The following publication M. Gao and K. -F. C. Yiu, "Asymptotic Behaviors and Confidence Intervals for the Number of Operating Sensors in a Sensor Network," in IEEE Transactions on Information Theory, vol. 69, no. 1, pp. 548-560, Jan. 2023 is available at https://doi.org/10.1109/TIT.2022.3222617.

Asymptotic behaviors and confidence intervals for the number of operating sensors in a sensor network

Mingjie Gao
School of Economics and Statistics, Guangzhou University,
Guangzhou, China
mjgao@gzhu.edu.cn

Ka-Fai Cedric Yiu
Department of Applied Mathematics
The Hong Kong Polytechnic University
Hunghom, Kowloon, Hong Kong, China
cedric.yiu@polyu.edu.hk

Abstract

In this paper, we study asymptotic behaviors of an estimator of the number of operating sensors in a sensor network based on the Good-Turing estimator. The asymptotic normality, some moderate deviations and deviation inequalities of the estimator are obtained. Our approach is based on the tail probability estimates and moderate deviations for occupancy problems. Applying these asymptotic behaviors, we give a performance analysis for the estimator of the number N of operating nodes when the deviations of the estimator are in $(\sqrt{N}, o(N))$. These estimates also provide a method to build confidence interval of N.

Keywords: Asymptotic behavior, confidence interval, Good-Turing estimator, asymptotic normality, moderate deviation, sensor network.

1 Introduction

In large-scale sensor networks, knowing the number N of operating nodes of a sensor network is crucial to network operation since the number of operating sensors can vary with time and the density of operating sensors may affect some network operations significantly. Budianu, Ben-David and Tong first used the Good-Turing algorithm to estimate the number of operating sensors and presented a performance analysis in [1] [2] based on large deviations for occupancy problems. The sensor network considered has operating sensors, each having an ID that is an element of finite set $\mathcal{N} = \{x_1, \cdots, x_N\}$ with $|\mathcal{N}| = N$. The vector sample $\mathbf{X} := (X_1, \cdots, X_n)$ received ID scan be modeled as an i.i.d. sampling with uniform distribution in a finite $\mathcal{N}$, i.e., $p_x := P(X_i = x) = \frac{1}{N}$, for any $x \in \mathcal{N}$. For each $k \geq 0$, $\mathcal{S}_k$ is the set of all the elements of $\mathcal{N}$ that appear in the vector sample $\mathbf{X}$ exactly k times and $S_k = |\mathcal{S}_k|$ denotes the total number of $\mathcal{S}_k$. Set $S = N - S_0$, $P_k = \sum_{x\in\mathcal{S}_k} p_x$. The probability P_0 is called the missing mass and $1 - P_0$ the coverage of the sample $\mathbf{X}$. An estimator of N was defined in [1] [2] by

$$\hat{N} = \frac{S}{1-\hat{P}_0} = \frac{N(1-\Gamma_{0,N})}{1-\hat{P}_0},$$

where $\Gamma_{0,N} = S_0/N$, and $\hat{P}_0 = S_1/n$ is the Good-Turing estimator for the missing mass P_0 which was proposed in [3]. Since the publication of the Good-Turing estimators, these estimators have been used extensively in language modeling applications. Their theoretical properties have been studied widely in the recent years. For example, see [4] for consistency and convergence rates, [5], [6], [7] and [8] for concentration inequalities, [9], [10], [11], [12] and [13] for asymptotic normality and large deviations, [14] and [15] for optimality and minimax properties.

It is natural to consider asymptotic behaviors for $\hat{N}$, for example, estimates of the order of magnitude of the bias, the asymptotic normality, the large deviations, etc. The large deviations have been considered in [1] [2] and the rate function of large deviations is obtained by solving a numerical optimization problem. Since the denominator of the estimator $\hat{N}$ can be close to zero, there is no estimate for the order of magnitude of the bias.

In this paper, we consider the asymptotic normality, deviation inequalities and moderate deviations of $\hat{N}$. The main results are two moderate deviation results which provide a precise exponential convergence rate or precise tail probabilities for deviations probabilities:

$$P\left(|\hat{N}-N| \geq rN^{\nu}\right) = P\left(\left|\frac{\hat{N}}{N}-1\right| \geq rN^{\nu-1}\right), \quad r>0, \tag{1.1}$$

where $\nu \in (1/2, 1)$. When $\nu = 1$, this is large deviation problem. We consider two types of moderate deviations. The first one is the following moderate deviations: for any sequence of positive numbers with $b_N \to \infty$ and $b_N = o(\sqrt{N})$,

$$\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left|\frac{\hat{N}}{N}-1\right| > x\right) = -\frac{x^2}{2\sigma^2}, \quad x>0, \tag{1.2}$$

where $\sigma > 0$ is a constant defined by in (3.2). The moderate deviations only give us a precise exponential convergence rate. Another one is Cramér-type moderate deviations, i.e.,

$$\begin{aligned} &\frac{P\left(\sqrt{N}\left(\frac{\hat{N}}{N}-1\right) > x\sigma\right)}{1-\Phi(x)} \to 1, \\ &\frac{P\left(\sqrt{N}\left(\frac{\hat{N}}{N}-1\right) < -x\sigma\right)}{\Phi(-x)} \to 1, \text{ for any } 0<x=o(b_N), \end{aligned} \tag{1.3}$$

where Φ is the standard normal distribution function, and b_N is a sequence of positive numbers with $b_N \to \infty$ and $b_N = O(\sqrt{N})$. We prove that the Cramér-type moderate deviations hold for $b_N = N^{1/6}$. In i.i.d case, the $N^{1/6}$ is the best order (see [16]). The Cramér-type moderate deviations tell us that the relative errors of the tail probabilities of $\sqrt{N}\left(\frac{\hat{N}}{N}-1\right)$ and the standard normal distribution convergence 0 uniformly in $(0, o(N^{1/6}))$. In particular, (1.3) implies the asymptotic normality. The method of cumulants (cf. [17]) is an important method of the Cramér-type moderate deviations. Recently, Feray, Meliot and Nikeghbali (see

[18]) developed the mod-ϕ convergence to study precise deviations, including the Cramér-type moderate deviations.

We will apply our results to give a performance analysis for the number N of operating nodes when the deviations of the estimator are in $(\sqrt{N}, o(N))$. These estimates also provide a method to build confidence interval of N. Our approach is based on tail probability estimates and moderate deviations for occupancy problems. Our analysis provides a more refined performance analysis for the number N of operating nodes.

The study of moderate deviations in information theory started with the work by Altuğ-Wagner [19], [20] and Polyanskiy-Verdú [21] on channel coding. We refer to [22], [23], [24], [25], [26] for some progress.

The paper is organized as follows. In section 2, we introduce the model and the Good-Turing estimator. The main results in asymptotic behaviors are stated in Section 3. In Section 4, we give a performance analysis and build confidence intervals for the number N of operating nodes by the Cramér-type moderate deviations. Some simulations and numerical results are also demonstrated in Section 4. The proofs of the theoretical results are given in Section 5.

2 Sensor network with mobile access and Good-Turing estimator

2.1 Sensor network with mobile access

In this paper, the model considered is the Sensor Network with Mobile Access (SENMA) (cf.[1]). A key feature of SENMA is the presence of the mobile access points (APs) that have high processing power and act as mobile base stations for sensors. In SENMA, the sensors may transmit the collected data to the mobile access points in the form of packets, and each packet may contain the ID of the transmitting sensor. The sensor network considered has operating sensors, each having an ID that is an element of set $\mathcal{N}$, with $|\mathcal{N}| = N$. The mobile AP collects n packets, each of them containing the ID of the transmitting sensor. We denote by $X_i \in \mathcal{N}$ the ID in the ith received packet and by $\mathbf{X} := (X_1, \cdots, X_n)$ the vector sample of received IDs. The unknown N is assumed to remain constant during collection.

For SENMA, packet collection can be modeled as an i.i.d. sampling with uniform distribution, i.e.,

$$p_x := P(X_i = x) = \frac{1}{N}, \text{ for any } x \in \mathcal{N}. \tag{2.1}$$

This model is identical to an urn model with replacement. Note that some of the received packets may come from the same sensor. The set of received IDs is denoted by

$$\mathcal{S} := \{x \in \mathcal{N}; X_k = x \text{ for some } 1 \le k \le n\}. \tag{2.2}$$

2.2 An estimator of the number of operating sensors

Consider a finite set $\mathcal{N}$, a uniform distribution P on $\mathcal{N}$ and a sample $\mathbf{X} := (X_1, \cdots, X_n)$, where $X_i \in \mathcal{N}$ are i.i.d random variables with distribution P. For each $x \in \mathcal{N}$, set $p_x = P(X_i = x) = 1/N$. For each k, the set $\mathcal{S}_k$ is composed of all the elements of $\mathcal{N}$ that appear in the vector sample $\mathbf{X}$ exactly k times and $S_k = |\mathcal{S}_k|$ denotes the total number of $\mathcal{S}_k$. Set

$$P_k = \sum_{x \in \mathcal{S}_k} p_x. \tag{2.3}$$

The probability P_0 is called the missing mass. An estimator for the missing mass is Good-Turing estimator proposed in [3]:

$$\hat{P}_0 = \frac{S_1}{n}. \tag{2.4}$$

The Good-Turing estimator can be used to estimate the number of operating sensors. Under the uniform distribution assumption, the missing mass is given by

$$P_0 = 1 - \frac{S}{N},$$

where $S = |\mathcal{S}|$. Then $S_0 = N - S$. Using the estimated value of P_0, we have the following estimator for N:

$$\hat{N} = \frac{S}{1 - \hat{P}_0} = \frac{S}{1 - S_1/n}. \tag{2.5}$$

The number of operating sensors is explained using urns and balls. The number of operating sensors corresponds to the number of available urns, and the number of available samples to the number of balls. There are n balls that are thrown (one by one) in N urns, each ball falling in any of the urns with equal probability. The number of balls in an urn corresponds to the number of packets received from a sensor. Assume that all urns are initially empty. Set

$$\Gamma_{i,N} := \frac{S_i}{N}, \quad i = 0, 1, 2, \cdots, \tag{2.6}$$

the fraction of urns that contain exactly i balls (or sensors that appear i times in the current sample). Then

$$\frac{\hat{N}}{N} = \frac{S}{N} \frac{1}{1 - \frac{S_1}{N}\frac{N}{n}} = \frac{\frac{n}{N}(1 - \Gamma_{0,N})}{\frac{n}{N} - \Gamma_{1,N}},$$

and so, the relative error $\frac{\hat{N}-N}{N}$ can be written as

$$\frac{\hat{N}}{N} - 1 = \frac{\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}}{\frac{n}{N} - \Gamma_{1,N}}. \tag{2.7}$$

Therefore, in order to study the estimator of the number of operating sensors in a sensor network, we need to consider asymptotic behaviors of $\Gamma_{0,N}$, $\Gamma_{1,N}$ and $\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}$.

3 Asymptotic behaviors for the estimator of number of operating sensors

In this section, we state our main results. Their proof are postponed to Section 5. We assume that for some $\beta \in (0, \infty)$,

$$\left| \frac{n}{N} - \beta \right| = O\left(\frac{1}{\sqrt{N}} \right). \tag{3.1}$$

For example, $n = [\beta N]$ satisfies (3.1).

Define

$$\sigma^2(\beta) = \frac{e^{-\beta}(1 - e^{-\beta} + \beta)}{\beta(1 - e^{-\beta})^2}. \tag{3.2}$$

Theorem 3.1 (Moderate deviations)**.** *Let $b_N, N \geq 1$ be a sequence of positive numbers with $b_N \to \infty$ and $b_N = o(\sqrt{N})$. Then for any $x > 0$,*

$$\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left(\frac{\hat{N}}{N} - 1\right) > x\right) = -\frac{x^2}{2\sigma^2(\beta)}, \tag{3.3}$$

and

$$\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left(\frac{\hat{N}}{N} - 1\right) < -x\right) = -\frac{x^2}{2\sigma^2(\beta)}. \tag{3.4}$$

Remark 3.1. *Theorem 3.1 yields that $\{\frac{\hat{N}}{N} - 1, N \geq 1\}$ satisfies the moderate deviation principle, that is, for any measurable set $B \subset \mathbb{R}$,*

$$-\inf_{x\in B^o} \frac{x^2}{2\sigma^2(\beta)} \leq \lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left(\frac{\hat{N}}{N} - 1\right) \in B\right)$$
$$\leq -\inf_{x\in \bar{B}} \frac{x^2}{2\sigma^2(\beta)}.$$

where B^o and $\bar{B}$ denote the interior and the closure of the set B, respectively. (see [27]).

Theorem 3.2 (Cramér-type moderate deviations)**.** *(1). For $0 \leq x \leq o(N^{1/6})$,*

$$\frac{P\left(\sqrt{N}\left(\frac{\hat{N}}{N} - 1\right) \geq \sigma(\beta)x\right)}{1 - \Phi(x)} = 1 + O\left(\frac{1 + x^2}{N^{1/3}}\right), \tag{3.5}$$

and

$$\frac{P\left(\sqrt{N}\left(\frac{\hat{N}}{N} - 1\right) \leq -\sigma(\beta)x\right)}{\Phi(-x)} = 1 + O\left(\frac{1 + x^2}{N^{1/3}}\right). \tag{3.6}$$

where Φ is the standard normal distribution function.

(2). For $0 \leq x \leq o(N^{1/6})$,

$$\frac{P\left(\frac{1}{\sqrt{\hat{N}}}\left(\hat{N} - N\right) \geq \sigma(\beta)x\right)}{1 - \Phi(x)} = 1 + O\left(\frac{1 + x^2}{N^{1/3}}\right), \tag{3.7}$$

and

$$\frac{P\left(\frac{1}{\sqrt{\hat{N}}}\left(\hat{N} - N\right) \leq -\sigma(\beta)x\right)}{\Phi(-x)} = 1 + O\left(\frac{1 + x^2}{N^{1/3}}\right). \tag{3.8}$$

Remark 3.2. *Theorem 3.2 (2) is a self-normalized Cramér-type moderate deviation. It plays an important role in statistical inferences. In the paper, we apply it to build confidence intervals for the number of operating sensors.*

In particular, Theorem 3.2 yields the asymptotic normality.

Corollary 3.1 (Asymptotic normality)**.** *For* $x \in \mathbb{R}$,

$$\lim_{N\to\infty} P\left(\sqrt{N}\left(\frac{\hat{N}}{N} - 1\right) \leq \sigma(\beta)x\right) = \Phi(x), \tag{3.9}$$

and

$$\lim_{N\to\infty} P\left(\frac{1}{\sqrt{\hat{N}}}\left(\hat{N} - N\right) \leq \sigma(\beta)x\right) = \Phi(x). \tag{3.10}$$

Remark 3.3. *Gao [13] studied the moderate deviation principle of the Good estimator. In general occupation models, the result is just a moderate deviation principle for* $\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}$ *which is a linear statistics for the occupation problem. The Cramér-type moderate deviations for the occupation problem were studied in [28]. In this paper, we concern the estimator* $\hat{N}$ *of the number of operating sensors that is a nonlinear functional of some statistics in the occupation problem. Our study in this paper is based on the moderate deviation estimates for the statistics in occupation problem (cf. Theorem 2.1 in [13], Theorem 2 in [28], also see Lemma 5.9 and Lemma 5.3 for some special cases). In order to derive the moderate deviation principle and the Cramér-type moderate deviations of the estimator* $\hat{N}$ *from that of some statistics in occupation problem, we need to establish some comparison theorems and concentration inequalities of* $\Gamma_{m,N}$*. Lemma 5.5 gives such a comparison theorem on the Cramér-type moderate deviations. Proposition 3.1 presents a concentration inequality of* $\Gamma_{m,N}$*. Our main results provide a complete description of convergence of the estimator* $\hat{N}$ *in the regime* $(\sqrt{N}, o(N))$*.*

Our results on moderate deviations and performance analysis for the number of operating sensors are based on the tail probability estimates of $\Gamma_{m,N}$. The tail probability estimates have had many works. Kamath, Motwani, Palem and Spirakis (Theorem 2 in [29]) established the following precise tail probability estimate: for any $\theta > 0$,

$$P\left(|\Gamma_{0,N} - \gamma_{0,N}| \geq \theta\gamma_{0,N}\right) \leq 2\exp\left\{-\frac{\theta^2\gamma_{0,N}^2(N-1/2)}{1-\gamma_{0,N}^2}\right\}. \tag{3.11}$$

The concentration inequalities of the missing mass (i.e., $m = 1$) were first considered by McAllester and Schapire [4], also see [5], [6], [7] and [8]. Rencntly, Ben-Hamou, Boucheron and Ohannessian considered general cases $m \geq 0$ in [7] using the entropy method. In this paper, for our applications, we present a different form concentration inequality of $\Gamma_{m,N}$ for any $m \geq 0$ using negative dependence. Our rate functions are represented by the entropies of Bernoulli distributions.

Define

$$I_m^N(x) = \begin{cases} (x+\mu_{m,N})\log\frac{x+\mu_{m,N}}{\mu_{m,N}} \\ +(1-(x+\mu_{m,N}))\log\frac{1-(x+\mu_{m,N})}{1-\mu_{m,N}}, \\ \qquad\qquad \text{if } x \in [-\mu_{m,N}, 1-\mu_{m,N}], \\ \\ +\infty, \\ \qquad\qquad \text{if } x \notin [-\mu_{m,N}, 1-\mu_{m,N}], \end{cases} \tag{3.12}$$

where

$$\mu_{m,N} = \sum_{l=0}^{m} \gamma_{l,N}, \quad \gamma_{l,N} = \frac{n!}{(n-l)!l!} \frac{1}{N^l} \left(1 - \frac{1}{N}\right)^{n-l}. \tag{3.13}$$

Remark 3.4. *The Kullback-Leibler (KL) divergence between two Bernoulli distributions with parameters $q \in [0,1]$ and $p \in [0,1]$ is defined by*

$$D_{KL}(q||p) = q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}.$$

Thus, for $x \in [-\mu_{m,N}, 1 - \mu_{m,N}]$, $I_m^N(x)$ is just the KL divergence between two Bernoulli distributions with parameters $x + \mu_{m,M}$ and $\mu_{m,N}$.

Proposition 3.1. *Let $0 \le m \le n$ be given. Then for any $x > 0$,*

$$P\left(|\Gamma_{0,N} - \gamma_{0,N}| \ge x\right) \le 2 \exp\left\{-N J_0^N(x)\right\}, \tag{3.14}$$

and for any $m \ge 1$,

$$\begin{aligned} P\left(|\Gamma_{m,N} - \gamma_{m,N}| \ge x\right) \le & 2 \exp\left\{-N J_{m-1}^N(x/2)\right\} \\ & + 2 \exp\left\{-N J_m^N(x/2)\right\}, \end{aligned} \tag{3.15}$$

where

$$J_m^N(x) = \min\{I_m^N(x), I_m^N(-x)\}. \tag{3.16}$$

By the definition of J_m, for any $m \ge 0$, $x > 0$, $\lim_{N\to\infty} J_m^N(x) > 0$. By Poisson's approximation theorem, as $N \to \infty$,

$$\gamma_{m,N} \to \frac{\beta^m}{m!} e^{-\beta}, \quad \mu_{m,N} \to \sum_{l=0}^{m} \frac{\beta^l}{l!} e^{-\beta}.$$

4 Performance analysis and confidence intervals for the number of operating sensors

In this section, we give a performance analysis and build confidence intervals for the number of operating sensors using our main results.

4.1 Performance analysis

Firstly, we use the identity

$$\frac{\hat{N}}{N} - 1 = \frac{\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}}{\frac{n}{N} - \Gamma_{1,N}},$$

and apply Proposition 3.1 to give non-asymptotic bounds of tail probabilities of $\hat{N} - N$.

Let $\epsilon \in [0,1)$ satisfy $(1+\epsilon)e^{-\beta} < 1$ and let $0 \le \epsilon_N \to \epsilon$ as $N \to \infty$. Then

$$\begin{aligned}
&\left\{\left|\frac{\frac{n}{N}\Gamma_{0,N} - \Gamma_{1,N}}{\frac{n}{N} - \Gamma_{1,N}}\right| \ge x\right\} \\
\subset &\{|\Gamma_{1,N} - \gamma_{1,N}| \ge \epsilon_N \gamma_{1,N}\} \\
&\cup \left\{\left|\frac{\frac{n}{N}\Gamma_{0,N} - \Gamma_{1,N}}{\frac{n}{N} - \Gamma_{1,N}}\right| \ge x, |\Gamma_{1,N} - \gamma_{1,N}| \le \epsilon_N \gamma_{1,N}\right\} \\
\subset &\{|\Gamma_{1,N} - \gamma_{1,N}| \ge \epsilon_N \gamma_{1,N}\} \\
&\cup \left\{\left|\frac{n}{N}\Gamma_{0,N} - \Gamma_{1,N}\right| \ge x\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right)\right\} \\
\subset &\{|\Gamma_{1,N} - \gamma_{1,N}| \ge \epsilon_N \gamma_{1,N}) \\
&\cup \left\{|\Gamma_{0,N} - \gamma_{0,N}| \ge \frac{xN}{2n}\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right)\right\} \\
&\cup \left\{|\Gamma_{1,N} - \frac{n}{N}\gamma_{0,N}| \ge \frac{x}{2}\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right)\right\}.
\end{aligned}$$

Since $\gamma_{0,N} = \frac{N-1}{n}\gamma_{1,N}$, we have

$$\begin{aligned}
&\left\{\left|\frac{\frac{n}{N}\Gamma_{0,N} - \Gamma_{1,N}}{\frac{n}{N} - \Gamma_{1,N}}\right| \ge x\right\} \\
\subset &\{|\Gamma_{1,N} - \gamma_{1,N}| \ge \epsilon_N \gamma_{1,N}\} \\
&\cup \left\{|\Gamma_{0,N} - \gamma_{0,N}| \ge \frac{xN}{2n}\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right)\right\} \\
&\cup \left\{|\Gamma_{1,N} - \gamma_{1,N}| \ge \frac{x}{2}\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right) - \frac{1}{N}\gamma_{1,N}\right\}.
\end{aligned}$$

Thus

$$\begin{aligned}
&P\left(\left|\frac{\hat{N}}{N} - 1\right| \ge x\right) \\
=&P\left(\left|\frac{\frac{n}{N}\Gamma_{0,N} - \Gamma_{1,N}}{\frac{n}{N} - \Gamma_{1,N}}\right| \ge x\right) \\
\le&P\left(|\Gamma_{1,N} - \gamma_{1,N}| \ge \epsilon_N \gamma_{1,N}\right) \\
&+ P\left(|\Gamma_{0,N} - \gamma_{0,N}| \ge \frac{xN}{2n}\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right)\right) \\
&+ P\left(|\Gamma_{1,N} - \gamma_{1,N}| \ge \frac{x}{2}\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right) - \frac{1}{N}\gamma_{1,N}\right).
\end{aligned}$$

Therefore, from Proposition 3.1, we obtain that for any $x > 0$,

$$P\left(\left|\hat{N} - N\right| \ge xN\right) \le 10\exp\left\{-NJ_N(x)\right\}, \tag{4.1}$$

which gives non-asymptotic bounds of tail probabilities of $\hat{N} - N$, where

$$J_N(x) = \min\Big\{ J_1^N\left(\epsilon_N \gamma_{1,N}/2\right), J_0^N\left(\epsilon_N \gamma_{1,N}/2\right), \\ J_0^N\left(\frac{xN}{2n}\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right)\right)^+, \\ J_1^N\left(\frac{x}{4}\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right) - \frac{\gamma_{1,N}}{2N}\right)^+, \\ J_0^N\left(\frac{x}{4}\left(\frac{n}{N} - (1+\epsilon_N)\gamma_{1,N}\right) - \frac{\gamma_{1,N}}{2N}\right)^+ \Big\}.$$

Since $\lim_{N\to\infty} J_N(r) \in (0,\infty)$ for any $r > 0$, for any $b_N = o(\sqrt{N})$,

$$\limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(\left|\frac{\hat{N}}{N} - 1\right| \geq r\right) = -\infty. \tag{4.2}$$

Then, by (3.3) and (3.4) in Theorem 3.1, when the order of deviation is in $(\sqrt{N}, o(N))$, then exponential convergence rate is

$$\begin{aligned} &\frac{1}{N^{2\nu-1}} \log P\left(|\hat{N} - N| \geq N^\nu r\right) \\ \approx &-\frac{\beta(1-e^{-\beta})^2 r^2}{2e^{-\beta}\left(1 - e^{-\beta} + \beta\right)}. \end{aligned} \tag{4.3}$$

Finally, (3.5) and (3.6) in Theorem 3.2 give us a precise estimate when the order of deviation is in $(\sqrt{N}, o(N^{2/3}))$:

$$\begin{aligned} &P\left(|\hat{N} - N| \geq N^\nu r\right) \\ \approx &2\left(1 - \Phi\left(N^{\nu-1/2} r/\sigma(\beta)\right)\right). \end{aligned} \tag{4.4}$$

Remark 4.1. *When the order of deviation is in* $(\sqrt{N}, o(N^{2/3}))$, (4.4) *gives a precise convergence rate of the estimate* (1.1).

4.2 Confidence intervals

Now, by (4.2) and the Cramér-type moderate deviations (3.7) and (3.8), for $x = o(N^{1/6})$ uniformly, the relative error

$$\frac{P\left(|\hat{N} - N| \geq x\sigma(\beta)\sqrt{\hat{N}}\right)}{2(1 - \Phi(x))} - 1 \to 0.$$

Therefore, for $\alpha \in (0,1)$, take $u_{\alpha/2}$ such that $1 - \Phi(u_{\alpha/2}) = \alpha/2$, then,

$$P\left(\frac{1}{\sqrt{\hat{N}}}|\hat{N} - N| \geq \sigma(\beta) u_{\alpha/2}\right) \approx \alpha.$$

Thus, we obtain an approximate $1-\alpha$ confidence interval of N:

$$[\hat{N} - \sqrt{\hat{N}}\sigma(\beta)u_{\alpha/2}, \hat{N} + \sqrt{\hat{N}}\sigma(\beta)u_{\alpha/2}].$$

4.3 Simulations and numerical results

In this section, we give some simulations and numerical results for the tail probability

$$P\left(|\hat{N} - N| \geq N^{\nu} r\right), \quad r > 0.$$

Set

$$p_{\nu,\beta}(r) = 2\left(1 - \Phi\left(N^{\nu-1/2} r/\sigma(\beta)\right)\right),$$

and

$$\tilde{p}_{\nu,\beta}(r) = \exp\left\{-\frac{\beta(1-e^{-\beta})^2 r^2 N^{2\nu-1}}{2e^{-\beta}\left(1-e^{-\beta}+\beta\right)}\right\}.$$

Case I: N fixed. We take $\mathcal{N} = \{1, 2, \cdots, 1000\}$, and consider the following four cases: $\nu = 1/2, \beta = 3/4$; $\nu = 1/2, \beta = 5/6$; $\nu = 5/8, \beta = 3/4$; $\nu = 5/8, \beta = 5/6$. Ten thousand samples of size $n := [\beta N]$ are generated from the uniform distribution on $\mathcal{N}$. For each sample, we compute $\hat{N}$ and denote respectively by $\hat{N}_k$, $k = 1, ..., 10000$.

For each $\nu = 1/2, \beta = 3/4$; $\nu = 1/2, \beta = 5/6$; $\nu = 5/8, \beta = 3/4$; $\nu = 5/8, \beta = 5/6$, and $r = 0.3 + 0.1j, j = 0, 1, 2, \cdots, 10$, we compute $p_{\nu,\beta}(r)$, $\tilde{p}_{\nu,\beta}(r)$ and the frequencies $\hat{p}_{\nu,\beta}(r)$ of $\left\{|\hat{N}_k - N| \geq N^{\nu} r\right\}$, $\quad k = 1, ..., 10000$. The numerical results are summarized in the following Figures. In Figure 1, Figure 2, Figure 3 and Figure 4, the red circle denotes $\hat{p}_{\nu,\beta}(r)$, the blue star denotes $p_{\nu,\beta}(r)$ and the green diamond denotes $\tilde{p}_{\nu,\beta}(r)$. From the Figures, we see that $|p_{\nu,\beta}(r) - \hat{p}_{\nu,\beta}(r)|$ is smaller than $|\tilde{p}_{\nu,\beta}(r) - \hat{p}_{\nu,\beta}(r)|$, the numerical results also demonstrate that the Cramér-type moderate deviations (Theorem 3.2) are more precise than the moderate deviations (Theorem 3.1). Moreover, the red circles and the blue stars almost overlap, indicating that the simulation results are highly consistent with the theorem results(Theorem 3.2).

Case II: $N \to \infty$. We take $\mathcal{N} = \{1, 2, \cdots, N\}$, where N is from 1000 to 16000. We consider $\nu = 3/4, \beta = 3/4$ and $r = 0.5$. Another case we take $\mathcal{N} = \{1, 2, \cdots, N\}$, where N is from 1000 to 20000.and consider $\nu = 8/9$, $\beta = 3/4$, and $r = 0.1$. Ten thousand samples of size $n := [\beta N]$ are generated from the uniform distribution on $\mathcal{N}$. For different N, we compute $\hat{N}$ and denote respectively by $\hat{N}_k$, $k = 1, ..., 10000$. We compute $p_{\nu,\beta}(r)$, $\tilde{p}_{\nu,\beta}(r)$ and the frequencies $\hat{p}_{\nu,\beta}(r)$ of $\left\{|\hat{N}_k - N| \geq N^{\nu} r\right\}$, $\quad k = 1, ..., 10000$. In Figure 5 and Figure 6, the red circle denotes $\hat{p}_{\nu,\beta}(r)$, the blue star denotes $p_{\nu,\beta}(r)$ and the green diamond denotes $\tilde{p}_{\nu,\beta}(r)$. From Figure 5 and Figure 6, we can see the larger the N, the closer the three estimates are. Numerical results demonstrate the validity of the estimates.

5 Proofs of main results

In this section, we prove the main results. In the first subsection, we consider tail probabilities of $\Gamma_{i,N} - \gamma_{i,N}$, and prove Proposition 3.1. Proofs of Theorem 3.1 and Theorem 3.2 are given in the second subsection.

5.1 Tail probability estimates

The estimates of the tail probabilities and the proof of Proposition 3.1 are based on some results of the negatively associated random variables ([30]). The key is to represent $N \sum_{l=0}^{m} \Gamma_{l,N}$

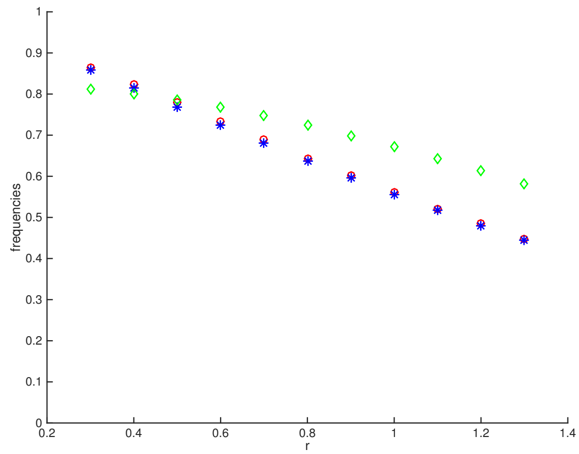

Figure 1: $\nu = 1/2, \beta = 3/4, N = 1000$

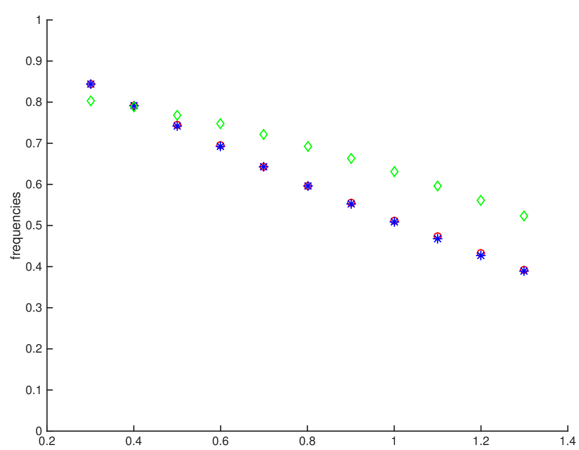

Figure 2: $\nu = 1/2, \beta = 5/6, N = 1000$

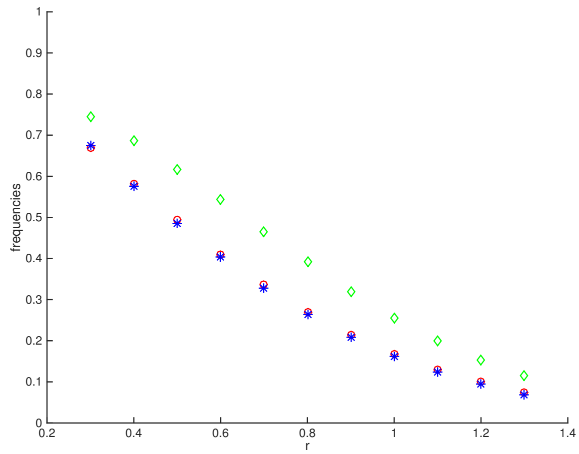

Figure 3: $\nu = 5/8, \beta = 3/4, N = 1000$

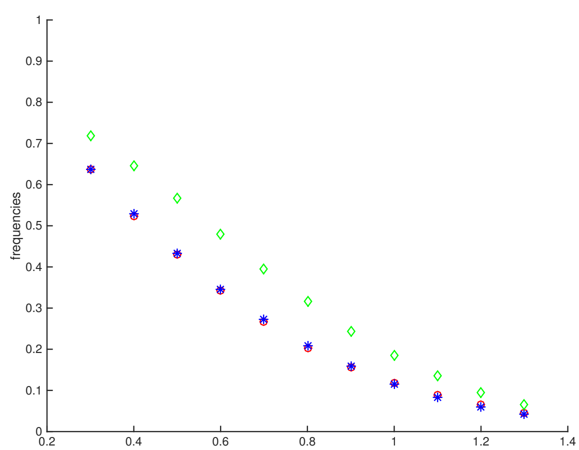

Figure 4: $\nu = 5/8, \beta = 5/6, N = 1000$

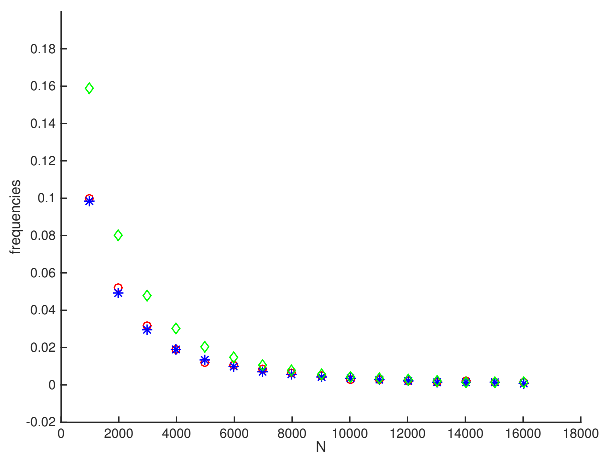

Figure 5: $\nu = 3/4, \beta = 3/4, r = 0.5$

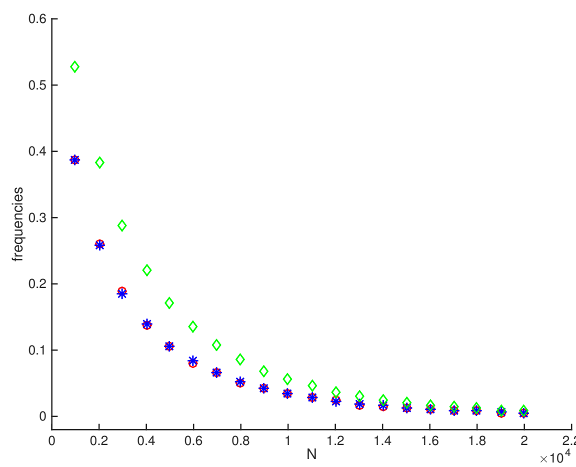

Figure 6: $\nu = 8/9, \beta = 3/4, r = 0.1$

into sum of some N negatively associated random variables. Then, the estimates of the tail probabilities and Proposition 3.1 are obtained by Chebyshev's inequality.

Firstly, let us give a representation of sum of N negatively associated random variables for $N\sum_{l=0}^{m}\Gamma_{l,N}$, and establish exponential moment estimates of $N\sum_{l=0}^{m}\Gamma_{l,N}$, $1\le m\le n$. Define the indicator random variables $\xi_{i,k}$, $1\le i\le N$, $1\le k\le n$ as follows:

$$\xi_{i,k}=\begin{cases}1, & \text{if ball } k \text{ goes into urn } i;\\ \\ 0, & \text{otherwise.}\end{cases} \tag{5.1}$$

Set

$$X_i^N:=\sum_{k=1}^{n}\xi_{i,k},\ 1\le i\le N,$$

i.e., X_i^N denotes the number of balls in urn i. Then for any $l\ge 0$,

$$N\sum_{l=0}^{m}\Gamma_{l,N}=\sum_{i=1}^{N}\delta_{\{X_i^N\le m\}},$$

where δ_A denotes the indicator function of set A.

Then by Theorem 14 in [30], $\{X_i^N, 1\le i\le N\}$ is negatively associated, that is, for every two disjoint index finite sets $\Lambda_1,\Lambda_2\subset\{1,2,\cdots,N\}$,

$$\begin{aligned}&E\left(f(X_k^N,k\in\Lambda_1)g(X_k^N,k\in\Lambda_2)\right)\\ \le&E\left(f(X_k^N,k\in\Lambda_1)\right)E\left(g(X_k^N,k\in\Lambda_2)\right)\end{aligned}$$

for all nonnegative function $f:R^{\Lambda_1}\to R$ and $g:R^{\Lambda_2}\to R$ that are both non-decreasing or both non-increasing.

Furthermore, since $f(x)=I_{(-\infty,m]}(x)$ is a decreasing function, by Proposition 8 in Dubhashi and Ranjan [30], we obtain that $\{\delta_{\{X_i^N\le m\}}, 1\le i\le N\}$ is negatively associated.

Lemma 5.1. *For any $r\in R$, and for any $0\le m\le n$,*

$$E\left(\exp\left\{rN\sum_{l=0}^{m}\Gamma_{l,N}\right\}\right)\le\left((e^r-1)\mu_{m,N}+1\right)^N. \tag{5.2}$$

Proof. For any $r\in R$ given, set $f(x)=e^{rx}$, $x\in R$. Then, when $r\ge 0$, all $f(x)$ is nonnegative and increasing; when $r<0$, $f(x)$ is nonnegative and decreasing. Since $\{\delta_{\{X_i^N\le m\}}, 1\le i\le N\}$ is negatively associated and

$$E\left(\delta_{\{X_i^N\le m\}}\right)=P(X_i^N\le m)=\mu_{m,N},$$

we have that

$$\begin{aligned}&E\left(\exp\left\{rN\sum_{l=0}^{m}\Gamma_{l,N}\right\}\right)\\ \le&\left(E\left(\exp\left\{r\delta_{\{X_i^N\le m\}}\right\}\right)\right)^N\\ =&\left(P(X_i^N>m)+e^rP(X_i^N\le m)\right)^N\\ =&\left((e^r-1)\mu_{m,N}+1\right)^N.\end{aligned}$$

□

Lemma 5.2. *Let $0 \le m \le n$ be given. Then for any $x \ge 0$,*

$$P\left(\sum_{l=0}^{m}\Gamma_{l,N} - \mu_{m,N} \ge x\right) \le \exp\left\{-NI_m^N(x)\right\}, \tag{5.3}$$

and for any $x \le 0$,

$$P\left(\sum_{l=0}^{m}\Gamma_{l,N} - \mu_{m,N} \le x\right) \le \exp\left\{-NI_m^N(x)\right\}. \tag{5.4}$$

Proof. Since $0 \le \sum_{l=0}^{m}\Gamma_{l,N} \le 1$, for any $x > 1-\mu_{m,N}$,

$$P\left(\sum_{l=0}^{m}\Gamma_{l,N} - \mu_{m,N} \ge x\right) = 0 = \exp\left\{-NI_m^N(x)\right\},$$

and for any $x < -\mu_{m,N}$,

$$P\left(\sum_{l=0}^{m}\Gamma_{l,N} - \mu_{m,N} \le x\right) = 0 = \exp\left\{-NI_m^N(x)\right\}.$$

Thus, we only consider $x \in [-\mu_{m,N}, 1-\mu_{m,N}]$. It is easy to get (3.12). By Chebyshev's inequality and Lemma 5.1, for any $0 < x < 1-\mu_{m,N}$, for any $\lambda > 0$,

$$\begin{aligned}
&P\left(\sum_{l=0}^{m}\Gamma_{l,N} - \mu_{m,N} \ge x\right)\\
=&P\left(\exp\left\{\lambda N\sum_{l=0}^{m}\Gamma_{l,N} - \lambda N\mu_{m,N}\right\} \ge e^{\lambda N x}\right)\\
\le& e^{-\lambda N x}E\left(\exp\left\{\lambda N\sum_{l=0}^{m}\Gamma_{l,N} - \lambda N\mu_{m,N}\right\}\right)\\
\le& \exp\left\{-N\left(\lambda(x+\mu_{m,N}) - \log\left((e^{\lambda}-1)\mu_{m,N}+1\right)\right)\right\}.
\end{aligned}$$

Set

$$f(\lambda) = \lambda(x+\mu_{m,N}) - \log\left(\mu_{m,N}e^{\lambda} + 1 - \mu_{m,N}\right).$$

Then $f'(\lambda) = x + \mu_{m,N} - 1 + \frac{1-\mu_{m,N}}{\mu_{m,N}e^{\lambda}+1-\mu_{m,N}}$, and $f'(\lambda) = 0$ has a unique solution $\lambda_0 = \log\frac{(x+\mu_{m,N})(1-\mu_{m,N})}{\mu_{m,N}(1-\mu_{m,N}-x)} \in (0,\infty)$. Therefore,

$$\begin{aligned}
P\left(\sum_{l=0}^{m}\Gamma_{l,N} - \mu_{m,N} \ge x\right) \le& e^{-N\sup_{\lambda>0}f(\lambda)}\\
=& e^{-Nf(\lambda_0)} = e^{-NI_m^N(x)}.
\end{aligned}$$

Thus, (5.3) is valid. Similarly, for any $-\mu_{m,N} < x < 0$, for any $\lambda < 0$,

$$\begin{aligned}
&P\left(\sum_{l=0}^{m}\Gamma_{l,N}-\mu_{m,N}\le x\right)\\
=&P\left(\exp\left\{\lambda N\sum_{l=0}^{m}\Gamma_{l,N}-\lambda N\mu_{m,N}\right\}\ge e^{\lambda N x}\right)\\
\le&e^{-\lambda N x}E\left(\exp\left\{\lambda N\sum_{l=0}^{m}\Gamma_{l,N}-\lambda N\mu_{m,N}\right\}\right)\\
\le&\exp\left\{-N\left(\lambda(x+\mu_{m,N})-\log\left((e^{\lambda}-1)\mu_{m,N}+1\right)\right)\right\},
\end{aligned}$$

and

$$P\left(\sum_{l=0}^{m}\Gamma_{l,N}-\mu_{m,N}\le x\right)\le e^{-N\sup_{\lambda<0}f(\lambda)}=e^{-NI_m^N(x)}.$$

□

Proof of Proposition 3.1. By Lemma 5.2, for any $x > 0$,

$$P\left(\left|\sum_{l=0}^{m}\Gamma_{l,N}-\mu_{m,N}\right|\ge x\right)\le 2\exp\left\{-NJ_m^N(x)\right\}. \tag{5.5}$$

Since

$$|\Gamma_{m,N}-\gamma_{m,N}|\le\left|\sum_{l=0}^{m}\Gamma_{l,N}-\mu_{m,N}\right|+\left|\sum_{l=0}^{m-1}\Gamma_{l,N}-\mu_{m-1,N}\right|,$$

we have

$$\begin{aligned}
&P\left(|\Gamma_{m,N}-\gamma_{m,N}|\ge x\right)\\
\le&P\left(\left|\sum_{l=0}^{m}\Gamma_{l,N}-\mu_{m,N}\right|+\left|\sum_{l=0}^{m-1}\Gamma_{l,N}-\mu_{m-1,N}\right|\ge x\right)\\
\le&P\left(\left|\sum_{l=0}^{m}\Gamma_{l,N}-\mu_{m,N}\right|\ge x/2\right)\\
&+P\left(\left|\sum_{l=0}^{m-1}\Gamma_{l,N}-\mu_{m-1,N}\right|\ge x/2\right)\\
\le&2\exp\left\{-NJ_m^N(x/2)\right\}+2\exp\left\{-NJ_{m-1}^N(x/2)\right\}.
\end{aligned}$$

Thus we obtain the tail probability estimates (3.14) and (3.15).

□

5.2 Proofs of Theorem 3.1 and Theorem 3.2

By the identity (2.7), i.e.,

$$\frac{\hat{N}}{N}-1=\frac{\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}}{\frac{n}{N}-\Gamma_{1,N}},$$

in order to prove Theorem 3.1 and Theorem 3.2, we need to study the Cramér moderate deviations of $\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}$, and the deviation estimates of $\frac{n}{N} - \Gamma_{1,N}$.

We will show the Cramér moderate deviations of $\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}$ in Lemma 5.6. We first write $\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}$ into randomized decomposable statistics in multinomial scheme via the urn mode in the section 2, and then apply Theorem 2 in [28] and the comparison theorem for Cramér moderate deviations which will be established in Lemma 5.5.

The deviation estimates of $\frac{n}{N} - \Gamma_{1,N}$ are obtained easily from Proposition 3.1.

We will complete the proofs of Theorem 3.1 and Theorem 3.2 by dividing it into seven steps.

Step 1. A representation of $\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}$ and its moderate deviations. We still consider the urn model in the section 2. There are n balls that are thrown (one by one) in N urns, each ball falling in any of the urns with equal probability. Let after n throwing the number of thrown balls of the different urns be denoted by $\eta_1, \cdots, \eta_N$. Then $(\eta_1, \cdots, \eta_N)$ have a multinomial distribution $M(n, 1/N, \cdots, 1/N)$. It is known that

$$\begin{aligned} & P\left(\eta_m = x_m; m = 1, \cdots, N\right) \\ =& P\left(\xi_m = x_m; m = 1, \cdots, N \middle| \sum_{m=1}^{N} \xi_m = n\right), \end{aligned}$$

where $\xi_m, m \geq 1$ are independent random variables and ξ_m is Poisson distributed with mean $\frac{n}{N}$.

Set

$$f(x) = I_{\{1\}}(x) - \frac{n}{N} I_{\{0\}}(x),$$

where I_A is the indicator function of set A.

Then $\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}$ has the following representation:

$$N\left(\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}\right) = \sum_{m=1}^{N} f(\eta_m).$$

Define

$$A_N = \sum_{m=1}^{N} E(f(\xi_m)) = 0,$$

$$\gamma_N = \frac{1}{n}\sum_{m=1}^{N} \mathrm{cov}(f(\xi_m), \xi_m) = \frac{N}{n} P(\xi_1 = 1) = e^{-n/N},$$

and

$$\begin{aligned} B_N^2 =& \sum_{m=1}^{N} \left(E((f(\xi_m))^2) - \frac{n}{N}\gamma_N^2\right) \\ =& N\left(\frac{n}{N} e^{-n/N} + \left(\frac{n}{N}\right)^2 e^{-n/N} - \frac{n}{N} e^{-2n/N}\right) \\ =& n e^{-n/N}\left(1 + \frac{n}{N} - e^{-n/N}\right). \end{aligned}$$

By (3.1),

$$\left|\frac{B_N^2}{N} - \beta e^{-\beta}(1-e^{-\beta}+\beta)\right| = O\left(\frac{1}{\sqrt{N}}\right).$$

Note that f is bounded and $\beta e^{-\beta}(1-e^{-\beta}+\beta) > 0$. We have that for any $H > 0$ and $\alpha > 0$ given,

$$\liminf_{N\to\infty} \frac{B_N^2}{N} > 0,$$

$$E(e^{hf(\xi_m)}) \le |H|(1+\frac{n}{N}) \text{ for all } |h| \le H, \quad m = 1, \cdots, N,$$

and

$$E(e^{hf(\xi_m(\alpha))}) \le |H|(1+\frac{n}{N}) \text{ for all } |h| \le H, \quad m = 1, \cdots, N,$$

where $\xi_m(\alpha)$ is a Poisson distribution with parameter α for each $m = 1, \cdots, N$. Therefore, we can apply Theorem 2 and Theorem 3 in [28] with $f_m = f_{mN} = f$, $m = 1, \cdots, N$ to obtain the following two lemmas.

Lemma 5.3. *For $0 \le x \le o(N^{1/6})$,*

$$\frac{P\left(\frac{N}{B_N}\left(\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}\right) \ge x\right)}{1-\Phi(x)} = 1 + O\left(\frac{1+x^3}{\sqrt{N}}\right)$$

and

$$\frac{P\left(\frac{N}{B_N}\left(\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}\right) \le -x\right)}{\Phi(-x)} = 1 + O\left(\frac{1+x^3}{\sqrt{N}}\right).$$

Lemma 5.4. *For $0 \le x \le o(\sqrt{N})$,*

$$\begin{aligned}&\frac{P\left(\frac{N}{B_N}\left(\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}\right) \ge x\right)}{1-\Phi(x)}\\ &= \left(1 + O\left(\frac{1+x}{\sqrt{N}}\right)\right)\exp\left\{\frac{x^3}{\sqrt{N}}O\left(\frac{x}{\sqrt{N}}\right)\right\}\end{aligned}$$

and

$$\begin{aligned}&\frac{P\left(\frac{N}{B_N}\left(\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}\right) \le -x\right)}{\Phi(-x)}\\ &= \left(1 + O\left(\frac{1+x}{\sqrt{N}}\right)\right)\exp\left\{\frac{x^3}{\sqrt{N}}O\left(\frac{x}{\sqrt{N}}\right)\right\}.\end{aligned}$$

Step 2. A comparison theorem for Cramér moderate deviations. In order to obtain the Cramér moderate deviations of $\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}$, we establish a comparison theorem for Cramér moderate deviations.

Lemma 5.5. *Let $\{Y_N, N \ge 1\}$ and $\{Z_N, N \ge 1\}$ be two sequences of random variables and let $\{D_N, n \ge 1\}$ be a sequence of positive numbers. Assume that $\{Y_N/Z_N, N \ge 1\}$ satisfies* *the following Cramér-type moderate deviations: for* $0 \le x \le o(N^{1/6})$,

$$\begin{cases} \dfrac{P\left(\frac{Y_N}{Z_N} \ge x\right)}{1-\Phi(x)} = 1 + O\left(\dfrac{1+x^3}{\sqrt{N}}\right) \\ \dfrac{P\left(\frac{Y_N}{Z_N} \le -x\right)}{\Phi(-x)} = 1 + O\left(\dfrac{1+x^3}{\sqrt{N}}\right). \end{cases} \tag{5.6}$$

(1). If there exists $L > 0$ *such that*

$$\limsup_{N\to\infty} \frac{1}{N^{1/3}} \log P\left(\left|\frac{D_N}{Z_N} - 1\right| \ge L/\sqrt{N}\right) < 0, \tag{5.7}$$

then for $0 \le x \le o(N^{1/6})$,

$$\begin{cases} \dfrac{P\left(\frac{Y_N}{D_N} \ge x\right)}{1-\Phi(x)} = 1 + O\left(\dfrac{1+x^3}{\sqrt{N}}\right) \\ \dfrac{P\left(\frac{Y_N}{D_N} \le -x\right)}{\Phi(-x)} = 1 + O\left(\dfrac{1+x^3}{\sqrt{N}}\right). \end{cases} \tag{5.8}$$

(2). If there exists $L > 0$ *such that*

$$\limsup_{N\to\infty} \frac{1}{N^{1/3}} \log P\left(\left|\frac{D_N}{Z_N} - 1\right| \ge L/N^{1/3}\right) < 0, \tag{5.9}$$

then for $0 \le x \le o(N^{1/6})$,

$$\begin{cases} \dfrac{P\left(\frac{Y_N}{D_N} \ge x\right)}{1-\Phi(x)} = 1 + O\left(\dfrac{1+x^2}{N^{1/3}}\right) \\ \dfrac{P\left(\frac{Y_N}{D_N} \le -x\right)}{\Phi(-x)} = 1 + O\left(\dfrac{1+x^2}{N^{1/3}}\right). \end{cases} \tag{5.10}$$

Proof. Let us first give some estimates for $\Phi(x)$. Since

$$\begin{aligned} 1-\Phi(x) &= \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &\le \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} \frac{y}{x} e^{-\frac{y^2}{2}} dy \\ &= \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \quad x > 0. \end{aligned}$$

Let $g(x) = 1 - \Phi(x) - \frac{1}{\sqrt{2\pi}} \frac{x}{1+x^2} e^{-x^2/2}$, $x \ge 0$. Then $g(x)$ is a decreasing function on $[0, \infty)$, $g(0) = 1/2 > 0$ and $g(x) \to 0$ as $x \to \infty$. Thus we have the following estimates:

$$\frac{1}{\sqrt{2\pi}} \frac{x}{1+x^2} e^{-x^2/2} \le 1 - \Phi(x) \le \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}, \ x > 0. \tag{5.11}$$

We write

$$\frac{1-\Phi(x+\epsilon)}{1-\Phi(x)} = 1 + \frac{1}{1-\Phi(x)}\int_{x+\epsilon}^{x}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy.$$

Note that for $|\epsilon| < x$,

$$\begin{aligned} x|\epsilon|e^{-(2x|\epsilon|+\epsilon^2)/2} \leq & \frac{1}{1-\Phi(x)}\left|\int_{x+\epsilon}^{x}\frac{1}{\sqrt{2\pi}}e^{-y^2/2}dy\right| \\ \leq & \frac{|\epsilon|(1+x^2)}{x}e^{(2x|\epsilon|+\epsilon^2)/2}. \end{aligned}$$

Take $\epsilon = O(x/\sqrt{N})$. Then for $0 \leq x \leq o(N^{1/6})$, we have that $\epsilon = o(N^{-1/3})$, $\frac{|\epsilon|}{x} = O(1/\sqrt{N})$ and $e^{(2x|\epsilon|+\epsilon^2)/2} = O(1)$, and so

$$\frac{|\epsilon|(1+x^2)}{x}e^{-(2x|\epsilon|+\epsilon^2)/2} = O\left(\frac{1+x^2}{\sqrt{N}}\right),$$

and

$$x|\epsilon|e^{(2x|\epsilon|+\epsilon^2)/2} = O\left(\frac{1+x^2}{\sqrt{N}}\right).$$

Thus, for $0 \leq x \leq o(N^{1/6})$,

$$\frac{1-\Phi((1+O(1/\sqrt{N}))x)}{1-\Phi(x)} = 1 + O\left(\frac{1+x^2}{\sqrt{N}}\right). \tag{5.12}$$

Similarly, we also have that for $0 \leq x \leq o(N^{1/6})$,

$$\frac{1-\Phi((1+O(1/N^{1/3}))x)}{1-\Phi(x)} = 1 + O\left(\frac{1+x^2}{N^{1/3}}\right). \tag{5.13}$$

Now, let us show (1). We only prove the first one of (5.8). Without loss of generality, we assume $Z_N > 0$. We write

$$\begin{aligned} & P\left(\frac{Y_N}{D_N} \geq x\right) \\ =& P\left(\frac{Y_N}{Z_N} \geq x\left(1+\frac{D_N}{Z_N}-1\right), \left|\frac{D_N}{Z_N}-1\right| < L/\sqrt{N}\right) \\ & + P\left(\frac{Y_N}{D_N} \geq x, \left|\frac{D_N}{Z_N}-1\right| \geq L/\sqrt{N}\right). \end{aligned}$$

It is obvious that

$$\begin{aligned} & P\left(\frac{Y_N}{Z_N} \geq x\left(1+L/\sqrt{N}\right)\right) - P\left(\left|\frac{D_N}{Z_N}-1\right| \geq L/\sqrt{N}\right) \\ \leq & P\left(\frac{Y_N}{Z_N} \geq x\left(1+\frac{D_N}{Z_N}-1\right), \left|\frac{D_N}{Z_N}-1\right| < L/\sqrt{N}\right) \\ \leq & P\left(\frac{Y_N}{Z_N} \geq x\left(1-L/\sqrt{N}\right)\right). \end{aligned}$$

By the condition (5.7), there exists $C > 0$ such that for N large enough, and for all $0 \le x \le o(N^{1/6})$,

$$\frac{P\left(\left|\frac{D_N}{Z_N} - 1\right| \ge L/\sqrt{N}\right)}{1-\Phi(x)} = O\left(e^{-CN^{1/3}}\right).$$

By (5.6) and (5.12) ,

$$\begin{aligned}
&\frac{P\left(\frac{Y_N}{Z_N} \ge x\left(1 \pm L/\sqrt{N}\right)\right)}{1-\Phi(x)} \\
=&\frac{P\left(\frac{Y_N}{Z_N} \ge x\left(1 \pm L/\sqrt{N}\right)\right)}{1-\Phi(x(1\pm L/\sqrt{N}))}\frac{1-\Phi(x(1\pm L/\sqrt{N}))}{1-\Phi(x)} \\
=&\left(1+O\left(\frac{1+x^3}{\sqrt{N}}\right)\right)\left(1+O\left(\frac{1+x^2}{\sqrt{N}}\right)\right) \\
=&1+O\left(\frac{1+x^3}{\sqrt{N}}\right).
\end{aligned}$$

Thus, by (5.7) and

$$\begin{aligned}
&\frac{P\left(\frac{Y_N}{Z_N} \ge x\left(1 + L/\sqrt{N}\right)\right)}{1-\Phi(x)} - \frac{P\left(\left|\frac{D_N}{Z_N} - 1\right| \ge L/\sqrt{N}\right)}{1-\Phi(x)} \\
\le&\frac{P\left(\frac{Y_N}{D_N} \ge x\right)}{1-\Phi(x)} \\
\le&\frac{P\left(\frac{Y_N}{Z_N} \ge x\left(1 - L/\sqrt{N}\right)\right)}{1-\Phi(x)} + \frac{P\left(\left|\frac{D_N}{Z_N} - 1\right| \ge L/\sqrt{N}\right)}{1-\Phi(x)},
\end{aligned}$$

the first one of (5.8) holds.

The proof of (2) is similar to (1). We only notice that by condition (5.9), there exists $C > 0$ such that for N large enough, and for all $0 \le x \le o(N^{1/6})$,

$$\frac{P\left(\left|\frac{D_N}{Z_N} - 1\right| \ge L/N^{1/3}\right)}{1-\Phi(x)} = O\left(e^{-CN^{1/3}}\right).$$

By (5.6) and (5.13) ,

$$\begin{aligned}
&\frac{P\left(\frac{Y_N}{Z_N} \ge x\left(1 \pm L/N^{1/3}\right)\right)}{1-\Phi(x)} \\
=&\frac{P\left(\frac{Y_N}{Z_N} \ge x\left(1 \pm L/N^{1/3}\right)\right)}{1-\Phi(x(1\pm L/N^{1/3}))}\frac{1-\Phi(x(1\pm L/N^{1/3}))}{1-\Phi(x)} \\
=&\left(1+O\left(\frac{1+x^3}{\sqrt{N}}\right)\right)\left(1+O\left(\frac{1+x^2}{N^{1/3}}\right)\right) \\
=&1+O\left(\frac{1+x^2}{N^{1/3}}\right).
\end{aligned}$$

Thus, the first one of (5.10) holds.

□

Step 3. Cramér moderate deviations for $\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}$. Since

$$\left|\frac{B_N^2}{N}-\beta e^{-\beta}(1-e^{-\beta}+\beta)\right|=O\left(\frac{1}{\sqrt{N}}\right),$$

we have

$$\left|\frac{B_N}{\sqrt{N}\sqrt{\beta e^{-\beta}(1-e^{-\beta}+\beta)}}-1\right|=O\left(\frac{1}{\sqrt{N}}\right).$$

Take $Y_n=\sqrt{N}(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N})$, $Z_n=\frac{B_N}{\sqrt{N}}$ and $D_N=\sqrt{\beta e^{-\beta}(1-e^{-\beta}+\beta)}$ in Lemma 5.5 (1). Then noting that Z_n and D_N are non-random, the following Cramér moderate deviations is a consequence of Lemma 5.3 and Lemma 5.5 (1).

Lemma 5.6. *For $0\le x\le o(N^{1/6})$,*

$$\frac{P\left(\frac{\sqrt{N}(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N})}{\sqrt{\beta e^{-\beta}(1-e^{-\beta}+\beta)}}\ge x\right)}{1-\Phi(x)}=1+O\left(\frac{1+x^3}{\sqrt{N}}\right), \tag{5.14}$$

$$\frac{P\left(\frac{\sqrt{N}(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N})}{\sqrt{\beta e^{-\beta}(1-e^{-\beta}+\beta)}}\le -x\right)}{\Phi(-x)}=1+O\left(\frac{1+x^3}{\sqrt{N}}\right). \tag{5.15}$$

Step 4. Deviation estimates for $\Gamma_{m,N}-\gamma_{m,N}$. By Proposition 3.1, we have the following deviation estimates for $\Gamma_{m,N}-\gamma_{m,N}$.

Lemma 5.7. *For each $m\ge 0$, $L>0$,*

$$\limsup_{N\to\infty}\frac{1}{N^{1/3}}\log P\left(|\Gamma_{m,N}-\gamma_{m,N}|\ge\frac{L}{N^{1/3}}\right)<0, \tag{5.16}$$

and

$$\limsup_{N\to\infty}\frac{1}{N^{1/3}}\log P\left(\left|\Gamma_{m,N}-\frac{\beta^m}{m!}e^{-\beta}\right|\ge\frac{L}{N^{1/3}}\right)<0. \tag{5.17}$$

Proof. Since $\left|\gamma_{m,N}-\frac{\beta^m}{m!}e^{-\beta}\right|=O(1/\sqrt{N})$ we only need to show (5.16).

Using $\log(1+x)=x-\frac{x^2}{2}+O(x^3)$ as $x\to 0$, and noting that for any $m\ge 0$, as $N\to\infty$,

$\mu_{m,N} \to \sum_{l=0}^{m} \frac{\beta^l}{l!} e^{-\beta}$, we have that

$$
\begin{aligned}
& I_m^N\left(\frac{x}{N^{1/3}}\right) \\
=& \left(\frac{x}{N^{1/3}} + \mu_{m,N}\right) \log\left(1 + \frac{x}{N^{1/3}\mu_{m,N}}\right) \\
& + \left(1 - \mu_{m,N} - \frac{x}{N^{1/3}}\right) \log\left(1 - \frac{x}{(1-\mu_{m,N})N^{1/3}}\right) \\
=& \left(\frac{x}{N^{1/3}} + \mu_{m,N}\right) \\
& \times \left(\frac{x}{N^{1/3}\mu_{m,N}} - \frac{x^2}{2(N^{1/3}\mu_{m,N})^2} + O\left(\frac{|x|^3}{N}\right)\right) \\
& + \left(1 - \mu_{m,N} - \frac{x}{N^{1/3}}\right) \\
& \quad \times \left(-\frac{x}{N^{1/3}(1-\mu_{m,N})}\right. \\
& \qquad \left. + \frac{x^2}{2(N^{1/3}(1-\mu_{m,N}))^2} + O\left(\frac{|x|^3}{N}\right)\right) \\
=& \frac{x^2}{2N^{2/3}\mu_{m,N}} + \frac{x^2}{2N^{2/3}(1-\mu_{m,N})} + O\left(\frac{|x|^3}{N}\right) \\
=& \frac{x^2}{2N^{2/3}\mu_{m,N}(1-\mu_{m,N})} + O\left(\frac{|x|^3}{N}\right).
\end{aligned}
$$

Therefore, for any $m \geq 0$, $L > 0$,

$$
\begin{aligned}
J_m^N(L/N^{1/3}) =& \min\{I_m^N(L/N^{1/3}), I_m^N(-L/N^{1/3}))\} \\
=& \frac{L^2}{2N^{2/3}\mu_{m,N}(1-\mu_{m,N})} + O\left(\frac{L^3}{N}\right),
\end{aligned}
$$

which yields

$$
\begin{aligned}
& \liminf_{N\to\infty} \frac{1}{N^{1/3}} N J_m^N(L/N^{1/3}) \\
=& \frac{L^2}{2\sum_{l=0}^{m} \frac{\beta^l}{l!} e^{-\beta}(1 - \sum_{l=0}^{m} \frac{\beta^l}{l!} e^{-\beta})} > 0.
\end{aligned}
$$

By (3.14) and (3.15), we have that for any $\epsilon > 0$,

$$
\begin{aligned}
& \log P\left(|\Gamma_{m,N} - \gamma_{m,N}| \geq \epsilon\right) \\
\leq & \log(2\exp\left\{-NJ_{m-1}^N(\epsilon/2)\right\} + 2\exp\left\{-NJ_m^N(\epsilon/2)\right\}) \\
=& \log 2 + \log(\exp\left\{-NJ_{m-1}^N(\epsilon/2)\right\} + \exp\left\{-NJ_m^N(\epsilon/2)\right\}) \\
\leq & 2\log 2 + \max\{-NJ_{m-1}^N(\epsilon/2), -NJ_m^N(\epsilon/2)\}.
\end{aligned}
\tag{5.18}
$$

Thus, (5.16) holds.

□

The following lemma is also obtained easily from Proposition 3.1.

Lemma 5.8. *Let $b_N, N \geq 1$ be a sequence of positive numbers with $b_N \to \infty$ and $b_N = o(\sqrt{N})$. Then for each $m \geq 0$, for any $\epsilon > 0$,*

$$\limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(|\Gamma_{m,N} - \gamma_{m,N}| \geq \epsilon\right) = -\infty, \tag{5.19}$$

and

$$\limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(\left|\Gamma_{m,N} - \frac{\beta^m}{m!}e^{-\beta}\right| \geq \epsilon\right) = -\infty. \tag{5.20}$$

Step 5. Moderate deviation principle for $\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}$. The following lemma is a corollary of Lemma 5.4. It is also obtained from Theorem 2.1 in [13].

Lemma 5.9. *Let $b_N, N \geq 1$ be a sequence of positive numbers with $b_N \to \infty$ and $b_N = o(\sqrt{N})$. Then for any $x > 0$,*

$$\begin{aligned}&\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}\right) > x\right)\\ &= -\frac{x^2}{2\beta e^{-\beta}(1-e^{-\beta}+\beta)},\end{aligned} \tag{5.21}$$

and

$$\begin{aligned}&\lim_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}\right) < -x\right)\\ &= -\frac{x^2}{2\beta e^{-\beta}(1-e^{-\beta}+\beta)}.\end{aligned} \tag{5.22}$$

Proof. We only show the first one. Note that for $x > 0$,

$$\begin{aligned}&\lim_{N\to\infty} \frac{1}{b_N^2} \log\left(1 - \Phi\left(\frac{\sqrt{N}b_N x}{B_N}\right)\right)\\ &= -\frac{x^2}{2\beta e^{-\beta}(1-e^{-\beta}+\beta)}.\end{aligned}$$

By Lemma 5.4

$$\begin{aligned}&P\left(\sqrt{N}\left(\Gamma_{1,N} - \frac{n}{N}\Gamma_{0,N}\right) \geq b_N x\right)\\ &= \frac{P\left(\frac{N\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)}{B_N} \geq \frac{\sqrt{N}b_N x}{B_N}\right)}{1-\Phi\left(\frac{\sqrt{N}b_N x}{B_N}\right)}\left(1 - \Phi\left(\frac{\sqrt{N}b_N x}{B_N}\right)\right)\\ &= \left(1 + O\left(\frac{1+\frac{\sqrt{N}b_N x}{B_N}}{\sqrt{N}}\right)\right)\exp\left\{\frac{(\sqrt{N}b_N x)^3}{B_N^3\sqrt{N}} O\left(\frac{b_N x}{B_N}\right)\right\}\\ &\quad\times\left(1 - \Phi\left(\frac{\sqrt{N}b_N x}{B_N}\right)\right),\end{aligned}$$

We have

$$\lim_{N\to\infty}\frac{1}{b_N^2}\log P\left(\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)>x\right)$$
$$=-\frac{x^2}{2\beta e^{-\beta}(1-e^{-\beta}+\beta)}.$$

□

Step 6. Proof of Theorem 3.1.

Proof of Theorem 3.1. We only show (1). For any $\epsilon>0$, when N is large enough such that $|\frac{n}{N}-\beta|<\epsilon/2$,

$$\begin{aligned}
&\left\{\frac{\sqrt{N}}{b_N}\left(\frac{\hat{N}}{N}-1\right)>x\right\}\\
=&\left\{\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)>x(\frac{n}{N}-\Gamma_{1,N})\right\}\\
\subset&\left\{\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)>x(\frac{n}{N}-\Gamma_{1,N}),\right.\\
&\left.|\frac{n}{N}-\beta-\Gamma_{1,N}+\beta e^{-\beta}|\le\epsilon\right\}\\
&\cup\left\{|\frac{n}{N}-\beta-\Gamma_{1,N}+\beta e^{-\beta}|\ge\epsilon\right\}\\
\subset&\left\{\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)>x(\beta(1-e^{-\beta})-\epsilon)\right\}\\
&\cup\left\{|\frac{n}{N}-\beta-\Gamma_{1,N}+\beta e^{-\beta}|\ge\epsilon\right\}\\
\subset&\left\{\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)>x(\beta(1-e^{-\beta})-\epsilon)\right\}\\
&\cup\left\{|\Gamma_{1,N}-\beta e^{-\beta}|\ge\epsilon/2\right\},
\end{aligned}$$

and

$$\begin{aligned}
&\left\{\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)>x\beta(1-e^{-\beta})+x\epsilon\right\}\\
\subset&\left\{\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)>x(\frac{n}{N}-\Gamma_{1,N}),\right.\\
&\left.|\frac{n}{N}-\beta-\Gamma_{1,N}+\beta e^{-\beta}|<\epsilon\right\}\\
&\cup\left\{|\frac{n}{N}-\beta-\Gamma_{1,N}+\beta e^{-\beta}|\ge\epsilon\right\}\\
\subset&\left\{\frac{\sqrt{N}}{b_N}\left(\frac{\hat{N}}{N}-1\right)>x\right\}\cup\left\{|\Gamma_{1,N}-\beta e^{-\beta}|\ge\epsilon/2\right\}.
\end{aligned}$$

Therefore, by Lemma 5.8 and 5.9 we have

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left(\frac{\hat{N}}{N}-1\right) > x\right)\\
\leq &\max\Bigg\{\limsup_{N\to\infty} \frac{1}{b_N^2} \log P\bigg(\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)\\
&\qquad > x(\beta(1-e^{-\beta})-\epsilon)\bigg),\\
&\quad \limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(|\Gamma_{1,N}-\beta e^{-\beta}| \geq \epsilon/2\right)\Bigg\}\\
= &-\frac{x^2(\beta(1-e^{-\beta})-\epsilon)^2}{2\beta e^{-\beta}(1-e^{-\beta}+\beta)},
\end{aligned}$$

and

$$\begin{aligned}
&\liminf_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left(\frac{\hat{N}}{N}-1\right) > x\right)\\
= &\max\Bigg\{\liminf_{N\to\infty} \frac{1}{b_N^2} \log P\left(\frac{\sqrt{N}}{b_N}\left(\frac{\hat{N}}{N}-1\right) > x\right),\\
&\quad \limsup_{N\to\infty} \frac{1}{b_N^2} \log P\left(|\Gamma_{1,N}-\beta e^{-\beta}| \geq \epsilon/2\right)\Bigg\}\\
\geq &\liminf_{N\to\infty} \frac{1}{b_N^2} \log P\bigg(\frac{\sqrt{N}}{b_N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)\\
&\qquad > x(\beta(1-e^{-\beta})+\epsilon)\bigg)\\
= &-\frac{x^2(\beta(1-e^{-\beta})+\epsilon)^2}{2\beta e^{-\beta}(1-e^{-\beta}+\beta)}.
\end{aligned}$$

Finally, letting $\epsilon \to 0$, we obtain (1).

□

Step 7. Proof of Theorem 3.2.

Proof of Theorem 3.2. (1). Since

$$\frac{\sqrt{N}\left(\frac{\hat{N}}{N}-1\right)}{\sigma(\beta)} = \frac{\sqrt{N}\left(\Gamma_{1,N}-\frac{n}{N}\Gamma_{0,N}\right)}{\sqrt{\beta e^{-\beta}(1-e^{-\beta}+\beta)}\frac{\frac{n}{N}-\Gamma_{1,N}}{\beta(1-e^{-\beta})}},$$

by Lemma 5.5 (2), and Lemma 5.6, it suffices to prove that for some $L > 0$,

$$\limsup_{N\to\infty} \frac{1}{N^{1/3}} \log P\left(\left|\frac{\beta(1-e^{-\beta})}{\frac{n}{N}-\Gamma_{1,N}}-1\right| \geq \frac{L}{N^{1/3}}\right) < 0. \tag{5.23}$$

When N is large enough,

$$\begin{aligned}
&\left\{\left|\frac{\beta(1-e^{-\beta})}{\frac{n}{N}-\Gamma_{1,N}}-1\right|\geq\frac{L}{N^{1/3}}\right\}\\
=&\left\{\frac{|\beta(1-e^{-\beta})-(\frac{n}{N}-\Gamma_{1,N})|}{|\frac{n}{N}-\Gamma_{1,N}|}\geq\frac{L}{N^{1/3}}\right\}\\
\subset&\left\{\frac{|\beta-\frac{n}{N}|+|\beta e^{-\beta}-\Gamma_{1,N}|}{|\frac{n}{N}-\Gamma_{1,N}|}\geq\frac{L}{N^{1/3}}\right\}\\
=&\left\{|\frac{n}{N}-\Gamma_{1,N}|\leq\frac{N^{1/3}}{L}(|\beta-\frac{n}{N}|+|\beta e^{-\beta}-\Gamma_{1,N}|)\right\}\\
\subset&\bigg\{|\frac{n}{N}-\Gamma_{1,N}|\leq\frac{N^{1/3}}{L}(|\beta-\frac{n}{N}|+|\beta e^{-\beta}-\Gamma_{1,N}|),\\
&\quad|\beta e^{-\beta}-\Gamma_{1,N}|\leq\frac{L|\frac{n}{N}-\gamma_{1,N}|}{2N^{1/3}}\bigg\}\\
&\cup\left\{|\beta e^{-\beta}-\Gamma_{1,N}|\geq\frac{L|\frac{n}{N}-\gamma_{1,N}|}{2N^{1/3}}\right\}\\
\subset&\left\{|\beta e^{-\beta}-\Gamma_{1,N}|\geq\frac{L|\frac{n}{N}-\gamma_{1,N}|}{2N^{1/3}}\right\}\\
&\cup\left\{|\frac{n}{N}-\Gamma_{1,N}|\leq|\frac{n}{N}-\gamma_{1,N}|/2+\frac{N^{1/3}}{L}|\beta-\frac{n}{N}|\right\}\\
\subset&\left\{|\beta e^{-\beta}-\Gamma_{1,N}|\geq\frac{L|\frac{n}{N}-\gamma_{1,N}|}{2N^{1/3}}\right\}\\
&\cup\left\{|\gamma_{1,N}-\Gamma_{1,N}|\geq|\frac{n}{N}-\gamma_{1,N}|/2-\frac{N^{1/3}}{L}|\beta-\frac{n}{N}|\right\}\\
\subset&\left\{|\beta e^{-\beta}-\Gamma_{1,N}|\geq\frac{L\beta(1-e^{-\beta})}{4N^{1/3}}\right\}\\
&\cup\left\{|\gamma_{1,N}-\Gamma_{1,N}|\geq\beta(1-e^{-\beta})/4)\right\},
\end{aligned}$$

since $|\frac{n}{N}-\gamma_{1,N}|\to\beta-\beta e^{-\beta}$ and $\frac{N^{1/3}}{L}|\beta-\frac{n}{N}|\to 0$ as $N\to\infty$.

By Lemma 5.7, we obtain (5.23).

(2). By the proof of Lemma 5.5 (2), and Lemma 5.6, it suffices to prove that for some $L>0$,

$$\limsup_{N\to\infty}\frac{1}{N^{1/3}}\log P\left(\left|\frac{\sqrt{\hat{N}}}{\sqrt{N}}-1\right|\geq\frac{L}{N^{1/3}}\right)<0. \tag{5.24}$$

Note that

$$\left|\frac{\sqrt{\hat{N}}}{\sqrt{N}}-1\right|=\frac{|N-\hat{N}|}{\sqrt{N}(\sqrt{N}+\sqrt{\hat{N}})}\leq\frac{|N-\hat{N}|}{N}.$$

Take $b_N = N^{1/6}$ in Theorem 3.1, then

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N^{1/3}} \log P\left(\left| \frac{\sqrt{\hat{N}}}{\sqrt{N}} - 1 \right| \geq \frac{L}{N^{1/3}} \right) \\
\leq &\limsup_{N\to\infty} \frac{1}{N^{1/3}} \log P\left(\frac{\sqrt{N}}{N^{1/6}} \left| \frac{\hat{N}}{N} - 1 \right| \geq L \right) \\
\leq &- \frac{L^2}{2\sigma^2(\beta)} < 0.
\end{aligned}$$

□

6 Conclusion

In large-scale sensor networks, estimating the number N of operating nodes of a sensor network is crucial to network operation. In [2], Budianu, Ben-David and Tong presented a performance analysis of the Good-Turing estimator $\hat{N}$ of the number N of operating nodes where the deviations are $O(N)$ based on large deviations. In this paper, we establish asymptotic normality, Cramér-type moderate deviations, moderate deviation principle and some tail probability estimatesof the estimator of the number N of operating nodes and give a performance analysis based on tail probability estimates and moderate deviations. Our approach provides a characterization of convergence rate of the Good-Turing estimator $\hat{N}$ where the deviations of $\hat{N} - N$ are N^ν, $\nu \in (1/2, 1)$. These estimates also provide a method to build confidence interval of N.

Acknowledgment

The authors thank the editor and two anonymous referees for helpful comments and suggestions, which greatly improves the quality of the paper.

References

[1] C. Budianu and L. Tong, Estimation of the number of operating sensors in a sensor network, *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, 2 (2003), 1728–1732.

[2] C. Budianu, S. Ben-David and L. Tong, Estimation of the number of operating sensors in large-scale sensor networks with mobile access, *IEEE Transactions on Signal Processing*, 54 (2006), 1703–1715.

[3] I. Good, The population frequencies of species and the estimation of population parameters, *Biometrika*, 40 (1953), 237–264.

[4] D. McAllester, R. E. Schapire, On the convergence rate of Good-Turing estimators, In *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory*, (2000), 1-6.

[5] D. McAllester, L. Ortiz, Concentration inequalities for the missing mass and for histogram rule error, *J. Mach. Learn. Res.*, 4 (2003), 895-911.

[6] D. Berend and A. Kontorovich, On the concentration of the missing mass, *Electronic Communications in Probability*, 18 (2013), no.3, 1-7.

[7] A. Ben-Hamou, S. Boucheron and M. I. Ohannessian, Concentration inequalities in the infinite urn scheme for occupancy counts and the missing mass, with applications, *Bernoulli*, 23 (2017), 249-287.

[8] B. Y. S. Khanloo and G. Haffari, Novel Bernstein-like concentration inequalities for the missing mass, arXiv preprint arXiv:1503.02768, 2015.

[9] W. W. Esty, A normal limit law for a nonparametric estimator of the coverage of a random sample, *Annals of Statistics*, 11 (1983), 905-912.

[10] C. H. Zhang, Z. Zhang, Asymptotic normality of a nonparametric estimator of sample coverage, *Annals of Statistics*, 37 (2009), 2582-2595.

[11] M. Grabchak and Z. Zhang, Asymptotic properties of turing's formula in relative error, *Machine Learning*, 106 (2017), 1771-1785.

[12] P. Dupuis, C. Nuzman, and P. Whiting, Large deviations asymptotics for occupancy problems, *Annals of Probability*, 32 (2004), 2765-2818.

[13] F. Q. Gao, Moderate deviations for a nonparametric estimator of sample coverage, *Annals of Statistics*, 41 (2013),641–669.

[14] A. Orlitsky, N. Santhanam, and J. Zhang, Always Good-Turing: Asymptotically optimal probability estimation, *Science*, 302 (2003), 427–431.

[15] N. Rajaraman, A. Thangaraj, A. T. Suresh, Minimax Risk for Missing Mass Estimation, In *Proceedings of the IEEE International Symposium on Information Theory* , 2017.

[16] V. V. Petrov, Sums of Independent Random Variables, *Springer*, New York, 1975.

[17] L. Saulis and V. A. Statulevicius, Limit theorems for large deviations, *Springer*, 1991.

[18] V. Feray, P. L. Meliot and A. Nikeghbali, Mod-ϕ convergence I: Normality zones and precise deviations, *SpringerBriefs in Probability and Mathematical Statistics*, Springer, 2016.

[19] Y. Altuğ and A. B. Wagner, Moderate deviation analysis of channel coding: Discrete memoryless case, in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, Jun. 2010, 265–269.

[20] Y. Altuğ and A. B. Wagner, Moderate deviations in channel coding, *IEEE Trans. Inf. Theory*, 60 (2014), 4417–4426.

[21] Y. Polyanskiy and S. Verdú, Channel dispersion and moderate deviations limits for memoryless channels, in *Proc. 48th Annu. Allerton Conf. Commun. Control, Comput.*, Monticello, IL, USA, 2010,1334-1339.

[22] I. Sason, Moderate deviations analysis of binary hypothesis testing, In *Proceedings of the International Symposium on Information Theory*, Cambridge, MA, Jul 2012, 821–825.

[23] V. Y. F. Tan, Moderate-deviations of lossy source coding for discrete and Gaussian sources, In *Proceedings of the International Symposium on Information Theory*, Cambridge, MA, Jul 2012, 920–924.

[24] V. Y. F. Tan, S. Watanabe, and M. Hayashi, Moderate deviations for joint source-channel coding of systems with Markovian memory, In *Proceedings of the International Symposium on Information Theory*, Honolulu, HI, Jul 2014,1687–1691.

[25] S. H. Lee, Vincent Y. F. Tan and A. Khisti, Exact Moderate Deviations Asymptotics for Streaming Data Transmission, *IEEE Transactions on Information Theory*, 63 (2017), 2726–2736.

[26] L. Zhou, Vincent Y. F. Tan and M. Motani, Second-Order and Moderate Deviation Asymptotics for Successive Refinement, *IEEE Transactions on Information Theory*, 63 (2017), 2896–2921.

[27] A. Dembo and O. Zeitouni, Large deviations techniques and applications, 2nd ed., *Springer Verlag*, 1998.

[28] S. H. Sirazhdinov, Sh. A. Mirakhmedov, and Sh. A. Ismatullaev, Probabilities of large deviations for randomized divisible statistics in the multinomial scheme. *Theory Probab. Appl.*, 34 (1990), 645–657.

[29] A. Kamath, R. Motwani, K. Palem and P. Spirakis, Tail bounds for occupancy problems and the satisability threshold conjecture, *Random Structures and Algorithm*, 7 (1995), 59–80.

[30] D. Dubhashi, D. Ranjan, Balls and bins: A study in negative dependence, *Random Structures Algorithms*, 13 (1998), 99–124.