NON-PARAMETRIC ESTIMATION OF INFORMATION-THEORETIC QUANTITIES IN ENTROPIC PERSPECTIVE

by

Jialin 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

2019

Approved by:

____

Dr. Zhiyi Zhang

____

Dr. Jiancheng Jiang

____

Dr. Jun Song

____

Dr. Arindam Mukherjee

ABSTRACT

JIALIN ZHANG. NON-PARAMETRIC ESTIMATION OF INFORMATION-THEORETIC QUANTITIES IN ENTROPIC PERSPECTIVE. (Under the direction of DR. ZHIYI ZHANG)

Introduced by Shannon [1], mutual information is a fundamental brick of information theory for its essential role in measuring association on non-ordinal alphabets. Mutual information being zero is a golden property as it indicates a probabilistic independent between the distributions. This article offers asymptotic chi-square distributions for the plug-in estimator and a non-parametric estimator of mutual information. The established distributions allow new tests of independence in entropic perspective.

ACKNOWLEDGEMENTS

The author would like to express the deepest appreciation to his committee chair Professor Zhiyi Zhang, who has continually and convincingly conveyed a spirit of bold adventure in regard to research. Without his uncountable, persistent, and invaluable guidance this dissertation would not have been possible. Special thanks also go to the members in the committee, Prof. Jiancheng Jiang, Prof. Jun Song, and Prof. Arindam Mukherjee, for their insightful and abundant assistance during this study.

In addition, the author would like to thank the University of North Carolina at Charlotte for financial assistance via GASP.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Introduction

1.1 Challenges in Big Data

Recent advancement in information technology increases the capability to obtain, exchange, and store data, hence the term big data. In analyzing these data, at least two fundamental issues immediately present themselves: 1) high dimensionality, and 2) discrete and non-ordinal nature.

Specifically, the vastly complex data space suggests that a data observation can only be appropriately registered in a very high-dimensional space, so much so that the dimensionality could essentially be infinite. On such spaces, the usual statistical methodologies quickly run into estimation and fundamental conceptual problems. Besides, the generality of the data space suggests that possible data values may not have an inherent order among themselves, for example, different gene types in the human genome, different words in a text, and various species in an ecological population.

The absence of inherent order immediately challenges the concept of a random variable. Namely, random variable, a function from the sample space to the real space, does not naturally exist on non-ordinal data spaces. Consequently, many familiar and fundamental concepts of Statistics and Probability no longer exist, for example, moments, correlation, tail, characteristic function. As a result, to characterize the probability distributions on non-ordinal alphabets, inter-disciplinary research between Statistics, Probability, and Information Theory is needed.

1.2 Entropy, and Information-theoretic Quantities

In 1948, Shannon introduced the concept of entropy and mutual information in his landmark paper [1], where he defined Shannon's entropy as:

$$H = \sum_k p_k \ln p_k.$$

Compared to classical concepts (*e.g.*, moments), entropy is calculated by probabilities (or ordered probabilities), and it does not rely on metric information, and thus it could exist in non-ordinal alphabets. Without using any metric information, entropy describes the level of dispersion in the probability distribution. In general, the larger the entropy, the heavier the dispersion. For example, a probability distribution with an effective cardinality $K = 4$ can produce a maximum possible entropy of $\ln 4$, and the maximum is achieved when the population distribution is uniform. And thus entropy can be considered as the moments on non-ordinal alphabets.

Based on entropy, various information-theoretic quantities were proposed, for example, mutual information, Kullback-Leibler divergence, and entropic moments. These quantities characterize the information from different non-ordinal perspectives, and they exist under much broader conditions.

For example, mutual information can be considered as covariance in classical Statistics as it could capture the associations among both non-ordinal and ordinal random elements (or random variables). Furthermore, no matter if the relationship between ordinal random variables is linear or nonlinear, mutual information could capture it.

Because covariance is not bounded from above, it is inconvenient to interpret the degree of correlation between the random variables. As a result, correlation coefficient was introduced to bound the measurement within $[-1, 1]$ for a convenient interpretation, where 0 stands for no linear correlation and an absolute value of 1 stands for perfect linearity. The same issue is with mutual information: it is not bounded from above. Under a very similar spirit, standardized mutual information, another information-theoretic quantity, was introduced to bound the measurement with $[0, 1]$, where 0 stands for probabilistic independence and 1 stands for a one-to-one correspondence.

1.3 Entropy Estimation

Most of the existing information-theoretic quantities are linear functions of entropy. As a result, the estimation of entropy plays a central role in practice in Information Theory. However, the estimation of entropy is technically difficult problems due to the curse of "High Dimensionality" and "Discrete and Non-ordinal Nature". For about 50 years since [1], advances in this area have been slow to come. Naively, people have been using the plug-in estimator (or the maximum likelihood estimator)

$$\hat{H} = \sum \hat{p}_k \ln \hat{p}_k.$$

When the K is finite, [2] showed that the bias of $\hat{H}$ is

$$\mathrm{E}(\hat{H}) - H = -\frac{K-1}{2n} + \frac{1}{12n^2}\left(1 - \sum_{k=1}^{K} \frac{1}{p_k}\right) + \mathcal{O}\left(n^{-3}\right). \tag{1.3.1}$$

Plentiful estimators tried to add bias corrections to the plug-in estimators when K is finite. Such attempts include Miller-Madow (1955) estimator ($\hat{H}_{MM}$) [3] and the Jackknife (1977) estimator ($\hat{H}_{JK}$) [4]. let $\hat{K}$ be the number of categories observed in the sample, and

$$\hat{H}_{MM} = \hat{H} + \frac{\hat{K}-1}{2n}. \tag{1.3.2}$$

It can be shown that, for finite K, the bias of $\hat{H}_{MM}$ is

$$\mathrm{E}(\hat{H}_{MM}) - H = \frac{1}{12n^2}\left(1 - \sum_{k=1}^{K} \frac{1}{p_k}\right) + \mathcal{O}\left(n^{-3}\right).$$

$\hat{H}_{JK}$ is calculated in three steps:

1. for each $i \in \{1, 2, \ldots, n\}$ construct $\hat{H}^{(i)}$, which is a plug-in estimator based on a sub-sample of size $n-1$ obtained by leaving the ith observation out;

2. obtain $\hat{H}_{(i)} = n\hat{H} - (n-1)\hat{H}^{(i)}$ for $i = 1, \cdots, n$; and then

3. compute the jackknife estimator

$$\hat{H}_{JK} = \frac{\sum_{i=1}^{n} \hat{H}_{(i)}}{n}. \tag{1.3.3}$$

Equivalently, (1.3.3) can be written as

$$\hat{H}_{JK} = n\hat{H} - (n-1)\frac{\sum_{i=1}^{n} \hat{H}^{(i)}}{n}.$$

When $K < \infty$, it can be shown that the bias of $\hat{H}_{JK}$ is

$$\mathrm{E}\left(\hat{H}_{JK}\right) - H = \mathcal{O}\left(n^{-2}\right).$$

$\hat{H}_{MM}$ and $\hat{H}_{JK}$ reduce the rate of bias to a higher order power-decaying. While researchers were seeking for unbiased estimators, [5] proved that for finite K, an unbiased estimator for entropy does not exist. As a result, it is only possible to reduce the bias to a smaller extent.

1.4 Entropic Perspective

Entropic perspective In this section, the origin of entropic perspective is introduced. It is followed by the estimation of entropic moments, which is the core of entropic perspective. Some properties of entropic moments are then discussed.

1.4.1 Origin

Shannon defined entropy as

$$
\begin{aligned}
H &= -\sum_{k\geq 1} p_k \ln p_k \\
&= \sum_{k\geq 1} p_k(-\ln p_k).
\end{aligned}
$$

By Taylor's expansion

$$
-\ln p_k = \sum_{v=1}^{\infty} \frac{(1-p_k)^v}{v},
$$

hence

$$
H = \sum_{k\geq 1} p_k \sum_{v=1}^{\infty} \frac{(1-p_k)^v}{v}.
$$

By Fubini's lemma, a finite entropy has the following alternative representation:

$$
H = \sum_{v=1}^{\infty} \frac{1}{v} \sum_{k\geq 1} p_k(1-p_k)^v.
$$

Definition 1 (Entropic Moments) *For any probability distribution $\{p_k\}$, the entropic moments, ζ_v, are defined as*

$$
\zeta_v = \sum_{k\geq 1} p_k(1-p_k)^v, \qquad v = 1, 2, \cdots .
$$

Therefore

$$
H = \sum_{v=1}^{\infty} \frac{1}{v}\, \zeta_v,
$$

which indicates H is a harmonically weighted linear combination of the entropic moments, ζ_v.

1.4.2 Estimation of Entropic Moments

[6] proposed an unbiased estimator for ζ_v for any $v \le n-1$:

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

Moreover, if K is finite, $Z_{1,v}$ is a Uniformly Minimum Variance Unbiased Estimator (UMVUE) of ζ_v for any $v \le n-1$. The asymptotic property of $Z_{1,v}$ were also established:

$$\frac{\sqrt{n}(Z_{1,v}-\zeta_v)}{(v+1)\hat{\sigma}(1,v)} \xrightarrow{L} N(0,1)$$

where

$$\hat{\sigma}^2(u,v) = \left(\frac{v}{u+v}\right)^2\left(Z_{2u-1,2v-2} + \frac{u^2+v^2}{v^2}Z_{2u-1,2v}\right) - \left(\frac{v}{u+v}\right)^2\left[2Z_{2u-1,2v-1} + \frac{2u}{v}Z_{2u,2v-1} + \left(\frac{u}{v}Z_{u,v} - Z_{u+1,v-1}\right)^2\right].$$

Based on these UMVUEs, $\hat{H}_z$ [7], the state-of-the-art entropy estimator, was developed as

$$\hat{H}_z(\cdot) = \sum_{v=1}^{n-1}\frac{1}{v}\frac{n^{1+v}[n-(1+v)]!}{n!}\sum_k\left[\hat{p}_k\prod_{j=0}^{v-1}\left(1-\hat{p}_k-\frac{j}{n}\right)\right]. \qquad (1.4.1)$$

The bias of $\hat{H}_z$ is

$$\mathrm{E}(\hat{H}_z) - H = \mathcal{O}\left(\frac{(1-p_\wedge)^n}{n}\right), \qquad (1.4.2)$$

where $p_\wedge = \min\{p_k > 0\}$. When K is finite, $\hat{H}_z$ reduces the bias of entropy estimation from power decaying to exponentially decaying. Compared to $\hat{H}$ and $\hat{H}_{MM}$, of which the biases are infinity with a finite sample, $\hat{H}_z$ has an exponentially decaying bias under the same condition.

The established results provided massive potentials on entropic perspective. Some properties of entropic moments are introduced in the following part.

1.4.3 Properties of Entropic Moments

One of the most important properties of entropic moments, proved in [8], is

$$\{\zeta_{1,v}; v \geq 0\} \quad \text{and} \quad \{p_k; k \geq 1\}$$

uniquely determine each other up to a permutation on the index set $\{k : k \geq 1\}$.

As we have discussed in Chapter 1.1, characteristics function, one of the fundamental concepts in Statistics, does not exist on non-ordinal alphabets. As a result, to characterize different distributions on non-ordinal alphabets, a new concept is urgently needed, hence the entropic moments.

Furthermore, it was also proved in [8] that numerous quantities could be represented as functions of entropic moments. Some of these quantities are summarized in Table 1.1. All these quantities can be estimated (or partially estimated) from entropic perspective. Besides, all information-theoretic quantities that are functions of entropy can also be estimated from entropic perspective. An example of the advantage in estimation from entropic perspective is in [9].

Table 1.1: Example quantities that can be represented as functions of entropic moments.

Simpson's index:	λ	=	$\sum_{k\geq 1} p_k^2$	=	$\zeta_{1,0} - \zeta_{1,1}$
Gini-Simpson index:	$1-\lambda$	=	$\sum_{k\geq 1} p_k(1-p_k)$	=	$\zeta_{1,1}$
Shannon's entropy:	H	=	$-\sum_{k\geq 1} p_k \ln(p_k)$	=	$\sum_{v=1}^{\infty} \frac{1}{v}\, \zeta_{1,v}$
Rényi equiv. entropy:	h_r	=	$\sum_{k\geq 1} p_k^r$	=	$\zeta_{1,0} + \sum_{v=1}^{\infty} \prod_{i=1}^{v} \left(\frac{i-r}{i}\right) \zeta_{1,v}$
Emlen's index:	D	=	$\sum_{k\geq 1} p_k e^{-p_k}$	=	$\sum_{v=0}^{\infty} \frac{e^{-1}}{v!} \zeta_{1,v}$
Richness index:	K	=	$\sum_{k\geq 1} 1[p_k > 0]$	=	$\sum_{v=0}^{\infty} \zeta_{1,v}$
Gen. Simpson's index:	$\zeta_{u,\,m}$	=	$\sum_{k\geq 1} p_k^u (1-p_k)^m$	=	$\sum_{v=0}^{u-1} (-1)^v \binom{u-1}{v} \zeta_{1,\,m+v}$

1.5 Summary

In summary, the advantages of entropy on non-ordinal alphabets are demonstrated; different estimators of entropy are compared; entropic perspective and entropic moments are introduced.

In addition, for finite K, the asymptotic normalities for $\hat{H}, \hat{H}_{MM}$, and $\hat{H}_{JK}$ are straight forward, and the asymptotic normality for $\hat{H}_z$ was provided in [10]. When K is countable infinite, asymptotic distributions with certain conditions for $\hat{H}, \hat{H}_{MM}$, $\hat{H}_{JK}$, and $\hat{H}_z$ are established recently in [11], [12], and [13]. Additional discussions on entropy estimation could also be found in [14] and [15].

As discussed in Chapter 1.2, mutual information is an important concept on non-ordinal alphabets for its ability to capture association and detect independence. By far, asymptotic properties of mutual information when it is zero is missing from the literature. It is of interest as the asymptotic properties of mutual information being zero could induce a new test of independence. Recent progress in entropy estimation has made it even more attractive to investigate the asymptotic properties of mutual information. In Chapter 2, two mutual information estimators and their asymptotic properties are discussed.

CHAPTER 2: Asymptotic Distributions for Mutual Information Estimators

Let $\mathscr{X} = \{x_i; i = 1, \cdots, K_1\}$ and $\mathscr{Y} = \{y_j; j = 1, \cdots, K_2\}$ be two finite alphabets with cardinalities $K_1 < \infty$ and $K_2 < \infty$ respectively. Consider the Cartesian product $\mathscr{X} \times \mathscr{Y}$ with a joint probability distribution $\mathbf{p} = \{p_{i,j}\}$. Let the two marginal distributions be respectively denoted by $\mathbf{p}_x = \{p_{i,\cdot}\}$ and $\mathbf{p}_y = \{p_{\cdot,j}\}$ where $p_{i,\cdot} = \sum_j p_{i,j}$ and $p_{\cdot,j} = \sum_i p_{i,j}$. Assume that $p_{i,\cdot} > 0$ and $p_{\cdot,j} > 0$ for all $1 \le i \le K_1$ and $1 \le j \le K_2$, and that there are $K = \sum_{i,j} 1[p_{i,j} > 0]$ non-zero entries in $\{p_{i,j}\}$. We re-enumerate these K positive probabilities in one sequence and denote it as $\{p_k; k = 1, \cdots, K\}$.

Shannon's entropies for $\mathscr{X}$, $\mathscr{Y}$, and $\mathscr{X} \times \mathscr{Y}$, and mutual information between $\mathscr{X}$ and $\mathscr{Y}$ are defined as

$$\begin{aligned} H(X) &= -\textstyle\sum_i p_{i,\cdot} \ln p_{i,\cdot}, \\ H(Y) &= -\textstyle\sum_j p_{\cdot,j} \ln p_{\cdot,j}, \\ H(X,Y) &= -\textstyle\sum_i \sum_j p_{i,j} \ln p_{i,j} = -\sum_{k=1}^{K} p_k \ln p_k, \\ MI(X,Y) &= H(X) + H(Y) - H(X,Y). \end{aligned} \tag{2.0.1}$$

For every pair of i and j, let $f_{i,j}$ be the observed frequency of the random pair (X, Y) taking value (x_i, y_j), where $i = 1, \cdots, K_1$ and $j = 1, \cdots, K_2$, in an *iid* sample of size n from $\mathscr{X} \times \mathscr{Y}$ under $\mathbf{p}$; and let $\hat{p}_{i,j} = f_{i,j}/n$ be the corresponding relative frequency. Consequently, let $\hat{\mathbf{p}} = \{\hat{p}_{i,j}\}$, $\hat{\mathbf{p}}_x = \{\hat{p}_{i,\cdot}\}$ and $\hat{\mathbf{p}}_y = \{\hat{p}_{\cdot,j}\}$ as the sets of observed joint and marginal relative frequencies. The objective of interest is to estimate the mutual information MI.

Let

$$\widehat{MI} = \widehat{MI}(X,Y) = \hat{H}(X) + \hat{H}(Y) - \hat{H}(X,Y) \tag{2.0.2}$$

where $\hat{H}(X) = -\sum_i \hat{p}_{i,\cdot} \ln \hat{p}_{i,\cdot}$, $\hat{H}(Y) = -\sum_j \hat{p}_{\cdot,j} \ln \hat{p}_{\cdot,j}$, and $\hat{H}(X,Y) = -\sum_{i,j} \hat{p}_{i,j} \ln \hat{p}_{i,j}$. $\widehat{MI}$ is the so-called plugin estimator of mutual information, or maximum likelihood estimator when K is finite.

Let

$$\begin{aligned}
\widehat{MI}_z = \widehat{MI}_z(X,Y) &= \hat{H}_z(X) + \hat{H}_z(Y) - \hat{H}_z(X,Y) \\
&= \sum_{v=1}^{n-1} \frac{1}{v} \left\{ \frac{n^{v+1}[n-(v+1)]!}{n!} \sum_{i=1}^{K_1} \left[\hat{p}_{i,\cdot} \prod_{k=0}^{v-1} \left(1 - \hat{p}_{i,\cdot} - \frac{k}{n}\right) \right] \right\} \\
&\quad + \sum_{v=1}^{n-1} \frac{1}{v} \left\{ \frac{n^{v+1}[n-(v+1)]!}{n!} \sum_{j=1}^{K_2} \left[\hat{p}_{\cdot,j} \prod_{k=0}^{v-1} \left(1 - \hat{p}_{\cdot,j} - \frac{k}{n}\right) \right] \right\} \\
&\quad - \sum_{v=1}^{n-1} \frac{1}{v} \left\{ \frac{n^{v+1}[n-(v+1)]!}{n!} \sum_{i=1}^{K_1} \sum_{j=1}^{K_2} \left[\hat{p}_{i,j} \prod_{k=0}^{v-1} \left(1 - \hat{p}_{i,j} - \frac{k}{n}\right) \right] \right\}.
\end{aligned} \tag{2.0.3}$$

$\widehat{MI}_z$ is the mutual information estimator in Turing's perspective. It is showed in [16] that $\widehat{MI}$ has a power decaying bias and $\widehat{MI}_z$ has an exponentially decaying bias. The asymptotic properties for the two mutual information estimators are introduced in the following section. The demonstration is given in two parts, depending on if the underlying mutual information is zero.

2.1 When Mutual Information is Not Zero

When $MI \neq 0$, [16] provided the asymptotic normality of $\widehat{MI}$ and $\widehat{MI}_z$. Let $g(v)$ and $\Sigma(v)$ be as defined in [16], their main theoretical results are summarized into the well-proven Proposition 1 and Theorem 1 below.

Proposition 1 *Provided that* $g^\tau(v)\Sigma(v)g(v) > 0$,

$$\sqrt{n}\left(\widehat{MI} - MI\right)\left[g^\tau(v)\Sigma(v)g(v)\right]^{-\frac{1}{2}} \xrightarrow{L} N(0,1). \tag{2.1.1}$$

Theorem 1 *Provided that* $g^{\tau}(v)\Sigma(v)g(v) > 0$,

$$\sqrt{n}\left(\widehat{MI}_z - MI\right)\left[g^{\tau}(v)\Sigma(v)g(v)\right]^{-\frac{1}{2}} \xrightarrow{L} N(0,1). \tag{2.1.2}$$

The normality of Proposition 1 and Theorem 1 are useful in providing confidence intervals for $MI > 0$ or in testing $H_0 : MI = c > 0$ for any $c > 0$. It is much more often of interest in the practice to test $H_0 : MI = 0$. However, Proposition 1 and Theorem 1 cannot be used to test the hypothesis $H_0 : MI = 0$. The normality of Proposition 1 is based on a first-order delta method which requires $g^{\tau}(v)\Sigma(v)g(v) > 0$. It is demonstrated in Proposition 3 that this condition does not hold when $MI = 0$. In the following sub-section, two chi-square tests for $H_0 : MI = 0$ based on $\widehat{MI}$ and $\widehat{MI}_z$, respectively, are offered to complement what is not covered by the normality of Proposition 1 and Theorem 1.

2.2 When Mutual Information is Zero

The said tests are summarized in Proposition 2 and Theorem 2 below.

Proposition 2 *Provided that* $MI = 0$,

$$\chi_1^2 = 2n\widehat{MI} \xrightarrow{L} \chi^2\left((K_1 - 1)(K_2 - 1)\right) \tag{2.2.1}$$

Theorem 2 *Provided that* $MI = 0$,

$$\chi_2^2 = 2n\widehat{MI}_z + (K_1 - 1)(K_2 - 1) \xrightarrow{L} \chi^2\left((K_1 - 1)(K_2 - 1)\right) \tag{2.2.2}$$

The chi-square test in Proposition 2 is a well-known test, and its proof is a direct application of the log likelihood ratio test established by [17]. The proof is given in Chapter 3.

The chi-square test in Theorem 2 is the focal point of this article. The proof of Theorem 2 is non-trivial and is a part of the main results given in Chapter 3.

CHAPTER 3: Main Results

3.1 Proof of the Assumption in First Order Delta Method does not Hold when $MI = 0$

This section is to show that the assumption of the first order delta method is violated when $MI = 0$.

Proposition 3 *If $MI = 0$, then $g^{\tau}(v)\Sigma(v)g(v) = 0$.*

To prove Proposition 3, it is necessary to recall several notations in [16],

1. a re-enumeration of $\{p_{i,j}; i = 1, \cdots, K_1 \text{ and } j = 1, \cdots, K_2\}$ in the form of $\{p_k; k = 1, \cdots, K\}$, where $K = K_1 \times K_2$ (note that if $MI = 0$, $K = K_1 \times K_2$ is equivalent to $K = \sum_{i,j} 1[p_{i,j} > 0]$), and

2. a partition of the index set $\{(i,j); i = 1, \cdots, K_1 \text{ and } j = 1, \cdots, K_2\}$, denoted as $\{S_1, \cdots, S_{K_1}\}$ and $\{T_1, \cdots, T_{K_2}\}$, where

   (a) $S_s = \{k; p_k \in \{p_{s,j}; j = 1, \cdots, K_2\}\}$ for each s, $s = 1, \cdots, K_1$; and

   (b) $T_t = \{k; p_k \in \{p_{i,t}; i = 1, \cdots, K_1\}\}$ for each t, $t = 1, \cdots, K_2$.

Let $v = (p_1, \cdots, p_{K-1})^{\tau}$, $G(v) = MI = H(X) + H(Y) - H(X,Y)$ and $g(v) = \nabla G(v) = (\partial G(v)/\partial p_1, \cdots, \partial G(v)/\partial p_{K-1})^{\tau}$, it was shown in [16] that

$$\frac{\partial}{\partial p_k} G(v) = \begin{cases} \ln[(p_{K_1,\cdot})(p_{\cdot,K_2})(p_k)] - \ln[(p_{i,\cdot})(p_{\cdot,j})(p_K)], & \text{if } k \in S_i \neq S_{K_1} \text{ and } k \in T_j \neq T_{K_2} \\ \ln[(p_{\cdot,K_2})(p_k)] - \ln[(p_{\cdot,j})(p_K)], & \text{if } k \in S_{K_1} \text{ and } k \in T_j \neq T_{K_2} \\ \ln[(p_{K_1,\cdot})(p_k)] - \ln[(p_{i,\cdot})(p_K)], & \text{if } k \in S_i \neq S_{K_1} \text{ and } k \in T_{K_2} \end{cases} \tag{3.1.1}$$

where $p_K = 1 - \sum_{k \neq K} p_k$.

Proof of Proposition 3. If $\mathit{MI} = 0$ then X and Y are independent, *i.e.*, $p_{i,j} = p_{i,\cdot}p_{\cdot,j}$ for all (i,j). Consider the three cases of (3.1.1) separately. If $k \in S_i \neq S_{K_1}$ and $k \in T_j \neq T_{K_2}$, then $p_k = p_{i,\cdot}p_{\cdot,j}$ and $p_K = p_{K_1,\cdot}p_{\cdot,K_2}$, and therefore $\partial G(v)/\partial p_k = 0$. If $k \in S_{K_1}$ and $k \in T_j \neq T_{K_2}$, then $p_K = p_{K_1,\cdot}p_{\cdot,K_2}$ and $p_k = p_{\cdot,j}p_{K_1,j}$, and therefore $\partial G(v)/\partial p_k = 0$. If $k \in S_i \neq S_{K_1}$ and $k \in T_{K_2}$, then $p_K = p_{K_1,\cdot}p_{\cdot,K_2}$ and $p_k = p_{i,\cdot}p_{\cdot,K_2}$, and therefore $\partial G(v)/\partial p_k = 0$. It follows that $g(v) = \nabla G(v) = 0$ and hence $g^\tau(v)\Sigma(v)g(v) = 0$. □

3.2 Proof of the Asymptotic Distribution for $\widehat{\mathit{MI}}$ when $\mathit{MI} = 0$

Proof of Proposition 2

Consider the test $H_0 : p_{i,j} = p_{i,\cdot}p_{\cdot,j}$; $\sum p_{i,\cdot} = 1, \sum p_{\cdot,j} = 1$. For a random sample of size n, let $f_{i,\cdot}$, $f_{\cdot,j}$, and f_k be the observed frequency of the i-th marginal category of X, the j-th marginal category of Y, and the k-th joint category, respectively. The generalized likelihood-ratio is

$$\begin{aligned} L &= \frac{\sup_{\theta\in\Theta_0} L(\theta|x)}{\sup_{\theta\in\Theta} L(\theta|x)} \\ &= \frac{\frac{n!}{f_1!\cdots f_K!}\left(\hat{p}_{1,\cdot}^{f_{1,\cdot}}\cdots\hat{p}_{K_1,\cdot}^{f_{K_1,\cdot}}\right)\left(\hat{p}_{\cdot,1}^{f_{\cdot,1}}\cdots\hat{p}_{\cdot,K_2}^{f_{\cdot,K_2}}\right)}{\frac{n!}{f_1!\cdots f_K!}\hat{p}_1^{f_1}\cdots\hat{p}_K^{f_K}} \\ &= \frac{\left(\hat{p}_{1,\cdot}^{f_{1,\cdot}}\cdots\hat{p}_{K_1,\cdot}^{f_{K_1,\cdot}}\right)\left(\hat{p}_{\cdot,1}^{f_{\cdot,1}}\cdots\hat{p}_{\cdot,K_2}^{f_{\cdot,K_2}}\right)}{\hat{p}_1^{f_1}\cdots\hat{p}_K^{f_K}} \end{aligned}$$

And

$$\begin{aligned} -2\ln L &= -2\ln\frac{\left(\hat{p}_{1,\cdot}^{f_{1,\cdot}}\cdots\hat{p}_{K_1,\cdot}^{f_{K_1,\cdot}}\right)\left(\hat{p}_{\cdot,1}^{f_{\cdot,1}}\cdots\hat{p}_{\cdot,K_2}^{f_{\cdot,K_2}}\right)}{\hat{p}_1^{f_1}\cdots\hat{p}_K^{f_K}} \\ &= -2\left(\sum_i f_{i,\cdot}\ln\hat{p}_{i,\cdot} + \sum_j f_{\cdot,j}\ln\hat{p}_{\cdot,j} - \sum_{i,j} f_{i,j}\ln\hat{p}_{i,j}\right) \\ &= 2n\left(-\sum_i \hat{p}_{i,\cdot}\ln\hat{p}_{i,\cdot} - \sum_j \hat{p}_{\cdot,j}\ln\hat{p}_{\cdot,j} + \sum_{i,j}\hat{p}_{i,j}\ln\hat{p}_{i,j}\right) \\ &= 2n\widehat{\mathit{MI}} \end{aligned}$$

By [17], $-2\ln L \sim \chi^2$ with degrees of freedom $(K_1-1)(K_2-1)$. □

3.3 Proof of the Asymptotic Distribution for $\widehat{MI}_z$ when $MI = 0$

The proof of Theorem 2 needs several additional notations and lemmas. Consider a single alphabet $\mathscr{X}$ and the associated probability distribution $\mathbf{p} = \{p_k; k = 1, \cdots, K\}$. Suppose an *iid* sample of size n results in letter frequencies $\{Y_k; k \geq 1\}$ and relative frequencies $\hat{\mathbf{p}} = \{\hat{p}_k; k \geq 1\}$. Let $\hat{H} = H(\hat{\mathbf{p}})$ and

$$\hat{H}_z = \hat{H}_z(\hat{\mathbf{p}}) = \sum_{v=1}^{n-1} \frac{1}{v} Z_v \tag{3.3.1}$$

where

$$Z_v = \sum_{k=1}^{\infty} \left[\hat{p}_k \prod_{j=0}^{v-1} \left(1 - \hat{p}_k - \frac{j}{n} \right) \right]. \tag{3.3.2}$$

For clarity of the presentation, the proof of Theorem 1 is given with several lemmas.

Lemma 1 *Given any iid sample of size $n > 1$ and $\hat{K} > 1$,*

$$\hat{H}_z > \hat{H}_{MM}. \tag{3.3.3}$$

To prove Lemma 1, noting that rewriting $\hat{H}_z$ of (3.3.1) in several algebraic steps gives

$$\hat{H}_z = \sum_{k \geq 1} \hat{p}_k \left\{ \frac{(n-Y_k)!}{(n-1)!} \sum_{v=1}^{n-Y_k} \left[\frac{1}{v} \frac{(n-1-v)!}{(n-Y_k-v)!} \right] \right\}. \tag{3.3.4}$$

By (1.3.2) and (3.3.4), Lemma 1 is equivalent to

$$\begin{aligned} \sum_{k \geq 1} \hat{p}_k \left\{ \frac{(n-Y_k)!}{(n-1)!} \sum_{v=1}^{n-Y_k} \left[\frac{1}{v} \frac{(n-1-v)!}{(n-Y_k-v)!} \right] \right\} &\geq -\sum_{k \geq 1} \hat{p}_k \ln \hat{p}_k + \frac{\hat{K}-1}{2n} \\ &= \sum_{k \geq 1} \left[-\hat{p}_k \ln \hat{p}_k + \frac{1}{2n} - \frac{\hat{p}_k}{2n} \right] \\ &= \sum_{k \geq 1} \left\{ \hat{p}_k \left[-\ln \hat{p}_k + \frac{1}{2n\hat{p}_k} - \frac{1}{2n} \right] \right\} \end{aligned}$$

$$= \sum_{k\geq 1} \left\{ \hat{p}_k \left[-\ln \frac{Y_k}{n} + \frac{1}{2Y_k} - \frac{1}{2n} \right] \right\}$$

Lemma 1 may be proved component-wisely for each k in the summation, hence Lemma 2.

Lemma 2 *For any positive integer $y \in \{1, 2, \cdots, n-1\}$,*

$$\frac{(n-y)!}{(n-1)!} \sum_{v=1}^{n-y} \left[\frac{1}{v} \frac{(n-1-v)!}{(n-y-v)!} \right] > -\ln \frac{y}{n} + \frac{1}{2y} - \frac{1}{2n}. \tag{3.3.5}$$

Let the left-hand-side and the right-hand-side of (3.3.5) be denoted as $L(y)$ and $R(y)$ respectively.

Lemma 3 *For any positive integer $y \in \{1, 2, \cdots, n-1\}$,*

1. *$L(y) - L(y+1) = 1/y$, and*
2. *$R(y) - R(y+1) = \ln(1 + 1/y) + 1/(2y(y+1))$.*

Lemma 4 (Hockey-Stick identity) *For any pair of positive integers, n and r, $r \leq n$,*

$$\sum_{j=r}^{n} \binom{j}{r} = \binom{n+1}{r+1}.$$

Proof of Lemma 3.

For part 1, by (3.3.5),

$$\begin{aligned} L(y) &= \frac{(n-y)!}{(n-1)!} \left[\sum_{v=1}^{n-y-1} \frac{1}{v} \frac{(n-v-1)!}{(n-y-v)!} + \frac{(y-1)!}{(n-y)} \right] \\ &= \frac{(n-y)!}{(n-1)!} \left[\sum_{v=1}^{n-y-1} \frac{1}{v} \frac{1}{n-y-v} \frac{(n-v-1)!}{(n-y-v-1)!} + \frac{(y-1)!}{(n-y)} \right] \\ &= \frac{(n-y)!}{(n-1)!} \left\{ \sum_{v=1}^{n-y-1} \frac{1}{n-y} \left[\frac{1}{v} + \frac{1}{(n-y-v)} \right] \frac{(n-v-1)!}{(n-y-v-1)!} + \frac{(y-1)!}{(n-y)} \right\} \end{aligned}$$

$$
\begin{aligned}
&= \frac{(n-y-1)!}{(n-1)!}\left\{\sum_{v=1}^{n-y-1}\left[\frac{1}{v}+\frac{1}{(n-y-v)}\right]\frac{(n-v-1)!}{(n-y-v-1)!}+(y-1)!\right\}\\
&= \frac{(n-y-1)!}{(n-1)!}\left\{\sum_{v=1}^{n-y-1}\frac{1}{v}\frac{(n-v-1)!}{(n-y-v-1)!}+\sum_{v=1}^{n-y-1}\frac{(n-v-1)!}{(n-y-v)!}+(y-1)!\right\}\\
&= L(y+1)+\frac{(n-y-1)!}{(n-1)!}\sum_{v=1}^{n-y}\frac{(n-v-1)!}{(n-v-y)!}\\
&= L(y+1)+\frac{(n-y-1)!(y-1)!}{(n-1)!}\sum_{v=1}^{n-y}\frac{(n-v-1)!}{(n-v-y)!(y-1)!}\\
&= L(y+1)+\frac{1}{y}\frac{1}{\binom{n-1}{y}}\sum_{v=1}^{n-y}\binom{n-v-1}{y-1}.
\end{aligned}
$$

By Lemma 4,

$$
L(y)-L(y+1)=\frac{1}{y}\binom{n-1}{y}^{-1}\binom{n-1}{y}=\frac{1}{y}.
$$

For part 2,

$$
\begin{aligned}
R(y)-R(y+1) &= \left(-\ln\frac{y}{n}+\frac{1}{2y}-\frac{1}{2n}\right)-\left(-\ln\frac{y+1}{n}+\frac{1}{2(y+1)}-\frac{1}{2n}\right)\\
&= \ln\frac{y+1}{y}+\left(\frac{1}{2y}-\frac{1}{2(y+1)}\right)\\
&= \ln\left(1+\frac{1}{y}\right)+\frac{1}{2y(y+1)}
\end{aligned}
$$

□

Proof of Lemma 2. By Lemma 3,

$$
(L(y)-R(y))-(L(y+1)-R(y+1))=\frac{1}{y}-\ln\left(1+\frac{1}{y}\right)-\frac{1}{2y(y+1)}=:f(y). \qquad (3.3.6)
$$

Note that

$$
\begin{aligned}
f'(y) &= -\frac{1}{y^2}-\frac{1}{1+1/y}\cdot\frac{(-1)}{y^2}-\frac{1}{2}\cdot\frac{(-1)}{y^2(y+1)^2}\cdot(2y+1)\\
&= -\frac{1}{y^2}+\frac{1}{y(y+1)}+\frac{2y+1}{2y^2(y+1)^2}\\
&= \frac{-2(y+1)^2+2y(y+1)+(2y+1)}{2y^2(y+1)^2}
\end{aligned}
$$

$$= \frac{-2(y^2+2y+1)+2y^2+2y+2y+1}{2y^2(y+1)^2} = \frac{-1}{2y^2(y+1)^2} < 0$$

Therefore $f(y)$ is a decreasing function of y on $\{1, 2, \cdots, n-1\}$. Since $f(1) = 1 - \ln 2 - 1/4 > 0$ and $\lim_{y\to\infty} f(y) = 0$, we have $f(y) > 0$ for each $y \in \{1, 2, \cdots, n-1\}$. Hence $L(y) - R(y)$ is strictly positive, *i.e.*, $L(y) > R(y)$, for any y on $\{1, 2, \cdots, n-1\}$.

□

Proof of Lemma 1. By Lemma 2,

$$\hat{H}_z - \hat{H}_{MM} = \sum_{k\geq 1} \hat{p}_k (L(Y_k) - R(Y_k)) = \sum_{k\geq 1} \hat{p}_k 1[Y_k \in \{1, \cdots, n-1\}](L(Y_k) - R(Y_k)) > 0.$$

□

Lemma 5 *For any distribution with* $1 < K < \infty$*,*

$$2n(\hat{H}_z - \hat{H}) \xrightarrow{p} K - 1.$$

Lemma 6 (Markov's inequality) *For any* $X \geq 0$ *and constant* $a > 0$*,*

$$\mathrm{P}(X \geq a) \leq \frac{\mathrm{E}(X)}{a}.$$

Proof of Lemma 5. Because of (1.4.2) and (1.3.1), we have

$$\lim_{n\to\infty} \left\{ 2n\, \mathrm{E}\left[\hat{H}_z - \hat{H}\right] \right\} = K - 1. \tag{3.3.7}$$

By Lemma 1 and (1.3.2),

$$\begin{aligned} \hat{H}_z &> \hat{H}_{MM} \\ &= \hat{H} + \frac{\hat{K} - 1}{2n} \end{aligned}$$

Thus

$$2n(\hat{H}_z - \hat{H}) > \hat{K} - 1,$$

and

$$2n(\hat{H}_z - \hat{H}) - (\hat{K} - 1) \geq 0 \quad \text{for any } n > 1.$$

Therefore, for any $\epsilon > 0$,

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{P}\left(2n(\hat{H}_z - \hat{H}) - (\hat{K} - 1) \geq \epsilon\right) &\leq \lim_{n\to\infty} \frac{\mathrm{E}\left[2n(\hat{H}_z - \hat{H}) - (\hat{K} - 1)\right]}{\epsilon} \\
&= \frac{\lim_{n\to\infty} \mathrm{E}\left[2n(\hat{H}_z - \hat{H})\right] - \lim_{n\to\infty} \mathrm{E}\left[\hat{K} - 1\right]}{\epsilon} \\
&= \frac{(K-1) - (K-1)}{\epsilon} = 0.
\end{aligned}$$

Hence

$$2n(\hat{H}_z - \hat{H}) - (\hat{K} - 1) \xrightarrow{p} 0. \tag{3.3.8}$$

Since

$$\hat{K} - 1 \xrightarrow{a.s.} K - 1, \tag{3.3.9}$$

by adding (3.3.8) and (3.3.9), we have

$$2n(\hat{H}_z - \hat{H}) \xrightarrow{p} K - 1.$$

□

Proof of Theorem 2. By Proposition 2 and noting that

$$\begin{aligned}
2n\widehat{MI} &= 2n\left(\hat{H}(X) + \hat{H}(Y) - \hat{H}(X,Y)\right) \\
&= 2n\left[-(\hat{H}_z(X) - \hat{H}_X) - (\hat{H}_z(Y) - \hat{H}(Y)) + (\hat{H}_z(X,Y) - \hat{H}(X,Y))\right] + 2n\widehat{MI}_z,
\end{aligned}$$

and applying Lemma 5, it follows that

$$2n\widehat{MI}_z + (K_1 - 1)(K_2 - 1) \sim 2n\widehat{MI} \xrightarrow{L} \chi^2((K_1 - 1)(K_2 - 1)).$$

□

3.4 Examples

[16] gave three examples involving evaluation of gene-to-gene association. These examples are re-worked to demonstrate the usage of the proposed new test. Three genes were under consideration, and they were coded as TMEM30A, MTCH2, and ENAH. In Example 1, readings of TMEM30A and MTCH2 were analyzed. In Examples 2 and 3, readings on two different probes, designed for the same gene ENAH, on the same microchip were analyzed.

Example 1 TMEM30A and MTCH2. Using the results of [16], $\widehat{MI} = 0.1459$ and $\widehat{MI}_z = 0.0552$, $\chi_1^2 = 2n\widehat{MI} = 55.7338$ and $\chi_2^2 = 2n\widehat{MI}_z + (K_1 - 1)(K_2 - 1) = 102.0864$. With degrees of freedom 81, the respective p-values are 0.9856 and 0.0567. At $\alpha = 0.05$, neither test rejects $H_0 : MI = 0$.

Example 2 ENAH and ENAH with $K = K_1 \times K_2$. Using the results of [16], $\widehat{MI} = 0.2060$ and $\widehat{MI}_z = 0.1157$, $\chi_1^2 = 2n\widehat{MI} = 78.692$ and $\chi_2^2 = 2n\widehat{MI}_z + (K_1 - 1)(K_2 - 1) = 125.1974$. With degrees of freedom 81, the respective p-values are 0.5519 and 0.0012. χ_2^2 detects an association with strong evidence and χ_1^2 fails to do so by far. Since in this case it is known a priori that an association exists, this example illustrates the added utility of χ_2^2.

Example 3 ENAH and ENAH with $K \le K_1 \times K_2$. In this example, [16] assumes that several cells in the join alphabet are associated with zero probabilities. This assumption is invalid since, if it were the case, then under the null hypothesis of $H_0 : MI = 0$ (X and Y are independent) either some of the marginal probabilities of X or Y would have to be zeros. However every (of the ten) marginal categories is covered by observations, that is to say, none of the marginal probabilities can be zero. Without the assumption of zero probabilities, the chi-square tests of Proposition 2 and Theorem 2 give identical results as in Example 2.

CHAPTER 4: Simulation Study

To further explore and study the property of $\widehat{MI}_z$, various tests of independence were compared in the simulation study to evaluate the size and power of the tests under different sample sizes.

Particularly, five tests of independence were compared:

1. Pearson Chi-square Test:

   Test Statistic:

$$\chi^2_{Pearson} = \sum_{k=1}^{K^*} \frac{(O_k - E_k)^2}{E_k} \sim \chi^2\left((K_1^* - 1)(K_2^* - 1)\right). \tag{4.0.1}$$

   Calculating $\chi^2_{Pearson}$ requires all the denominators E_k's to be positive, where each E_k is calculated by n multiplies the corresponding $\hat{p}_{i,\cdot}\hat{p}_{\cdot,j}$. When the sample size is not sufficiently large, many $\hat{p}_{i,\cdot}$'s and $\hat{p}_{\cdot,j}$'s could be zero, and it makes Pearson's test invalid. Moreover, it is commonly suggested that each cell should have at least five observations to use Pearson's test. In order to fairly compare other tests with Pearson's test under each situation, all the original samples in the simulation were adjusted. Particularly, in each bivariate sample data frequency table, all rows (and columns) with total frequencies less than 5 were combined to the row (and column) with the least total frequency among the rows (and columns) with at least five total frequency. If there are two or more such rows (or columns), one is selected randomly. For example, suppose the sample data frequency table is Table 4.1. The frequencies of $X = 3$ and 4 are less than 5, and the frequency of $X = 5$ is the category with the least frequency that is at least 5. As a result, the adjusted sample regarding the frequencies of X is described in Table 4.2. For the frequency of Y, the frequencies of $Y = 2$ and 4 are less than 5, and the frequencies of $Y = 1$ and 3 are the same. In the example, $Y = 1$ is randomly selected to be combined with low-frequency categories.

And the adjusted sample is as described in Table 4.3. Note that after the adjustment, the cardinality of X, Y, and $X \times Y$ are reduced from 5, 4, and 20 to 3, 2, and 6. Because of the possible adjustment, K^*, K_1^*, and K_2^* are used instead of K, K_1, and K_2 in (4.0.1). In the following simulation settings, the cardinality of X and Y are 10 and 15, and sample sizes are started from 100. Therefore it is guaranteed that there is always at least one category in X and Y with a frequency of 5 or more.

Table 4.1: Original Sample Frequency Table

		Y			
		1	2	3	4
X	1	4	0	6	0
	2	1	1	3	1
	3	2	2	0	0
	4	0	0	0	0
	5	3	0	1	1

Table 4.2: Partially Adjusted Sample (X categories with low frequencies combined)

		Y			
		1	2	3	4
X	1	4	0	6	0
	2	1	1	3	1
	5	5	2	1	1

Table 4.3: Adjusted Sample (X and Y categories with low frequencies combined)

		Y	
		1	3
X	1	4	6
	2	3	3
	5	8	1

2. Test of independence using $\widehat{MI}$ and Proposition 2:

   Test Statistic:

$$\chi^2_{MIhat} = 2n\widehat{MI}(\text{original sample}) \sim \chi^2\left((K_1 - 1)(K_2 - 1)\right),$$

and

$$\chi^{2*}_{\widehat{MI}hat} = 2n\widehat{MI}(\text{adjusted sample}) \sim \chi^2\left((K_1^* - 1)(K_2^* - 1)\right).$$

Two tests of independence using $\widehat{MI}$ are examined. $\widehat{MI}$ does not require non-zero marginal sample probabilities as that of Pearson's test. To make a complete comparison, tests using $\widehat{MI}$ on both original sample and adjusted sample are included.

3. Test of independence using $\widehat{MI}_z$ [1] and Theorem 2:

   Test Statistic:

$$\chi^2_{\widehat{MI}z} = 2n\widehat{MI}_z(\text{original sample}) + (K_1 - 1)(K_2 - 1) \sim \chi^2\left((K_1 - 1)(K_2 - 1)\right),$$

and

$$\chi^{2*}_{\widehat{MI}z} = 2n\widehat{MI}_z(\text{adjusted sample}) + (K_1^* - 1)(K_2^* - 1) \sim \chi^2\left((K_1^* - 1)(K_2^* - 1)\right).$$

For the same reason (to make a complete comparison), tests using $\widehat{MI}_z$ on both original sample and adjusted sample are included.

4.1 Simulation Settings

When evaluating the size of tests, the two marginal distributions and the joint distribution are

$$p_{i,\cdot} = \frac{1}{10}, \qquad i = 1, 2, \cdots, 10;$$
$$p_{\cdot,j} = \frac{16 - j}{120}, \qquad j = 1, 2, \cdots, 15;$$

and

$$p_{i,j} = p_{i,\cdot}p_{\cdot,j}.$$

When evaluating the power of tests,

$$p_{i,j} = \frac{150 - 15(i-1) - j + 1}{11325} \qquad \text{(triangle distribution)}.$$

[1] The computation of $\widehat{MI}_z$ has been implemented in R and can be found in R package "Entropy-Estimation".

For both evaluation, the simulations were conducted with sample size $n = 100$, 200, 300, ..., 14900, 15000. And for each sample size, the simulation was iterated for 50000 times.

4.2 Simulation Results

The simulation results are presented in Figure 4.1 and 4.2. To help understand the legend, pearson.reject_vec, miz.reject_vec, miz.reject_adj_vec, mihat.reject_vec, and mihat.reject_adj_vec stand for testing using $\chi^2_{Pearson}$, χ^2_{MIz}, χ^{2*}_{MIz}, χ^2_{MIhat}, and χ^{2*}_{MIhat}, respectively.

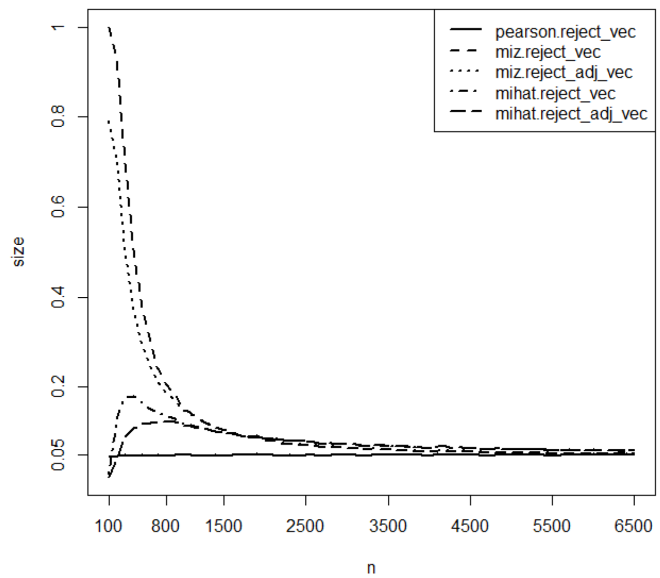

Figure 4.1: Size of different tests under different sample size when $\alpha = 0.05$.

The sizes of all tests reached the neighborhood of α when n is more than 4500; therefore Figure 4.1 did not include the simulation results when n is more than 6500. Based on Figure 4.1, Pearson's test on adjusted (combined) samples converged to α faster than other tests. The two sets of tests using estimators of mutual information converged to α at a similar rate. Although $\widehat{MI}_z$ has a smaller bias over $\widehat{MI}$, surprisingly that the simulation showed that testing independence using $\widehat{MI}$ has better size than that of using $\widehat{MI}_z$ when the sample size is relatively small.

The powers of all tests using mutual information estimators are consistently higher than

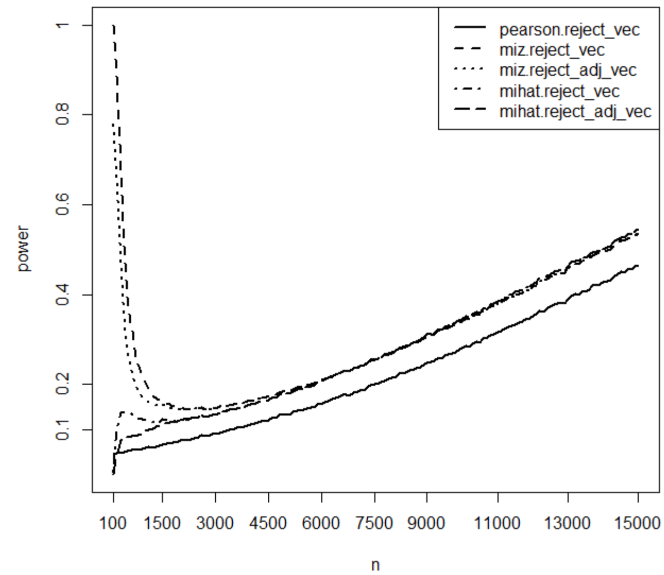

Figure 4.2: Power of different tests under different sample size when $\alpha = 0.05$ under a joint triangle distribution.

the power of Pearson's test. It suggests that when the sample size is sufficiently large, a test using either $\widehat{MI}$ or $\widehat{MI}_z$ should be adopted instead of Pearson's test. Moreover, when the sample size is moderate (n from 1500 to 4500 in the designed simulation), testing using $\widehat{MI}_z$ and $\widehat{MI}$ have similar size, but the power of using $\widehat{MI}_z$ is higher.

CHAPTER 5: Conclusion and Future Work

In conclusion, the asymptotic distributions for $\widehat{MI}$ and $\widehat{MI}_z$ are offered under the situation that $MI = 0$. Based on the simulation study, when the sample size is sufficiently large, testing independence using mutual information estimators are preferred because they have higher powers than Pearson's test. When the sample size is moderate, Pearson test's size has a faster convergence rate to α and is preferred. When the sample size is small, although Pearson's test also has a faster convergence rate of size, it is frequently incalculable without a merge of empty cells.

$\widehat{MI}_z$ is known to have a faster-decaying bias compared to $\widehat{MI}$, whereas the size of the test of independence with $\widehat{MI}$ has a faster converging rate when the sample size is relatively small. It leads to my conjecture that additional bias correction terms are needed in Theorem 2 to improve its performance under small samples, which will be future work. Furthermore, it is worthy of investigating why Pearson's test has such a fast converging rate in the size of test. Finally, Theorem 2 could be generated as a new test of goodness-of-fit using Kullback-Leibler divergence $\widehat{KL}_z$ [18], one could study its property and compare it with other goodness-of-fit tests.

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] B. Harris, "The statistical estimation of entropy in the non-parametric case," tech. rep., WISCONSIN UNIV-MADISON MATHEMATICS RESEARCH CENTER, 1975.

[3] G. A. Miller and W. G. Madow, *On the maximum likelihood estimate of the Shannon-Weiner measure of information*. Operational Applications Laboratory, Air Force Cambridge Research Center, 1954.

[4] S. Zahl, "Jackknifing an index of diversity," *Ecology*, vol. 58, no. 4, pp. 907–913, 1977.

[5] L. Paninski, "Estimation of entropy and mutual information," *Neural computation*, vol. 15, no. 6, pp. 1191–1253, 2003.

[6] Z. Zhang and J. Zhou, "Re-parameterization of multinomial distributions and diversity indices," *Journal of Statistical Planning and Inference*, vol. 140, no. 7, pp. 1731–1738, 2010.

[7] Z. Zhang, "Entropy estimation in turing's perspective," *Neural computation*, vol. 24, no. 5, pp. 1368–1389, 2012.

[8] Z. Zhang and M. Grabchak, "Entropic representation and estimation of diversity indices," *Journal of Nonparametric Statistics*, vol. 28, no. 3, pp. 563–575, 2016.

[9] Z. Zhang, C. Chen, and J. Zhang, "Estimation of population size in entropic perspective," *Communications in Statistics-Theory and Methods*, pp. 1–18, 2018.

[10] Z. Zhang, "Asymptotic normality of an entropy estimator with exponentially decaying bias," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 504–508, 2013.

[11] Z. Zhang and X. Zhang, "A normal law for the plug-in estimator of entropy," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2745–2747, 2012.

[12] X. Zhang, "Asymptotic normality of entropy estimators," 2013.

[13] C. Chen, M. Grabchak, A. Stewart, J. Zhang, and Z. Zhang, "Normal laws for two entropy estimators on infinite alphabets," *Entropy*, vol. 20, no. 5, p. 371, 2018.

[14] Z. Zhang, *Statistical Implications of Turing's Formula*. John Wiley & Sons, 2016.

[15] A. Antos and I. Kontoyiannis, "Convergence properties of functional estimates for discrete distributions," *Random Structures & Algorithms*, vol. 19, no. 3-4, pp. 163–193, 2001.

[16] Z. Zhang and L. Zheng, "A mutual information estimator with exponentially decaying bias," *Statistical applications in genetics and molecular biology*, vol. 14, no. 3, pp. 243–252, 2015.

[17] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, 1938.

[18] Z. Zhang and M. Grabchak, “Nonparametric estimation of küllback-leibler divergence,” *Neural computation*, vol. 26, no. 11, pp. 2570–2593, 2014.

APPENDIX A: Simulation Data

Table A.1: Simulation results: Size of tests, Part 1.

n	pearson.reject	miz.reject	miz.reject_adj	mihat.reject	mihat.reject_adj
100	0.04384	0.99914	0.79116	0.00612	0.00058
200	0.04648	0.91952	0.6986	0.13542	0.04492
300	0.04906	0.70278	0.51052	0.17822	0.08984
400	0.04774	0.50598	0.37642	0.17896	0.10908
500	0.04864	0.37804	0.29606	0.16544	0.11722
600	0.04688	0.29688	0.24558	0.15224	0.11928
700	0.04888	0.24028	0.20914	0.14158	0.12032
800	0.0482	0.20634	0.18756	0.13526	0.12204
900	0.04848	0.18076	0.16932	0.12986	0.1216
1000	0.04902	0.1545	0.14854	0.11876	0.1144
1100	0.0501	0.14218	0.1389	0.1163	0.11372
1200	0.04762	0.12826	0.12638	0.1098	0.10818
1300	0.0481	0.11858	0.1176	0.10566	0.10504
1400	0.04924	0.10998	0.10926	0.1018	0.10138
1500	0.04712	0.10362	0.10322	0.09792	0.0977
1600	0.04864	0.09862	0.09854	0.09652	0.09646
1700	0.04844	0.09258	0.09256	0.09206	0.09202
1800	0.05022	0.08828	0.08824	0.09012	0.09006
1900	0.05092	0.08764	0.0876	0.09084	0.09082
2000	0.0484	0.08288	0.08286	0.0857	0.0857
2100	0.04986	0.08014	0.08014	0.08508	0.08508
2200	0.04792	0.07618	0.07618	0.08214	0.08214
2300	0.04896	0.07358	0.07358	0.08026	0.08024
2400	0.04808	0.07198	0.07198	0.0797	0.0797
2500	0.04972	0.07146	0.07146	0.07934	0.07934
2600	0.04936	0.06972	0.06972	0.07758	0.07758
2700	0.04804	0.06666	0.06666	0.07406	0.07406
2800	0.0503	0.06768	0.06768	0.0761	0.0761
2900	0.0486	0.06546	0.06546	0.0732	0.0732
3000	0.04856	0.06422	0.06422	0.07284	0.07284
3100	0.04916	0.06366	0.06366	0.07236	0.07236
3200	0.0478	0.06182	0.06182	0.06946	0.06946
3300	0.05008	0.06308	0.06308	0.07168	0.07168
3400	0.04968	0.0611	0.0611	0.07038	0.07038
3500	0.04976	0.0606	0.0606	0.06914	0.06914
3600	0.05002	0.05894	0.05894	0.06754	0.06754
3700	0.04964	0.05912	0.05912	0.0677	0.0677
3800	0.04882	0.05862	0.05862	0.06694	0.06694
3900	0.04856	0.05756	0.05756	0.06632	0.06632
4000	0.0471	0.05584	0.05584	0.06376	0.06376
4100	0.05182	0.0596	0.0596	0.06846	0.06846
4200	0.05044	0.059	0.059	0.06726	0.06726
4300	0.04946	0.05688	0.05688	0.06458	0.06458
4400	0.05026	0.05794	0.05794	0.06618	0.06618
4500	0.04994	0.05638	0.05638	0.06442	0.06442
4600	0.0501	0.0556	0.0556	0.06404	0.06404
4700	0.04734	0.05336	0.05336	0.06052	0.06052
4800	0.0496	0.05526	0.05526	0.06332	0.06332
4900	0.04828	0.05366	0.05366	0.0614	0.0614
5000	0.04972	0.0539	0.0539	0.06208	0.06208

Table A.2: Simulation results: Size of tests, Part 2.

n	pearson.reject	miz.reject	miz.reject_adj	mihat.reject	mihat.reject_adj
5100	0.04998	0.05398	0.05398	0.06184	0.06184
5200	0.0498	0.05458	0.05458	0.06156	0.06156
5300	0.05108	0.0547	0.0547	0.06192	0.06192
5400	0.04998	0.05378	0.05378	0.06038	0.06038
5500	0.04812	0.05176	0.05176	0.0591	0.0591
5600	0.04936	0.05266	0.05266	0.05996	0.05996
5700	0.0503	0.05292	0.05292	0.06052	0.06052
5800	0.04834	0.0519	0.0519	0.05922	0.05922
5900	0.05014	0.05204	0.05204	0.05958	0.05958
6000	0.05098	0.05394	0.05394	0.06078	0.06078
6100	0.0492	0.0526	0.0526	0.05962	0.05962
6200	0.04954	0.05222	0.05222	0.05916	0.05916
6300	0.05036	0.053	0.053	0.0592	0.0592
6400	0.05076	0.05338	0.05338	0.06066	0.06066
6500	0.04864	0.05148	0.05148	0.05782	0.05782
6600	0.04878	0.05232	0.05232	0.05876	0.05876
6700	0.05076	0.05282	0.05282	0.05926	0.05926
6800	0.049	0.05202	0.05202	0.05796	0.05796
6900	0.04936	0.052	0.052	0.05828	0.05828
7000	0.04952	0.05156	0.05156	0.0579	0.0579
7100	0.0498	0.0519	0.0519	0.05782	0.05782
7200	0.04922	0.0507	0.0507	0.05634	0.05634
7300	0.04846	0.05026	0.05026	0.0554	0.0554
7400	0.04966	0.05168	0.05168	0.0576	0.0576
7500	0.04968	0.0515	0.0515	0.05702	0.05702
7600	0.05028	0.05202	0.05202	0.0578	0.0578
7700	0.05158	0.05278	0.05278	0.05876	0.05876
7800	0.05236	0.05464	0.05464	0.05976	0.05976
7900	0.05016	0.05234	0.05234	0.05748	0.05748
8000	0.04966	0.05166	0.05166	0.05672	0.05672
8100	0.05088	0.0525	0.0525	0.05786	0.05786
8200	0.0502	0.05158	0.05158	0.0569	0.0569
8300	0.04984	0.05098	0.05098	0.05604	0.05604
8400	0.05146	0.05262	0.05262	0.0578	0.0578
8500	0.04924	0.05146	0.05146	0.05608	0.05608
8600	0.05052	0.05296	0.05296	0.05788	0.05788
8700	0.05048	0.05144	0.05144	0.05628	0.05628
8800	0.04946	0.0506	0.0506	0.0556	0.0556
8900	0.05082	0.05246	0.05246	0.05642	0.05642
9000	0.0507	0.05176	0.05176	0.05664	0.05664
9100	0.0482	0.05022	0.05022	0.05478	0.05478
9200	0.05088	0.05246	0.05246	0.05672	0.05672
9300	0.05038	0.05186	0.05186	0.0565	0.0565
9400	0.04934	0.05064	0.05064	0.05536	0.05536
9500	0.04886	0.04936	0.04936	0.05382	0.05382
9600	0.04912	0.0506	0.0506	0.05494	0.05494
9700	0.05002	0.051	0.051	0.05524	0.05524
9800	0.04858	0.05012	0.05012	0.05404	0.05404
9900	0.05006	0.05192	0.05192	0.0564	0.0564
10000	0.0501	0.05056	0.05056	0.05554	0.05554

Table A.3: Simulation results: Size of tests, Part 3.

n	pearson.reject	miz.reject	miz.reject_adj	mihat.reject	mihat.reject_adj
10100	0.04988	0.05152	0.05152	0.05574	0.05574
10200	0.04942	0.05088	0.05088	0.0544	0.0544
10300	0.05058	0.05164	0.05164	0.05562	0.05562
10400	0.0493	0.04994	0.04994	0.05408	0.05408
10500	0.05106	0.05252	0.05252	0.0562	0.0562
10600	0.04914	0.04996	0.04996	0.05442	0.05442
10700	0.04992	0.05128	0.05128	0.05484	0.05484
10800	0.0489	0.04956	0.04956	0.05348	0.05348
10900	0.05018	0.05074	0.05074	0.05432	0.05432
11000	0.04996	0.05122	0.05122	0.05488	0.05488
11100	0.04936	0.04992	0.04992	0.05356	0.05356
11200	0.04944	0.05028	0.05028	0.05396	0.05396
11300	0.05016	0.05072	0.05072	0.05402	0.05402
11400	0.05218	0.05266	0.05266	0.05628	0.05628
11500	0.0505	0.05096	0.05096	0.05466	0.05466
11600	0.04972	0.05024	0.05024	0.054	0.054
11700	0.0502	0.0515	0.0515	0.05482	0.05482
11800	0.05038	0.0512	0.0512	0.05526	0.05526
11900	0.05014	0.0508	0.0508	0.05492	0.05492
12000	0.05018	0.05038	0.05038	0.05366	0.05366
12100	0.04794	0.04906	0.04906	0.05246	0.05246
12200	0.05152	0.05196	0.05196	0.05522	0.05522
12300	0.04848	0.04932	0.04932	0.05256	0.05256
12400	0.04926	0.05004	0.05004	0.05354	0.05354
12500	0.0487	0.0498	0.0498	0.05318	0.05318
12600	0.04998	0.0511	0.0511	0.05438	0.05438
12700	0.0502	0.0522	0.0522	0.0551	0.0551
12800	0.04946	0.05012	0.05012	0.05296	0.05296
12900	0.04882	0.05024	0.05024	0.05334	0.05334
13000	0.05038	0.0504	0.0504	0.05348	0.05348
13100	0.05002	0.05004	0.05004	0.05344	0.05344
13200	0.05052	0.05144	0.05144	0.0545	0.0545
13300	0.04916	0.05012	0.05012	0.0528	0.0528
13400	0.0491	0.05014	0.05014	0.05316	0.05316
13500	0.0494	0.04974	0.04974	0.05286	0.05286
13600	0.05046	0.05148	0.05148	0.05422	0.05422
13700	0.0502	0.0506	0.0506	0.05324	0.05324
13800	0.0503	0.05064	0.05064	0.05368	0.05368
13900	0.0493	0.04988	0.04988	0.05288	0.05288
14000	0.04962	0.05094	0.05094	0.0539	0.0539
14100	0.04966	0.04998	0.04998	0.0528	0.0528
14200	0.05048	0.05128	0.05128	0.05424	0.05424
14300	0.04922	0.05052	0.05052	0.05332	0.05332
14400	0.05086	0.0515	0.0515	0.05386	0.05386
14500	0.05056	0.05112	0.05112	0.05432	0.05432
14600	0.0501	0.05092	0.05092	0.05402	0.05402
14700	0.05	0.051	0.051	0.05386	0.05386
14800	0.0494	0.04934	0.04934	0.05234	0.05234
14900	0.0501	0.05082	0.05082	0.05388	0.05388
15000	0.0487	0.04936	0.04936	0.05206	0.05206

Table A.4: Simulation results: Power of tests, Part 1.

n	pearson.reject	miz.reject	miz.reject_adj	mihat.reject	mihat.reject_adj
100	0.04432	0.99876	0.77844	0.00346	0.00064
200	0.04618	0.90526	0.67478	0.09632	0.0392
300	0.04722	0.69538	0.48794	0.13604	0.07472
400	0.05044	0.51818	0.35972	0.13992	0.08392
500	0.05072	0.40226	0.28046	0.13408	0.08324
600	0.05348	0.32888	0.23512	0.13386	0.0853
700	0.05318	0.27872	0.20352	0.12822	0.08572
800	0.05422	0.24562	0.18536	0.1254	0.08712
900	0.05638	0.2192	0.17258	0.12228	0.09194
1000	0.0588	0.20584	0.16688	0.1213	0.09592
1100	0.06034	0.19124	0.16136	0.1191	0.09854
1200	0.05896	0.17786	0.15468	0.11662	0.09968
1300	0.06204	0.1707	0.1533	0.11802	0.1055
1400	0.06294	0.16498	0.15192	0.11662	0.10676
1500	0.06684	0.16528	0.15592	0.12178	0.11412
1600	0.06784	0.15832	0.1504	0.11996	0.1143
1700	0.06882	0.15482	0.14936	0.1198	0.11586
1800	0.0702	0.15046	0.14628	0.11876	0.11548
1900	0.07156	0.15006	0.147	0.12008	0.1178
2000	0.0752	0.14718	0.14512	0.12108	0.11946
2100	0.07476	0.14724	0.14578	0.12268	0.12134
2200	0.07522	0.14512	0.1437	0.1216	0.12058
2300	0.0788	0.14598	0.1453	0.1246	0.12398
2400	0.08074	0.14702	0.14636	0.12568	0.12524
2500	0.08232	0.14734	0.14688	0.12882	0.12842
2600	0.08636	0.14864	0.14812	0.13178	0.13146
2700	0.0847	0.14486	0.14456	0.12822	0.12804
2800	0.0858	0.14378	0.1437	0.12894	0.12888
2900	0.08898	0.14774	0.14764	0.13334	0.13328
3000	0.0893	0.14566	0.14558	0.13218	0.13214
3100	0.09244	0.15106	0.151	0.13766	0.1376
3200	0.09524	0.15164	0.15162	0.13924	0.13922
3300	0.09718	0.1515	0.15142	0.14044	0.14038
3400	0.09842	0.15334	0.15334	0.14302	0.14302
3500	0.09948	0.1552	0.15514	0.14486	0.1448
3600	0.10306	0.15762	0.15762	0.1462	0.1462
3700	0.1043	0.15806	0.15806	0.14808	0.14808
3800	0.10578	0.15806	0.15806	0.14876	0.14876
3900	0.10806	0.16218	0.16218	0.15318	0.15318
4000	0.1109	0.16314	0.16314	0.1547	0.1547
4100	0.11222	0.16456	0.16456	0.15654	0.15654
4200	0.11614	0.16824	0.16824	0.16006	0.16006
4300	0.11778	0.17028	0.17028	0.16212	0.16212
4400	0.11822	0.17046	0.17046	0.16326	0.16326
4500	0.1202	0.17366	0.17366	0.1657	0.16568
4600	0.12406	0.1748	0.1748	0.17	0.17
4700	0.12402	0.17668	0.17668	0.17	0.17
4800	0.13118	0.1824	0.1824	0.17668	0.17668
4900	0.13122	0.1834	0.1834	0.1778	0.1778
5000	0.13298	0.18422	0.18422	0.17952	0.17952

Table A.5: Simulation results: Power of tests, Part 2.

n	pearson.reject	miz.reject	miz.reject_adj	mihat.reject	mihat.reject_adj
5100	0.13314	0.18408	0.18408	0.17962	0.17962
5200	0.13886	0.18984	0.18984	0.18484	0.18484
5300	0.142	0.19334	0.19334	0.18894	0.18894
5400	0.14108	0.19286	0.19286	0.18804	0.18804
5500	0.1444	0.19728	0.19728	0.1928	0.1928
5600	0.1471	0.19764	0.19764	0.1943	0.1943
5700	0.14882	0.20034	0.20034	0.1956	0.1956
5800	0.1497	0.20252	0.20252	0.19944	0.19944
5900	0.15664	0.20874	0.20874	0.2055	0.2055
6000	0.15694	0.2094	0.2094	0.20744	0.20744
6100	0.1615	0.21206	0.21206	0.21012	0.21012
6200	0.16256	0.21544	0.21544	0.21402	0.21402
6300	0.16668	0.21968	0.21968	0.2177	0.2177
6400	0.17052	0.22314	0.22314	0.22166	0.22166
6500	0.17156	0.22418	0.22418	0.22314	0.22314
6600	0.17736	0.22926	0.22926	0.22934	0.22934
6700	0.17692	0.23072	0.23072	0.22894	0.22894
6800	0.17556	0.2297	0.2297	0.22922	0.22922
6900	0.18248	0.23544	0.23544	0.23566	0.23566
7000	0.18132	0.23708	0.23708	0.23556	0.23556
7100	0.18644	0.2398	0.2398	0.23914	0.23914
7200	0.18706	0.24244	0.24244	0.24266	0.24266
7300	0.19288	0.24756	0.24756	0.24692	0.24692
7400	0.19868	0.25346	0.25346	0.25386	0.25386
7500	0.20034	0.25544	0.25544	0.25564	0.25564
7600	0.20054	0.25696	0.25696	0.25762	0.25762
7700	0.20568	0.25806	0.25806	0.25916	0.25916
7800	0.20844	0.26374	0.26374	0.26576	0.26576
7900	0.21314	0.2676	0.2676	0.26832	0.26832
8000	0.21544	0.2702	0.2702	0.27216	0.27216
8100	0.21568	0.27248	0.27248	0.27344	0.27344
8200	0.2234	0.27876	0.27876	0.28088	0.28088
8300	0.22324	0.28146	0.28146	0.28302	0.28302
8400	0.22598	0.2835	0.2835	0.28534	0.28534
8500	0.23034	0.28726	0.28726	0.28876	0.28876
8600	0.23122	0.29026	0.29026	0.29246	0.29246
8700	0.2354	0.29362	0.29362	0.29606	0.29606
8800	0.23848	0.2971	0.2971	0.29966	0.29966
8900	0.24332	0.30158	0.30158	0.30408	0.30408
9000	0.24844	0.30734	0.30734	0.31056	0.31056
9100	0.24936	0.30954	0.30954	0.3128	0.3128
9200	0.2527	0.31242	0.31242	0.31476	0.31476
9300	0.25764	0.31798	0.31798	0.3212	0.3212
9400	0.25842	0.31736	0.31736	0.32084	0.32084
9500	0.26432	0.32338	0.32338	0.3272	0.3272
9600	0.26572	0.32698	0.32698	0.33098	0.33098
9700	0.26876	0.32948	0.32948	0.33294	0.33294
9800	0.27218	0.33316	0.33316	0.33686	0.33686
9900	0.27702	0.33678	0.33678	0.3415	0.3415
10000	0.28096	0.34032	0.34032	0.34396	0.34396

Table A.6: Simulation results: Power of tests, Part 3.

n	pearson.reject	miz.reject	miz.reject_adj	mihat.reject	mihat.reject_adj
10100	0.28656	0.34762	0.34762	0.35222	0.35222
10200	0.28294	0.3458	0.3458	0.35008	0.35008
10300	0.2901	0.35298	0.35298	0.35682	0.35682
10400	0.29408	0.35608	0.35608	0.36078	0.36078
10500	0.29268	0.35514	0.35514	0.35972	0.35972
10600	0.3019	0.36412	0.36412	0.3683	0.3683
10700	0.30556	0.36898	0.36898	0.37422	0.37422
10800	0.30886	0.37208	0.37208	0.377	0.377
10900	0.31172	0.37538	0.37538	0.38012	0.38012
11000	0.31552	0.37956	0.37956	0.38526	0.38526
11100	0.31904	0.38156	0.38156	0.38692	0.38692
11200	0.32238	0.38622	0.38622	0.39102	0.39102
11300	0.32878	0.3943	0.3943	0.39946	0.39946
11400	0.3302	0.39406	0.39406	0.4	0.4
11500	0.3333	0.39826	0.39826	0.4053	0.4053
11600	0.33758	0.40206	0.40206	0.40748	0.40748
11700	0.33904	0.40456	0.40456	0.41106	0.41106
11800	0.34462	0.40944	0.40944	0.41606	0.41606
11900	0.34698	0.41192	0.41192	0.41886	0.41886
12000	0.35176	0.4179	0.4179	0.42366	0.42366
12100	0.35598	0.42144	0.42144	0.42744	0.42744
12200	0.36368	0.43086	0.43086	0.43634	0.43634
12300	0.36226	0.42898	0.42898	0.4357	0.4357
12400	0.36818	0.43542	0.43542	0.4418	0.4418
12500	0.37398	0.43832	0.43832	0.44486	0.44486
12600	0.37822	0.44418	0.44418	0.45128	0.45128
12700	0.37912	0.44582	0.44582	0.45212	0.45212
12800	0.38036	0.4493	0.4493	0.45646	0.45646
12900	0.381	0.44692	0.44692	0.4542	0.4542
13000	0.38982	0.45722	0.45722	0.46338	0.46338
13100	0.39744	0.46408	0.46408	0.47112	0.47112
13200	0.40086	0.46702	0.46702	0.47464	0.47464
13300	0.39934	0.46664	0.46664	0.47436	0.47436
13400	0.4064	0.47514	0.47514	0.48178	0.48178
13500	0.40706	0.47494	0.47494	0.48236	0.48236
13600	0.41168	0.48118	0.48118	0.48812	0.48812
13700	0.41978	0.48636	0.48636	0.49392	0.49392
13800	0.42078	0.4903	0.4903	0.49706	0.49706
13900	0.4224	0.49094	0.49094	0.49928	0.49928
14000	0.42854	0.49472	0.49472	0.5018	0.5018
14100	0.43074	0.49786	0.49786	0.50576	0.50576
14200	0.43218	0.50094	0.50094	0.5081	0.5081
14300	0.43962	0.5086	0.5086	0.5161	0.5161
14400	0.44332	0.5138	0.5138	0.52106	0.52106
14500	0.45006	0.51884	0.51884	0.52658	0.52658
14600	0.44938	0.51888	0.51888	0.5267	0.5267
14700	0.4557	0.52532	0.52532	0.53314	0.53314
14800	0.4568	0.52892	0.52892	0.53704	0.53704
14900	0.46172	0.53192	0.53192	0.53972	0.53972
15000	0.4648	0.53682	0.53682	0.5446	0.5446