ASYMPTOTIC NORMALITY OF ENTROPY ESTIMATORS

by

Xing Zhang

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2013

Approved by:

_____

Dr. Zhiyi Zhang

_____

Dr. Yanqing Sun

_____

Dr. Jiancheng Jiang

_____

Dr. Yuliang Zheng

ABSTRACT

XING ZHANG. Asymptotic normality of entropy estimators.
(Under the direction of DR. ZHIYI ZHANG)

Shannon's entropy plays a central role in many fields of mathematics. In the first chapter, we present a sufficient condition for the asymptotic normality of the plug-in estimator of Shannon's entropy defined on a countable alphabet. The sufficient condition covers a range of cases with countably infinite alphabets, for which no normality results were previously known.

In the second chapter of this dissertation, we establish the asymptotic normality of a recently introduced non-parametric entropy estimator under another sufficient condition. The proposed estimator, developed in Turing's perspective, is known for its improved estimation accuracy.

ACKNOWLEDGMENTS

First and foremost, I want to thank my advisor Zhiyi Zhang. I thank him for inspiring and motivating me to do creative and original research. His intelligence, humor, patience, and flexibility make each discussion enlightening and enjoyable. He helped me overcome the major difficulties of this dissertation so I could finish it. He has been very understanding and supportive of my career goals. Without his encouragement and guidance, I would not have improved my English a lot and have found my current job which fits me well. I'm very grateful and thankful for everything he has done for me.

I sincerely thank other members of my doctoral committee, Dr. Yanqing Sun, Dr. Jiancheng Jiang and Dr. Yuliang Zheng for their time and insightful questions. I also thank the Department of Mathematics & Statistics for offering me the opportunity to study here. Without their academic, financial and technical support, I would not be able to afford to study here and receive my degree.

Meanwhile, I want to express my deepest gratitude to my parents and my sister. Their everlasting love, encouragement, and unwavering faith in me give me the strength and confidence that make me the best of myself.

Last, but by no means least, I sincerely thank my friends in China, America and elsewhere for their friendship and support which make my life fun and enjoyable.

TABLE OF CONTENTS

CHAPTER 1:   A NORMAL LAW FOR THE PLUG-IN ESTIMATOR

1.1   Introduction

Let $\{p_k\}$ be a probability distribution on an alphabet $\mathscr{X} = \{\ell_k; 1 \leq k \leq K\}$, where K denotes either a finite integer or ∞. Let P_X be a random variable such that $P(P_X = p_k) = p_k$. Entropy in the form of

$$H = E(-\ln P_X) = -\sum_k p_k \ln p_k$$

was introduced by Shannon (1948) and is often referred to as Shannon's Entropy. The estimations of entropy-like quantities have become growingly important for their wide applications in the fields of neural science and information theory, etc.

Let $X_1, \cdots, X_n$ be an *iid* sample from $\mathscr{X}$ according to the probability distribution $\{p_k\}$, and $\{y_{k,n} = \sum_{i=1}^n 1[X_i = \ell_k]\}$ be the sequence of observed counts of letters, and $\{\hat{p}_{k,n} = y_{k,n}/n\}$. The plug-in estimator for H, given by

$$\hat{H}_n = -\sum_k \hat{p}_{k,n} \ln \hat{p}_{k,n} \tag{1.1}$$

plays a central role in the literature. $\hat{H}_n$ is simple and intuitive; and it often serves as a reference estimator for other estimators, many of which were derived based on $\hat{H}_n$.

When K is fixed and finite,

$$\sqrt{n}(\hat{H}_n - H) \xrightarrow{D} N(0, \sigma^2)$$

where $\sigma^2 = Var(-\ln P_X) > 0$ has long been known. See Miller and Madow (1954) and Basharin (1959). In this case, it is also known that

$$E(\hat{H}_n - H) = -\frac{K-1}{2n} + \frac{1}{12n^2}\left(1 - \sum_{k=1}^K \frac{1}{p_k}\right) + O(n^{-3}). \qquad (1.2)$$

$$Var(\hat{H}_n) = \frac{1}{n}\left(\sum_{k=1}^K p_k \ln^2 p_k - H^2\right) + \frac{K-1}{2n^2} + O(n^{-3}). \qquad (1.3)$$

See Miller (1955), Basharin (1959) and Harris (1975).

When $K = K(n)$ is assumed to dynamically vary as the sample size n increases, i.e., $\{p_{k,n}; k = 1, 2, \cdots, K(n)\}$, Paninski (2003) established a normal law for $\hat{H}_n$, stated as Lemma 1.1 below.

When K is infinite, Antos and Kontoyiannis (2001) obtained different rates of convergence for $\hat{H}_n$ under a variety of tail conditions on $\{p_k\}$. However, no results regarding the asymptotic normality of $\hat{H}_n$ were known. We seek to lay down a pebble in that blank space by presenting a sufficient normality condition for $\hat{H}_n$ when the cardinality of $\mathscr{X}$ is countably infinite. More specifically, the sufficient condition is satisfied by distributions with tails decaying at the rate of $[k\ln(\ln k)]^{-2}(\ln k)^{-1}$, but not by those with tails decaying at the rate of $k^{-2}(\ln k)^{-1}$.

1.2 Main Results

Theorem 1.1. For any non-uniform distribution $\{p_k; k \geq 1\}$ satisfying $E(lnP_X)^2 < \infty$, if there exists an integer valued function $K(n)$ such that, $K(n) \to \infty$, $K(n) = o(\sqrt{n})$ and $\sqrt{n}\sum_{k \geq K(n)} p_k lnp_k \to 0$, as $n \to \infty$, then

$$\sqrt{n}(\hat{H}_n - H)/\sigma \xrightarrow{D} N(0, 1)$$

where $\sigma^2 = Var(-\ln P_X)$.

A proof of Theorem 1.1 requires Lemmas 1.1 and 1.2 below. Lemma 1.1 is due to Paninski (2003).

Lemma 1.1. Let $\{p_{k,n}; k = 1, \cdots, K(n)\}$ be a probability distribution, P_X be a random variable such that $P(P_X = p_{k,n}) = p_{k,n}$, and

$$\tau_n^2 \;=\; Var(-\ln P_X) = \sum_{k=1}^{K(n)} p_{k,n} \ln^2 p_{k,n} - \left(\sum_{k=1}^{K(n)} p_{k,n} \ln p_{k,n}\right)^2.$$

If $K(n) = o(\sqrt{n})$ and $\liminf_{n\to\infty} n^{1-\alpha}\tau_n^2 > 0$ for some $\alpha > 0$, then

$$\sqrt{n}(\hat{H}_n - H)/\tau_n \xrightarrow{D} N(0,1).$$

Lemma 1.2. For a probability distribution $\{p_k; k \geq 1\}$, if there exists an integer valued function $K(n)$ such that as $n \to \infty$, $K(n) \to \infty$, and $\sqrt{n}\sum_{k\geq K(n)} p_k \ln p_k \to 0$, then

$$\sqrt{n}\ln n \sum_{k\geq K(n)} p_k \to 0$$

.

Proof. Let $p_n^* = \sum_{k\geq K(n)} p_k$. Since $1 < -\ln p_n^*$ for a sufficiently large n,

$$0 \leq \sqrt{n}p_n^* \leq -\sqrt{n}p_n^* \ln p_n^* = -\sqrt{n}\sum_{k\geq K(n)} p_k \ln p_n^* \leq -\sqrt{n}\sum_{k\geq K(n)} p_k \ln p_k \to 0.$$

$$(1.4)$$

$\sqrt{n}p_n^* \to 0$ implies $p_n^* = \alpha_n n^{-1/2}$ where $\alpha_n = o(1)$. On the other hand, since $\alpha_n \ln \alpha_n \to 0$, $-\sqrt{n}p_n^* \ln p_n^* = \alpha_n(\ln\sqrt{n} - \ln\alpha_n) \to 0$ implies $\alpha_n = \beta_n/\ln\sqrt{n}$ where $\beta_n = o(1)$. Hence $\sqrt{n}\ln n \sum_{k\geq K(n)} p_k = 2\beta_n \to 0$. $\square$

Proof of Theorem 1.1.

Consider a modified probability distribution $\{p_{k,n}; k = 1, \cdots, K(n)\}$ based on $\{p_k\}$ as follows. Let

$$p_{k,n} = \begin{cases} p_k, & \text{for } 1 \le k \le K(n) - 1 \\ \sum_{k \ge K(n)} p_k \equiv p_n^*, & \text{for } k = K(n). \end{cases}$$

Since $E(\ln P_X)^2 = \sum_k p_k \ln^2 p_k < \infty$ implies $H = -\sum_k p_k \ln p_k < \infty$, we have

$$0 \le p_n^* \ln^2 p_n^* = \sum_{k \ge K(n)} p_k \ln^2 p_n^* \le \sum_{k \ge K(n)} p_k \ln^2 p_k \to 0,$$

and

$$0 \le -p_n^* \ln p_n^* = \sum_{k \ge K(n)} (-p_k \ln p_n^*) \le \sum_{k \ge K(n)} (-p_k \ln p_k) \to 0. \tag{1.5}$$

Let $\tau_n^2 = Var(-\ln P_X)$ under the modified distribution $\{p_{k,n}\}$. After a few algebraic steps,

$$\sigma^2 - \tau_n^2 = \left(\sum_{k \ge K(n)} p_k \ln^2 p_k - p_n^* \ln^2 p_n^*\right) - \left(-\sum_{k \ge K(n)} p_k \ln p_k + p_n^* \ln p_n^*\right) \tag{1.6}$$
$$\times \left(-\sum_{k \ge K(n)} p_k \ln p_k - p_n^* \ln p_n^* - 2 \sum_{k=1}^{K(n)-1} p_k \ln p_k\right).$$

It is clear that the first term in (1.6) converges to zero, that the first factor of the second term converges to zero, and that the second factor of the second term converges to $2H < \infty$. Therefore $\tau_n \to \sigma$, and hence by Lemma 1.1,

$$\sqrt{n} \sum_{k=1}^{K(n)} (-\hat{p}_{k,n} \ln \hat{p}_{k,n} + p_{k,n} \ln p_{k,n}) \xrightarrow{D} N(0, \sigma^2). \tag{1.7}$$

However,

$$\sqrt{n}(\hat{H}_n - H) - \sqrt{n} \sum_{k=1}^{K(n)} (-\hat{p}_{k,n} \ln \hat{p}_{k,n} + p_{k,n} \ln p_{k,n})$$
$$= \sqrt{n} \sum_{k \ge K(n)} (-\hat{p}_{k,n} \ln \hat{p}_{k,n}) - \sqrt{n} \sum_{k \ge K(n)} (-p_k \ln p_k) + \sqrt{n} \hat{p}_n^* \ln \hat{p}_n^* - \sqrt{n} \ p_n^* \ln p_n^* \tag{1.8}$$

where $\hat{p}_n^* = \sum_{k \ge K(n)} y_{k,n}/n$. The proof is complete if it is shown that the right hand

side of (1.8) is $o_p(1)$. Toward that end, it is to show that each of the four terms in the last expression of (1.8) is $o_p(1)$.

The second term converges to zero in probability by the condition of Theorem 1.1. The fact that the fourth term converges to zero is established in the proof of Lemma 1.2. For the first and third terms, we first observe $-\hat{p}_{k,n} \ln \hat{p}_{k,n} \leq \hat{p}_{k,n} \ln n$ and $-\hat{p}_n^* \ln \hat{p}_n^* \leq \hat{p}_n^* \ln n$, and then observe the following two inequalities

$$0 \leq \sqrt{n} \sum_{k \geq K(n)} (-\hat{p}_{k,n} \ln \hat{p}_{k,n}) \leq \sqrt{n}(\ln n)\hat{p}_n^* \tag{1.9}$$

and

$$0 \leq -\sqrt{n}\hat{p}_n^* \ln \hat{p}_n^* \leq \sqrt{n}(\ln n)\hat{p}_n^*. \tag{1.10}$$

Since, by Lemma 1.2,

$$E[\sqrt{n}(\ln n)\hat{p}_n^*] = \sqrt{n}(\ln n)p_n^* \to 0 \tag{1.11}$$

and, noting $\sqrt{n}(\ln n)\hat{p}_n^* \geq 0$, $\sqrt{n}(\ln n)\hat{p}_n^* = o_p(1)$. By (1.9) and (1.10), both the first and the third terms converge to zero in probability. The theorem follows by Slutsky's lemma. $\qquad\square$

Let $\hat{\sigma}_n^2 = \sum_k \hat{p}_{k,n} \ln^2 \hat{p}_{k,n} - (\sum_k -\hat{p}_{k,n} \ln \hat{p}_{k,n})^2$.

Corollary 1.1. Under the condition of Theorem 1.1,

$$\sqrt{n}(\hat{H}_n - H)/\hat{\sigma}_n \xrightarrow{D} N(0,1).$$

Corollary 1.1 provides a means of large sample inference on H. A proof of Corollary 1.1 requires the following lemma due to Devroye (1991).

Lemma 1.3. Let $X_1, \cdots, X_n$ be independent random variables on $\mathscr{X}$, and assume

that $\hat{F}_n \colon \mathscr{X}^n \to \mathscr{R}$ satisfies, for $1 \le i \le n$,

$$sup_{x_1,\cdots,x_n,x_i' \in \mathscr{X}} |\hat{F}_n(x_1,\cdots,x_n) - \hat{F}_n(x_1,\cdots,x_{i-1},x_i',x_{i+1},\cdots,x_n)| \le c_i.$$

Then $Var\{\hat{F}_n(X_1,\cdots,X_n)\} \le \frac{1}{4}\sum_{i=1}^n c_i^2$.

Proof of Corollary 1.1. Let

$$\hat{F}_n \equiv \hat{F}_n(X_1,\cdots,X_n) = \sum_k f(\hat{p}_{k,n}) = \sum_k \hat{p}_{k,n} \ln^2 \hat{p}_{k,n}.$$

We first want to show $\lim_{n\to\infty} E(\hat{F}_n - F)^2 = 0$ for $F \equiv \sum_k p_k \ln^2 p_k < \infty$.

For all integers $0 \le i < n$ and $n \ge 21 > e^3$, $\left|\frac{i+1}{n}\ln^2(\frac{i+1}{n}) - \frac{i}{n}\ln^2(\frac{i}{n})\right| \le \frac{\ln^2 n}{n}$. Therefore

$$sup_{x_1,\cdots,x_n,x_i' \in \mathscr{X}} \left|\hat{F}_n(x_1,\cdots,x_n) - \hat{F}_n(x_1,\cdots,x_{i-1},x_i',x_{i+1},\cdots,x_n)\right| \le \frac{2\ln^2 n}{n}.$$

By Lemma 1.3, $Var(\hat{F}_n) \le \frac{\ln^4 n}{n} \to 0$. For each k, $\hat{p}_{k,n} \overset{as}{\to} p_k$, $f(\hat{p}_{k,n}) \overset{as}{\to} f(p_k)$, $f(\hat{p}_{k,n}) \le e^{-2}\ln^2 e^{-2} = 4e^{-2}$, so $Ef(\hat{p}_{k,n}) \to f(p_k)$.

Since $0 \le f(\hat{p}_{k,n}) \le 4e^{-2}$, by Fatou's Lemma,

$$\limsup_{n\to\infty} \sum_k Ef(\hat{p}_{k,n}) \le \sum_k \limsup_{n\to\infty} Ef(\hat{p}_{k,n}) = \sum_k f(p_k) \text{ and}$$

$$\liminf_{n\to\infty} \sum_k Ef(\hat{p}_{k,n}) \ge \sum_k \liminf_{n\to\infty} Ef(\hat{p}_{k,n}) = \sum_k f(p_k); \text{ and therefore}$$

$$\lim_{n\to\infty} E\hat{F}_n = \lim_{n\to\infty} \sum_k Ef(\hat{p}_{k,n}) = \sum_k f(p_k) = F.$$

By Theorem 1.1, $\hat{H}_n^2 \overset{p}{\to} H^2$, and therefore $\hat{\sigma}_n^2 \overset{p}{\to} \sigma^2$. Finally the corollary follows by Theorem 1.1 and Slusky's lemma. $\square$

Corollary 1.2. If a probability distribution $\{p_k\}$ satisfying $p_k = C_\lambda k^{-\lambda}$ where $\lambda > 1$,

and $\sqrt{n}(\hat{H}_n - H) \xrightarrow{D} N(0, \sigma^2)$, as $n \to \infty$, then we have $\lambda \geq 2$

Proof. Antos and Kontoyiannis (2001) proved that

$$E[(\hat{H}_n - H)] \sim n^{-(\lambda-1)/\lambda} \quad \text{and} \quad Var(\hat{H}_n) \leq O(\frac{\ln^2(n)}{n}).$$

Assume to the contrary that $\lambda < 2$, then there exists a sequence $\{a_n\}$ converging to zero (for example, $a_n = -\ln^{-\alpha}(n)$ with $\alpha > 1$) such that $E[a_n\sqrt{n}(\hat{H}_n - H)] \to +\infty$, $Var(a_n\sqrt{n}\hat{H}_n) \to 0$ and hence $a_n\sqrt{n}(\hat{H}_n - E\hat{H}_n) \xrightarrow{p} 0$,

It leads to

$$a_n\sqrt{n}(\hat{H}_n - H) = a_n\sqrt{n}(\hat{H}_n - E\hat{H}_n) + a_n\sqrt{n}(E\hat{H}_n - H) \xrightarrow{p} +\infty,$$

which contradicts the assumption that $a_n\sqrt{n}(\hat{H}_n - H) \xrightarrow{p} 0$ implied by $\sqrt{n}(\hat{H}_n - H) \xrightarrow{D} N(0, \sigma^2)$. Therefore, λ must be greater or equal to 2.

$\square$

Example 1.1. If $p_k = C_\lambda k^{-\lambda}$ where $\lambda > 1$, the sufficient condition of Theorem 1.1 holds for $\lambda > 2$ but not for $1 < \lambda \leq 2$.

Note

$$\sqrt{n} \sum_{k \geq K(n)} (-p_k \ln p_k) \sim \sqrt{n} \int_{K(n)}^{\infty} \frac{C_\lambda}{x^\lambda} \ln\left(\frac{x^\lambda}{C_\lambda}\right) dx$$
$$= \frac{C_\lambda \lambda}{\lambda-1} \frac{\sqrt{n} \ln K(n)}{(K(n))^{\lambda-1}} + \left(\frac{C_\lambda \lambda}{(\lambda-1)^2} - \frac{C_\lambda \ln C_\lambda}{\lambda-1}\right) \frac{\sqrt{n}}{(K(n))^{\lambda-1}} \sim \frac{C_\lambda \lambda}{\lambda-1} \frac{\sqrt{n} \ln K(n)}{(K(n))^{\lambda-1}}.$$

If $\lambda > 2$, letting $K(n) \sim n^{1/\lambda}$, $\frac{C_\lambda \lambda}{\lambda-1} \frac{\sqrt{n} \ln K(n)}{(K(n))^{\lambda-1}} = \frac{C_\lambda}{\lambda-1} \frac{\ln n}{n^{1/2-1/\lambda}} \to 0$.

If $1 < \lambda \leq 2$, for any $K(n)$ satisfying $K(n) \sim o(\sqrt{n})$ and a sufficiently large n, $\frac{C_\lambda \lambda}{\lambda-1} \frac{\sqrt{n} \ln K(n)}{(K(n))^{\lambda-1}} \geq \frac{C_\lambda \lambda}{\lambda-1} \frac{\sqrt{n}}{n^{\lambda/2-1/2}} \geq \frac{C_\lambda \lambda}{\lambda-1} \geq 2C_\lambda > 0$.

Example 1.2. If $p_k = C_\lambda e^{-\lambda k}$ for any $\lambda > 0$, then the sufficient condition of Theorem 1.1 holds.

Letting $K(n) \sim \lambda^{-1} \ln n$, for a sufficiently large n,

$$\sqrt{n} \sum_{k \geq K(n)} (-p_k \ln p_k) \sim -\sqrt{n} \int_{\ln n^{1/\lambda}}^{\infty} C_\lambda e^{-\lambda x} \ln(C_\lambda e^{-\lambda x}) dx \sim \frac{C_\lambda}{\lambda} (\ln n) n^{-1/2} \to 0.$$

Example 1.3. If $p_k = C/(k^2 \ln^2 k)$, then the sufficient condition of Theorem 1.1 holds.

Letting $K(n) \sim \sqrt{n}/\ln \ln n$, for a sufficiently large n,

$$\sqrt{n} \sum_{k \geq K(n)} (-p_k \ln p_k) \sim \sqrt{n} C \int_{K(n)}^{\infty} \frac{2 \ln x + 2 \ln \ln x - \ln C}{x^2 \ln^2 x} dx$$
$$\sim 2\sqrt{n} C \int_{K(n)}^{\infty} \frac{1}{x^2 \ln x} dx \leq \frac{2C\sqrt{n}}{K(n) \ln K(n)} \to 0.$$

Example 1.4. If $p_k = \frac{C}{k^2 \ln k \ln^2 (\ln k)}$, the sufficient condition of Theorem 1.1 holds.

First, note that

$$\int_{K(n)}^{\infty} \frac{\ln x - 1}{x^2} dx = \int_{K(n)}^{\infty} \frac{\ln x}{x^2} dx - \int_{K(n)}^{\infty} \frac{1}{x^2} dx = \int_{K(n)}^{\infty} \frac{\ln x}{x^2} dx - \int_{K(n)}^{\infty} \frac{1}{x^2} dx$$
$$= [-\frac{\ln x}{x} |_{K(n)}^{\infty} + \int_{K(n)}^{\infty} \frac{1}{x^2} dx] - \int_{K(n)}^{\infty} \frac{1}{x^2} dx = \frac{\ln K(n)}{K(n)}.$$

Then, for sufficient large n, we have

$$\frac{1}{2C} \sqrt{n} \sum_{k \geq K(n)} (-p_k \ln p_k) = \frac{1}{2} \sqrt{n} \sum_{k \geq K(n)} \frac{\ln(k^2 \ln k \ln^2 (\ln k)) - \ln C}{k^2 \ln k \ln^2 (\ln k)}$$
$$\sim \sqrt{n} \sum_{k \geq K(n)} \frac{1}{k^2 \ln^2 (\ln k)} \sim \sqrt{n} \int_{K(n)}^{\infty} \frac{\ln k d[k \ln(\ln k)] - \ln k \ln(\ln k) dk}{k^2 \ln^2 (\ln k)}$$
$$= \sqrt{n} \int_{K(n)}^{\infty} \frac{\ln k d[k \ln(\ln k)]}{k^2 \ln^2 (\ln k)} - \sqrt{n} \int_{K(n)}^{\infty} \frac{\ln k \ln(\ln k)}{k^2 \ln^2 (\ln k)} dk$$
$$= \sqrt{n} [-\frac{\ln k}{k \ln(\ln k)} |_{K(n)}^{\infty} + \int_{K(n)}^{\infty} \frac{1}{k^2 \ln(\ln k)} dk] - \sqrt{n} \int_{K(n)}^{\infty} \frac{\ln k}{k^2 \ln(\ln k)} dk$$
$$= \sqrt{n} [\frac{\ln K(n)}{K(n) \ln(\ln K(n))} + \int_{K(n)}^{\infty} \frac{1}{k^2 \ln(\ln k)} dk] - \sqrt{n} \int_{K(n)}^{\infty} \frac{\ln k}{k^2 \ln(\ln k)} dk$$
$$= \frac{\sqrt{n} \ln K(n)}{K(n) \ln(\ln K(n))} - \sqrt{n} \int_{K(n)}^{\infty} \frac{\ln k - 1}{k^2 \ln(\ln k)} dk$$
$$\leq \frac{\sqrt{n} \ln K(n)}{K(n) \ln(\ln K(n))} - \sqrt{n} \int_{K(n)}^{\sqrt{n} \ln n} \frac{\ln x - 1}{x^2 \ln(\ln x)} dx$$
$$\leq \frac{\sqrt{n} \ln K(n)}{K(n) \ln(\ln K(n))} - \sqrt{n} \frac{\ln K(n) - 1}{\ln(\ln K(n))} \int_{K(n)}^{\sqrt{n} \ln n} \frac{1}{x^2} dx$$
$$= \frac{\sqrt{n} \ln K(n)}{K(n) \ln(\ln K(n))} - \frac{\sqrt{n} (\ln K(n) - 1)}{K(n) \ln(\ln K(n))} + \frac{\sqrt{n} (\ln K(n) - 1)}{\sqrt{n} \ln n \ln(\ln K(n))}$$
$$= \frac{\sqrt{n}}{K(n) \ln(\ln K(n))} + \frac{\ln K(n) - 1}{\ln n \ln(\ln K(n))}$$

If we choose $K(n) = \frac{\sqrt{n}}{\ln(\ln(\ln n))}$, the above expression approaches 0.

Example 1.5. If $p_k = C/(k^2 \ln k)$, the sufficient condition of Theorem 1.1 does not hold.

For any $K(n)$ satisfying with $K(n) \sim o(\sqrt{n})$, for a sufficiently large n,

$$\sqrt{n} \sum_{k \geq K(n)}(-p_k \ln p_k) \sim \sqrt{n} C \int_{K(n)}^{\infty} \frac{2\ln x + \ln\ln x - \ln C}{x^2 \ln x} dx$$
$$\sim \sqrt{n} C \int_{K(n)}^{\infty} \frac{2}{x^2} dx = \frac{2C\sqrt{n}}{K(n)} \to \infty.$$

1.3 Remarks

Under distributions $p_k = C/k^\lambda$, a necessary condition for $\sqrt{n}(\hat{H}_n - H)$ to hold asymptotic normality is $\lambda \geq 2$, because bias terms $E[(\hat{H}_n - H)]$ has a rate of $n^{-(\lambda-1)/\lambda}$, no faster than $n^{-1/2}$ if $\lambda \in (1, 2]$, *e.g.*, see Theorem 7, Antos and Kontoyiannis (2001) . On the other hand, as shown in Example 4, $\sqrt{n}(\hat{H}_n - H)$ does have asymptotic normality when $p_k = C/(k^2 \ln k \ln^2(\ln k))$. Even though Theorem 1.1 gives only a sufficient condition, the band of distributions which are not covered by the sufficient condition but may still support asymptotic normality of $\hat{H}_n$ must be, if existed, very narrow.

CHAPTER 2:   ASYMPTOTIC NORMALITY OF A NEW ESTIMATOR

2.1   Introduction

The plug-in estimator $\hat{H}_n$ is known for its large bias in an undersampled regime. We can see from (1.2) that when K is finite, the first bias term is hardly negligible for an often unknown large K and a small sample size n. Many have provided various ways to adjust bias terms based on (1.2), for examples, see Treves and Panzeri (1995, 1996), Paninski (2003) and Schürmann (2004); some of these procedures were able to greatly reduce the bias at a moderate expense of an increase in variance.

When K is infinite, Antos and Kontoyiannis (2001) showed that no universal rate of convergence exists for any sequence of estimators, and specifically, $\hat{H}_n$ can approach H at an arbitrarily slow rate. They also obtained different rates of convergence for $\hat{H}_n$ under a variety of tail conditions on $\{p_k\}$.

Other popular estimators include the jackknifed version of the plug-in estimator proposed by Strong *et al.* (1998) , the NSB estimator proposed by Nemenman, *et al.* (2002), and the coverage-adjusted entropy estimator (CAE) proposed by Chao and Chen (2003). The jackknife estimator evaluates entropy through an extrapolation procedure which utilizes the dependence of $\hat{H}_n$ on the sample size. The NSB method counts coincidences in letters and introduces a Bayeisan prior to correct the bias. The CAE estimator recognized the loss of information on the uncovered letters of alphabet, and hence incorporated Turing's formula (proposed by Good (1953) but largely credited to Alan Turing) to adjust the bias. Vu, Yu & Kass (2007) later proved several convergence properties of CAE, but the revealing convergence rate were quite discouraging. All these estimators are all claimed to remove bias effectively

in simulation study, but there seems to be lack of rigorous analysis of their rates of convergence. Also, see Paninski (2003) and Panzeri, Senatore, Montemurro & Petersen (2007) for a comprehensive review and comparison of various estimators.

Zhang (2012) proposed a non-parametric estimator of Shannon's entropy on a countable alphabet $\mathscr{X}$.

$$\hat{H}_z = \sum_{v=1}^{n-1} \frac{1}{v} \left\{ \frac{n^{v+1}[n-(v+1)]!}{n!} \sum_k \left[\hat{p}_{k,n} \prod_{j=0}^{v-1} \left(1 - \hat{p}_{k,n} - \frac{j}{n} \right) \right] \right\} \tag{2.1}$$

This new estimator, constructed in Turing's perspective, is fundamentally different than previous ones. Through Turing's formula, it recovers some distributional characteristics on the uncovered subset of $\mathscr{X}$, and thus significantly improves the estimation accuracy; it is worth mentioning that it has a bias decaying at a rate of $O[(1-p_0)^n/n]$ where $p_0 = \min\{p_k > 0; 1 \leq k \leq K\}$ on a finite alphabet where K is the cardinality. Also, because $\hat{H}_z$ is a weighted sum of U-statistic, $\hat{H}_z$ are more analytically tractable and its rates of convergence can be obtained under a wide range of distribution subclasses, see Zhang (2012). Simulation results also show that $\hat{H}_z$ and its bias-adjusted versions, are quite competitive among existing estimators.

Zhang (2013) established the asymptotic normality of $\sqrt{n}(\hat{H}_z - H)$ on any finite alphabet. This paper extends the normality results of Zhang (2013) to include certain cases of alphabets with countably infinite cardinality, as stated in the following theorem and corollary.

2.2 Main Results

Let $F = E[-\ln(p_X)]^2 = \sum_k p_k \ln^2(p_k)$ and

$$\hat{F}_z = \sum_{v=1}^{n-1} \left\{ \left(\sum_{i=1}^{v-1} \frac{1}{i(v-i)} \right) \left\{ \frac{n^{v+1}[n-(v+1)]!}{n!} \sum_k \left[\hat{p}_{k,n} \prod_{j=0}^{v-1} \left(1 - \hat{p}_{k,n} - \frac{j}{n} \right) \right] \right\} \right\}. \tag{2.2}$$

Theorem 2.1. For a non-uniform distribution $\{p_k; k \geq 1\}$ satisfying $E(lnP_X)^2 < \infty$, if there exists an integer valued function $K(n)$ such that, $K(n) \to \infty$, $K(n) = o(\sqrt{n}/\ln n)$ and $\sqrt{n} \sum_{k \geq K(n)} p_k lnp_k \to 0$, as $n \to \infty$. Then for $\hat{H}_z$ as in (2.1), it has

$$\sqrt{n}\left(\hat{H}_z - H\right) \xrightarrow{D} N(0, \sigma^2)$$

where $\sigma^2 = Var\left[-\ln(p_X)\right] = F - H^2$.

Corollary 2.1. Let $\{p_k; k \geq 1\}$ be a probability distribution on an alphabet satisfying the condition of Theorem 2.1, $\hat{H}_z$ be as in (2.1), and $\hat{F}_z$ be as in (2.2). Then

$$\sqrt{n}\left(\frac{\hat{H}_z - H}{\sqrt{\hat{F}_z - \hat{H}_z^2}}\right) \xrightarrow{D} N(0, 1).$$

Remark 2.1. The condition of Theorem 2.1 is slightly stronger than that of Theorem 1.1 therefore, there will be less probability distributions satisfying the condition of Theorem 2.1 than that of Theorem 1.1. We can show that, the sufficient condition of Theorem 2.1 still holds for $p_k = C_\lambda k^{-\lambda}$ where $\lambda > 2$, but not for $p_k = C/(k^2 \ln^2 k)$.

Example 2.1. If $p_k = C/(k^2 \ln^2 k)$, then the sufficient condition of Theorem 2.1 doesn't hold.

For any $K(n) = o(\sqrt{n}/\ln n)$, and a sufficiently large n,

$$\sqrt{n} \sum_{k \geq K(n)} (-p_k \ln p_k) \sim \sqrt{n} C \int_{K(n)}^{\infty} \frac{2\ln x + 2\ln\ln x - \ln C}{x^2 \ln^2 x} dx$$

$$\sim 2\sqrt{n} C \int_{K(n)}^{\infty} \frac{1}{x^2 \ln x} dx \geq 2\sqrt{n} C \int_{K(n)}^{n} \frac{1}{x^2 \ln x} dx$$

$$\geq \frac{2\sqrt{n} C}{\ln n} \int_{K(n)}^{n} \frac{1}{x^2} dx = \frac{2\sqrt{n} C}{\ln n}\left(\frac{1}{K(n)} - \frac{1}{n}\right) \geq \frac{2\sqrt{n} C}{\ln n} \frac{1}{K(n)} \to \infty.$$

To prove Theorem 2.1 and its Corollary, we define

$$\begin{aligned}
\zeta_{1,v} &= \sum_k p_k (1-p_k)^v, \; C_v = \sum_{i=1}^{v-1} \frac{1}{i(v-i)}, \\
Z_{1,v} &= \frac{n^{1+v}[n-(1+v)]!}{n!} \sum_k \left[\hat{p}_{k,n} \prod_{j=0}^{v-1} \left(1 - \hat{p}_{k,n} - \frac{j}{n}\right) \right],
\end{aligned}$$

and we have,

$$H = \sum_{v=1}^{\infty} \frac{1}{v} \zeta_{1,v}, \; \hat{H}_z = \sum_{v=1}^{n-1} \frac{1}{v} Z_{1,v}, F = \sum_{v=1}^{\infty} C_v \zeta_{1,v} \text{ and } \hat{F}_z = \sum_{v=1}^{n-1} C_v Z_{1,v}.$$

Note that

$$Z_{1,v} = \frac{n^{v+1}[n-(v+1)]!}{n!} \sum_k \left[\hat{p}_{k,n} \prod_{j=0}^{v-1} \left(1 - \hat{p}_{k,n} - \frac{j}{n}\right) \right]$$

$$= \sum_k \frac{n^{v+1}[n-(v+1)]!}{n!} \hat{p}_{k,n} \prod_{j=0}^{v-1} \frac{n - y_{k,n} - j}{n} = \sum_k \hat{p}_{k,n} \prod_{j=0}^{v-1} \frac{n - y_{k,n} - j}{n-j-1}$$

$$= \sum_k \hat{p}_{k,n} \prod_{j=1}^{v} \left(1 - \frac{y_{k,n} - 1}{n-j}\right),$$

and therefore,

$$\hat{H}_z = \sum_{v=1}^{n-1} \frac{1}{v} Z_{1,v} = \sum_{v=1}^{n-1} \frac{1}{v} \sum_k \hat{p}_{k,n} \prod_{j=1}^{v} \left(1 - \frac{y_{k,n} - 1}{n-j}\right)$$

$$= \sum_k \hat{p}_{k,n} \sum_{v=1}^{n-1} \frac{1}{v} \prod_{j=1}^{v} \left(1 - \frac{y_{k,n} - 1}{n-j}\right)$$

$$= \sum_k \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \frac{1}{v} \prod_{j=1}^{v} \left(1 - \frac{y_{k,n} - 1}{n-j}\right)$$

and

$$\hat{F}_z = \sum_{v=1}^{n-1} C_v Z_{1,v} = \sum_k \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{v} \left(1 - \frac{y_{k,n} - 1}{n - j}\right).$$

Let $p_{k,n}$, $\hat{p}_{k,n}$, p_n^*, y_n^* and $\hat{p}_n^*$ be defined the same way as in the proof of Theorem 1.1, and accordingly, we have

$$
\begin{aligned}
\zeta_{1,v}^* &= \sum_{k=1}^{K(n)} p_{k,n}(1 - p_{k,n})^v, \\
Z_{1,v}^* &= \frac{n^{1+v}[n-(1+v)]!}{n!} \sum_{k=1}^{K(n)} \left[\hat{p}_{k,n} \prod_{j=0}^{v-1} \left(1 - \hat{p}_{k,n} - \frac{j}{n}\right)\right], \\
H_n^* &= \sum_{k=1}^{K(n)} (-p_{k,n} \ln p_{k,n}) = \sum_{v=1}^{\infty} \frac{1}{v} \zeta_{1,v}^*, \\
\hat{H}_z^* &= \sum_{v=1}^{n-1} \frac{1}{v} Z_{1,v}^* = \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \frac{1}{v} \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}), \\
\hat{H}_n^* &= \sum_{k=1}^{K(n)} (-\hat{p}_{k,n} \ln \hat{p}_{k,n}), \\
F_n^* &= \sum_{v=1}^{\infty} C_v \zeta_{1,v}^*, \\
\hat{F}_z^* &= \sum_{v=1}^{n-1} C_v Z_{1,v}^* = \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}), \\
\hat{F}_n^* &= \sum_{k=1}^{K(n)} (\hat{p}_{k,n} \ln^2 \hat{p}_{k,n}).
\end{aligned}
\tag{2.3}
$$

Also, we will need the following facts in our proofs: Zhang and Zhou (2010) established $E(Z_{1,v}) = \zeta_{1,v}$ and $E(Z_{1,v}^*) = \zeta_{1,v}^*$; also, $C_v = \sum_{i=1}^{v-1} \frac{1}{i(v-i)} = \frac{1}{v} \sum_{i=1}^{v-1} (\frac{1}{i} + \frac{1}{v-i}) \leq \frac{2(\ln v + 1)}{v}$.

Lemma 2.1. Under the condition of Theorem 2.1, we have $\sqrt{n}(\hat{H}_z - \hat{H}_z^*) = o_p(1)$.

Proof.

Noting that for any $k \geq K(n)$, we have $y_{k,n} \leq y_n^*$ and

$$
0 \leq \sum_{v=1}^{n-y_{k,n}} \frac{1}{v} \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}) - \sum_{v=1}^{n-y_n^*} \frac{1}{v} \prod_{j=1}^{v} (1 - \frac{y_n^*-1}{n-j})
$$

$$
\leq \sum_{v=1}^{n-y_{k,n}} \frac{1}{v} \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}) \leq \sum_{v=1}^{n-y_{k,n}} \frac{1}{v} \leq \ln n + 1,
$$

therefore,

$$0 \leq \sqrt{n}(\hat{H}_z - \hat{H}_z^*)$$

$$= \sqrt{n} \sum_{k \geq K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \frac{1}{v} \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}) - \sqrt{n} \hat{p}_n^* \sum_{v=1}^{n-y_n^*} \frac{1}{v} \prod_{j=1}^{v} (1 - \frac{y_n^*-1}{n-j})$$

$$= \sqrt{n} \sum_{k \geq K(n)} \hat{p}_{k,n} \left[\sum_{v=1}^{n-y_{k,n}} \frac{1}{v} \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}) - \sum_{v=1}^{n-y_n^*} \frac{1}{v} \prod_{j=1}^{v} (1 - \frac{y_n^*-1}{n-j}) \right]$$

$$\leq \sqrt{n}(\ln n + 1) \sum_{k \geq K(n)} \hat{p}_{k,n}.$$

By lemma 1.2, $\sqrt{n}(\ln n + 1) E \sum_{k \geq K(n)} \hat{p}_{k,n} = \sqrt{n}(\ln n + 1) \sum_{k \geq K(n)} p_{k,n} \to 0$, hence $\sqrt{n}(\hat{H}_z - \hat{H}_z^*) = o_p(1)$ follows by Markov's Inequality. $\qquad \square$

Lemma 2.2. As $n \to \infty$, under the condition of Theorem 2.1, we have :

$$\sqrt{n}(E\hat{H}_z^* - H_n^*) \to 0$$

.

Proof.

$$0 \leq \sqrt{n}(H_n^* - E\hat{H}_z^*) = \sqrt{n} \sum_{v=1}^{\infty} \frac{1}{v} \zeta_{1,v}^* - \sqrt{n} \sum_{v=1}^{n-1} \frac{1}{v} \zeta_{1,v}^* = \sqrt{n} \sum_{v=n}^{\infty} \frac{1}{v} \zeta_{1,v}^*$$

$$= \sqrt{n} \sum_{v=n}^{\infty} \frac{1}{v} \sum_{k=1}^{K(n)} p_{k,n}(1 - p_{k,n})^v = \sqrt{n} \sum_{k=1}^{K(n)} p_{k,n} \sum_{v=n}^{\infty} \frac{1}{v}(1 - p_{k,n})^v$$

$$\leq \frac{1}{\sqrt{n}} \sum_{k=1}^{K(n)} p_{k,n} \sum_{v=n}^{\infty} (1 - p_{k,n})^v \leq \frac{1}{\sqrt{n}} \sum_{k=1}^{K(n)} p_{k,n} \frac{(1-p_{k,n})^n}{p_{k,n}}$$

$$= \frac{1}{\sqrt{n}} \sum_{k=1}^{K(n)} (1 - p_{k,n})^n \leq \frac{K(n)}{\sqrt{n}} \to 0.$$

Therefore, the lemma follows. $\qquad \square$

Lemma 2.3. As $n \to \infty$, under the condition of Theorem 2.1, we have

$$\sqrt{n}(E\hat{H}_n^* - H_n^*) \to 0$$

.

Proof. Because $f(x) = -x\ln(x)$ is a concave function, by Jensen's inequality, we have

$\sqrt{n}\sum_{k=1}^{K(n)} E(-\hat{p}_{k,n}\ln\hat{p}_{k,n} + p_{k,n}\ln p_{k,n}) \le 0.$

Also, according to (1.2),

$$\sqrt{n}\sum_{k=1}^{K(n)}(E\hat{H}_n^* - H_n^*)$$
$$= \sqrt{n}\sum_{k=1}^{K(n)} E(-\hat{p}_{k,n}\ln\hat{p}_{k,n} + p_{k,n}\ln p_{k,n})1_{[p_{k,n}\ge\frac{1}{n}]}$$
$$+ \sqrt{n}\sum_{k=1}^{K(n)} E(-\hat{p}_{k,n}\ln\hat{p}_{k,n} + p_{k,n}\ln p_{k,n})1_{[p_{k,n}<\frac{1}{n}]}$$
$$\ge \sqrt{n}\left[-\frac{K(n)-1}{2n} + \frac{1}{12n^2}\left(1 - \sum_{k=1}^{K(n)}\frac{1}{p_{k,n}}1_{[p_{k,n}\ge\frac{1}{n}]}\right) + O(n^{-3})\right]$$
$$+ \sqrt{n}\sum_{k=1}^{K(n)}(p_{k,n}\ln p_{k,n})1_{[p_{k,n}<\frac{1}{n}]}$$
$$\ge -\frac{\sqrt{n}K(n)}{2n} - \frac{\sqrt{n}K(n)n}{12n^2} - \frac{\sqrt{n}K(n)\ln n}{n} \to 0.$$

Therefore, $\sqrt{n}(E\hat{H}_n^* - H_n^*) \to 0$. □

Lemma 2.4. If a and b are such that $0 < a < b < 1$, then for any integer $m \ge 0$,

$$b^m - a^m \le mb^{m-1}(b-a).$$

Proof. Noting that $f(x) = x^m$ is convex on interval $(0,1)$ and $f'(b) = mb^{m-1}$, the results follows immediately. □

Lemma 2.5. Under the condition of Theorem 2.1, we have $\sqrt{n}(\hat{H}_z^* - \hat{H}_n^*) = o_p(1)$.

Proof.

$$\sqrt{n}(\hat{H}_z^* - \hat{H}_n^*) = \sqrt{n}\left(\hat{H}_z^* + \sum_{k=1}^{K(n)} \hat{p}_{k,n} \ln \hat{p}_{k,n}\right)$$

$$= \sqrt{n}\left(\hat{H}_z^* - \sum_{k=1}^{K(n)} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}\hat{p}_{k,n}(1 - \hat{p}_{k,n})^v\right)$$

$$- \sqrt{n}\sum_{k=1}^{K(n)} \sum_{v=n-y_{k,n}+1}^{\infty} \tfrac{1}{v}\hat{p}_{k,n}(1 - \hat{p}_{k,n})^v$$

$$:= \mathcal{A}_1 - \mathcal{A}_2.$$

Since

$$0 \le \mathcal{A}_2 = \sqrt{n}\sum_{k=1}^{K(n)} \sum_{v=n-y_{k,n}+1}^{\infty} \tfrac{1}{v}\hat{p}_{k,n}(1 - \hat{p}_{k,n})^v$$

$$\le \sum_{k=1}^{K(n)} \tfrac{\sqrt{n}}{n-y_{k,n}+1}\hat{p}_{k,n} \sum_{v=n-y_{k,n}+1}^{\infty}(1 - \hat{p}_{k,n})^v$$

$$= \sum_{k=1}^{K(n)} \tfrac{\sqrt{n}}{n-y_{k,n}+1}(1 - \hat{p}_{k,n})^{n-y_{k,n}+1} \le \tfrac{1}{\sqrt{n}}\sum_{k=1}^{K(n)} \tfrac{1}{1-\hat{p}_{k,n}+1/n}(1 - \hat{p}_{k,n})^{n-y_{k,n}+1}$$

$$= \tfrac{1}{\sqrt{n}}\sum_{k=1}^{K(n)} \tfrac{1}{1-\hat{p}_{k,n}+1/n}(1 - \hat{p}_{k,n})^{n-y_{k,n}+1}1[\hat{p}_{k,n} < 1]$$

$$\le \tfrac{1}{\sqrt{n}}\sum_{k=1}^{K(n)} \tfrac{1}{1-\hat{p}_{k,n}}(1 - \hat{p}_{k,n})^{n-y_{k,n}+1}1[\hat{p}_{k,n} < 1]$$

$$= \tfrac{1}{\sqrt{n}}\sum_{k=1}^{K(n)}(1 - \hat{p}_{k,n})^{n-y_{k,n}}1[\hat{p}_{k,n} < 1] \le \tfrac{K(n)}{\sqrt{n}},$$

$\mathcal{A}_2 \xrightarrow{a.s.} 0$ and therefore $\mathcal{A}_2 \xrightarrow{p} 0$.

Before considering $\mathcal{A}_1$, we first note the facts that

$$\left(1 - \tfrac{y_{k,n}-1}{n-j}\right) \ge \left(1 - \tfrac{y_{k,n}}{n}\right) = (1 - \hat{p}_{k,n}) \text{ if and only if } 0 \le j \le \tfrac{n}{y_{k,n}+1[y_{k,n}=0]} := \tfrac{1}{\hat{p}_{k,n}^*}$$

$$(2.4)$$

and that, after a few algebraic steps, $Z_{1,v}^*$ may be expressed as

$$Z_{1,v}^* = \sum_{k=1}^{K(n)} \hat{p}_{k,n} \prod_{j=1}^{v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right)$$

$$= \sum_{k=1}^{K(n)} \hat{p}_{k,n} \prod_{j=1}^{J_k \wedge v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right) \prod_{j=J_k \wedge v+1}^{v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right)$$

where $J_k = \lfloor 1/\hat{p}^*_{k,n} \rfloor$ and $\prod_{v=a}^b (\cdot) = 1$ if $a > b$.

$$
\begin{aligned}
\mathcal{A}_1 &= \sqrt{n}\left(\hat{H}^*_z - \sum_{k=1}^{K(n)} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}\hat{p}_{k,n}(1-\hat{p}_{k,n})^v \right) \\
&= \sqrt{n}\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}\left[\prod_{j=1}^{v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right) - (1-\hat{p}_{k,n})^v \right] \\
&= \sqrt{n}\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}\left[\prod_{j=1}^{J_k \wedge v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right)\prod_{j=J_k \wedge v+1}^{v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right)\right. \\
&\quad\quad \left. - \prod_{j=1}^{J_k \wedge v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right)(1-\hat{p}_{k,n})^{0\vee(v-J_k)} \right] \\
&\quad + \sqrt{n}\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}\left[\prod_{j=1}^{J_k \wedge v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right)(1-\hat{p}_{k,n})^{0\vee(v-J_k)} - (1-\hat{p}_{k,n})^v \right] \\
&= \mathcal{A}_{1,1} + \mathcal{A}_{1,2}.
\end{aligned}
$$

$$(2.5)$$

By (2.4), we have $\mathcal{A}_{1,1} \le 0$ and $\mathcal{A}_{1,2} \ge 0$. We want to show that $\mathrm{E}(\mathcal{A}_{1,1}) \to 0$ and $\mathrm{E}(\mathcal{A}_{1,2}) \to 0$ respectively.

$$
\begin{aligned}
\mathcal{A}_{1,1} &= \sqrt{n}\left[\hat{H}^*_z - H^*_n \right] \\
&\quad - \sqrt{n}\left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}\prod_{j=1}^{J_k \wedge v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right)(1-\hat{p}_{k,n})^{0\vee(v-J_k)} - H^*_n \right] \\
&= \sqrt{n}\left[\hat{H}^*_z - H^*_n \right] \\
&\quad - \sqrt{n}\left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}\prod_{j=1}^{J_k \wedge v}\left(1 - \tfrac{y_{k,n}-1}{n-j}\right)(1-\hat{p}_{k,n})^{0\vee(v-J_k)} \right. \\
&\quad\quad \left. - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}(1-\hat{p}_{k,n})^v \right] \\
&\quad - \sqrt{n}\left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}(1-\hat{p}_{k,n})^v - H^*_n \right] \\
&:= \mathcal{A}_{1,1,1} - \mathcal{A}_{1,1,2} - \mathcal{A}_{1,1,3}.
\end{aligned}
$$

$\mathrm{E}(\mathcal{A}_{1,1,1}) \to 0$ follows by Lemma 2.2. Then,

$$
\begin{aligned}
\mathcal{A}_{1,1,3} &= \sqrt{n}\left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v}(1-\hat{p}_{k,n})^v - H^*_n \right] \\
&= \sqrt{n}\left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{\infty} \tfrac{1}{v}(1-\hat{p}_{k,n})^v - H^*_n \right] \\
&\quad - \sqrt{n}\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=n-y_{k,n}+1}^{\infty} \tfrac{1}{v}(1-\hat{p}_{k,n})^v \\
&= \sqrt{n}\left[\hat{H}^*_n - H^*_n \right] - \sqrt{n}\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=n-y_{k,n}+1}^{\infty} \tfrac{1}{v}(1-\hat{p}_{k,n})^v \\
&:= \mathcal{A}_{1,1,3,1} - \mathcal{A}_2.
\end{aligned}
$$

$E(\mathcal{A}_{1,1,3,1}) \to 0$ by Lemma 2.3, $E(\mathcal{A}_2) \to 0$ is established above, and therefore $E(\mathcal{A}_{1,1,3}) \to 0$.

$$
\begin{aligned}
\mathcal{A}_{1,1,2} &= \sqrt{n} \left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} \prod_{j=1}^{J_k \wedge v} \left(1 - \tfrac{y_{k,n}-1}{n-j} \right) (1 - \hat{p}_{k,n})^{0 \vee (v-J_k)} \right. \\
&\qquad \left. - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} (1 - \hat{p}_{k,n})^v \right] \\
&\le \sqrt{n} \left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} \prod_{j=1}^{J_k \wedge v} \left(1 - \tfrac{y_{k,n}-1}{n-1} \right) (1 - \hat{p}_{k,n})^{0 \vee (v-J_k)} \right. \\
&\qquad \left. - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} (1 - \hat{p}_{k,n})^v \right] \\
&= \sqrt{n} \left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} \left(1 - \tfrac{y_{k,n}-1}{n-1} \right)^{J_k \wedge v} (1 - \hat{p}_{k,n})^{0 \vee (v-J_k)} \right. \\
&\qquad \left. - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} (1 - \hat{p}_{k,n})^{J_k \wedge v} (1 - \hat{p}_{k,n})^{0 \vee (v-J_k)} \right] \\
&= \sqrt{n} \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} \left[\left(1 - \tfrac{y_{k,n}-1}{n-1} \right)^{J_k \wedge v} - (1 - \hat{p}_{k,n})^{J_k \wedge v} \right] (1 - \hat{p}_{k,n})^{0 \vee (v-J_k)} \\
&\le \sqrt{n} \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} \left[\left(1 - \tfrac{y_{k,n}-1}{n-1} \right)^{J_k \wedge v} - (1 - \hat{p}_{k,n})^{J_k \wedge v} \right] \\
&\le \sqrt{n} \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} \left[(J_k \wedge v) \left(1 - \tfrac{y_{k,n}-1}{n-1} \right)^{J_k \wedge v - 1} \tfrac{n-y_{k,n}}{n(n-1)} \right] \text{ (by Lemma 2.4)} \\
&\le \tfrac{\sqrt{n}}{n-1} \sum_{k=1}^{K(n)} \hat{p}_{k,n} (1 - \hat{p}_{k,n}) \left[\sum_{v=1}^{n-y_{k,n}} \tfrac{1}{v} (J_k \wedge v) \right] \\
&= \tfrac{\sqrt{n}}{n-1} \sum_{k=1}^{K(n)} \hat{p}_{k,n} (1 - \hat{p}_{k,n}) \left[\sum_{v=1}^{J_k} \tfrac{1}{v} (J_k \wedge v) + \sum_{v=J_k+1}^{n-y_{k,n}} \tfrac{1}{v} (J_k \wedge v) \right] \\
&= \tfrac{\sqrt{n}}{n-1} \sum_{k=1}^{K(n)} \hat{p}_{k,n} (1 - \hat{p}_{k,n}) \left[J_k + J_k \sum_{v=J_k+1}^{n-y_{k,n}} \tfrac{1}{v} \right] \\
&\le \tfrac{\sqrt{n}}{n-1} \sum_{k=1}^{K(n)} \hat{p}_{k,n} \left(J_k + J_k \sum_{v=1}^{n} \tfrac{1}{v} \right) \\
&\le \tfrac{\sqrt{n}}{n-1} \sum_{k=1}^{K(n)} \tfrac{y_{k,n}}{n} \tfrac{n}{y_{k,n}+1[y_{k,n}=0]} (\ln n + 2) \le \tfrac{\sqrt{n} K(n)(\ln n+2)}{n-1}
\end{aligned}
$$

Therefore,

$$
E(\mathcal{A}_{1,1,2}) = \mathcal{O} \left(\tfrac{\sqrt{n} K(n) \ln n}{n} \right) \to 0.
$$

Finally $E(\mathcal{A}_{1,2}) = E(\mathcal{A}_{1,1,2}) \to 0$. It follows that $\sqrt{n}(\hat{H}_z^* - \hat{H}_n^*) = o_p(1)$.

Proof of Theorem 2.1.

Note that

$$\sqrt{n}(\hat{H}_z - H) - \sqrt{n}(\hat{H}_n^* - H_n^*) = \sqrt{n}(\hat{H}_z - \hat{H}_n^*) - \sqrt{n}(H - H_n^*)$$

$$= \sqrt{n}(\hat{H}_z - \hat{H}_z^*) + \sqrt{n}(\hat{H}_z^* - \hat{H}_n^*) - \sqrt{n}(H - H_n^*) \qquad (2.6)$$

$$= \sqrt{n}(\hat{H}_z - \hat{H}_z^*) + \sqrt{n}(\hat{H}_z^* - \hat{H}_n^*) + \sqrt{n} \sum_{k \geq K(n)} (p_k \ln p_k) - \sqrt{n} \, p_n^* \ln p_n^*.$$

We proved $\sqrt{n}(\hat{H}_n^* - H_n^*) \xrightarrow{D} N(0, \sigma^2)$ in (1.7). The proof is complete if we can show each of the four terms in the right hand side of (2.6) is $o_p(1)$.

The first two terms are $o_p(1)$ by Lemma 2.1 and Lemma 2.5 respectively, and the third term goes to 0 by the condition of Theorem 2.1, and the fourth term goes to 0 by (1.4). Therefore, by Slutsky's lemma, we conclude the theorem. □

To prove the Corollary 2.1, we need a few lemmas as follows:

Lemma 2.6. Under the condition of Theorem 2.1, we have $\hat{F}_n - \hat{F}_n^* = o_p(1)$.

Proof.
$$0 \leq \hat{F}_n - \hat{F}_n^* = \sum_{k \geq K(n)} \hat{p}_{k,n} \ln^2 \hat{p}_{k,n} - \hat{p}_n^* \ln^2 \hat{p}_n^*$$
$$= \sum_{k \geq K(n)} \hat{p}_{k,n} \ln^2 \hat{p}_{k,n} - \sum_{k \geq K(n)} \hat{p}_{k,n} \ln^2 \hat{p}_n^*$$
$$= \sum_{k \geq K(n)} \hat{p}_{k,n} (\ln^2 \hat{p}_{k,n} - \ln^2 \hat{p}_n^*) \leq \sum_{k \geq K(n)} \hat{p}_{k,n} \ln^2 \hat{p}_{k,n}$$
$$\leq \ln^2 n \sum_{k \geq K(n)} \hat{p}_{k,n}$$

By lemma 1.2, $\ln^2 n E \sum_{k \geq K(n)} \hat{p}_{k,n} = \ln^2 n \sum_{k \geq K(n)} p_{k,n} \to 0$, $\hat{F}_n - \hat{F}_n^* = o_p(1)$ follows by Markov's Inequality. □

Lemma 2.7. Under the condition of Theorem 2.1, we have $\hat{F}_z - \hat{F}_z^* = o_p(1)$.

Proof. Noting that for any $k \geq K(n)$, we have $y_{k,n} \leq y_n^*$ and

$$0 \leq \sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}) - \sum_{v=1}^{n-y_n^*} C_v \prod_{j=1}^{v} (1 - \frac{y_n^*-1}{n-j})$$

$$\leq \sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}) \leq \sum_{v=1}^{n-y_{k,n}} C_v \leq \sum_{v=1}^{n} \frac{2(\ln v+1)}{v} \leq 2(\ln n + 1)^2,$$

therefore,

$$0 \leq \hat{F}_z - \hat{F}_z^*$$

$$= \sum_{k \geq K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}) - \hat{p}_n^* \sum_{v=1}^{n-y_n^*} C_v \prod_{j=1}^{v} (1 - \frac{y_n^*-1}{n-j})$$

$$= \sum_{k \geq K(n)} \hat{p}_{k,n} \left[\sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{v} (1 - \frac{y_{k,n}-1}{n-j}) - \sum_{v=1}^{n-y_n^*} C_v \prod_{j=1}^{v} (1 - \frac{y_n^*-1}{n-j}) \right]$$

$$\leq 2(\ln n + 1)^2 \sum_{k \geq K(n)} \hat{p}_{k,n}.$$

By lemma 1.2, $(\ln n + 1)^2 E \sum_{k \geq K(n)} \hat{p}_{k,n} = (\ln n + 1)^2 \sum_{k \geq K(n)} p_{k,n} \to 0$, therefore, $\hat{F}_z - \hat{F}_z^* = o_p(1)$ follows by Markov's Inequality. $\square$

Lemma 2.8. As $n \to \infty$, under the condition of Theorem 2.1, we have :

$$E(\hat{F}_z^*) - F_n^* \to 0$$

.

Proof.

$$0 \le F_n^* - E(\hat{F}_z^*) = \sum_{v=1}^{\infty} C_v \zeta_{1,v}^* - \sum_{v=1}^{n-1} C_v \zeta_{1,v}^* = \sum_{v=n}^{\infty} C_v \zeta_{1,v}^*$$

$$= \sum_{v=n}^{\infty} C_v \sum_{k=1}^{K(n)} p_{k,n} (1 - p_{k,n})^v = \sum_{k=1}^{K(n)} p_{k,n} \sum_{v=n}^{\infty} C_v (1 - p_{k,n})^v$$

$$\le \frac{2(\ln n + 1)}{n} \sum_{k=1}^{K(n)} p_{k,n} \sum_{v=n}^{\infty} (1 - p_{k,n})^v \le \frac{2(\ln n + 1)}{n} \sum_{k=1}^{K(n)} p_{k,n} \frac{(1 - p_{k,n})^n}{p_{k,n}}$$

$$= \frac{2(\ln n + 1)}{n} \sum_{k=1}^{K(n)} (1 - p_{k,n})^n \le \frac{2(\ln n + 1) K(n)}{n} \to 0.$$

$\square$

Lemma 2.9. Under the condition of Theorem 2.1, we have $\hat{F}_z^* - \hat{F}_n^* = o_p(1)$.

Proof.

$$\hat{F}_z^* - \hat{F}_n^* = \hat{F}_z^* - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \ln^2 \hat{p}_{k,n}$$

$$= \left[\hat{F}_z^* - \sum_{k=1}^{K(n)} \sum_{v=1}^{n-y_{k,n}} C_v \hat{p}_{k,n} (1 - \hat{p}_{k,n})^v \right]$$

$$- \sum_{k=1}^{K(n)} \sum_{v=n-y_{k,n}+1}^{\infty} C_v \hat{p}_{k,n} (1 - \hat{p}_{k,n})^v$$

$$:= \mathcal{B}_1 - \mathcal{B}_2.$$

Since

$$0 \le \mathcal{B}_2 = \sum_{k=1}^{K(n)} \sum_{v=n-y_{k,n}+1}^{\infty} C_v \hat{p}_{k,n} (1 - \hat{p}_{k,n})^v$$

$$\le \sum_{k=1}^{K(n)} \frac{2(\ln n + 1)}{n - y_{k,n} + 1} \hat{p}_{k,n} \sum_{v=n-y_{k,n}+1}^{\infty} (1 - \hat{p}_{k,n})^v$$

$$= \sum_{k=1}^{K(n)} \frac{2(\ln n + 1)}{n - y_{k,n} + 1} (1 - \hat{p}_{k,n})^{n-y_{k,n}+1}$$

$$\le \frac{2(\ln n + 1)}{n} \sum_{k=1}^{K(n)} \frac{1}{1 - \hat{p}_{k,n} + 1/n} (1 - \hat{p}_{k,n})^{n-y_{k,n}+1}$$

$$= \frac{2(\ln n + 1)}{n} \sum_{k=1}^{K(n)} \frac{1}{1 - \hat{p}_{k,n} + 1/n} (1 - \hat{p}_{k,n})^{n-y_{k,n}+1} 1[\hat{p}_{k,n} < 1]$$

$$\le \frac{2(\ln n + 1)}{n} \sum_{k=1}^{K(n)} \frac{1}{1 - \hat{p}_{k,n}} (1 - \hat{p}_{k,n})^{n-y_{k,n}+1} 1[\hat{p}_{k,n} < 1]$$

$$= \frac{2(\ln n + 1)}{n} \sum_{k=1}^{K(n)} (1 - \hat{p}_{k,n})^{n-y_{k,n}} 1[\hat{p}_{k,n} < 1] \le \frac{2(\ln n + 1) K(n)}{n},$$

$\mathcal{B}_2 \overset{a.s.}{\to} 0$ and therefore $\mathcal{B}_2 \overset{p}{\to} 0$. Next,

$$
\begin{aligned}
\mathcal{B}_1 &= \hat{F}_z^* - \sum_{k=1}^{K(n)} \sum_{v=1}^{n-y_{k,n}} C_v \hat{p}_{k,n} (1 - \hat{p}_{k,n})^v \\
&= \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \left[\prod_{j=1}^{v} \left(1 - \frac{y_{k,n}-1}{n-j}\right) - (1 - \hat{p}_{k,n})^v \right] \\
&= \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \left[\prod_{j=1}^{J_k \wedge v} \left(1 - \frac{y_{k,n}-1}{n-j}\right) \prod_{j=J_k \wedge v+1}^{v} \left(1 - \frac{y_{k,n}-1}{n-j}\right) \right. \\
&\quad \left. - \prod_{j=1}^{J_k \wedge v} \left(1 - \frac{y_{k,n}-1}{n-j}\right) (1 - \hat{p}_{k,n})^{0 \vee (v - J_k)} \right] \\
&\quad + \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \left[\prod_{j=1}^{J_k \wedge v} \left(1 - \frac{y_{k,n}-1}{n-j}\right) (1 - \hat{p}_{k,n})^{0 \vee (v - J_k)} - (1 - \hat{p}_{k,n})^v \right] \\
&= \mathcal{B}_{1,1} + \mathcal{B}_{1,2}.
\end{aligned}
$$

$$(2.7)$$

By (2.4), we have $\mathcal{B}_{1,1} \leq 0$ and $\mathcal{B}_{1,2} \geq 0$. We want to show that $\mathrm{E}(\mathcal{B}_{1,1}) \to 0$ and $\mathrm{E}(\mathcal{B}_{1,2}) \to 0$ respectively.

$$
\begin{aligned}
\mathcal{B}_{1,1} &= \left(\hat{F}_z^* - F_n^* \right) \\
&\quad - \left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{J_k \wedge v} \left(1 - \frac{y_{k,n}-1}{n-j}\right) (1 - \hat{p}_{k,n})^{0 \vee (v - J_k)} - F_n^* \right] \\
&= \left(\hat{F}_z^* - F_n^* \right) \\
&\quad - \left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{J_k \wedge v} \left(1 - \frac{y_{k,n}-1}{n-j}\right) (1 - \hat{p}_{k,n})^{0 \vee (v - J_k)} \right. \\
&\quad \left. - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v (1 - \hat{p}_{k,n})^v \right] \\
&\quad - \left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v (1 - \hat{p}_{k,n})^v - F_n^* \right] \\
&:= \mathcal{B}_{1,1,1} - \mathcal{B}_{1,1,2} - \mathcal{B}_{1,1,3}.
\end{aligned}
$$

$\mathrm{E}(\mathcal{B}_{1,1,1}) \to 0$ follows by Lemma 2.8. Next,

$$
\begin{aligned}
\mathcal{B}_{1,1,3} &= \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v (1 - \hat{p}_{k,n})^v - F_n^* \\
&= \left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{\infty} C_v (1 - \hat{p}_{k,n})^v - F_n^* \right] \\
&\quad - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=n-y_{k,n}+1}^{\infty} C_v (1 - \hat{p}_{k,n})^v \\
&= \left(\hat{F}_n^* - F_n^* \right) - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=n-y_{k,n}+1}^{\infty} C_v (1 - \hat{p}_{k,n})^v \\
&:= \mathcal{B}_{1,1,3,1} - \mathcal{B}_2.
\end{aligned}
$$

As we showed $\lim_{n\to\infty} E\hat{F}_n = F$ in Corollary 1.1, $\mathrm{E}(\mathcal{B}_{1,1,3,1}) \to 0$ can be proved in the same way. Also, $\mathrm{E}(\mathcal{B}_2) \to 0$ is established above, therefore, $\mathrm{E}(\mathcal{B}_{1,1,3}) \to 0$.

$$
\begin{aligned}
\mathcal{B}_{1,1,2} &= \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{J_k \wedge v} \left(1 - \frac{y_{k,n}-1}{n-j}\right) (1-\hat{p}_{k,n})^{0\vee(v-J_k)} \\
&\quad - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v (1-\hat{p}_{k,n})^v \\
&\leq \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \prod_{j=1}^{J_k \wedge v} \left(1 - \frac{y_{k,n}-1}{n-1}\right) (1-\hat{p}_{k,n})^{0\vee(v-J_k)} \\
&\quad - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v (1-\hat{p}_{k,n})^v \\
&= \left[\sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \left(1 - \frac{y_{k,n}-1}{n-1}\right)^{J_k \wedge v} (1-\hat{p}_{k,n})^{0\vee(v-J_k)} \right. \\
&\quad \left. - \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v (1-\hat{p}_{k,n})^{J_k \wedge v} (1-\hat{p}_{k,n})^{0\vee(v-J_k)} \right] \\
&= \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \left[\left(1 - \frac{y_{k,n}-1}{n-1}\right)^{J_k \wedge v} - (1-\hat{p}_{k,n})^{J_k \wedge v}\right] (1-\hat{p}_{k,n})^{0\vee(v-J_k)} \\
&\leq \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \left[\left(1 - \frac{y_{k,n}-1}{n-1}\right)^{J_k \wedge v} - (1-\hat{p}_{k,n})^{J_k \wedge v}\right] \\
&\leq \sum_{k=1}^{K(n)} \hat{p}_{k,n} \sum_{v=1}^{n-y_{k,n}} C_v \left[(J_k \wedge v)\left(1 - \frac{y_{k,n}-1}{n-1}\right)^{J_k \wedge v-1} \frac{n-y_{k,n}}{n(n-1)}\right] \quad \text{(by Lemma 2.4)} \\
&\leq \frac{1}{n-1} \sum_{k=1}^{K(n)} \hat{p}_{k,n}(1-\hat{p}_{k,n}) \left[\sum_{v=1}^{n-y_{k,n}} C_v(J_k \wedge v)\right] \\
&\leq \frac{1}{n-1} \sum_{k=1}^{K(n)} \hat{p}_{k,n}(1-\hat{p}_{k,n}) J_k \sum_{v=1}^{n-y_{k,n}} C_v \\
&\leq \frac{1}{n-1} \sum_{k=1}^{K(n)} \hat{p}_{k,n}(1-\hat{p}_{k,n}) J_k \sum_{v=1}^{n} \frac{2(\ln n+1)}{v} \\
&\leq \frac{2(\ln n+1)}{n-1} \sum_{k=1}^{K(n)} \hat{p}_{k,n} J_k (\ln n + 1) \\
&\leq \frac{2(\ln n+1)^2}{n-1} \sum_{k=1}^{K(n)} \frac{y_{k,n}}{n} \frac{n}{y_{k,n}+1[y_{k,n}=0]} \leq \frac{2(\ln n+1)^2 K(n)}{n-1}
\end{aligned}
$$

Therefore,

$$
\mathrm{E}(\mathcal{B}_{1,1,2}) = \mathcal{O}\left(\frac{K(n)\ln^2 n}{n}\right) \to 0.
$$

Finally $\mathrm{E}(\mathcal{B}_{1,2}) = \mathrm{E}(\mathcal{B}_{1,1,2}) \to 0$. It follows that $\hat{F}_z^* - \hat{F}_n^* = o_p(1)$.

Proof of Corollary 2.1.

$$
\hat{F}_z - F = (\hat{F}_z - \hat{F}_z^*) + (\hat{F}_z^* - \hat{F}_n^*) + (\hat{F}_n^* - \hat{F}_n) + (\hat{F}_n - F)
$$

Each of the first three terms in above equation is $o_p(1)$ by Lemma 2.7, Lemma

2.9, and Lemma 2.6 respectively. Also, we showed in the proof of Corollary 1.1 that $\hat{F}_n - F = o_p(1)$. Therefore, $\hat{F}_z - F = o_p(1)$.

By Theorem 2.1, $\hat{H}_z^2 \xrightarrow{p} H^2$ and hence, $\hat{F}_z - \hat{H}_z^2 \xrightarrow{p} F - H^2 = \sigma^2$. Finally,we conclude the corollary by Theorem 2.1 and Slusky's Theorem. $\square$

2.3   Remarks

In conclusion, the sufficient condition of the new estimator supports less distribution class than that of the plug-in estimator.However, simulations showed that $\hat{H}_z$ always outperforms $\hat{H}_n$ under various distributions; also, Zhang (2012) showed that the upper bound of the variance of $\hat{H}_z$ decays faster than that of $\hat{H}_n$ by a factor of $\ln(n)$ at a rate $O(\ln(n)/n)$ for all distributions with finite entropy, they lead to my conjecture that the sufficient condition of Theorem 2.1 can be further relaxed. Next question that one would naturally ask is whether there exists a probability distribution under which the normality of one estimator holds but does not for the other.To answer this question completely, the directions of the future research should aim to establish the necessary and sufficient conditions of both estimators.

REFERENCES

Antos, A. and Kontoyiannis, I. (2001). Convergence properties of functional estimates for discrete distributions, Random Structures and Algorithm, 19, 163-193.

Basharin, G. (1959). On a statisical estimate for the entropy of a sequence of independent random variables, Theory of Probablity and Its Applications, 4, 333-336.

Beirlant, J., Dudewicz, E.J., Györfi, L. & Meulen, E.C. (2001). Nonparametric Entropy Estimation: An Overview, International Journal of the Mathematical Statistics Sciences, 6, 17-39.

Chao, A. and Shen, T.J. (2003). Nonparametric estimation of Shannon' index of diversity when there are unseen species in sample, Environmental and Ecological Statistics, 10, 429-443.

Devroye, L. (1991). Expontential inequalities in nonparametric estimation, In Nonparametirc Functional Estimation and Related Topics, ed. G. Roussas, NATO ASI Series, Kluwer Academic, Dordrecht, 31-44.

Good, I.J.(1953). The population frequencies of species and the estimation of population parameters, Biometrika, 40, pp. 237-264.

Grassberger, P. (1988). Finite sample corrections to entropy and dimension estimates, Physics Letters A, 128 (6-7), 369-373.

Harris, B. (1975). The Statistical estimation of entropy in the non-parametric case, Topics in Information Theory, edited by I. Csiszar, Amsterdam: North-Holland, 323-355.

Miller, G.A. and Madow, W.G. (1954). On the maximum likelihood estimate of the Shannon-Wiener measure of information, Air Force Cambridge Research Center Technical Report 54-75.

Miller, G.A. (1955). Note on the bias of information estimates, Information theory in psychology II-B, ed. H. Quastler, Glencoe, IL: Free Press, 95-100.

Nemenman, I., Shafee, F. and Bialek, W. (2002). Entropy and inference, revisited. Advances in Neural Information Processing Systems 14, Cambridge, MA, 2002. MIT Press.

Nemenman, I., Bialek, W. and de Ruyter van Steveninck R. (2004). Entropy and information in neural spike trains: Progress on the sampling problem. Physical Review E, 69(5), 056111.

Paninski, L. (2003). Estimation of entropy and mutual information, Neural Comp. 15, 1191-1253.

Panzeri, S., Senatore, R., Montemurro, M.A., and Petersen, R.S. (2007). Correcting for the sampling bias problem in spike train information measures, J. Neurophysiology 98, 1064-1072.

Panzeri, S. and Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures, Network: Computation in Neural Systems 7, 87-107.

Shannon, C.E. (1948). A Mathematical Theory of Communications, Bell Syst. Tech. J., 27, 379-423 and 623-656.

Strong, S.P., Koberle, R., de Ruyter van Steveninck, R.R., & Bialek, W. (1998). Entropy and information in neural spike trains. Physical Review Letters, 80 (1), 197-200.

Sch$\ddot{u}$rmann, T. (2004). Bias analysis in entropy estimation, J. Phys. A: Math. Gen. 37, 295-301.

Treves, A. and Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples, Neural Computation 7, 399-407.

Victor, J.D. (2000). Asymptotic bias in information estimates and the exponential (bell) polynomials, Neural Computation 12, 2797-2804.

Vu, V.Q., Yu, B., and Kass, R.E. (2007). Coverage-adjusted entropy estimation, Statist. Med., 26, 4039-4060.

Zhang, Z. (2012). Entropy estimation in Turing's perspective, Neural Computation, Vol. 24, No. 5, pp.1368-1389.

Zhang, Z. (2013). Asymptotic normality of an entropy estimator with exponentially decaying bias, IEEE Transactions on Information Theory, Vol. 59, No. 1, pp.504-508.

Zhang, Z. and Zhang, X. (2012). A normal law for the plug-in estimator of entropy, IEEE Transactions on Information Theory, Vol. 58, No. 5, pp.2745-2747.

Zhang, Z. and Zhou, J. (2010). Re-parameterization of multinomial distribution and diversity indices, J. of Statistical Planning and Inference, Vol. 140, No. 7, pp.1731-1738.