_i = n$. Another way to extend our work is to deal with the case where the variables $X_i$ are independent but not i.i.d. This case occurs when counting from a random permutation the number of cycles of a fixed size see, for example, [1], Chapter 1.

## References


[1] Arratia, R., Barbour, A.D. and Tavaré, S. (2003). *Logarithmic Combinatorial Structures*: *A Probabilistic Approach*. *EMS Monographs in Mathematics*. Zürich: European Mathematical Society (EMS). MR2032426

[2] Azencott, R. (1980). Grandes déviations et applications. In *Eighth Saint Flour Probability Summer School* – 1978 (*Saint Flour*, 1978). *Lecture Notes in Math*. **774** 1–176. Berlin: Springer. MR0590626

[3] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*. *Wiley Series in Probability and Mathematical Statistics*. Chichester: Wiley. MR0489333

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

[5] Boucheron, S., Gamboa, F. and Léonard, C. (2002). Bins and balls: Large deviations of the empirical occupancy process. *Ann*. *Appl*. *Probab*. **12** 607–636. MR1910642

[6] Csiszár, I. (1984). Sanov property, generalized $I$-projection and a conditional limit theorem. *Ann*. *Probab*. **12** 768–793. MR0744233

[7] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. New York: Springer. MR1619036

[8] Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. *Monographs on Statistics and Applied Probability* **57**. New York: Chapman & Hall. MR1270903

[9] Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, *Vol*. 1, 3rd ed. New York: Willey.

[10] Flajolet, P., Poblete, P. and Viola, A. (1998). On the analysis of linear probing hashing. *Algorithmica* **22** 490–515.

[11] Hald, A. (1998). *A History of Mathematical Statistics from* 1750 *to* 1930. *Wiley Series in Probability and Statistics*: *Texts and References Section*. New York: Wiley. MR1619032

[12] Hipp, C. (1984). Asymptotic expansions for conditional distributions: The lattice case. *Probab*. *Math*. *Statist*. **4** 207–219. MR0792786

[13] Hollander, F.D. (2008). *Large Deviation*. *Fields Institute Monographs*. Providence, RI: Amer. Math. Soc.

[14] Holst, L. (1979). Two conditional limit theorems with applications. *Ann*. *Statist*. **7** 551–557. MR0527490

[15] Janson, S. (2001). Asymptotic distribution for the cost of linear probing hashing. *Random Structures Algorithms* **19** 438–471. MR1871562

[16] Janson, S. (2001). Moment convergence in conditional limit theorems. *J*. *Appl*. *Probab*. **38** 421–437. MR1834751

[17] Jing, B.Y., Shao, Q.M. and Wang, Q. (2003). Self-normalized Cramér-type large deviations for independent random variables. *Ann*. *Probab*. **31** 2167–2215. MR2016616

[18] Kolchin, V.F. (1984). *Random Mappings*. Moscou: Nauka (in Russian). English translation: New York: Optimization Software, 1986.

[19] Kudlaev, È.M. (1984). Conditional limit distributions of sums of random variables. *Teor*. *Veroyatnost*. *i Primenen*. **29** 743–752. English translation: *Theory Probab. Appl.* **29** (1984) 776–786. MR0773443

[20] Pavlov, J.L. (1977). Limit theorems for the number of trees of a given size in a random forest. *Mat*. *Sb*. (*N.S.*) **103** 392–403, 464. English translation: *Math. USSR Sb.* **32** 335–345. MR0453573

[21] Pavlov, Y.L. (1996). *Random Forest*. Petrozavodsk: Karelin Research Institut Centre, Russian Academy of Science (in Russian). English translation: Zeist: VSP, 2000.

[22] Robinson, J., Höglund, T., Holst, L. and Quine, M.P. (1990). On approximating probabilities for small and large deviations in $\mathbf{R}^d$. *Ann*. *Probab*. **18** 727–753. MR1055431

[23] Rockafellar, R.T. (1970). *Convex Analysis*. Princeton, NJ: Princeton Univ. Press.

[24] Steck, G.P. (1957). Limit theorems for conditional distributions. *Univ. California Publ. Statist.* **2** 237–284. MR0091552

[25] Van Campenhout, J.M. and Cover, T.M. (1981). Maximum entropy and conditional probability. *IEEE Trans. Inform. Theory* **27** 483–489. MR0635527

*Received September 2010*