Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 27 (2024), No. 5, pp. 959–969
DOI : 10.47974/JSMS-1220

Large deviations : Estimating the number of positive individuals in a Bernoulli sample through sample pooling

Thomas Katsekpor [†]
Department of Mathematics
University of Ghana
Legon
Ghana

Kwabena Doku-Amponsah [*]
Department of Statistics & Actuarial Science
University of Ghana
Legon
Ghana

Abstract

We define the empirical positives measure as a stochastic measure that quantifies the number of infections in a Bernoulli sample. Similarly, we define the empirical pool measure as a stochastic measure that enumerates the number of infected pool samples. A joint large deviation principle (LDP) is established for both empirical measures derived from the Bernoulli samples. Using this joint LDP, we establish a correlation between the asymptotic ratio of infected individuals to the sample size and the asymptotic ratio of infected pool samples to the total number of pool samples.

Subject Classification: *60F10, 11B68, 11P84, 68Q25.*

Keywords: *Large deviations, Infected pool sample, Empirical pool measure, Empirical positives measure, Integer partition, Entropy, Relative entropy.*

1. Introduction

Sample pooling in medical research involves testing multiple individuals for infection simultaneously, whereas pool testing entails combining samples from individuals and examining them for the presence

[†] *E-mail:* `tkatsekpor@ug.edu.gh`
[*] *E-mail:* `kdoku-amponsah@ug.edu.gh (Corresponding Author)`

of infectious diseases. For example, during the initial stages of the outbreak of Covid-19, the Noguchi Memorial Institute for Medical Research (NMIMR) grouped samples together to enhance testing capacity in Ghana, particularly employing asymptomatic diagnostic tests grouping up to fifteen individuals per pool. This strategy proved beneficial given the country's scarce resources, such as tools and test kit availability.

This approach is particularly effective when the prevalence rate is low, resulting in more negative results. However, a significant challenge arises when specificity is exceptionally high, potentially leading to all pool samples testing positive. Identifying the number of asymptotically infected individuals within the sample through infected pool samples could significantly impact the efficient management of the pandemic.

Researchers have already identified relationships in COVID-19 data, effectively aiding pandemic management. For instance, significant statistical relationships have been observed between gender and patient transmission type. See, [2]. Additionally, relationships in time series data through support vector machines have been utilized to manage the evolving number of infections from the pandemic. See, [8].

In this article, we found a function between the asymptotic proportion of infected cases and proportion of infected grouped individuals as the number of individuals to be tested increases significantly. Specifically, we define empirical positive and pooled measures for any sample pool and derive a function between the number of positive individuals in the sample and the number of positive grouped samples using Large deviation techniques. We employ methodologies similar to those used by previous researchers. i.e. combinatorial ideas via the method of types, see [6], [3], [5], [4] and method of mixtures, see [1].

The remaining sections of the article are outlined as follows: Section Section 2 provides background information and presents the main results. Section 3 presents the proofs of the main results. while Section 5 presents the derivation of Corollary 2.2 and the research conclusion.

2. Statement of Main Results

2.1 *Empirical Measures of the Bernoulli Samples*

Denote by $Y=(Y_1,Y_2,Y_3,...,Y_n)$ a sequences of i.i.d bernoulli variables each with mean p_n. By $(Z_1,Z_2,Z_3,...,Z_{m(n)})$, we designate a random uniform sample taken from a collection of integer partitions of n into $m(n)$ numbers. Let members of Y be arranged as $(Y_1^{(1)}, ..., Y_{Z_1}^{(1)}), ..., (Y_1^{(k)}, ...,$

$Y_{Z_m}^{(m)}$) such that $Y_r^{(u)}$ and $Y_s^{(v)}$ are always independent random variables for all $u \neq v$, and $r \neq s$, $r, s = 1, 2, 3, \ldots$. We write $X_j := (Y_1^{(j)}, \ldots, Y_{Z_j}^{(j)})$, for $j = 1, 2, \ldots m(n)$ and refer to $(X_1, \ldots, X_{m(n)})$ as pool samples formed on the members of Y. We shall consider the grouped individuals under the conditional law of the grouped individuals given the integer partitions of n:

$$\mathbb{P}_n\{X_1, \ldots, X_{m(n)}\} := \mathbb{P}\{X_1, \ldots, X_m(n) | Z_1 \neq 0, \ldots, Z_{m(n)} \neq 0, N = n\},$$

where

$$Z_1 + Z_2 + Z_3 + \ldots Z_{m(n)-1} + Z_{m(n)} = N.$$

We write $\Omega = \{0,1\}$ and define for each sample Y, the *empirical individual infection measure*, $M_1^Y \in \mathcal{P}(\Omega)$, by

$$M_1^Y(q) := \sum_{i=1}^{n} \delta_{Y_i}(q) / n$$

and the empirical pooled infection measure $M_2^Y \in \mathcal{P}(\mathbb{N} \cup \{0\} \times \mathcal{M}(\Omega))$, by

$$M_2^{Y,N}(z,l) := \sum_{r=1}^{m(n)} \delta_{(Z_r, \mathcal{L}_r)}(z,l) / n a_n,$$

where $\mathcal{L}_r = (\frac{1}{Z_r} \sum_{i=1}^{Z_r} \delta_{Y_i^{(r)}}(y), y \in \Omega)$ and $a_n = m(n) n^{-1}$. Note that $\| M_1^Y \| = \| M_2^{Y,N} \| = \mathbb{1}$. Furthermore, we note that $M_1^Y(y) = a_n \sum_{(z,l) \in \mathbb{N} \cup \{0\} \times \mathcal{M}(\Omega)} z l(y)$ $M_2^Y(z,l)$. Denote by $\mathcal{P}(\Omega)$ space of probability distributions on the polish space Ω endowed with the weak topology. We represent space of counting measures on space $\mathcal{X}$ by $\mathcal{M}(\mathcal{X})$ and by convention, for any pair (z,l) we have $l = (0,0)$ iff $m = 0$. We present the first theorem of article, Theorem 2.1 as follows:

2.2 *Main Results*

Throughout the remaining sections, we take $m(n) \to \infty$, $m(n)/n := a_n \to a \in (0,1]$ as $n \to \infty$ and $np_n / m(n) \to p$. Note, under above conditions,we obtain $a(p(0) + p(1)) = 1$. We define a rate function J_a : $\mathcal{P}(\Omega) \times \mathcal{P}(\mathbb{N} \cup \{0\} \times \mathcal{M}(\Omega)) \to (0, \infty]$ by

$$J_a(\sigma, \nu) = \begin{cases} H(\sigma / a \| p) + H(\nu \| \varpi_a^{\sigma}), & \text{if} \quad \langle \nu \rangle = \sigma / a, \\ \infty & \text{otherwise,} \end{cases} \tag{2.1}$$

where

$$\langle \nu \rangle(y) = \sum_{(z,l)\in\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega)} zl(y)\nu(z,l)$$

and the tri-variate probability distribution

$$\varpi_a^{\sigma}(z,l) = \frac{1}{(1-e^{-1/a})}\left[\frac{[\sigma(0)/a]^{[zl(0)]}e^{-[\sigma(0)/a]}}{[zl(0)]!}\right]\left[\frac{[\sigma(1)/a]^{[zl(1)]}e^{-[\sigma(1)/a]}}{[zl(1)]!}\right],$$
$$z\in\mathbb{N}\cup\{0\}, l\in\mathcal{M}(\Omega). \quad (2.2)$$

Theorem 2.1: *Let* $Y=(Y_1,Y_2,\ldots,Y_n)^T$ *be a succession of i.i.d bernoulli numbers each having mean* p_n *that satisfies* $np_n(x)\,/\,m(n)\to p(x),$ *for all* $x\in\Omega.$ *Suppose* $X_1,\ldots,X_{m(n)}$ *is pool individuals formed from terms of Y. Then, as* $n\to\infty,$ $(M_1^Y, M_2^{Y,N})$ *obeys an LDP on* $\mathcal{P}(\Omega)\times\mathcal{P}(\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega))$ *with good rate function* $aJ_a(\sigma,\nu).$

We denote by I the the proportion of positive individuals in the sample and state the following corollary:

Corollary 2.2: *Suppose* $Y=(Y_1,Y_2,\ldots,Y_n)^T$ *is a succession of i.i.d bernoulli random numbers each with mean* p_n *that satisfies* $np_n(y)\,/\,m(n)\to p(y),$ *for all* $y\in\Omega.$ *Then, I, satisfies*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}_n\{I=x,S=\sigma\} = -\left[(1-x)\log\left(\frac{(1-x)}{(1-xp(1))}\right)+x\log\frac{x}{ap(1)}\right], \quad (2.3)$$

$$x=-a\log[1-(1-e^{-1/a})\sigma].$$

The next LDP, Theorem 2.3 will be instrumental in our proof of the main result.

Theorem 2.3: Let $Y=(Y_1,Y_2,\ldots,Y_n)^T$ be a succession of i.i.d bernoulli numbers each having mean such that $np_n(y)\,/\,m(n)\to p(y).$ Suppose $X_1,\ldots,X_{m(n)}$ is pool individuals formed from terms of Y conditionally on the space $\{M_1^Y=\sigma\}.$ Then, as $n\to\infty,$ $M_2^{Y,N}$ obeys an LDP on $\mathcal{P}(\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega))$ with good rate function $aJ_\sigma^a(\nu),$ where

$$J_\sigma^a(\nu)=\begin{cases} H(\nu\,\|\,\varpi_a^{\sigma}), & \text{if} \quad \langle\nu\rangle=\sigma/a. \\ \infty & \text{otherwise.}\end{cases} \quad (2.4)$$

The next LDP, Theorem 2.4 will also be instrumental in the proof of the main Theorem. We write

$$I^a(\sigma) := H(\sigma / a \| p). \tag{2.5}$$

Theorem 2.4: *Let* $Y = (Y_1, Y_2, ..., Y_n)^T$ *be a succession of i.i.d bernoulli numbers each having mean that satisfies* $np_n(x) \ / \ m(n) \to q(x)$. *Let* $X_1, ..., X_{m(n)}$ *be pool individuals formed from terms of Y. Then, as* $n \to \infty$, M_1^Y *obeys an LDP in* $\mathcal{P}(\Omega\})$ *on scale n and with rate* $aI^a(\sigma)$.

3. Proof of The Auxiliary Large Deviation Principles

3.1 *Proof of Theorem 2.4*

Lemma 3.1 will be crucial for proving the second auxiliary LDP.

Lemma 3.1: *Let* $Y = (Y_1, Y_2, ..., Y_n)^T$ *is a succession of i.i.d bernoulli numbers each having mean* $np_n(x)/m(n) \to p(x)$. *Let* $X_1, ..., X_{m(n)}$ *be grouped samples formed from terms of Y. Then,*

$$\lim_{\lambda \to \infty} \frac{1}{n} \log \mathbb{E}\{e^{n\langle f, M_1^Y \rangle}\} = -\sum_{y=0}^{1}([1 - e^{f(y)}]ap(y)) = -\langle 1 - e^f, ap \rangle.$$

Proof:

$$\begin{aligned}
\mathbb{E}\{e^{n\langle f, M_1^Y \rangle}\} &= \mathbb{E}\{\prod_{i=1}^{n} e^{f(Y_i)}\} \\
&= \prod_{i=1}^{n} \mathbb{E}(e^{f(Y_i)}) \\
&= \prod_{x \in \Omega} (1 - p_n(y) + e^{f(x)} p_n(y))^n \\
&= \prod_{y \in \Omega} e^{-(1 - e^{f(y)})p(y)m(n) + o(n)}
\end{aligned}$$

We take limit of the normalized logarithm to obtain

$$\lim_{\lambda \to \infty} \frac{1}{n} \log \mathbb{E}\{e^{n\langle f, M_1^Y \rangle}\} = -\textstyle\sum_{y=0}^{1}([1 - e^{f(y)}]ap(y)).$$

By Gartner-Ellis Theorem, law of M_1^Y satisfies an LDP on the scale n and good rate function

$$I(\sigma) = \sup_f \{\langle f, \sigma \rangle + \langle (1 - e^f), pa \rangle\}.$$

We solve the variational problem to obtain the relative entropy

$$I(\sigma) = aH(\sigma / a \,\| p)$$

This proves Theorem 2.4

3.2 *Proof of Theorem 2.3*

Write down two main sets:

$$\mathcal{P}_n(\Omega):=\{\sigma_n\in\mathcal{P}(\Omega): n\sigma_n(0), n\sigma_n(1)\in\mathbb{N}\cup\{0\}\}$$

and

$$\mathcal{P}_n(\mathbb{N}\times\mathcal{M}(\Omega)):=\{\nu_n\in\mathcal{P}(\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega)): n\nu_n(z,l)\in\mathbb{N}\cup\{0\},\ \text{for all}(z,l)\in\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega)\}.$$

Lemma 3.2 is important for the first auxiliary Theorem.

Lemma 3.2: *Suppose* $Y=(Y_1,Y_2,...,Y_n)^T$ *is a succession of i.i.d bernoulli numbers each having mean* p_n. *Let* $X_1,...,X_{m(n)}$ *be grouped individuals formed from terms of Y. Let* $(\sigma_n,\nu_n)\to(\sigma,\nu)\in\mathcal{P}(\Omega)\times\mathcal{P}(\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega))$. *Then, we have*

$$e^{-na_nH(\nu_n|\varpi_{a_n}^{\sigma_n})+n\beta_1(n)}\le\mathbb{P}_n\{M_2^{Y,N}=\nu_n\mid M_1^Y=\sigma_n\}\le e^{-na_nH(\nu_n|\varpi_{a_n}^{\sigma_n})+n\beta_2(n)},$$

$$\lim_{n\to\infty}\beta_1(n)=\lim_{n\to\infty}\beta_2(n)=0.$$

Proof: Note that the conditionally on the event $\{M_1^X=\sigma_n\}$ the probability of $M_2^{Y,N}=\nu_n$ is

$$\begin{aligned}&\mathbb{P}_n\{M_2^{Y,N}=\nu_n\mid M_1^Y=\sigma_n\}\\&=\binom{n}{na_n\nu_n(z,l),(z,l)\in\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega)}\times\prod_{y\in\Omega}\binom{n\sigma_n(y)}{m_1l_1(y),m_2l_2(y),...,m_{m(n)}l_{m(n)}(y)}\left(\frac{1}{n}\right)^{n\sigma_n(y)}\\&\times\prod_{(z,l)\in\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega)}\left[1-\prod_{y\in\Omega}\left(1-\frac{1}{m(n)}\right)^{n\sigma(z)}\right]^{-na_n\nu_n(z,l)}\end{aligned}$$

Note, for each $m\in\mathbb{N}$, the modified Stirling's Formula is

$$m^me^{-m}\le m!\le(2\pi m)^{-1/2}m^me^{-m+1/(12m)}.$$

Applying the above Formula we obtain

$$\begin{aligned}&-n\sum_{(z,l)\in\mathbb{N}\times\mathcal{M}(\Omega)}a_n\nu_n(z,l)\log a_n\nu_n(z,l)-\sum_{(z,l)\in\mathbb{N}\times\mathcal{M}(\Omega)}\frac{1}{12na_n\nu_n(z,l)}\\&-n\sum_{(z,l)\in\mathbb{N}\times\mathcal{M}(\Omega)}a_n\nu_n(z,l)\log(1-e^{-1/a_n})-o(1)\\&\le\log\binom{n}{a_n\nu_n(z,l),(z,l)\in\mathbb{N}\times\mathcal{M}(\Omega)}\prod_{(z,l)\in\mathbb{N}\times\mathcal{M}(\Omega)}\left[1-\prod_{x\in\Omega}\left(1-\frac{1}{m(n)}\right)^{n\sigma(z)}\right]^{-na_n\nu_n(z,l)}\end{aligned}\tag{3.1}$$

$$\leq -n\sum_{(z,l)\in\mathbb{N}\times\mathcal{M}(\Omega)} a_n\nu_n(z,l)\log a_n\nu_n(z,l)+\frac{1}{12n}+\frac{1}{2}\sum_{(z,l)\in\mathbb{N}\times\mathcal{M}(\Omega)}\log 2\pi n a_n\nu_n(z,l)$$
$$-\sum_{(z,l)\in\mathcal{M}(\Omega)}\frac{1}{12na_n\nu_n(z,l)}-n\sum_{(z,l)\in\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega)} a_n\nu_n(z,l)\log(1-e^{-1/a_n})+o(1)$$

Furthermore,

$$\sum_{x\in\Omega} n\sigma_n(y)\log\sigma_n(y)-n\sum_{x\in\Omega}\sigma_n(y)-na_n\sum_{m\in\mathbb{N}}\sum_{l\mathcal{M}(\Omega)}\log l(x)!\nu_n(z,l)$$
$$\leq \log\prod_{x\in\Omega}\binom{n\sigma_n(y)}{m_1l_1(y),m_2l_2(x),\ldots,m_{m(n)}l_{m(n)}(x)}\left(\frac{1}{n}\right)^{n\sigma_n(y)} \tag{3.2}$$
$$\leq \sum_{x\in\Omega} n\sigma_n(y)\log\sigma_n(y)-n\sum_{x\in\Omega}\sigma_n(z)-na_n\sum_{m\in\mathbb{N}}\sum_{l\in\mathcal{M}(\Omega)}\log l(y)!\nu_n(z,l)+na_n\sum_{x\in\Omega}\frac{1}{n\sigma(z)}.$$

We combine 3.1, 3.2 and take

$$\beta_1(n):=-\sum_{(z,l)\in\mathbb{N}\times\mathcal{M}(\Omega)}\frac{1}{12na_n\nu_n(z,l)}+o(1)\quad\text{and}\quad \beta_2(n):=\sum_{x\in\Omega}\frac{1}{n\sigma(z)}+o(1).$$

This gives

$$-na_nH(\nu_n|\varpi_{a_n}^{\sigma_n})+n\beta_1(n)\leq\log\mathbb{P}_n\{M_2^{Y,N}=\nu_n\mid M_1^Y=\sigma_n\}$$
$$\leq -na_nH(\nu_n|\varpi_{a_n}^{\sigma_n})+n\beta_2(n),$$

where

$$\varpi_a^{\sigma}(z,l)=\frac{1}{(1-e^{-1/a})}\left[\frac{[\sigma(0)/a]^{[zl(0)]}e^{-[\sigma(0)/a]}}{[zl(0)]!}\right]\left[\frac{[\sigma(1)/a]^{[zl(1)]}e^{-[\sigma(1)/a]}}{[zl(1)]!}\right],$$
$$z\in\mathbb{N}\cup\{0\}\text{ and }l\in\mathcal{M}(\Omega) \tag{3.3}$$

This concludes the proof.

Lemma 3.3: *Let* $(\sigma_n,\nu_n)\to(\sigma,\nu)\in\mathcal{P}(\Omega)\times\mathcal{P}(\mathbb{N}\times\mathcal{M}(\Omega))$. *Then, we have*

$$\lim_{n\to\infty}|H(\nu_n|\varpi_{a_n}^{\sigma_n})-H(\nu|\varpi_{a_n}^{\sigma})|=0.$$

Proof: Using the triangle inequality we obtain

$$|H(\nu_n|\varpi_{a_n}^{\sigma_n})-H(\nu|\varpi_a^{\sigma})|\leq|H(\nu_n|\varpi_{a_n}^{\sigma_n})-H(\nu_n|\varpi_a^{\sigma})|+|H(\nu_n|\varpi_a^{\sigma})-H(\nu|\varpi_a^{\sigma})|$$

Note here that, $|H(\nu_n|\varpi_a^{\sigma})-H(\nu|\varpi_a^{\sigma})|=0$ using the continuity of relative entropy and also note

$$|H(\nu_n \,|\, \varpi_{a_n}^{\sigma_n)}) - H(\nu_n \,|\, \varpi_a^{\sigma})|$$
$$= \sum_{y\in\Omega}\sigma_n(y)\log\sigma_n(y) - \sum_{y\in\Omega}\sigma(y)\log\sigma(y) - \sum_{y\in\Omega}\sigma_n(y) + \sum_{x\in\Omega}\sigma(z)$$
$$+\log(1-e^{-1/a_n}) + a_n \sum_{(z,l)\in\mathbb{N}\times\mathcal{M}(\Omega)}\sum_{z\in\Omega}\log ml(z)![\nu_n(z,l)-\nu(z,l)] - \log(1-e^{-1/a})$$

$$|H(\nu_n \,|\, \varpi_{a_n}^{\sigma_n}) - H(\nu_n \,|\, \varpi_a^{\sigma})|$$
$$\le \sum_{y\in\Omega}\sigma_n(y)\log\sigma_n(y) - \sum_{y\in\Omega}\sigma(y)\log\sigma(y)$$
$$+\log(1-e^{-1/a_n}) + \frac{1}{na_n}\log n![m(n)-m(n))] - \log(1-e^{-1/a})$$

We take limit as $n\to\infty$ to obtain $|H(\nu_n \,|\, \varpi_{a_n}^{\sigma_n)}) - H(\nu \,|\, \varpi_a^{\sigma})| \to 0$, which ends the proof.

Using similar Integer partitions arguments as in [5], the method of types, see [6, Proof of Theorem 1.1.10] and Lemma 3.3 above, we can infer that, empirical group measure conditionally on the empirical measure obeys an LDP on the scale n and rate $aH(\nu \,|\, \varpi_a^{\sigma})$. This ends the proof of Theorem.

4. Proof of main Result

Let $n\in\mathbb{N}$ and define

$$\mathcal{P}_n(\Omega) := \{\sigma_n \in \mathcal{P}(\Omega) : n\sigma_n(t)\in\mathbb{N}, \text{for all}, t\in\Omega\},$$
$$\mathcal{P}_n(\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega)) := \{na_n\nu_n \in \mathcal{P}(\mathbb{N}\cup\{0\}\times\mathcal{M}(\Omega)) : na_n\nu_n(s,t)\in\mathbb{N},$$
$$\text{for all } s,t\in\Omega\}.$$

Write $\Theta_n := \mathcal{P}_n(\Omega)$ and $\Theta := \mathcal{P}(\Omega)$ and

$$P_n(\sigma_n) := \mathbb{P}\{M_1^Y = \sigma_n\},$$

$$\mathbb{P}_{\sigma_n}(\nu_n) := \mathbb{P}\{M_2^{Y,N} = \nu_n \,|\, M_1^Y = \sigma_n\}$$

the distribution of $(M_1^Y, M_2^{Y,N})$ is given by the mixtures:

$$d\tilde{\mathbb{P}}_n(\sigma_n,\nu_n) := d\mathbb{P}_{\sigma_n}(\nu_n)\, d\mathbb{P}_n(\sigma_n). \tag{4.1}$$

The validity of LDPs for $\tilde{\mathbb{P}}_n$ and the goodness of its rate function depends on LDP for $\mathbb{P}_{\sigma_n}$ and $\mathbb{P}_n$. See, Biggins [1, Theorem 5(b)]. As group of laws $(\mathbb{P}_n : n \in \mathbb{N})$ are exponentially tight on Θ, the following three lemmas will ensure validity of $\tilde{\mathbb{P}}_n$.

Lemma 4.1: $(\tilde{\mathbb{P}}_n : n \in \mathbb{N})$ *is exponentially tight family on* $\Theta \times \mathcal{P}(\Omega \times \mathcal{M}(\Omega))$.

Proof: The proof of this Lemma is as a results of the fact that $\tilde{\mathbb{P}}_n$ is mixture combination of two exponentially tight families $(\mathbb{P}_{\sigma_n}, \ \mathbb{P}_n)$, See, example [5], [7, Lemma 4.3], the Prokhov's Theorem and a concentration inequality.

Define $I : \Theta \times \mathcal{P}(\Omega \times \mathcal{M}(\Omega)) \to [0, \infty]$, by

$$J_\sigma^a(\nu) = \begin{cases} H(\nu \| \varpi_a^\sigma), & \text{if} \quad \langle \nu \rangle = \sigma / a. \\ \infty & \text{otherwise.} \end{cases} \tag{4.2}$$

and observe from Theorem 2.4 we have

$$I^a(\nu) = H(\sigma / a \| q). \tag{4.3}$$

Lemma 4.2: J_σ^a *and* J_a *are both lower semi-continuous.*

Proof: Note, as J_σ^a and J_a are functions of entropy functions, we may infer that both J_σ^a and J_a are lower semi-continuous.

Now, using [1, Theorem 5(b)], Lemmas and and the LDPs; Theorem 2.3 and Theorem 2.4, we obtain Theorem 2.1. which ends the proof of the main Theorem.

5. Proof of the Corollary and Conclusion

5.1 *Proof of Corollary 2.2.*

We obtain this Corollary from the main Theorem by the application of the contraction Principle, see example [6], to the continuous map $(\sigma, \nu) \to \sigma(1)$. Note from Theorem 2.1, $(M_1^X, M_2^{Y,N})$ satisfies an LDP on the scale n and with good rate function $J_\sigma(\sigma, \nu)$. By contraction principle, we can infer from that $I = M_1^Y(1)$ satisfies an LDP on the scale n and with good rate function

$$\varpi_\sigma(y) = \inf\{J(\sigma, \nu) : \sigma(1) = y, 1 - \nu_2((0,1)) = \sigma\}, \tag{5.1}$$

where

$$\nu_2((0,1)) = \sum_{z=1}^{\infty} \varpi_a^\sigma(z, (0,1)) = \frac{(1 - e^{-(1-\sigma(1))/a})}{(1 - e^{-1/a})} e^{-\sigma(1)/a}.$$

Solving the problem in 5.1 by optimization techniques, we obtain the good rate function in 2.3 which conclude the proof. We note that here $0 \leq \sigma \leq 1$.

5.2 *Conclusion*

We have found a bound on the number of infected persons in a bernoulli sequence via the number of infected group individuals in this article. The asymptotic relationship in Corollary 2.2 may be interpreted as folows:

- $\sigma = 0$: The absence of infected group samples and hence there is no likelihood of infected persons in sample Y.
- $\sigma = 1$: Nearly all pool samples formed from Y are infected, suggesting that there is a positive chance for every individual in sample Y to be infected
- $0 < \sigma < 1$: There exist infected pool samples formed from Y, implying that there is a positive probability for certain individuals in sample Y to be infected

In summary, it can be deduced that there is a positive probability that nI individuals within sample Y are infected. In particular, the article has indirectly shown that $nI \geq na\sigma$, with probability 1.

This article has already been published as a pre-print in the arxiv database. See, [4].

References

[1] Biggins, J. D. Large deviations for mixtures In Electron. Comm. Probab.9 60-71 (2004).

[2] Bhatnagar, V., Poonia, R. C., Nagar, P., Kumar, S. ,Singh, V. , Raja, L. and Dass, P. Descriptive analysis of COVID-19 patients in the context of India. *Journal of Interdisciplinary Mathematics* (2020), DOI: 10.1080/09720502.2020.1761635.

[3] Doku-Amponsah, K. Large Deviations and Basic Information Theory for Hierarchical and Networked Data Structures Ph.D Thesis, Bath (2006).

[4] Doku-Amponsah, K. (2006). Large deviations, asymptotic bounds on the number of positive individuals in a Bernoulli sample via the

number of positive pool samples drawn on the bernoulli sample https://doi.org/10.48550/arXiv.2008.02151.

[5] Doku-Amponsah, K. and Moeters. P. Large deviation principle for empirical measures of coloured random graphs. Ann. Appl. Prob. 20(6), 1089-2021 (2010).

[6] Dembo, A. and Zeitouni, O. Large deviations Techniques and Applications. Springer (1998).

[7] Sakyi-Yeboah, E., Asiedu, L. and Doku-Amponsah, K. Local large deviation principle, large deviation principle and information theory for the signal-to-interference-plus-noise ratio graph models. *Journal of Information and Optimization Science*, 42(1), 249-273 (2020).

[8] Singh, V., Poonia, R. C., Kumar, S., Dass, P., Agarwal, P., Bhatnagar, V. and Raja, L. Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine. *Journal of Discrete Mathematical Sciences and Cryptography* (2020), DOI: 10.1080/09720529.2020.1784535.

Received August, 2020