JENSEN-SHANNON DIVERGENCE: ESTIMATION AND HYPOTHESIS TESTING

by

Ann Marie Stewart

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. Jianjiang Jiang

____

Dr. Eliana Christou

____

Dr. Craig Depken

Abstract

ANN MARIE STEWART. Jensen-Shannon Divergence: Estimation and Hypothesis Testing. (Under the direction of DR. ZHIYI ZHANG)

Jensen-Shannon divergence is one reasonable solution to the problem of measuring the level of difference or "distance" between two probability distributions on a multinomial population. If one of the distributions is assumed to be known *a priori*, estimation is a one-sample problem; if the two probability distributions are both assumed to be unknown, estimation becomes a two-sample problem. In both cases, the simple plug-in estimator has a bias that is $O(1/N)$, and hence bias reduction is explored in this dissertation. Using the well-known the jackknife method for both the one-sample and two-sample cases, an estimator with a bias of $O(1/N^2)$ is achieved. The asymptotic distributions of the estimators are determined to be chi-squared when the two distributions are equal, and normal when the two distributions are different. Then, hypothesis tests for the equality of the two multinomial distributions in both cases are established using test statistics based upon the jackknifed estimators. Finally, simulation studies are shown to verify the results numerically, and then the results are applied to real-world datasets.

DEDICATION

I dedicate my dissertation firstly to my PhD advisor, Zhiyi Zhang. He saw my intellectual potential when I didn't see it myself. To my parents who taught me how to succeed academically from a young age. To Sean, who always encouraged me in my PhD work.

ACKNOWLEDGEMENTS

I would like to thank my advisor Zhiyi Zhang, and his other two students Jialin Zhang and Chen Chen, with whom much theoretical discussion transpired. Additionally, I would like to thank my friend and colleague Ali Mahzarnia who spent hours helping me work through the details of this dissertation.

Contents

CHAPTER 1: INTRODUCTION

1.1 Problem Statement

Suppose we have a population that follows the multinomial distribution with a finite, but possibly unknown, number of classes K and that the classes are labeled with the corresponding letters $\mathscr{L} = \{\ell_1, \ldots, \ell_K\}$. Suppose there are two possible probability distributions on this population under consideration, defined by the $K-1$ dimensional vectors

$$\mathbf{p} = \{p_1, \ldots, p_{K-1}\}$$

and

$$\mathbf{q} = \{q_1, \ldots, q_{K-1}\}$$

Assume throughout the paper that p_K and q_K refer to

$$p_K = 1 - \sum_{k=1}^{K-1} p_k \tag{1.1}$$

and

$$q_K = 1 - \sum_{k=1}^{K-1} q_k \tag{1.2}$$

where the ordering of the elements is fixed. Furthermore, suppose that

$$\sum_{k=1}^{K} I[p_k > 0] = \sum_{k=1}^{K} I[q_k > 0] = K$$

so that all letters have positive probability for both distributions. Often in practice it may be desirable to have a measure of “distance” or “divergence” between the two probability distributions. From [6], such a measure is defined and is known as Kullback-Leibler divergence.

1.2 Kullback-Leibler Divergence

Definition 1. *For two probability distributions* $\mathbf{p}$ *and* $\mathbf{q}$ *on the same alphabet* $\mathscr{L}$ *of cardinality* K*, the relative entropy or the Kullback-Leibler divergence of* $\mathbf{p}$ *and* $\mathbf{q}$ *is defined as*

$$D(\mathbf{p}||\mathbf{q}) = \textstyle\sum_{k=1}^{K} p_k \ln\left(\frac{p_k}{q_k}\right) \tag{1.3}$$

observing that, for each summand $p \ln(p/q)$*,*

1) *If* $p = 0$*,* $p \ln\left(\frac{p}{q}\right) = 0$*, and*

2) *If* $p > 0$ *and* $q = 0$*, then* $p \ln\left(\frac{p}{q}\right) = +\infty$*.*

This measure has some notable advantageous qualities, one of which is described in the following theorem.

Theorem 1. *Given two probability distributions* $\mathbf{p}$ *and* $\mathbf{q}$ *on the same alphabet* $\mathscr{L}$*,*

$$D(\mathbf{p}||\mathbf{q}) \geq 0 \tag{1.4}$$

Moreover, the equality holds if and only if $\mathbf{p} = \mathbf{q}$*.*

However, Kullback-Leibler divergence is not symmetric with respect to $\mathbf{p}$ and $\mathbf{q}$, nor does it necessarily always take finite value. A remedy for these potential concerns is to use a different measure called Jensen-Shannon divergence, from [7].

1.3 Jensen-Shannon Divergence and Interpretation

Definition 2. *For two probability distributions* $\mathbf{p}$ *and* $\mathbf{q}$ *on the same alphabet* $\mathscr{L}$*, the Jensen-Shannon divergence of* $\mathbf{p}$ *and* $\mathbf{q}$ *is defined as*

$$JS(\mathbf{p}||\mathbf{q}) = \frac{1}{2}\left(D\left(\mathbf{p}\middle|\middle|\frac{\mathbf{p}+\mathbf{q}}{2}\right) + D\left(\mathbf{q}\middle|\middle|\frac{\mathbf{p}+\mathbf{q}}{2}\right)\right) \tag{1.5}$$

These measures are closely related to that of Shannon's Entropy, given in [15], which is defined loosely as a measure of the dispersion or "variance" of the individual distribution populations $\mathbf{p}$, $\mathbf{q}$. The more technical definition is as follows.

Definition 3. *For a probability distribution* $\mathbf{p}$ *on an alphabet* $\mathscr{L}$*, Shannon's entropy is defined as*

$$H(\mathbf{p}) = -\sum_{k}^{K} p_k \ln p_k \tag{1.6}$$

Using this definition, we can write Jensen-Shannon divergence in a more practically useful form.

Theorem 2. *Jensen-Shannon divergence for probability distributions* $\mathbf{p}$ *and* $\mathbf{q}$ *on alphabet* $\mathscr{L}$ *is equivalent to*

$$= -\frac{1}{2}(H(\mathbf{p}) + H(\mathbf{q})) + H\left(\frac{\mathbf{p}+\mathbf{q}}{2}\right) =: A + B$$

where H *is the entropy defined in* (1.6)*.*

Proof.

$$JS(\mathbf{p}||\mathbf{q}) = \frac{1}{2}\left(\sum_{k=1}^{K} p_k \ln\left(\frac{p_k}{(p_k+q_k)/2}\right) + \sum_{k=1}^{K} q_k \ln\left(\frac{q_k}{(p_k+q_k)/2}\right)\right)$$

$$= \frac{1}{2}\left(\sum_{k=1}^{K} p_k \ln(p_k) + \sum_{k=1}^{K} q_k \ln(q_k)\right) - \sum_{k=1}^{K} \frac{p_k+q_k}{2} \ln\left(\frac{p_k+q_k}{2}\right)$$

□

An intuitive interpretation of Jensen-Shannon Divergence may therefore be understood in this way: it is the difference between the entropy of the average and the average of the entropies for distributions $\mathbf{p}$ and $\mathbf{q}$. In other words, it is the "entropy" leftover from the interaction between $\mathbf{p}$ and $\mathbf{q}$ when the "entropy" from the individual distributions is subtracted out. Taking the difference leaves only that "entropy" which is accounted for by the interaction between $\mathbf{p}$ and $\mathbf{q}$ in the average of the distributions. The more "entropy" or "chaos" caused by the interaction between $\mathbf{p}$ and $\mathbf{q}$, the more "distance" between the two distributions.

1.4 Properties

Our natural understanding of the notion of "distance" is that it should be nonnegative, and if the elements are the same, the "distance" should be 0.

Theorem 3. *The Jensen-Shannon divergence of* $\mathbf{p}$ *and* $\mathbf{q}$ *is nonnegative, and equal to 0 if and only if* $\mathbf{p} = \mathbf{q}$.

Proof. By Theorem 1, $JS(\mathbf{p}||\mathbf{q})$ is nonnegative as the sum of nonnegative terms. Because both terms in $JS(\mathbf{p}||\mathbf{q})$ are nonnegative, if the sum is 0 then each term must be 0. Thus, $JS(\mathbf{p}||\mathbf{q})$ =0 if and only if

$$D\left(\mathbf{p}\middle|\middle|\frac{\mathbf{p}+\mathbf{q}}{2}\right) = D\left(\mathbf{q}\middle|\middle|\frac{\mathbf{p}+\mathbf{q}}{2}\right) = 0 \tag{1.7}$$

Since by Theorem 1, $D(\mathbf{p}||\mathbf{q}) = 0$ if and only if $\mathbf{p} = \mathbf{q}$, then (1.7) is true if and only if

$$2\mathbf{q} = 2\mathbf{p} = \mathbf{p} + \mathbf{q}$$

if and only if $\mathbf{p} = \mathbf{q}$.

$\square$

Although the notion of "distance" does not imply the concept of an upper bound, Jensen-Shannon divergence does happen to have an upper bound, as shown in [4].

Theorem 4. *For any two distributions* $\mathbf{p}$, $\mathbf{q}$

$$JS(\mathbf{p}||\mathbf{q}) \leq \frac{1}{2}\ln\left(\frac{2}{1+\exp\{-D(\mathbf{p}||\mathbf{q})\}}\right) + \frac{1}{2}\ln\left(\frac{2}{1+\exp\{-D(\mathbf{q}||\mathbf{p})\}}\right) < \ln(2)$$

Proof.

$$JS(\mathbf{p}||\mathbf{q}) = \frac{1}{2}\sum_{k=1}^{K} p_k \ln\left(\frac{2p_k}{p_k+q_k}\right) + \frac{1}{2}\sum_{k=1}^{K} q_k \ln\left(\frac{2q_k}{p_k+q_k}\right)$$

$$= \frac{1}{2}\sum_{k=1}^{K} p_k \ln\left(\frac{2}{1+\exp\{\ln(\frac{p_k}{q_k})\}}\right) + \frac{1}{2}\sum_{k=1}^{K} q_k \ln\left(\frac{2}{1+\exp\{\ln(\frac{q_k}{p_k})\}}\right)$$

$$\leq \frac{1}{2}\ln\left(\frac{2}{1+\exp\{-D(\mathbf{p}||\mathbf{q})\}}\right) + \frac{1}{2}\ln\left(\frac{2}{1+\exp\{-D(\mathbf{q}||\mathbf{p})\}}\right) < \ln(2)$$

where the inclusive inequality in the last line is due to Jensen's inequality. □

Note that the line derived from Jensen's inequality reaches equality if and only if $\mathbf{p} = \mathbf{q}$, in which case $JS(\mathbf{p}||\mathbf{q})$ collapses into 0. Otherwise we have all strict inequalities:

$$JS(\mathbf{p}||\mathbf{q}) = \frac{1}{2}\sum_{k=1}^{K} p_k \ln\left(\frac{2}{1+\exp\{\ln(\frac{p_k}{q_k})\}}\right) + \frac{1}{2}\sum_{k=1}^{K} q_k \ln\left(\frac{2}{1+\exp\{\ln(\frac{q_k}{p_k})\}}\right)$$

$$< \frac{1}{2}\ln\left(\frac{2}{1+\exp\{-D(\mathbf{p}||\mathbf{q})\}}\right) + \frac{1}{2}\ln\left(\frac{2}{1+\exp\{-D(\mathbf{q}||\mathbf{p})\}}\right) < \ln(2)$$

Note that

$$\frac{1}{2}\ln\left(\frac{2}{1+\exp\{-D(\mathbf{p}||\mathbf{q})\}}\right) + \frac{1}{2}\ln\left(\frac{2}{1+\exp\{-D(\mathbf{q}||\mathbf{p})\}}\right) \tag{1.8}$$

approaches $\ln(2)$ as $D(\mathbf{p}||\mathbf{q})$ and $D(\mathbf{q}||\mathbf{p})$ increase, and therefore $JS(\mathbf{p}||\mathbf{q})$ approaches

$\ln(2)$ as $\mathbf{p}$ and $\mathbf{q}$ get "further apart," as expected. The value in (1.8) will never reach $\ln(2)$ because $\exp\{-D(\mathbf{q}||\mathbf{p})\}$ can never be 0. Therefore $\ln(2)$ is an upper bound for $JS(\mathbf{p}||\mathbf{q})$, but there will never be equivalence. $JS(\mathbf{p}||\mathbf{q})$ approaches, but does not reach its upper bound.

There are two common scenarios which may arise where Jensen-Shannon divergence would be of use in practice: one may be interested in the comparison of an unknown distribution against a known one, or in estimating the divergence between two unknown distributions. The first case would necessitate only one sample, and the second two samples. Clearly there are different theoretical implications, so we tackle each problem separately in each of the following chapters on estimation and asymptotic distributions.

CHAPTER 2: PLUG-IN ESTIMATORS AND BIAS

2.1 One-Sample

Assume that the distribution $\mathbf{p}$ is known, and we are trying to estimate $\mathbf{q}$. Suppose that we have a sample from $\mathbf{q}$ of size N from the alphabet $\mathscr{L} = \{\ell_1, \ell_2, \ldots, \ell_K\}$ that is represented by the observations $\{\omega_1, \ldots, \omega_N\}$. Define the sequences of observed frequencies as:

$$Y_1 = \sum_{j=1}^{N} I[\omega_j = \ell_1], \ldots, Y_K = \sum_{j=1}^{N} I[\omega_j = \ell_K]$$

Additionally, denote the vector of plug-in estimates for the probabilities as

$$\hat{\mathbf{q}} = \{\hat{q}_1, \ldots, \hat{q}_{K-1}\}$$

with

$$\hat{q}_K = 1 - \sum_{k=1}^{K-1} \hat{q}_k$$

where, for each k from 1 to $K-1$,

$$\hat{q}_k = \frac{Y_k}{N}$$

Using these, we can directly estimate the Jensen-Shannon Divergence between a known distribution $\mathbf{p}$ and the estimated one $\mathbf{q}$.

Definition 4. *Define the one-sample plug-in estimator for Jensen-Shannon Divergence as*

$$\begin{aligned}
\widehat{JS}_1(\mathbf{p}||\mathbf{q}) &= -\frac{1}{2}\left(H(\mathbf{p}) + H(\hat{\mathbf{q}})\right) + H\left(\frac{\mathbf{p}+\hat{\mathbf{q}}}{2}\right) \\
&= \frac{1}{2}\left(\sum_{k=1}^{K} p_k \ln(p_k) + \sum_{k=1}^{K} \hat{q}_k \ln(\hat{q}_k)\right) - \sum_{k=1}^{K} \frac{p_k+\hat{q}_k}{2} \ln\left(\frac{p_k+\hat{q}_k}{2}\right) \qquad (2.1) \\
&=: \hat{A}_1^0 + \hat{B}_1^0
\end{aligned}$$

We shall proceed to find the bias of this estimator and then propose a way to mitigate it, tackling each part $\hat{A}_1^0$ and $\hat{B}_1^0$ separately. Before doing so, it must be noted that [5] showed that the bias of the plug-in estimator of entropy, $\hat{H}$ is

$$-\frac{K-1}{2N} + \frac{1}{12N^2}\left(1 - \sum_{k=1}^{K}\frac{1}{p_k}\right) + O(N^{-3}) \qquad (2.2)$$

which implies that the bias of the plug-in of Jensen-Shannon Divergence is also $O(N^{-1})$.

Theorem 5. *Assuming a sample of size N from an unknown distribution $\mathbf{q}$, the bias of the one-sample plug-in estimator $\hat{A}_1^0$ is*

$$\frac{K-1}{4N} - \frac{1}{24N^2}\left(1 - \sum_{k=1}^{K}\frac{1}{q_k}\right) + O(N^{-3}) \qquad (2.3)$$

Proof. Using (2.2) we have

$$E(\hat{A}_1^0) - A = -\frac{1}{2}\left(E(H(\hat{\mathbf{q}})) - H(\mathbf{q})\right)$$

$$= -\frac{1}{2}\left(-\frac{K-1}{2N} + \frac{1}{12N^2}\left(1 - \sum_{k=1}^{K}\frac{1}{q_k}\right) + O(N^{-3})\right)$$

$$= \frac{K-1}{4N} - \frac{1}{24N^2}\left(1 - \sum_{k=1}^{K}\frac{1}{q_k}\right) + O(N^{-3})$$

□

Theorem 6. *Assuming a sample of size N for an unknown distribution $\mathbf{q}$, the bias of the one-sample plug-in estimator $\hat{B}_1^0$ is*

$$-\frac{1}{4}\left(\frac{1}{p_K+q_K}\left(\sum_{k=1}^{K-1}\frac{q_k(1-q_k)}{N} - \sum_{m\neq n}\frac{q_m q_n}{N}\right) + \sum_{k=1}^{K-1}\frac{q_k(1-q_k)}{N(p_k+q_k)}\right) + O(N^{-2}) \tag{2.4}$$

$$= \frac{c}{N} + \frac{\gamma}{N^2} + O(N^{-3})$$

where

$$c = -\frac{1}{4}\left(\sum_{k=1}^{K-1} q_k(1-q_k)\left(\frac{1}{p_K+q_K} + \frac{1}{p_k+q_k}\right) - \sum_{m\neq n}\frac{q_m q_n}{p_K+q_K}\right) \tag{2.5}$$

Proof. By Taylor series expansion, we have

$$\hat{B}_1^0 - B = B(\hat{\mathbf{q}}) - B(\mathbf{q})$$

$$= (\hat{\mathbf{q}} - \mathbf{q})^\tau \nabla B(\mathbf{q}) + \frac{1}{2}\Big((\hat{\mathbf{q}} - \mathbf{q})^\tau \nabla^2 B(\mathbf{q})(\hat{\mathbf{q}} - \mathbf{q})\Big) + R_N$$

where $\nabla B(\mathbf{q})$ is the gradient of $B(\mathbf{q})$ and $\nabla^2 B(\mathbf{q})$ is the Hessian matrix of $B(\mathbf{q})$. The expected value of the first term is clearly 0, and $E(R_N) = \frac{\gamma}{N^2} + O(N^{-3})$ for some constant γ. Thus we only have to contend with the term

$$\frac{1}{2}(\hat{\mathbf{q}} - \mathbf{q})^\tau \nabla^2 B(\mathbf{q})(\hat{\mathbf{q}} - \mathbf{q})$$

Note that

$$\nabla^2 B(\mathbf{q})$$

$$= -\frac{1}{2}\begin{pmatrix} \frac{1}{p_1+q_1} + \frac{1}{p_K+q_K} & \frac{1}{p_K+q_K} & \cdots & \frac{1}{p_K+q_K} \\ \frac{1}{p_K+q_K} & \frac{1}{p_2+q_2} + \frac{1}{p_K+q_K} & \cdots & \frac{1}{p_K+q_K} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{p_K+q_K} & \frac{1}{p_K+q_K} & \cdots & \frac{1}{p_{K-1}+q_{K-1}} + \frac{1}{p_K+q_K} \end{pmatrix}$$

$$=: -\frac{1}{2}\Omega \tag{2.6}$$

And so

$$\frac{1}{2}(\hat{\mathbf{q}}-\mathbf{q})^\tau\left(-\frac{1}{2}\right)\Omega(\hat{\mathbf{q}}-\mathbf{q}) = -\frac{1}{4}\left(\frac{\left(\sum_{k=1}^{K-1}\hat{q}_k - q_k\right)^2}{p_K+q_K} + \sum_{k=1}^{K-1}\frac{(\hat{q}_k-q_k)^2}{p_k+q_k}\right)$$

Taking the expected value of both sides and using Lemma 15 yields

$$-\frac{1}{4}\left(\frac{E\left(\sum_{k=1}^{K-1}\hat{q}_k - q_k\right)^2}{p_K+q_K} + \sum_{k=1}^{K-1}\frac{E\left(\hat{q}_k-q_k\right)^2}{p_k+q_k}\right)$$

$$\begin{aligned} &= -\frac{1}{4}\left(\frac{1}{p_K+q_K}\left(\sum_{k=1}^{K-1}\frac{q_k(1-q_k)}{N} - \sum_{j\neq k}\frac{q_jq_k}{N}\right) + \sum_{k=1}^{K-1}\frac{q_k(1-q_k)}{N(p_k+q_k)}\right) \\ &= -\frac{1}{4N}\left(\sum_{k=1}^{K-1}q_k(1-q_k)\left(\frac{1}{p_K+q_K}+\frac{1}{p_k+q_k}\right) - \sum_{j\neq k}\frac{q_jq_k}{p_K+q_K}\right) \end{aligned}$$

□

Theorems 5 and 6 taken together yield the following.

Theorem 7. *The bias of the plug-in estimator of Jensen-Shannon Divergence in the one-sample case is* $O(N^{-1})$*:*

$$\frac{K-1}{4N} - \frac{1}{24N^2}\left(1-\sum_{k=1}^{K}\frac{1}{q_k}\right) + \frac{c}{N} + \frac{\gamma}{N^2} + O(N^{-3})$$

for some constant γ *and where* c *is as in* (2.5)*.*

2.2 Two-Sample

For the two-sample case, assume there exist two independent samples of sizes $N_{\mathbf{p}}$ and $N_{\mathbf{q}}$, according to unknown distributions $\mathbf{p}$ and $\mathbf{q}$; both on the same alphabet $\mathscr{L} = \{\ell_1, \ell_2, \ldots, \ell_K\}$. Let the $\mathbf{p}$ sample be represented by $\{\upsilon_1, \ldots, \upsilon_{N_{\mathbf{p}}}\}$ and the $\mathbf{q}$ sample by $\{\omega_1, \ldots, \omega_{N_{\mathbf{q}}}\}$. Similar to the one-sample case, define the sequences of observed frequencies as

$$X_1 = \sum_{i=1}^{N_{\mathbf{p}}} I[\upsilon_i = \ell_1], \ldots, X_K = \sum_{i=1}^{N_{\mathbf{p}}} I[\upsilon_i = \ell_K]$$

and

$$Y_1 = \sum_{j=1}^{N_{\mathbf{q}}} I[\omega_j = \ell_1], \ldots, Y_K = \sum_{j=1}^{N_{\mathbf{q}}} I[\omega_j = \ell_K]$$

Also denote the plug-in estimators as

$$\hat{\mathbf{p}} = \{\hat{p}_1, \ldots, \hat{p}_{K-1}\}$$

and

$$\hat{\mathbf{q}} = \{\hat{q}_1, \ldots, \hat{q}_{K-1}\}$$

with

$$\hat{p}_K = 1 - \sum_{k=1}^{K-1} \hat{p}_k$$

and

$$\hat{q}_K = 1 - \sum_{k=1}^{K-1} \hat{q}_k$$

where, for each k from 1 to $K-1$,

$$\hat{p}_k = \frac{X_k}{N_{\mathbf{p}}}$$

and

$$\hat{q}_k = \frac{Y_k}{N_{\mathbf{q}}}$$

For notational simplicity in the two-sample case, define $\mathbf{v}$ and $\hat{\mathbf{v}}$ as the $2K-2$

dimensional vectors

$$\mathbf{v} = (\mathbf{p}, \mathbf{q}) = \{p_1, \ldots, p_{K-1}, q_1, \ldots, q_{K-1}\} \tag{2.7}$$

and

$$\hat{\mathbf{v}} = (\hat{\mathbf{p}}, \hat{\mathbf{q}}) = \{\hat{p}_1, \ldots, \hat{p}_{K-1}, \hat{q}_1, \ldots, \hat{q}_{K-1}\} \tag{2.8}$$

Additionally, we impose the following condition on the asymptotic behavior of the sample sizes.

Condition 1. *The probability distributions* $\mathbf{p}$ *and* $\mathbf{q}$ *and the observed sample distribution* $\hat{\mathbf{p}}$ *and* $\hat{\mathbf{q}}$ *satisfy*

- *There exists a constant* $\lambda \in (0, \infty)$ *such that* $N_{\mathbf{p}}/N_{\mathbf{q}} \to \lambda$ *as* $N_{\mathbf{p}}, N_{\mathbf{q}} \to \infty$

Under Condition 1, for any $x \in \mathbb{R}$, $O(N_{\mathbf{p}}^x) = O(N_{\mathbf{q}}^x)$ and will be heretofore notated more generally as $O(N^x)$.

Definition 5. *Define the two-sample plug-in estimator for Jensen-Shannon Divergence as*

$$\begin{aligned}
\widehat{JS}_2(\mathbf{p}||\mathbf{q}) &= -\frac{1}{2}\left(H(\hat{\mathbf{p}}) + H(\hat{\mathbf{q}})\right) + H\left(\frac{\hat{\mathbf{p}} + \hat{\mathbf{q}}}{2}\right) \\
&= \frac{1}{2}\left(\sum_{k=1}^{K} \hat{p}_k \ln(\hat{p}_k) + \sum_{k=1}^{K} \hat{q}_k \ln(\hat{q}_k)\right) - \sum_{k=1}^{K} \frac{\hat{p}_k + \hat{q}_k}{2} \ln\left(\frac{\hat{p}_k + \hat{q}_k}{2}\right) \\
&=: \hat{A}_2^0 + \hat{B}_2^0
\end{aligned} \tag{2.9}$$

Theorem 8. *Assuming sample sizes* $N_{\mathbf{p}}$, $N_{\mathbf{q}}$ *for* $\mathbf{p}$ *and* $\mathbf{q}$, *the bias of the two-sample* *plug-in estimator $\hat{A}_2^0$ is*

$$\frac{K-1}{4}\left(\frac{1}{N_{\mathbf{p}}}+\frac{1}{N_{\mathbf{q}}}\right)-\frac{1}{24N_{\mathbf{p}}^2}\left(1-\sum_{k=1}^{K}\frac{1}{p_k}\right)-\frac{1}{24N_{\mathbf{q}}^2}\left(1-\sum_{k=1}^{K}\frac{1}{q_k}\right)+O(N^{-3}) \tag{2.10}$$

Proof. Using (2.2) we have

$$E(\hat{A}_2^0)-A=-\frac{1}{2}\left(E(H(\hat{\mathbf{p}}))-H(\mathbf{p})\right)-\frac{1}{2}\left(E(H(\hat{\mathbf{q}}))-H(\mathbf{q})\right)$$

$$=-\frac{1}{2}\left(-\frac{K-1}{2N_{\mathbf{p}}}+\frac{1}{12N_{\mathbf{p}}^2}\left(1-\sum_{k=1}^{K}\frac{1}{p_k}\right)+O(N_{\mathbf{p}}^{-3})\right)$$

$$-\frac{1}{2}\left(-\frac{K-1}{2N_{\mathbf{q}}}+\frac{1}{12N_{\mathbf{q}}^2}\left(1-\sum_{k=1}^{K}\frac{1}{q_k}\right)+O(N_{\mathbf{q}}^{-3})\right)$$

$$=\frac{K-1}{4}\left(\frac{1}{N_{\mathbf{p}}}+\frac{1}{N_{\mathbf{q}}}\right)-\frac{1}{24N_{\mathbf{p}}^2}\left(1-\sum_{k=1}^{K}\frac{1}{p_k}\right)-\frac{1}{24N_{\mathbf{q}}^2}\left(1-\sum_{k=1}^{K}\frac{1}{q_k}\right)$$

$$+O(N^{-3})$$

□

Theorem 9. *Assuming sample sizes $N_{\mathbf{p}}$, $N_{\mathbf{q}}$ for $\mathbf{p}$ and $\mathbf{q}$, the bias of the two-sample plug-in estimator $\hat{B}_2^0$ is*

$$- \frac{1}{4N_{\mathbf{p}}} \left(\sum_{k=1}^{K-1} p_k(1-p_k) \left(\frac{1}{p_K+q_K} + \frac{1}{p_k+q_k} \right) - \sum_{j \neq k} \frac{p_j p_k}{p_K+q_K} \right)$$

$$- \frac{1}{4N_{\mathbf{q}}} \left(\sum_{k=1}^{K-1} q_k(1-q_k) \left(\frac{1}{p_K+q_K} + \frac{1}{p_k+q_k} \right) - \sum_{j \neq k} \frac{q_j q_k}{p_K+q_K} \right) \tag{2.11}$$

$$+ \frac{\alpha}{N_{\mathbf{p}}^2} + \frac{\gamma}{N_{\mathbf{q}}^2} + O(N^{-3})$$

$$= \frac{a}{N_{\mathbf{p}}} + \frac{c}{N_{\mathbf{q}}} + \frac{\alpha}{N_{\mathbf{p}}^2} + \frac{\gamma}{N_{\mathbf{q}}^2} + O(N^{-3})$$

where

$$a = -\frac{1}{4} \left(\sum_{k=1}^{K-1} p_k(1-p_k) \left(\frac{1}{p_K+q_K} + \frac{1}{p_k+q_k} \right) - \sum_{j \neq k} \frac{p_j p_k}{p_K+q_K} \right) \tag{2.12}$$

and

$$c = -\frac{1}{4} \left(\sum_{k=1}^{K-1} q_k(1-q_k) \left(\frac{1}{p_K+q_K} + \frac{1}{p_k+q_k} \right) - \sum_{j \neq k} \frac{q_j q_k}{p_K+q_K} \right) \tag{2.13}$$

Proof. By two variable Taylor series expansion, we have

$$\hat{B}_2^0 - B = B(\hat{\mathbf{v}}) - B(\mathbf{v})$$

$$= (\hat{\mathbf{v}} - \mathbf{v})^{\tau} \nabla B(\mathbf{v}) + \frac{1}{2} (\hat{\mathbf{v}} - \mathbf{v})^{\tau} \nabla^2 B(\mathbf{v}) (\hat{\mathbf{v}} - \mathbf{v}) + R_N$$

Taking the expected value of both sides yields the bias. For the first and third terms of the right hand side, we have

$$E\left((\hat{\mathbf{v}} - \mathbf{v})^\tau \nabla B(\mathbf{v})\right) = 0$$

and

$$E(R_N) = \frac{\alpha}{N_{\mathbf{p}}^2} + \frac{\gamma}{N_{\mathbf{q}}^2} + O(N^{-3})$$

This leaves us only to contend with the middle term

$$\frac{1}{2}(\hat{\mathbf{v}} - \mathbf{v})^\tau \nabla^2 B(\mathbf{v})(\hat{\mathbf{v}} - \mathbf{v})$$

Note that

$$\nabla^2 B(\mathbf{v}) = -\frac{1}{2}\begin{pmatrix} \Omega & \Omega \\ \Omega & \Omega \end{pmatrix}$$

where Ω is defined as in (2.6). Thus

$$\frac{1}{2}(\hat{\mathbf{v}} - \mathbf{v})^\tau \nabla^2 B(\mathbf{v})(\hat{\mathbf{v}} - \mathbf{v}) = -\frac{1}{4}\left((\hat{\mathbf{p}} - \mathbf{p})^\tau, (\hat{\mathbf{q}} - \mathbf{q})^\tau\right)\begin{pmatrix} \Omega & \Omega \\ \Omega & \Omega \end{pmatrix}\begin{pmatrix} \hat{\mathbf{p}} - \mathbf{p} \\ \hat{\mathbf{q}} - \mathbf{q} \end{pmatrix}$$

$$= -\frac{1}{4}(\hat{\mathbf{p}} - \mathbf{p})^\tau \Omega(\hat{\mathbf{q}} - \mathbf{q}) - \frac{1}{4}(\hat{\mathbf{q}} - \mathbf{q})^\tau \Omega(\hat{\mathbf{p}} - \mathbf{p})$$

$$-\frac{1}{4}(\hat{\mathbf{p}} - \mathbf{p})^\tau \Omega(\hat{\mathbf{p}} - \mathbf{p}) - \frac{1}{4}(\hat{\mathbf{q}} - \mathbf{q})^\tau \Omega(\hat{\mathbf{q}} - \mathbf{q})$$

Clearly the expected values of the terms in the first line are both 0, since $\mathbf{p}$ and $\mathbf{q}$ are independent. The expected values of the terms in the second line are derived in a similar manner to those in the proof of Theorem 6.

□

Theorems 8 and 9 immediately yield the following Theorem.

Theorem 10. *The bias of the plug-in estimator of Jensen-Shannon Divergence is* $O(N^{-1})$*:*

$$\frac{K-1}{4}\left(\frac{1}{N_{\mathbf{p}}}+\frac{1}{N_{\mathbf{q}}}\right)-\frac{1}{24N_{\mathbf{p}}^2}\left(1-\sum_{k=1}^{K}\frac{1}{p_k}\right)-\frac{1}{24N_{\mathbf{q}}^2}\left(1-\sum_{k=1}^{K}\frac{1}{q_k}\right)$$

$$+\frac{a}{N_{\mathbf{p}}}+\frac{c}{N_{\mathbf{q}}}+\frac{\alpha}{N_{\mathbf{p}}^2}+\frac{\gamma}{N_{\mathbf{q}}^2}+O(N^{-3})$$

where a *and* c *are defined as in* (2.12) *and* (2.13)*.*

Now that we have the precise forms of the biases in the one and two-sample cases given in Theorems 7 and 10, a method for mitigating them is developed in the following chapter.

CHAPTER 3: BIAS REDUCED ESTIMATORS

3.1 One-Sample

First we consider correcting the bias of $\hat{A}_1^0$ using the well known jackknife resampling technique. The idea is, for each datum j, $1 \leq j \leq N$, leave that observation out and compute the plug-in estimator from the corresponding sub-sample of size $N-1$, then find the average of these calculations. Denote $\hat{\mathbf{q}}^{(-j)}$ as the vector of plug-in estimates of $\mathbf{q}$ with the jth observation omitted,

$$\hat{A}_{1\mathbf{q}}^0 = -\frac{1}{2}H\left(\hat{\mathbf{q}}\right) \tag{3.1}$$

$$\hat{A}_{1\mathbf{q}(-j)} = -\frac{1}{2}H\left(\hat{\mathbf{q}}^{(-j)}\right) \tag{3.2}$$

The computation of the one-sample jackknife estimator is as follows:

$$\hat{A}_{JK_{1\mathbf{q}}} = N\hat{A}_{1\mathbf{q}}^0 - \frac{N-1}{N}\sum_{j=1}^{N}\hat{A}_{1\mathbf{q}(-j)} \tag{3.3}$$

And finally,

$$\hat{A}_{JK_1} = -\frac{1}{2}H(\mathbf{p}) + \hat{A}_{JK_{1\mathbf{q}}} \tag{3.4}$$

Theorem 11. *The one-sample jackknife estimator from* (3.4) *has a bias of order* $O(N^{-2})$*:*

$$E(\hat{A}_{JK_1}) - A = -\frac{1}{24N(N-1)}\left(1 - \sum_{k=1}^{K}\frac{1}{q_k}\right) + O(N^{-3}) = O(N^{-2})$$

Proof. Using Theorem 5, we have

$$E(\hat{A}_{JK_1}) = NE(\hat{A}_1) - (N-1)E(\hat{A}_{1(-j)})$$

$$= N\left(A + \frac{K-1}{N} - \frac{1}{24N^2}\left(1 - \sum_{k=1}^{K}\frac{1}{q_k}\right) + O(N^{-3})\right)$$

$$-(N-1)\left(A + \frac{K-1}{N-1} - \frac{1}{24(N-1)^2}\left(1 - \sum_{k=1}^{K}\frac{1}{q_k}\right) + O((N-1)^{-3})\right)$$

$$= A - \frac{1}{24N}\left(1 - \sum_{k=1}^{K}\frac{1}{q_k}\right) + \frac{1}{24(N-1)}\left(1 - \sum_{k=1}^{K}\frac{1}{q_k}\right) + O(N^{-3})$$

$$= A - \frac{1}{24N(N-1)}\left(1 - \sum_{k=1}^{K}\frac{1}{q_k}\right) + O(N^{-3}) = O(N^{-2})$$

□

Again use the jackknife approach with $\hat{B}_1^0$. Denote

$$\hat{B}_{1(-j)} = H\left(\frac{\mathbf{p} + \hat{\mathbf{q}}^{(-j)}}{2}\right) \tag{3.5}$$

as the corresponding plug-in estimator of B. Then, compute the jackknife estimator as

$$\hat{B}_{JK_1} = N\hat{B}_1^0 - \frac{N-1}{N}\sum_{j=1}^{N}\hat{B}_{1(-j)} \tag{3.6}$$

As will be shown, this procedure reduces the order of the bias, as desired.

Theorem 12. *The one-sample jackknife estimator from* (3.6) *has a bias of order* $O(N^{-2})$*:*

$$E(\hat{B}_{JK_1}) - B = \frac{\gamma}{N(N-1)} + O(N^{-3}) = O(N^{-2})$$

where γ *is as in Theorem* 6*.*

Proof. Using Theorem 6, we have

$$E(\hat{B}_{JK_1}) = NE(\hat{B}_1^0) - (N-1)E(\hat{B}_{1(-j)})$$

$$= N\left(B + \frac{c}{N} + \frac{\gamma}{N^2} + O(N^{-3})\right)$$

$$- (N-1)\left(B + \frac{c}{N-1} + \frac{\gamma}{(N-1)^2} + O(N^{-3})\right)$$

$$= B + \frac{\gamma}{N} - \frac{\gamma}{N-1} + O(N^{-3})$$

$$= B + \frac{\gamma}{N(N-1)} + O(N^{-3}) = O(N^{-2})$$

□

Definition 6. *Define the new, bias-adjusted estimator for Jensen-Shannon Divergence in the one-sample context as*

$$\widehat{JS}_{BA_1} = \hat{A}_{JK_1} + \hat{B}_{JK_1} \tag{3.7}$$

The next corollary follows immediately from Theorems 11 and 12.

Corollary 1. *The bias of the adjusted estimator* $\widehat{JS}_{BA_1}$ *is asymptotically* $O(N^{-2})$*.*

Now that the bias has been reduced in the one-sample case, we turn toward the two-sample case.

3.2 Two-Sample

To correct the bias of $\hat{A}_2^0$, we use a method similar to that of the one-sample case. First, denote

$$\hat{A}_2^0 = \hat{A}_{2\mathbf{p}}^0 + \hat{A}_{2\mathbf{q}}^0 = \left(-\frac{1}{2}H\left(\hat{\mathbf{p}}\right)\right) + \left(-\frac{1}{2}H\left(\hat{\mathbf{q}}\right)\right) \tag{3.8}$$

as the original plug-in estimator for $A = -\frac{1}{2}\left(H\left(\mathbf{p}\right) + H\left(\mathbf{q}\right)\right)$. Let $\hat{\mathbf{p}}^{(-i)}$ and $\hat{\mathbf{q}}^{(-j)}$ be the samples without the ith observation for $\mathbf{p}$ and without the jth observation for $\mathbf{q}$, respectively. Also, let

$$\hat{A}_{2\mathbf{p}}^{(-i)} = -\frac{1}{2}H\left(\hat{\mathbf{p}}^{(-i)}\right) \tag{3.9}$$

$$\hat{A}_{2\mathbf{q}(-j)} = -\frac{1}{2}H\left(\hat{\mathbf{q}}^{(-j)}\right) \tag{3.10}$$

Similar to the one-sample case, compute the jackknife estimators as

$$\hat{A}_{JK_{2\mathbf{p}}} = N_{\mathbf{p}}\hat{A}_{2\mathbf{p}}^0 - \frac{N_{\mathbf{p}}-1}{N_{\mathbf{p}}}\sum_{i=1}^{N_{\mathbf{p}}}\hat{A}_{2\mathbf{p}}^{(-i)} \tag{3.11}$$

and

$$\hat{A}_{JK_{2\mathbf{q}}} = N_{\mathbf{q}}\hat{A}_{2\mathbf{q}}^0 - \frac{N_{\mathbf{q}}-1}{N_{\mathbf{q}}}\sum_{j=1}^{N_{\mathbf{q}}}\hat{A}_{2\mathbf{q}(-j)} \tag{3.12}$$

Put them together to obtain

$$\hat{A}_{JK_2} = \hat{A}_{JK_{2\mathbf{p}}} + \hat{A}_{JK_{2\mathbf{q}}} \tag{3.13}$$

It can easily be shown using a proof similar to that of Theorem 11 that the bias of (3.13) is $O(N^{-2})$.

Theorem 13.

$$E(\hat{A}_{JK_2}) - A$$

$$= -\frac{1}{24N_{\mathbf{p}}(N_{\mathbf{p}}-1)}\left(1-\sum_{k=1}^{K}\frac{1}{p_k}\right) - \frac{1}{24N_{\mathbf{q}}(N_{\mathbf{q}}-1)}\left(1-\sum_{k=1}^{K}\frac{1}{q_k}\right) + O(N^{-3})$$

$$= O(N^{-2})$$

Next, a method for correcting the bias of $\hat{B}_2^0$ is explored. A process for two-sample jackknifing was introduced in [13], and will be used here. It is a two step procedure. In the first step, a jackknifed estimator is computed by deleting one datum from the $\mathbf{p}$ sample at a time. In the second step, the jackknifed estimator from the first step is further jackknifed by deleting one datum from the $\mathbf{q}$ sample at a time to produce the final estimator. Denote

$$\hat{B}_2^0 = H\left(\frac{\hat{\mathbf{p}}+\hat{\mathbf{q}}}{2}\right) \tag{3.14}$$

as the original plug-in estimator for $B = H\left(\frac{\mathbf{p}+\mathbf{q}}{2}\right)$. Let

$$\hat{B}_2^{(-i)} = H\left(\frac{\hat{\mathbf{p}}^{(-i)}+\hat{\mathbf{q}}}{2}\right) \tag{3.15}$$

$$\hat{B}_{2(-j)} = H\left(\frac{\hat{\mathbf{p}}+\hat{\mathbf{q}}^{(-j)}}{2}\right) \tag{3.16}$$

and

$$\hat{B}_{2(-j)}^{(-i)} = H\left(\frac{\hat{\mathbf{p}}^{(-i)}+\hat{\mathbf{q}}^{(-j)}}{2}\right) \tag{3.17}$$

For the first step, we let

$$\hat{B}_{2\mathbf{p}} = N_{\mathbf{p}}\hat{B}_2^0 - \frac{N_{\mathbf{p}} - 1}{N_{\mathbf{p}}}\sum_{i=1}^{N_{\mathbf{p}}}\hat{B}_2^{(-i)} \tag{3.18}$$

Then, the second and final step is obtained by jackknifing $\hat{B}_{2\mathbf{p}}$:

$$\hat{B}_{JK_2} = N_{\mathbf{q}}\hat{B}_{2\mathbf{p}} - \frac{N_{\mathbf{q}} - 1}{N_{\mathbf{q}}}\sum_{j=1}^{N_{\mathbf{q}}}\hat{B}_{2\mathbf{p}(-j)} \tag{3.19}$$

where

$$\hat{B}_{2\mathbf{p}(-j)} = N_{\mathbf{p}}\hat{B}_{2(-j)} - \frac{N_{\mathbf{p}} - 1}{N_{\mathbf{p}}}\sum_{i=1}^{N_{\mathbf{p}}}\hat{B}_{2(-j)}^{(-i)} \tag{3.20}$$

Note that (3.19) can also be written as

$$\begin{aligned}\hat{B}_{JK_2} = N_{\mathbf{p}}N_{\mathbf{q}}\hat{B}_2^0 - \frac{N_{\mathbf{q}}(N_{\mathbf{p}} - 1)}{N_{\mathbf{p}}}\sum_{i=1}^{N_{\mathbf{p}}}\hat{B}_2^{(-i)} \\ - \frac{N_{\mathbf{p}}(N_{\mathbf{q}} - 1)}{N_{\mathbf{q}}}\sum_{j=1}^{N_{\mathbf{q}}}\hat{B}_{2(-j)} + \frac{(N_{\mathbf{p}} - 1)(N_{\mathbf{q}} - 1)}{N_{\mathbf{p}}N_{\mathbf{q}}}\sum_{i=1}^{N_{\mathbf{p}}}\sum_{j=1}^{N_{\mathbf{q}}}\hat{B}_{2(-j)}^{(-i)}\end{aligned} \tag{3.21}$$

We will now show that the order of the bias of $\hat{B}_{JK_2}$ is reduced by one from that of the plug-in estimator.

Lemma 1.

$$E(\hat{B}_{2\mathbf{p}}) = B + \frac{c}{N_{\mathbf{q}}} + \frac{\alpha}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)} + \frac{\gamma}{N_{\mathbf{q}}^2} + O(N^{-3})$$

Proof. Using Theorem 9 and (3.18), we have

$$E(\hat{B}_{2\mathbf{p}}) = N_{\mathbf{p}} E(\hat{B}_2^0) + (N_{\mathbf{p}} - 1) E(\hat{B}_2^{(-i)})$$

$$= N_{\mathbf{p}} \left(\frac{a}{N_{\mathbf{p}}} + \frac{c}{N_{\mathbf{q}}} + \frac{\alpha}{N_{\mathbf{p}}^2} + \frac{\gamma}{N_{\mathbf{q}}^2} + O(N^{-3}) \right)$$

$$- (N_{\mathbf{p}} - 1) \left(\frac{a}{N_{\mathbf{p}} - 1} + \frac{c}{N_{\mathbf{q}}} + \frac{\alpha}{(N_{\mathbf{p}} - 1)^2} + \frac{\gamma}{N_{\mathbf{q}}^2} + O(N^{-3}) \right)$$

$$= B + \frac{N_{\mathbf{p}} c}{N_{\mathbf{q}}} - \frac{(N_{\mathbf{p}} - 1)c}{N_{\mathbf{q}}} + \frac{\alpha}{N_{\mathbf{p}}} - \frac{\alpha}{N_{\mathbf{p}} - 1} + \frac{N_{\mathbf{p}} \gamma}{N_{\mathbf{q}}^2} - \frac{(N_{\mathbf{p}} - 1)\gamma}{N_{\mathbf{q}}^2}$$

$$= B + \frac{c}{N_{\mathbf{q}}} + \frac{\alpha}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)} + \frac{\gamma}{N_{\mathbf{q}}^2} + O(N^{-3})$$

□

Theorem 14.

$$E(\hat{B}_{JK_2}) - B = \frac{\alpha}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)} + \frac{\gamma}{N_{\mathbf{q}}(N_{\mathbf{q}} - 1)} + O(N^{-3})$$

In other words, $\hat{B}_{JK_2}$ *is* $O(N^{-2})$.

Proof. Using (3.19) and Lemma 1,

$$E(\hat{B}_{JK_2}) = N_{\mathbf{q}}E(\hat{B}_{2\mathbf{p}}) - (N_{\mathbf{q}} - 1)E(\hat{B}_{2\mathbf{p}(-j)})$$

$$= N_{\mathbf{q}}\left(B + \frac{c}{N_{\mathbf{q}}} + \frac{\alpha}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)} + \frac{\gamma}{N_{\mathbf{q}}^2} + O(N^{-3})\right)$$

$$- (N_{\mathbf{q}} - 1)\left(B + \frac{c}{N_{\mathbf{q}} - 1} + \frac{\alpha}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)} + \frac{\gamma}{(N_{\mathbf{q}} - 1)^2} + O(N^{-3})\right)$$

$$= B + \frac{\alpha N_{\mathbf{q}}}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)} + \frac{\gamma}{N_{\mathbf{q}}} - \frac{(N_{\mathbf{q}} - 1)\alpha}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)} - \frac{\gamma}{N_{\mathbf{q}} - 1} + O(N^{-3})$$

$$= B + \frac{\alpha}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)} + \frac{\gamma}{N_{\mathbf{q}}(N_{\mathbf{q}} - 1)} + O(N^{-3})$$

Therefore

$$E(\hat{B}_{JK_2}) - B = \frac{\alpha}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)} + \frac{\gamma}{N_{\mathbf{q}}(N_{\mathbf{q}} - 1)} + O(N^{-3}) = O(N^{-2})$$

□

Definition 7. *Define the new, bias-adjusted estimator for Jensen-Shannon divergence in the two-sample context as*

$$\widehat{JS}_{BA_2} = \hat{A}_{JK_2} + \hat{B}_{JK_2} \tag{3.22}$$

The next corollary follows immediately from Theorems 13 and 14.

Corollary 2. *The bias of the adjusted estimator* $\widehat{JS}_{BA_2}$ *is asymptotically* $O(N^{-2})$.

CHAPTER 4: ASYMPTOTIC PROPERTIES OF ESTIMATORS

4.1 One-Sample

For finite K, the asymptotic normality of the one-sample plug-in $\hat{A}_1^0 + \hat{B}_1^0$ is easily derived. Let

$$a(\mathbf{q}) = \nabla A(\mathbf{q}) = \left(\frac{\partial}{\partial q_1} A(\mathbf{q}), \ldots, \frac{\partial}{\partial q_{K-1}} A(\mathbf{q})\right)$$

and

$$b(\mathbf{q}) = \nabla B(\mathbf{q}) = \left(\frac{\partial}{\partial q_1} B(\mathbf{q}), \ldots, \frac{\partial}{\partial q_{K-1}} B(\mathbf{q})\right)$$

denote the gradients of $A(\mathbf{q})$ and $B(\mathbf{q})$ respectively, and let

$$(a+b)(\mathbf{q}) = \nabla (A+B)(\mathbf{q}) = \left(\frac{\partial}{\partial q_1} (A+B)(\mathbf{q}), \ldots, \frac{\partial}{\partial q_{K-1}} (A+B)(\mathbf{q})\right) \tag{4.1}$$

be the gradient of $(A+B)(\mathbf{q})$ where, for $1 \le k \le K-1$

$$\frac{\partial}{\partial q_k}(A+B)(\mathbf{q}) = \frac{1}{2}\left(\ln\left(\frac{q_k}{q_K}\right) - \ln\left(\frac{p_k+q_k}{p_K+q_K}\right)\right)$$

The partial derivatives are derived in the Appendix, Lemma 14.

We know that $\hat{\mathbf{q}} \xrightarrow{p} \mathbf{q}$ as $n \to \infty$ and so by the multivariate normal approximation to the multinomial distribution,

$$\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q}) \xrightarrow{L} MVN(0, \Sigma(\mathbf{q}))$$

where $\Sigma(\mathbf{q})$ is a $(K-1)\times(K-1)$ covariance matrix given by

$$\Sigma(\mathbf{q}) = \begin{pmatrix} q_1(1-q_1) & -q_1q_2 & \dots & -q_1q_{K-1} \\ -q_2q_1 & q_2(1-q_2) & \dots & -q_2q_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ -q_{K-1}q_1 & -q_{K-1}q_2 & \dots & q_{K-1}(1-q_{K-1}) \end{pmatrix} \tag{4.2}$$

Using the delta method, we obtain the following theorem.

Theorem 15. *Provided that* $(a+b)^\tau(\mathbf{q})\Sigma(\mathbf{q})(a+b)(\mathbf{q}) > 0$,

$$\frac{\sqrt{N}((\hat{A}_1^0 + \hat{B}_1^0) - (A+B))}{\sqrt{(a+b)^\tau(\mathbf{q})\Sigma(\mathbf{q})(a+b)(\mathbf{q})}} \xrightarrow{L} N(0,1) \tag{4.3}$$

Next we show that $\hat{A}_{JK_1}$ and $\hat{B}_{JK_1}$ are sufficiently close to $\hat{A}_1^0$ and $\hat{B}_1^0$ asymptotically, so that we can also show that the asymptotic normality of $\widehat{JS}_{BA_1}$ holds when $(a+b)^\tau(\mathbf{q})\Sigma(\mathbf{q})(a+b)(\mathbf{q}) > 0$. The following lemma is used toward proving that $\sqrt{N}(\hat{A}_{JK_1} - \hat{A}_1^0) \xrightarrow{p} 0$.

Lemma 2.

$$\hat{A}_{JK_{1\mathbf{q}}} - \hat{A}^0_{1\mathbf{q}}$$

$$= -\frac{1}{4(N-1)}\sum_{k=1}^{K-1}(q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times\left(\frac{1}{q_K + O(N^{-1/2})} + \frac{1}{q_k + O(N^{-1/2})}\right)$$

$$+ \frac{1}{4(N-1)}\sum_{m\neq n}(q_n + O(N^{-1/2}))(q_m + O(N^{-1/2}))\left(\frac{1}{q_K + O(N^{-1/2})}\right)$$

Proof. For any vector η_j between $\hat{\mathbf{q}}^{(-j)}$ and $\hat{\mathbf{q}}$, using Taylor Series expansion we have

$$A_{1\mathbf{q}}\left(\hat{\mathbf{q}}^{(-j)}\right) - A_{1\mathbf{q}}\left(\hat{\mathbf{q}}\right)$$

$$= \left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^{\tau} \nabla A_{1\mathbf{q}}\left(\hat{\mathbf{q}}\right) + \frac{1}{2}\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^{\tau} \nabla^2 A_{1\mathbf{q}}\left(\eta_j\right)\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)$$

For any j, we can write

$$\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^{\tau} = \left\{\left(\frac{Y_1 - NI[\omega_j = \ell_1]}{N(N-1)}\right), \ldots, \left(\frac{Y_{K-1} - NI[\omega_j = \ell_{K-1}]}{N(N-1)}\right)\right\} \tag{4.4}$$

$$= \frac{1}{N-1}\left\{\hat{q}_1 - I[\omega_j = \ell_1], \ldots, \hat{q}_{K-1} - I[\omega_j = \ell_{K-1}]\right\}$$

Note that $\nabla A_{1\mathbf{q}}\left(\hat{\mathbf{q}}\right)$ is a gradient vector equivalent to

$$\frac{1}{2}\left\{\ln\left(\frac{\hat{q}_1}{\hat{q}_K}\right), \ldots, \ln\left(\frac{\hat{q}_{K-1}}{\hat{q}_K}\right)\right\}$$

and so

$$\sum_{j=1}^{N} \left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^{\tau} \nabla A_{1\mathbf{q}}\left(\hat{\mathbf{q}}\right)$$

$$= \frac{1}{2} \sum_{k=1}^{K-1} \ln\left(\frac{\hat{q}_k}{\hat{q}_K}\right) \sum_{j=1}^{N} \frac{Y_k - NI[\omega_j = \ell_k]}{N(N-1)}$$

$$= \frac{1}{2(N-1)} \sum_{k=1}^{K-1} \ln\left(\frac{\hat{q}_k}{\hat{q}_K}\right) \sum_{j=1}^{N} (\hat{q}_k - I[\omega_j = \ell_k])$$

$$= \frac{1}{2(N-1)} \sum_{k=1}^{K-1} \ln\left(\frac{\hat{q}_k}{\hat{q}_K}\right) \left(N\hat{q}_k - \sum_{j=1}^{N} I[\omega_j = \ell_k]\right)$$

$$= \frac{1}{2(N-1)} \sum_{k=1}^{K-1} \ln\left(\frac{\hat{q}_k}{\hat{q}_K}\right) (Y_k - Y_k) = 0$$

Note that for any j, $1 \leq j \leq N$,

$$\nabla^2 A_{1\mathbf{q}}(\eta_{\mathbf{j}})$$

$$= \frac{1}{2} \begin{pmatrix} \left(\frac{1}{\eta_{j,1}} + \frac{1}{\eta_{j,K}}\right) & \frac{1}{\eta_{j,K}} & \cdots & \frac{1}{\eta_{j,K}} \\ \frac{1}{\eta_{j,K}} & \left(\frac{1}{\eta_{j,2}} + \frac{1}{\eta_{j,K}}\right) & \cdots & \frac{1}{\eta_{j,K}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\eta_{j,K}} & \frac{1}{\eta_{j,K}} & \cdots & \left(\frac{1}{\eta_{j,K-1}} + \frac{1}{\eta_{j,K}}\right) \end{pmatrix}_{(K-1)\times(K-1)}$$

where $\eta_{j,k}$ and $\eta_{j,K}$ are the corresponding elements of the η_j vector. This gives rise to

$$\frac{1}{2}\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^{\tau} \nabla^2 A_{1\mathbf{q}}\left(\eta_j\right)\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)$$

$$= \frac{1}{4(N-1)^2}\left(\frac{\left(\sum_{k=1}^{K-1}\hat{q}_k - I[\omega_j = \ell_k]\right)^2}{\eta_{j,K}} + \sum_{k=1}^{K-1}\frac{\left(\hat{q}_k^2 - I[\omega_j = \ell_k]\right)^2}{\eta_{j,k}}\right)$$

Recall the well known fact that

$$\left(\sum_{k=1}^{K-1}\hat{q}_k - I[\omega_j = \ell_k]\right)^2 = \sum_{k=1}^{K-1}\left(\hat{q}_k - I[\omega_j = \ell_k]\right)^2 + \sum_{m\neq n}(\hat{q}_n - I[\omega_j = \ell_n])(\hat{q}_m - I[\omega_j = \ell_m]) \tag{4.5}$$

Therefore we can write

$$\hat{A}_{JK_{1\mathbf{q}}} = \hat{A}^0_{1\mathbf{q}} - \frac{N-1}{N}\sum_{j=1}^{N}\left(\hat{A}_{1\mathbf{q}(-j)} - \hat{A}^0_{1\mathbf{q}}\right)$$

$$= \hat{A}^0_{1\mathbf{q}} - \frac{N-1}{N}\sum_{j=1}^{N}\left(\frac{1}{2}\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^{\tau} \nabla^2 A_{1\mathbf{q}}\left(\eta_j\right)\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)\right)$$

$$= \hat{A}^0_{1\mathbf{q}} - \frac{1}{4N(N-1)}\sum_{j=1}^{N}\frac{\sum_{k=1}^{K-1}\left(\hat{q}_k - I[\omega_j = \ell_k]\right)^2}{\eta_{j,K}}$$

$$- \frac{1}{4N(N-1)}\sum_{j=1}^{N}\frac{\sum_{m\neq n}(\hat{q}_n - I[\omega_j = \ell_n])(\hat{q}_m - I[\omega_j = \ell_m])}{\eta_{j,K}}$$

$$- \frac{1}{4N(N-1)}\sum_{j=1}^{N}\sum_{k=1}^{K-1}\frac{\left(\hat{q}_k^2 - I[\omega_j = \ell_k]\right)^2}{\eta_{j,k}}$$

$$= \hat{A}^0_{1\mathbf{q}} - \frac{1}{4N(N-1)}\sum_{k=1}^{K-1}\sum_{j=1}^{N}\left(\hat{q}_k^2 - I[\omega_j = \ell_k]\right)^2\left(\frac{1}{\eta_{j,K}} + \frac{1}{\eta_{j,k}}\right)$$

$$- \frac{1}{4N(N-1)}\sum_{m\neq n}\sum_{j=1}^{N}\frac{(\hat{q}_n - I[\omega_j = \ell_n])(\hat{q}_m - I[\omega_j = \ell_m])}{\eta_{j,K}}$$

$$= \hat{A}^0_{1\mathbf{q}} - \frac{1}{4N(N-1)}\sum_{k=1}^{K-1}\left(Y_k(\hat{q}_k - 1)^2 + (N - Y_k)\hat{q}_k^2\right)$$

$$\times\left(\frac{1}{q_K + O(N^{-1/2})} + \frac{1}{q_k + O(N^{-1/2})}\right)$$

$$- \frac{1}{4N(N-1)}\sum_{m\neq n}\left(Y_m(\hat{q}_m - 1)\hat{q}_n + Y_n(\hat{q}_n - 1)\hat{q}_m + (N - Y_m - Y_n)\hat{q}_n\hat{q}_m\right)$$

$$\times\left(\frac{1}{q_K + O(N^{-1/2})}\right)$$

Taking the $\frac{1}{N}$ inside yields

$$\hat{A}^0_{1\mathbf{q}} - \frac{1}{4(N-1)} \sum_{k=1}^{K-1} \left(\hat{q}_k(\hat{q}_k - 1)^2 + (1-\hat{q}_k)\hat{q}_k^2\right) \left(\frac{1}{q_K + O(N^{-1/2})} + \frac{1}{q_k + O(N^{-1/2})}\right)$$

$$- \frac{1}{4(N-1)} \sum_{m \neq n} ((\hat{q}_m - 1)\hat{q}_n\hat{q}_m + (\hat{q}_n - 1)\hat{q}_n\hat{q}_m + (1 - \hat{q}_m - \hat{q}_n)\hat{q}_n\hat{q}_m)$$

$$\times \left(\frac{1}{q_K + O(N^{-1/2})}\right)$$

$$= \hat{A}^0_{1\mathbf{q}} - \frac{1}{4(N-1)} \sum_{k=1}^{K-1} \hat{q}_k(1-\hat{q}_k) \left(\frac{1}{q_K + O(N^{-1/2})} + \frac{1}{q_k + O(N^{-1/2})}\right)$$

$$+ \frac{1}{4(N-1)} \sum_{m \neq n} \hat{q}_n\hat{q}_m \left(\frac{1}{q_K + O(N^{-1/2})}\right)$$

$$= \hat{A}^0_{1\mathbf{q}}$$

$$- \frac{1}{4(N-1)} \sum_{k=1}^{K-1} (q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{q_K + O(N^{-1/2})} + \frac{1}{q_k + O(N^{-1/2})}\right)$$

$$+ \frac{1}{4(N-1)} \sum_{m \neq n} (q_n + O(N^{-1/2}))(q_m + O(N^{-1/2})) \left(\frac{1}{q_K + O(N^{-1/2})}\right)$$

□

Lemma 3.

$$\sqrt{N}(\hat{A}_{JK_1} - \hat{A}_1^0) \xrightarrow{p} 0 \tag{4.6}$$

Proof.

$$\sqrt{N}(\hat{A}_{JK_1} - \hat{A}_1^0) = \sqrt{N}\left(\hat{A}_{JK_{1\mathbf{q}}} - \hat{A}_1^0\right)$$

From Lemma 2, we have

$$\sqrt{N}\left(\hat{A}_{JK_{1\mathbf{q}}} - \hat{A}_1^0\right)$$

$$= -\frac{\sqrt{N}}{4(N-1)}\sum_{k=1}^{K-1}(q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times\left(\frac{1}{q_K + O(N^{-1/2})} + \frac{1}{q_k + O(N^{-1/2})}\right)$$

$$+ \frac{\sqrt{N}}{4(N-1)}\sum_{m \neq n}(q_n + O(N^{-1/2}))(q_m + O(N^{-1/2}))\left(\frac{1}{q_K + O(N^{-1/2})}\right)$$

$$= O(N^{-1/2}) \to 0$$

□

The following lemma is used toward proving that $\sqrt{N}(\hat{B}_{JK_1} - \hat{B}_1^0) \xrightarrow{p} 0$.

Lemma 4.

$$\hat{B}_{JK_1} - \hat{B}_1^0 = \frac{1}{4(N-1)} \sum_{k=1}^{K-1} (q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{1}{4(N-1)} \sum_{m \neq n} (q_n + O(N^{-1/2}))(q_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right) \tag{4.7}$$

Proof. For any vector η_j between $\hat{\mathbf{q}}^{(-j)}$ and $\hat{\mathbf{q}}$, it is true that

$$B\left(\hat{\mathbf{q}}^{(-j)}\right) - B\left(\hat{\mathbf{q}}\right)$$

$$= \left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^\tau \nabla B\left(\hat{\mathbf{q}}\right) + \frac{1}{2}\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^\tau \nabla^2 B\left(\eta_j\right)\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)$$

Note that $\nabla B\left(\hat{\mathbf{q}}\right)$ is a gradient vector such that

$$-\frac{1}{2}\left\{\ln\left(\frac{p_1 + \hat{q}_1}{p_K + \hat{q}_K}\right), \ldots, \ln\left(\frac{p_{K-1} + \hat{q}_{K-1}}{p_K + \hat{q}_K}\right)\right\}$$

and so, again using (4.4),

$$\sum_{j=1}^{N} \left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^\tau \nabla B\left(\hat{\mathbf{q}}\right)$$

$$= -\frac{1}{2}\sum_{k=1}^{K-1} \ln\left(\frac{p_k + \hat{q}_k}{p_K + \hat{q}_K}\right)\sum_{j=1}^{N} \frac{Y_k - NI[\omega_j = \ell_k]}{N(N-1)}$$

$$= -\frac{1}{2(N-1)}\sum_{k=1}^{K-1} \ln\left(\frac{p_k + \hat{q}_k}{p_K + \hat{q}_K}\right)\sum_{j=1}^{N}(\hat{q}_k - I[\omega_j = \ell_k])$$

$$= -\frac{1}{2(N-1)}\sum_{k=1}^{K-1} \ln\left(\frac{p_k + \hat{q}_k}{p_K + \hat{q}_K}\right)\left(N\hat{q}_k - \sum_{j=1}^{N} I[\omega_j = \ell_k]\right)$$

$$= -\frac{1}{2(N-1)}\sum_{k=1}^{K-1} \ln\left(\frac{p_k + \hat{q}_k}{p_K + \hat{q}_K}\right)(Y_k - Y_k) = 0$$

Next, we see that for any j, $1 \leq j \leq N$,

$$\frac{1}{2}\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^{\tau} \nabla^2 B\left(\eta_j\right)\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)$$

$$= -\frac{1}{4(N-1)^2}\left(\frac{\left(\sum_{k=1}^{K-1} \hat{q}_k - I[\omega_j = \ell_k]\right)^2}{p_K + \eta_{j,K}} + \sum_{k=1}^{K-1} \frac{\left(\hat{q}_k^2 - I[\omega_j = \ell_k]\right)^2}{p_k + \eta_{j,k}}\right)$$

where $\eta_{j,k}$ and $\eta_{j,K}$ are the corresponding elements of the η_j vector. Again using the well known fact from (4.5),

$$\hat{B}_{JK_1} = \hat{B}_1^0 - \frac{N-1}{N}\sum_{j=1}^{N}\left(\hat{B}_{1(-j)} - \hat{B}_1^0\right)$$

$$= \hat{B}_1^0 - \frac{N-1}{N}\sum_{j=1}^{N}\left(\frac{1}{2}\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)^{\tau} \nabla^2 B\left(\eta_j\right)\left(\hat{\mathbf{q}}^{(-j)} - \hat{\mathbf{q}}\right)\right)$$

$$= \hat{B}_1^0 + \frac{1}{4N(N-1)} \sum_{j=1}^{N} \frac{\sum_{k=1}^{K-1} \left(\hat{q}_k - I[\omega_j = \ell_k]\right)^2}{p_K + \eta_{j,K}}$$

$$+ \frac{1}{4N(N-1)} \sum_{j=1}^{N} \frac{\sum_{m \neq n} (\hat{q}_n - I[\omega_j = \ell_n])(\hat{q}_m - I[\omega_j = \ell_m])}{p_K + \eta_{j,K}}$$

$$+ \frac{1}{4N(N-1)} \sum_{j=1}^{N} \sum_{k=1}^{K-1} \frac{\left(\hat{q}_k^2 - I[\omega_j = \ell_k]\right)^2}{p_k + \eta_{j,k}}$$

$$= \hat{B}_1^0 + \frac{1}{4N(N-1)} \sum_{k=1}^{K-1} \sum_{j=1}^{N} \left(\hat{q}_k^2 - I[\omega_j = \ell_k]\right)^2 \left(\frac{1}{p_K + \eta_{j,K}} + \frac{1}{p_k + \eta_{j,k}}\right)$$

$$+ \frac{1}{4N(N-1)} \sum_{m \neq n} \sum_{j=1}^{N} \frac{(\hat{q}_n - I[\omega_j = \ell_n])(\hat{q}_m - I[\omega_j = \ell_m])}{p_K + \eta_{j,K}}$$

$$= \hat{B}_1^0 + \frac{1}{4N(N-1)} \sum_{k=1}^{K-1} \left(Y_k(\hat{q}_k - 1)^2 + (N - Y_k)\hat{q}_k^2\right)$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})}\right)$$

$$+ \frac{1}{4N(N-1)} \sum_{m \neq n} (Y_m(\hat{q}_m - 1)\hat{q}_n + Y_n(\hat{q}_n - 1)\hat{q}_m + (N - Y_m - Y_n)\hat{q}_n\hat{q}_m)$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})}\right)$$

Taking the $\frac{1}{N}$ inside yields

$$\hat{B}_1^0 + \frac{1}{4(N-1)} \sum_{k=1}^{K-1} \left(\hat{q}_k(\hat{q}_k - 1)^2 + (1 - \hat{q}_k)\hat{q}_k^2 \right)$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$+ \frac{1}{4(N-1)} \sum_{m \neq n} ((\hat{q}_m - 1)\hat{q}_n\hat{q}_m + (\hat{q}_n - 1)\hat{q}_n\hat{q}_m + (1 - \hat{q}_m - \hat{q}_n)\hat{q}_n\hat{q}_m)$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$$= \hat{B}_1^0 + \frac{1}{4(N-1)} \sum_{k=1}^{K-1} \hat{q}_k(1 - \hat{q}_k) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{1}{4(N-1)} \sum_{m \neq n} \hat{q}_n\hat{q}_m \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$$= \hat{B}_1^0 + \frac{1}{4(N-1)} \sum_{k=1}^{K-1} (q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{1}{4(N-1)} \sum_{m \neq n} (q_n + O(N^{-1/2}))(q_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$\square$

Now that this is established, we use it to show the following.

Lemma 5.

$$\sqrt{N}(\hat{B}_{JK_1} - \hat{B}_1^0) \xrightarrow{p} 0 \tag{4.8}$$

Proof. From Lemma 4, we have

$$\sqrt{N}(\hat{B}_{JK_1} - \hat{B}_1^0)$$

$$= \frac{\sqrt{N}}{4(N-1)} \sum_{k=1}^{K-1} (q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{\sqrt{N}}{4(N-1)} \sum_{m \neq n} (q_n + O(N^{-1/2}))(q_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$$= O(N^{-1/2}) \to 0$$

□

Putting together Theorem 15, Lemmas 6 and 7 and Slutzky's theorem, the next theorem follows immediately to yield the asymptotic normality of $\widehat{JS}_{BA_1}$.

Theorem 16. *Provided that* $(a+b)^\tau(\mathbf{q})\Sigma(\mathbf{q})(a+b)(\mathbf{q}) > 0$,

$$\frac{\sqrt{N}((\hat{A}_{JK_1} + \hat{B}_{JK_1}) - (A+B))}{\sqrt{(a+b)^\tau(\mathbf{q})\Sigma(\mathbf{q})(a+b)(\mathbf{q})}} \xrightarrow{L} N(0,1) \tag{4.9}$$

Corollary 3. *For the vector defined as in (4.1),*

$$(a+b)(\mathbf{q}) = 0$$

if and only if $\mathbf{p} = \mathbf{q}$.

Proof. Note that $(a+b)(\mathbf{q}) = 0$ if and only if each component of the vector is zero, and so we proceed with the proof component-wise. From Lemma 14, for any k, $1 \leq k \leq K-1$,

$$\frac{\partial}{\partial q_k}(A+B)(\mathbf{q}) = \frac{1}{2}\left(\ln\left(\frac{q_k}{q_K}\right) - \ln\left(\frac{p_k+q_k}{p_K+q_K}\right)\right) \tag{4.10}$$

($\Rightarrow$) Suppose (4.16) is zero for all k, $1 \leq k \leq K-1$. Then we must have

$$\frac{q_k}{q_K} = \frac{p_k+q_k}{p_K+q_K}$$

for all k, $1 \leq k \leq K-1$. This implies

$$q_k(p_K+q_K) = q_K(p_k+q_k)$$

$$p_K q_k = p_k q_K \tag{4.11}$$

$$\frac{p_k}{p_K} = \frac{q_k}{q_K}$$

which implies

$$\sum_{k=1}^{K} \frac{p_k}{p_K} = \sum_{k=1}^{K} \frac{q_k}{q_K}$$

and so

$$\frac{1}{p_K} = \frac{1}{q_K}$$

which means $p_K = q_K$. Plugging that back into (4.11) yields $p_k = q_k$ for $1 \leq k \leq K-1$.

($\Leftarrow$) Now suppose that $p_k = q_k$ for all k. Then

$$\frac{p_k + q_k}{p_K + q_K} = \frac{2p_k}{2p_K} = \frac{p_k}{p_K}$$

which renders (4.1) zero.

$\square$

This means that the asymptotic normality of $\widehat{JS}_{BA_1}$ breaks down if and only if $\mathbf{p} = \mathbf{q}$. Thus we move toward finding the asymptotic behavior in this case. Throughout, recall that Jensen-Shannon Divergence is 0 when $\mathbf{p} = \mathbf{q}$. We begin with the plug-in estimator.

Theorem 17. *When* $\mathbf{p} = \mathbf{q}$,

$$N\left(\hat{A}_1^0 + \hat{B}_1^0\right) \xrightarrow{L} \frac{1}{8}\chi^2_{K-1}$$

Proof. By Taylor Series Expansion,

$$N\left(\hat{A}_1^0 + \hat{B}_1^0\right) = N(A+B)(\hat{\mathbf{q}})$$

$$= N(A+B)(\mathbf{q}) + N(\hat{\mathbf{q}}-\mathbf{q})^\tau \nabla(A+B)(\mathbf{q}) + \frac{1}{2}\sqrt{N}(\hat{\mathbf{q}}-\mathbf{q})^\tau \nabla^2(A+B)(\mathbf{q})\sqrt{N}(\hat{\mathbf{q}}-\mathbf{q}) + O(N^{-1/2})$$

Since $\mathbf{p} = \mathbf{q}$, $(A+B)(\mathbf{q}) = 0$ by Theorem 1, and $\nabla(A+B)(\mathbf{q}) = (a+b)(\mathbf{q}) = 0$ by

Corollary 3. Obviously the $O(N^{-1/2})$ term goes to 0 in probability. Thus the only term we are left to contend with is

$$\frac{1}{2}\sqrt{N}(\hat{\mathbf{q}}-\mathbf{q})^{\tau}\nabla^2((A+B)(\mathbf{q}))\sqrt{N}(\hat{\mathbf{q}}-\mathbf{q}) \tag{4.12}$$

Using the multivariate normal approximation to the multinomial distribution, we have

$$\sqrt{N}(\hat{\mathbf{q}}-\mathbf{q}) \xrightarrow{L} MVN(0,\Sigma(\mathbf{q})) \tag{4.13}$$

where $\Sigma(\mathbf{q})$ is as in (4.2). Putting together (4.13) and Slutsky's Theorem, we have

$$\sqrt{N}(\hat{\mathbf{q}}-\mathbf{q})\Sigma(\mathbf{q})^{-1/2} \xrightarrow{L} MVN(0,\mathbf{I}_{K-1}) := \mathbf{Z_1} \tag{4.14}$$

Noting this fact, we rewrite (4.12) as

$$\frac{1}{2}\sqrt{N}\left(\Sigma(\mathbf{q})^{-1/2}(\hat{\mathbf{q}}-\mathbf{q})\right)^{\tau}\Sigma(\mathbf{q})^{1/2}\nabla^2(A+B)(\mathbf{q})\Sigma(\mathbf{q})^{1/2}\sqrt{N}\left(\Sigma(\mathbf{q})^{-1/2}(\hat{\mathbf{q}}-\mathbf{q})\right)$$

Because we know (4.14), this leaves us with finding the asymptotic behavior of

$$\Sigma(\mathbf{q})^{1/2}\nabla^2(A+B)(\mathbf{q})\Sigma(\mathbf{q})^{1/2} \tag{4.15}$$

Let

$$\nabla^2(A+B)(\mathbf{q}) = \Theta(\mathbf{q})$$

where

$$\Theta(\mathbf{q}) = \frac{1}{4}\begin{pmatrix} \frac{1}{q_1}+\frac{1}{q_K} & \frac{1}{q_K} & \dots & \frac{1}{q_K} \\ \frac{1}{q_K} & \frac{1}{q_2}+\frac{1}{q_K} & \dots & \frac{1}{q_K} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{q_K} & \frac{1}{q_K} & \dots & \frac{1}{q_{K-1}}+\frac{1}{q_K} \end{pmatrix}_{(K-1)\times(K-1)}$$

First, we show that

$$\Sigma(\mathbf{q})^{1/2}\Theta(\mathbf{q})\Sigma(\mathbf{q})^{1/2} = \frac{1}{4}\mathbf{I}_{K-1}$$

This is equivalent to showing that

$$(4\Theta(\mathbf{q}))^{-1} = \Sigma(\mathbf{q})$$

To do this, we must use Lemma 16, written in the Appendix.

$$4\Theta(\mathbf{q}) = \begin{pmatrix} \frac{1}{q_1} & 0 & \dots & 0 \\ 0 & \frac{1}{q_2} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{1}{q_{K-1}} \end{pmatrix}_{(K-1)\times(K-1)} + \begin{pmatrix} \frac{1}{q_K} & \frac{1}{q_K} & \dots & \frac{1}{q_K} \\ \frac{1}{q_K} & \frac{1}{q_K} & \dots & \frac{1}{q_K} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{q_K} & \frac{1}{q_K} & \dots & \frac{1}{q_K} \end{pmatrix}_{(K-1)\times(K-1)} =: \mathbf{G}+\mathbf{H}$$

Because all of the rows in $\mathbf{H}$ are equivalent, $\mathbf{H}$ has rank 1. The inverse of $\mathbf{G}$ is clearly

$$\mathbf{G}^{-1} = \begin{pmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & q_{K-1} \end{pmatrix}_{(K-1)\times(K-1)}$$

which greatly simplifies things. Next we need to find $g = tr\{\mathbf{HG}^{-1}\}$ and verify that it can never be -1 so that $(A.10)$ is never undefined.

$$g = tr\{\mathbf{HG}^{-1}\} = tr\left(\frac{1}{q_K} \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}_{(K-1)\times(K-1)} \begin{pmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & q_{K-1} \end{pmatrix}_{(K-1)\times(K-1)} \right)$$

$$= tr\left(\frac{1}{q_K} \begin{pmatrix} q_1 & q_2 & \dots & q_{K-1} \\ q_1 & q_2 & \dots & q_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ q_1 & q_2 & \dots & q_{K-1} \end{pmatrix}_{(K-1)\times(K-1)} \right)$$

$$= \frac{1}{q_K}\left(\sum_{k=1}^{K-1} q_k\right) = \frac{1-q_K}{q_K}$$

which can never be -1. Using this value to further work towards calculating $(A.10)$, we have

$$\frac{1}{1+g} = q_K$$

Next we need to find $\mathbf{G}^{-1}\mathbf{H}\mathbf{G}$:

$$\begin{pmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & q_{K-1} \end{pmatrix}_{(K-1)\times(K-1)} \frac{1}{q_K}\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix}_{(K-1)\times(K-1)} \begin{pmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & q_{K-1} \end{pmatrix}_{(K-1)\times(K-1)}$$

$$= \frac{1}{q_K}\begin{pmatrix} q_1^2 & q_1q_2 & \dots & q_1q_{K-1} \\ q_2q_1 & q_2^2 & \dots & q_2q_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ q_{K-1}q_1 & 0 & \dots & q_{K-1}^2 \end{pmatrix}_{(K-1)\times(K-1)}$$

Thus

$$\Theta(\mathbf{q})^{-1} = \mathbf{G}^{-1} - \frac{1}{1+g}\mathbf{G}^{-1}\mathbf{H}\mathbf{G}^{-1}$$

$$= \begin{pmatrix} q_1 & 0 & \dots & 0 \\ 0 & q_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & q_{K-1} \end{pmatrix}_{(K-1)\times(K-1)} - \begin{pmatrix} q_1^2 & q_1 q_2 & \dots & q_1 q_{K-1} \\ q_2 q_1 & q_2^2 & \dots & q_2 q_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ q_{K-1} q_1 & 0 & \dots & q_{K-1}^2 \end{pmatrix}_{(K-1)\times(K-1)}$$

$$= \Sigma(\mathbf{q})$$

as desired. Therefore

$$\Sigma(\mathbf{q})^{1/2}\nabla^2(A+B)(\mathbf{q})\Sigma(\mathbf{q})^{1/2} = \mathbf{I}_{K-1}$$

Thus we have

$$(4.12) = \frac{1}{2}\left(\sqrt{N}\Sigma(\mathbf{q})^{-1/2}(\hat{\mathbf{q}} - \mathbf{q})\right)^{\tau} \frac{1}{4}\mathbf{I}_{K-1}\left(\sqrt{N}\Sigma(\mathbf{q})^{-1/2}(\hat{\mathbf{q}} - \mathbf{q})\right)$$

$$= \frac{1}{8}\left(\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q})\Sigma(\mathbf{q})^{-1/2}\right)^{\tau}\left(\sqrt{N}(\hat{\mathbf{q}} - \mathbf{q})\Sigma(\mathbf{q})^{-1/2}\right) \xrightarrow{L} \frac{1}{8}\sum_{i=1}^{K-1}\mathbf{Z}_{1i}^2$$

by the Continuous Mapping Theorem, where each $\mathbf{Z}_{1i} \sim N(0,1)$. Therefore

$$\frac{1}{2}\sqrt{N}(\hat{\mathbf{q}}-\mathbf{q})^{\tau}\nabla^2(A+B)(\mathbf{q})\sqrt{N}(\hat{\mathbf{q}}-\mathbf{q}) \xrightarrow{L} \frac{1}{8}\chi^2_{K-1}$$

as was to be shown.

□

Lemma 6. *For the one-sample case, when* $\mathbf{p}=\mathbf{q}$,

$$N(\hat{A}_{JK_1}-\hat{A}_1^0) \xrightarrow{p} -\frac{1}{4}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right)-\sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

Proof. Using Theorem 2, we have

$$N\left(\hat{A}_{JK_{1\mathbf{q}}}-\hat{A}_1^0\right)$$

$$= -\frac{N}{4(N-1)}\sum_{k=1}^{K-1}(q_k+O(N^{-1/2}))(1-q_k+O(N^{-1/2}))$$

$$\times\left(\frac{1}{q_K+O(N^{-1/2})}+\frac{1}{q_k+O(N^{-1/2})}\right)$$

$$+\frac{N}{4(N-1)}\sum_{m\neq n}(q_n+O(N^{-1/2}))(q_m+O(N^{-1/2}))\left(\frac{1}{q_K+O(N^{-1/2})}\right)$$

$$\to -\frac{1}{4}\left(\sum_{k=1}^{K-1} q_k(1-q_k)\left(\frac{1}{q_K}+\frac{1}{q_k}\right)-\sum_{m\neq n}\frac{q_n q_m}{q_K}\right)$$

Since $\mathbf{p}=\mathbf{q}$, this is equivalent to

$$-\frac{1}{4}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right)-\sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

□

Lemma 7. *For the one-sample case, when* $\mathbf{p} = \mathbf{q}$,

$$N(\hat{B}_{JK_1} - \hat{B}_1^0) \xrightarrow{p} \frac{1}{8}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K} + \frac{1}{p_k}\right) - \sum_{m \neq n} \frac{p_n p_m}{p_K}\right)$$

Proof. From Lemma 4, we have that

$$N(\hat{B}_{JK_1} - \hat{B}_1^0)$$

$$= \frac{N}{4(N-1)} \sum_{k=1}^{K-1} (q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})}\right)$$

$$- \frac{N}{4(N-1)} \sum_{m \neq n} (q_n + O(N^{-1/2}))(q_m + O(N^{-1/2}))\left(\frac{1}{p_K + q_K + O(N^{-1/2})}\right)$$

$$\xrightarrow{p} \frac{1}{4}\sum_{k=1}^{K-1} q_k(1-q_k)\left(\frac{1}{p_K + q_K} + \frac{1}{p_k + q_k}\right) - \frac{1}{4}\sum_{m \neq n} q_n q_m \left(\frac{1}{p_K + q_K}\right)$$

Since $\mathbf{p} = \mathbf{q}$, this is equivalent to

$$\frac{1}{8}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K} + \frac{1}{p_k}\right) - \sum_{m \neq n} \frac{p_n p_m}{p_K}\right)$$

as desired. □

Lemmas 6 and 7 directly yield the following Corollary.

Corollary 4. *When* $\mathbf{p} = \mathbf{q}$ *in the one-sample case,*

$$N((\hat{A}_{JK_1} + \hat{B}_{JK_1}) - (\hat{A}_1^0 + \hat{B}_1^0)) \xrightarrow{p} -\frac{1}{8}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K} + \frac{1}{p_k}\right) - \sum_{m \neq n} \frac{p_n p_m}{p_K}\right)$$

By Slutsky's Theorem, Theorem 17, and Corollary 4, we have the following conclusion.

Theorem 18. *When* $\mathbf{p} = \mathbf{q}$ *in the one-sample case,*

$$N(\hat{A}_{JK_1} + \hat{B}_{JK_1}) + \frac{1}{8}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K} + \frac{1}{p_k}\right) - \sum_{m \neq n} \frac{p_n p_m}{p_K}\right) \xrightarrow{L} \frac{1}{8}\chi^2_{K-1}$$

Two-Sample

In the two-sample case for finite K, the asymptotic normality of the plug-in $\hat{A}_2^0 + \hat{B}_2^0$ is also readily derived. Toward this end we let

$$a(\mathbf{v}) = \nabla A(\mathbf{v}) = \left(\frac{\partial}{\partial p_1} A(\mathbf{v}), \ldots, \frac{\partial}{\partial p_{K-1}} A(\mathbf{v}), \frac{\partial}{\partial q_1} A(\mathbf{v}), \ldots, \frac{\partial}{\partial q_{K-1}} A(\mathbf{v})\right)$$

and

$$b(\mathbf{v}) = \nabla B(\mathbf{v}) = \left(\frac{\partial}{\partial p_1} B(\mathbf{v}), \ldots, \frac{\partial}{\partial p_{K-1}} B(\mathbf{v}), \frac{\partial}{\partial q_1} B(\mathbf{v}), \ldots, \frac{\partial}{\partial q_{K-1}} B(\mathbf{v})\right)$$

Let their sum be notated as

$$(a+b)(\mathbf{v}) = \nabla(A+B)(\mathbf{v})$$
$$= \left(\frac{\partial}{\partial p_1}(A+B)(\mathbf{v}), \ldots, \frac{\partial}{\partial p_{K-1}}(A+B)(\mathbf{v}), \frac{\partial}{\partial q_1}(A+B)(\mathbf{v}), \ldots, \frac{\partial}{\partial q_{K-1}}(A+B)(\mathbf{v})\right) \tag{4.16}$$

where, for $1 \leq k \leq K-1$

$$\frac{\partial}{\partial p_k}(A+B)(\mathbf{v}) = \frac{1}{2}\left(\ln\left(\frac{p_k}{p_K}\right) - \ln\left(\frac{p_k+q_k}{p_K+q_K}\right)\right)$$

and

$$\frac{\partial}{\partial q_k}(A+B)(\mathbf{v}) = \frac{1}{2}\left(\ln\left(\frac{q_k}{q_K}\right) - \ln\left(\frac{p_k+q_k}{p_K+q_K}\right)\right)$$

The partial derivatives are derived in the Appendix, Lemma 14. Note that $\hat{\mathbf{v}} \xrightarrow{p} \mathbf{v}$ as $n \to \infty$. By the multivariate normal approximation to the multinomial distribution

$$\sqrt{N_{\mathbf{p}}}(\hat{\mathbf{v}} - \mathbf{v}) \xrightarrow{L} MVN(0, \Sigma(\mathbf{v}))$$

where $\Sigma(\mathbf{v})$ is a $(2K-2) \times (2K-2)$ covariance matrix given by

$$\Sigma(\mathbf{v}) = \begin{pmatrix} \Sigma_{\mathbf{p}}(\mathbf{v}) & 0 \\ 0 & \Sigma_{\mathbf{q}}(\mathbf{v}) \end{pmatrix} \tag{4.17}$$

Here $\Sigma_{\mathbf{p}}(\mathbf{v})$ and $\Sigma_{\mathbf{q}}(\mathbf{v})$ are $(K-1) \times (K-1)$ matrices given by

$$\Sigma_{\mathbf{p}}(\mathbf{v}) = \begin{pmatrix} p_1(1-p_1) & -p_1 p_2 & \ldots & -p_1 p_{K-1} \\ -p_2 p_1 & p_2(1-p_2) & \ldots & -p_2 p_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ -p_{K-1} p_1 & -p_{K-1} p_2 & \ldots & p_{K-1}(1-p_{K-1}) \end{pmatrix}$$

and

$$\Sigma_{\mathbf{q}}(\mathbf{v}) = \lambda \begin{pmatrix} q_1(1-q_1) & -q_1q_2 & \dots & -q_1q_{K-1} \\ -q_2q_1 & q_2(1-q_2) & \dots & -q_2q_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ -q_{K-1}q_1 & -q_{K-1}q_2 & \dots & q_{K-1}(1-q_{K-1}) \end{pmatrix}$$

The delta method immediately yields the following theorem.

Theorem 19. *Provided that* $(a+b)^{\tau}(\mathbf{v})\Sigma(\mathbf{v})(a+b)(\mathbf{v}) > 0$,

$$\frac{\sqrt{N_{\mathbf{p}}}((\hat{A}_2^0 + \hat{B}_2^0) - (A+B))}{\sqrt{(a+b)^{\tau}(\mathbf{v})\Sigma(\mathbf{v})(a+b)(\mathbf{v})}} \xrightarrow{L} N(0,1) \tag{4.18}$$

The proof for the following lemma is almost identical to that of Lemma 2 and is therefore omitted here.

Lemma 8.

$$\hat{A}_{JK_{2\mathbf{p}}} - \hat{A}^0_{2\mathbf{p}}$$

$$= -\frac{1}{4(N_{\mathbf{p}}-1)} \sum_{k=1}^{K-1} (p_k + O(N^{-1/2}))(1 - p_k + O(N^{-1/2})) \left(\frac{1}{p_K + O(N^{-1/2})} + \frac{1}{p_k + O(N^{-1/2})} \right)$$

$$+ \frac{1}{4(N_{\mathbf{p}}-1)} \sum_{m \neq n} (p_n + O(N^{-1/2}))(p_m + O(N^{-1/2})) \left(\frac{1}{p_K + O(N^{-1/2})} \right)$$

and

$$\hat{A}_{JK_{2\mathbf{q}}} - \hat{A}^0_{2\mathbf{q}}$$

$$= -\frac{1}{4(N_{\mathbf{q}}-1)}\sum_{k=1}^{K-1}(q_k+O(N^{-1/2}))(1-q_k+O(N^{-1/2}))\left(\frac{1}{q_K+O(N^{-1/2})}+\frac{1}{q_k+O(N^{-1/2})}\right)$$

$$+\frac{1}{4(N_{\mathbf{q}}-1)}\sum_{m\neq n}(q_n+O(N^{-1/2}))(q_m+O(N^{-1/2}))\left(\frac{1}{q_K+O(N^{-1/2})}\right)$$

We now use the asymptotic normality of the plug-in estimator to obtain that of the bias-adjusted estimator.

Lemma 9.

$$\sqrt{N_{\mathbf{p}}}(\hat{A}_{JK_2}-\hat{A}_2^0)\xrightarrow{p}0 \tag{4.19}$$

Proof. Using Lemma 8,

$$\sqrt{N_{\mathbf{p}}}(\hat{A}_{JK_2} - \hat{A}_2^0) = \sqrt{N_{\mathbf{p}}}\left(\hat{A}_{JK_{2\mathbf{p}}} - \hat{A}_{2\mathbf{p}} + \hat{A}_{JK_{2\mathbf{q}}} - \hat{A}_{2\mathbf{q}}\right)$$

$$= -\frac{\sqrt{N_{\mathbf{p}}}}{4(N_{\mathbf{p}}-1)} \sum_{k=1}^{K-1} (p_k + O(N^{-1/2}))(1 - p_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{p_K + O(N^{-1/2})} + \frac{1}{p_k + O(N^{-1/2})}\right)$$

$$+ \frac{\sqrt{N_{\mathbf{p}}}}{4(N_{\mathbf{p}}-1)} \sum_{m \neq n} (p_n + O(N^{-1/2}))(p_m + O(N^{-1/2})) \left(\frac{1}{p_K + O(N^{-1/2})}\right)$$

$$- \frac{\sqrt{\lambda N_{\mathbf{q}}}}{4(N_{\mathbf{q}}-1)} \sum_{k=1}^{K-1} (q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{q_K + O(N^{-1/2})} + \frac{1}{q_k + O(N^{-1/2})}\right)$$

$$+ \frac{\sqrt{\lambda N_{\mathbf{q}}}}{4(N_{\mathbf{q}}-1)} \sum_{m \neq n} (q_n + O(N^{-1/2}))(q_m + O(N^{-1/2})) \left(\frac{1}{q_K + O(N^{-1/2})}\right)$$

$$= O(N^{-1/2}) \to 0$$

□

Lemma 10.

$$\hat{B}_{2\mathbf{p}} - \hat{B}_2^0$$

$$= \frac{1}{4(N_{\mathbf{p}}-1)} \sum_{k=1}^{K-1} (p_k + O(N^{-1/2}))(1 - p_k + O(N^{-1/2}))$$
$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right) \tag{4.20}$$
$$- \frac{1}{4(N_{\mathbf{p}}-1)} \sum_{m \neq n} (p_n + O(N^{-1/2}))(p_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

Similarly,

$$\hat{B}_{JK_2} = \hat{B}_{2\mathbf{p}}$$

$$+ \frac{1}{4(N_{\mathbf{q}}-1)} \sum_{k=1}^{K-1} (q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$
$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right) \tag{4.21}$$
$$- \frac{1}{4(N_{\mathbf{q}}-1)} \sum_{m \neq n} (q_n + O(N^{-1/2}))(q_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

Proof. First, note that for any i,

$$\left(\hat{\mathbf{p}}^{(-i)} - \hat{\mathbf{p}}\right)^{\tau} = \left\{\left(\frac{X_1 - N_{\mathbf{p}} I[\upsilon_i = \ell_1]}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)}\right), \ldots, \left(\frac{X_{K-1} - N_{\mathbf{p}} I[\upsilon_i = \ell_{K-1}]}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)}\right)\right\}$$

$$= \frac{1}{N_{\mathbf{p}} - 1}\left\{\hat{p}_1 - I[\upsilon_i = \ell_1], \ldots, \hat{p}_{K-1} - I[\upsilon_i = \ell_{K-1}]\right\}$$

Then, for any vector ξ_i between $\hat{\mathbf{p}}^{(-i)}$ and $\hat{\mathbf{p}}$ and fixed $\hat{\mathbf{q}}$, we have

$$\hat{B}_2^{(-i)} - \hat{B}_2^0 = B\left(\hat{\mathbf{p}}^{(-i)}, \hat{\mathbf{q}}\right) - B\left(\hat{\mathbf{p}}, \hat{\mathbf{q}}\right)$$

$$= \left(\hat{\mathbf{p}}^{(-i)} - \hat{\mathbf{p}}\right)^{\tau} \nabla B\left(\hat{\mathbf{p}}, \hat{\mathbf{q}}\right) + \frac{1}{2}\left(\hat{\mathbf{p}}^{(-i)} - \hat{\mathbf{p}}\right)^{\tau} \nabla^2 B\left(\xi_i, \hat{\mathbf{q}}\right)\left(\hat{\mathbf{p}}^{(-i)} - \hat{\mathbf{p}}\right)$$

We have that $\nabla B\left(\hat{\mathbf{p}}, \hat{\mathbf{q}}\right)$ is a vector such that

$$\nabla B\left(\hat{\mathbf{p}}, \hat{\mathbf{q}}\right) = -\frac{1}{2}\left\{\ln\left(\frac{\hat{p}_1 + \hat{q}_1}{\hat{p}_K + \hat{q}_K}\right), \ldots, \ln\left(\frac{\hat{p}_{K-1} + \hat{q}_{K-1}}{\hat{p}_K + \hat{q}_K}\right)\right\}$$

and so

$$\sum_{i=1}^{N_{\mathbf{p}}} \left(\hat{\mathbf{p}}^{(-i)} - \hat{\mathbf{p}}\right)^{\tau} \nabla B\left(\hat{\mathbf{p}}, \hat{\mathbf{q}}\right)$$

$$= -\frac{1}{2}\sum_{k=1}^{K-1} \ln\left(\frac{\hat{p}_1 + \hat{q}_1}{\hat{p}_K + \hat{q}_K}\right) \sum_{i=1}^{N_{\mathbf{p}}} \frac{X_k - N_{\mathbf{p}} I[\upsilon_i = \ell_k]}{N_{\mathbf{p}}(N_{\mathbf{p}} - 1)}$$

$$= -\frac{1}{2(N_{\mathbf{p}} - 1)} \sum_{k=1}^{K-1} \ln\left(\frac{\hat{p}_1 + \hat{q}_1}{\hat{p}_K + \hat{q}_K}\right) \sum_{i=1}^{N_{\mathbf{p}}} (\hat{p}_k - I[\upsilon_i = \ell_k])$$

$$= -\frac{1}{2(N_{\mathbf{p}} - 1)} \sum_{k=1}^{K-1} \ln\left(\frac{\hat{p}_1 + \hat{q}_1}{\hat{p}_K + \hat{q}_K}\right) \left(N_{\mathbf{p}}\hat{p}_k - \sum_{i=1}^{N_{\mathbf{p}}} I[\upsilon_i = \ell_k]\right)$$

$$= -\frac{1}{2(N_{\mathbf{p}} - 1)} \sum_{k=1}^{K-1} \ln\left(\frac{\hat{p}_1 + \hat{q}_1}{\hat{p}_K + \hat{q}_K}\right) (X_k - X_k) = 0$$

Next, we see that

$$\frac{1}{2}\left(\hat{\mathbf{p}}^{(-i)} - \hat{\mathbf{p}}\right)^{\tau} \nabla^2 B\left(\xi_i, \hat{\mathbf{q}}\right) \left(\hat{\mathbf{p}}^{(-i)} - \hat{\mathbf{p}}\right)$$

$$= -\frac{1}{4(N_{\mathbf{p}} - 1)^2} \left(\frac{\left(\sum_{k=1}^{K-1} \hat{p}_k - I[\upsilon_i = \ell_k]\right)^2}{\xi_{i,K} + \hat{q}_K} + \sum_{k=1}^{K-1} \frac{\left(\hat{p}_k^2 - I[\upsilon_i = \ell_k]\right)^2}{\xi_{i,k} + \hat{q}_k} \right)$$

where $\xi_{i,k}$ and $\xi_{i,K}$ are the corresponding elements of the ξ_i vector. We know that

$$\left(\sum_{k=1}^{K-1} \hat{p}_k - I[\upsilon_i = \ell_k]\right)^2 = \sum_{k=1}^{K-1} \left(\hat{p}_k - I[\upsilon_i = \ell_k]\right)^2 + \sum_{m \neq n} (\hat{p}_n - I[\upsilon_i = \ell_n])(\hat{p}_m - I[\upsilon_i = \ell_m])$$

Thus

$$\hat{B}_{2\mathbf{p}} = \hat{B}_2^0 - \frac{N_\mathbf{p}-1}{N_\mathbf{p}} \sum_{i=1}^{N_\mathbf{p}} \left(\hat{B}_2^{(-i)} - \hat{B}_2^0\right)$$

$$= \hat{B}_2^0 - \frac{N_\mathbf{p}-1}{N_\mathbf{p}} \sum_{i=1}^{N_\mathbf{p}} \left(\frac{1}{2}\left(\hat{\mathbf{p}}^{(-i)} - \hat{\mathbf{p}}\right)^\tau \nabla^2 B\left(\xi_i, \hat{\mathbf{q}}\right)\left(\hat{\mathbf{p}}^{(-i)} - \hat{\mathbf{p}}\right)\right)$$

$$= \hat{B}_2^0 + \frac{1}{4N_\mathbf{p}(N_\mathbf{p}-1)} \sum_{i=1}^{N_\mathbf{p}} \frac{\sum_{k=1}^{K-1}\left(\hat{p}_k - I[\upsilon_i = \ell_k]\right)^2}{\xi_{i,K} + \hat{q}_K}$$

$$+ \frac{1}{4N_\mathbf{p}(N_\mathbf{p}-1)} \sum_{i=1}^{N_\mathbf{p}} \frac{\sum_{m\neq n}(\hat{p}_n - I[\upsilon_i = \ell_n])(\hat{p}_m - I[\upsilon_i = \ell_m])}{\xi_{i,K} + \hat{q}_K}$$

$$+ \frac{1}{4N_\mathbf{p}(N_\mathbf{p}-1)} \sum_{i=1}^{N_\mathbf{p}} \sum_{k=1}^{K-1} \frac{\left(\hat{p}_k^2 - I[\upsilon_i = \ell_k]\right)^2}{\xi_{i,k} + \hat{q}_k}$$

$$= \hat{B}_2^0 + \frac{1}{4N_\mathbf{p}(N_\mathbf{p}-1)} \sum_{k=1}^{K-1} \sum_{i=1}^{N_\mathbf{p}} \left(\hat{p}_k^2 - I[\upsilon_i = \ell_k]\right)^2 \left(\frac{1}{\xi_{i,K} + \hat{q}_K} + \frac{1}{\xi_{i,k} + \hat{q}_k}\right)$$

$$+ \frac{1}{4N_\mathbf{p}(N_\mathbf{p}-1)} \sum_{m\neq n} \sum_{i=1}^{N_\mathbf{p}} \frac{(\hat{p}_n - I[\upsilon_i = \ell_n])(\hat{p}_m - I[\upsilon_i = \ell_m])}{\xi_{i,K} + \hat{q}_K}$$

$$= \hat{B}_2^0 + \frac{1}{4N_\mathbf{p}(N_\mathbf{p}-1)} \sum_{k=1}^{K-1} \left(X_k(\hat{p}_k - 1)^2 + (N_\mathbf{p} - X_k)\hat{p}_k^2\right)$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})}\right)$$

$$+ \frac{1}{4N_\mathbf{p}(N_\mathbf{p}-1)} \sum_{m\neq n} \left(X_m(\hat{p}_m - 1)\hat{p}_n + X_n(\hat{p}_n - 1)\hat{p}_m + (N_\mathbf{p} - X_m - X_n)\hat{p}_n\hat{p}_m\right)$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

Taking the $\dfrac{1}{N_{\mathbf{p}}}$ inside yields

$$\hat{B}_2^0 + \frac{1}{4(N_{\mathbf{p}} - 1)} \sum_{k=1}^{K-1} \left(\hat{p}_k(\hat{p}_k - 1)^2 + (1 - \hat{p}_k)\hat{p}_k^2 \right)$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$+ \frac{1}{4(N_{\mathbf{p}} - 1)} \sum_{m \neq n} ((\hat{p}_m - 1)\hat{p}_n\hat{p}_m + (\hat{p}_n - 1)\hat{p}_n\hat{p}_m + (1 - \hat{p}_m - \hat{p}_n)\hat{p}_n\hat{p}_m)$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$$= \hat{B}_2^0 + \frac{1}{4(N_{\mathbf{p}} - 1)} \sum_{k=1}^{K-1} \hat{p}_k(1 - \hat{p}_k) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{1}{4(N_{\mathbf{p}} - 1)} \sum_{m \neq n} \hat{p}_n\hat{p}_m \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$$= \hat{B}_2^0 + \frac{1}{4(N_{\mathbf{p}} - 1)} \sum_{k=1}^{K-1} (p_k + O(N^{-1/2}))(1 - p_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{1}{4(N_{\mathbf{p}}-1)} \sum_{m \neq n} (p_n + O(N^{-1/2}))(p_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

The proof for (4.21) follows analogously. □

Lemma 11.

$$\sqrt{N_{\mathbf{p}}}(\hat{B}_{2\mathbf{p}} - \hat{B}_2^0) \xrightarrow{p} 0 \tag{4.22}$$

and

$$\sqrt{N_{\mathbf{p}}}(\hat{B}_{JK_2} - \hat{B}_{2\mathbf{p}}) \xrightarrow{p} 0 \tag{4.23}$$

and therefore

$$\sqrt{N_{\mathbf{p}}}(\hat{B}_{JK_2} - \hat{B}_2^0) \xrightarrow{p} 0$$

Proof. From Lemma 10, we have

$$\sqrt{N_{\mathbf{p}}}(\hat{B}_{2\mathbf{p}} - \hat{B}_2^0)$$

$$= \frac{\sqrt{N_{\mathbf{p}}}}{4(N_{\mathbf{p}}-1)} \sum_{k=1}^{K-1} (p_k + O(N^{-1/2}))(1 - p_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{\sqrt{N_{\mathbf{p}}}}{4(N_{\mathbf{p}}-1)} \sum_{m \neq n} (p_n + O(N^{-1/2}))(p_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$$= O(N^{-1/2}) \to 0$$

Similarly,

$$\sqrt{N_{\mathbf{p}}}(\hat{B}_{JK_2} - \hat{B}_{2\mathbf{p}})$$

$$\approx \frac{\sqrt{\lambda N_{\mathbf{q}}}}{4(N_{\mathbf{q}}-1)} \sum_{k=1}^{K-1} (q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{\sqrt{\lambda N_{\mathbf{q}}}}{4(N_{\mathbf{q}}-1)} \sum_{m \neq n} (q_n + O(N^{-1/2}))(q_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$$= O(N^{-1/2}) \to 0$$

□

Given Theorem 19, Lemmas 9 and 11 along with Slutzky's theorem, the next theorem follows immediately to yield the asymptotic normality of $\widehat{JS}_{BA_2}$.

Theorem 20. *Provided that* $(a+b)^{\tau}(\mathbf{v})\Sigma(\mathbf{v})(a+b)(\mathbf{v}) > 0$,

$$\frac{\sqrt{N_{\mathbf{p}}}((\hat{A}_{JK_2} + \hat{B}_{JK_2}) - (A+B))}{\sqrt{(a+b)^{\tau}(\mathbf{v})\Sigma(\mathbf{v})(a+b)(\mathbf{v})}} \xrightarrow{L} N(0,1) \tag{4.24}$$

Using Corollary 3 and the symmetry of the partial derivatives, the asymptotic normality of the plug-in $\hat{A}_2^0 + \hat{B}_2^0$ and hence also $\widehat{JS}_{BA_2}$ falls through when $\mathbf{p} = \mathbf{q}$. The following theorem is stated toward finding the asymptotic behavior of $\widehat{JS}_{BA_2} = \hat{A}_{JK_2} + \hat{B}_{JK_2}$ when $\mathbf{p} = \mathbf{q}$.

Theorem 21. *When* $\mathbf{p} = \mathbf{q}$,

$$N_{\mathbf{p}} \left(\hat{A}_2^0 + \hat{B}_2^0 \right) \xrightarrow{L} \frac{1}{8} (1 + \lambda) \chi^2_{K-1}$$

where λ *is as in Condition* 1*. If* $\lambda = 1$*, this becomes*

$$N_{\mathbf{p}} \left(\hat{A}_2^0 + \hat{B}_2^0 \right) \xrightarrow{L} \frac{1}{4} \chi^2_{K-1}$$

Proof. Since $\mathbf{p} = \mathbf{q}$, we have $\mathbf{v}$ defined as

$$\mathbf{v} = \{p_1, \ldots, p_{K-1}, p_1, \ldots, p_{K-1}\}$$

Additionally, assume throughout the proof that λ is as in Condition 1. By Taylor Series Expansion,

$$N_{\mathbf{p}} \left(\hat{A}_2^0 + \hat{B}_2^0 \right) = N_{\mathbf{p}} (A + B)(\hat{\mathbf{v}})$$

$$= N_{\mathbf{p}} (A + B)(\mathbf{v}) + N_{\mathbf{p}} (\hat{\mathbf{v}} - \mathbf{v})^\tau \nabla (A + B)(\mathbf{v})$$

$$+ \frac{1}{2} \sqrt{N_{\mathbf{p}}} (\hat{\mathbf{v}} - \mathbf{v})^\tau \nabla^2 (A + B)(\mathbf{v}) \sqrt{N_{\mathbf{p}}} (\hat{\mathbf{v}} - \mathbf{v}) + O(N^{-1/2})$$

Since $\mathbf{p} = \mathbf{q}$, $(A+B)(\mathbf{v}) = 0$ by Theorem 1, and $\nabla(A+B)(\mathbf{v}) = (a+b)(\mathbf{v}) = 0$ by Corollary 3. Obviously the $O(N^{-1/2})$ term goes to 0 in probability. Thus the only term we are left to contend with is

$$\frac{1}{2}\sqrt{N_{\mathbf{p}}}(\hat{\mathbf{v}} - \mathbf{v})^{\tau}\nabla^2((A+B)(\mathbf{v}))\sqrt{N_{\mathbf{p}}}(\hat{\mathbf{v}} - \mathbf{v}) \tag{4.25}$$

Using the multivariate normal approximation to the multinomial distribution, we have

$$\sqrt{N_{\mathbf{p}}}(\hat{\mathbf{v}} - \mathbf{v}) \xrightarrow{L} MVN(0, \Sigma(\mathbf{v})) \tag{4.26}$$

where $\Sigma(\mathbf{v})$ is as in (4.17), except we note that

$$\Sigma_{\mathbf{q}}(\mathbf{v}) = \lambda\Sigma_{\mathbf{q}}(\mathbf{v}) = \lambda \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_{K-1} \\ -p_2p_1 & p_2(1-p_2) & \dots & -p_2p_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ -p_{K-1}p_1 & -p_{K-1}p_2 & \dots & p_{K-1}(1-p_{K-1}) \end{pmatrix}$$

since $\mathbf{p} = \mathbf{q}$. Putting together (4.26) and Slutsky's Theorem, we have

$$\sqrt{N_{\mathbf{p}}}(\hat{\mathbf{v}} - \mathbf{v})\Sigma(\mathbf{v})^{-1/2} \xrightarrow{L} MVN(0, \mathbf{I}_{2K-2}) := \mathbf{Z_2} \tag{4.27}$$

Noting this fact, we rewrite (4.25) as

$$\frac{1}{2}\sqrt{N_{\mathbf{p}}}\left(\Sigma(\mathbf{v})^{-1/2}(\hat{\mathbf{v}} - \mathbf{v})\right)^{\tau}\left(\Sigma(\mathbf{v})^{1/2}\right)^{\tau}\nabla^2((A+B)(\mathbf{v}))\Sigma(\mathbf{v})^{1/2}\sqrt{N_{\mathbf{p}}}\left(\Sigma(\mathbf{v})^{-1/2}(\hat{\mathbf{v}} - \mathbf{v})\right)$$

Because we know (4.27), this leaves us with finding the asymptotic behavior of

$$\left(\Sigma(\mathbf{v})^{1/2}\right)^{\tau}\nabla^2((A+B)(\mathbf{v}))\Sigma(\mathbf{v})^{1/2} \tag{4.28}$$

First, note that

$$\Sigma(\mathbf{v}) = \begin{pmatrix} \Sigma_{\mathbf{p}}(\mathbf{v}) & 0 \\ 0 & \Sigma_{\mathbf{p}}(\mathbf{v}) \end{pmatrix} \begin{pmatrix} \mathbf{I}_{K-1} & 0 \\ 0 & \lambda \mathbf{I}_{K-1} \end{pmatrix}$$

and so we can rewrite (4.28) as

$$diag\{\mathbf{I}_{K-1}, \sqrt{\lambda}\mathbf{I}_{K-1}\} \left(\Sigma(\mathbf{v})^{1/2}_{-\lambda}\right)^{\tau} \nabla^2((A+B)(\mathbf{v}))\Sigma(\mathbf{v})^{1/2}_{-\lambda} diag\{\mathbf{I}_{K-1}, \sqrt{\lambda}\mathbf{I}_{K-1}\}$$

We first find the value of

$$\left(\Sigma(\mathbf{v})^{1/2}_{-\lambda}\right)^{\tau} \nabla^2((A+B)(\mathbf{v}))\Sigma(\mathbf{v})^{1/2}_{-\lambda} \tag{4.29}$$

Let

$$\nabla^2(A+B)(\mathbf{v}) = \begin{pmatrix} \Theta(\mathbf{v}) & -\Theta(\mathbf{v}) \\ -\Theta(\mathbf{v}) & \Theta(\mathbf{v}) \end{pmatrix}_{(2K-2)\times(2K-2)}$$

where, since $\mathbf{p} = \mathbf{q}$,

$$\Theta(\mathbf{v}) = \frac{1}{4} \begin{pmatrix} \frac{1}{p_1} + \frac{1}{p_K} & \frac{1}{p_K} & \cdots & \frac{1}{p_K} \\ \frac{1}{p_K} & \frac{1}{p_2} + \frac{1}{p_K} & \cdots & \frac{1}{p_K} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{p_K} & \frac{1}{p_K} & \cdots & \frac{1}{p_{K-1}} + \frac{1}{p_K} \end{pmatrix}_{(K-1)\times(K-1)}$$

First, we show that

$$\Sigma_{\mathbf{p}}(\mathbf{v})^{1/2}\Theta(\mathbf{v})\Sigma_{\mathbf{p}}(\mathbf{v})^{1/2} = \frac{1}{4}\mathbf{I}_{K-1}$$

This is equivalent to showing that

$$(4\Theta(\mathbf{v}))^{-1} = \Sigma_{\mathbf{p}}(\mathbf{v})$$

An analogous proof of this fact is given in the proof of Theorem 17 and is therefore omitted here. Assuming the veracity of this fact, we have

$$\left(\Sigma(\mathbf{v})_{-\lambda}^{1/2}\right)^{\tau} \nabla^2((A+B)(\mathbf{v}))\Sigma(\mathbf{v})_{-\lambda}^{1/2}$$

$$= \begin{pmatrix} \Sigma_{\mathbf{p}}(\mathbf{v})^{1/2} & 0 \\ 0 & \Sigma_{\mathbf{p}}(\mathbf{v})^{1/2} \end{pmatrix} \begin{pmatrix} \Theta(\mathbf{v}) & -\Theta(\mathbf{v}) \\ -\Theta(\mathbf{v}) & \Theta(\mathbf{v}) \end{pmatrix} \begin{pmatrix} \Sigma_{\mathbf{p}}(\mathbf{v})^{1/2} & 0 \\ 0 & \Sigma_{\mathbf{p}}(\mathbf{v})^{1/2} \end{pmatrix}$$

$$= \begin{pmatrix} \Sigma_{\mathbf{p}}(\mathbf{v})^{1/2}\Theta(\mathbf{v})\Sigma_{\mathbf{p}}(\mathbf{v})^{1/2} & -\Sigma_{\mathbf{p}}(\mathbf{v})^{1/2}\Theta(\mathbf{v})\Sigma_{\mathbf{p}}(\mathbf{v})^{1/2} \\ -\Sigma_{\mathbf{p}}(\mathbf{v})^{1/2}\Theta(\mathbf{v})\Sigma_{\mathbf{p}}(\mathbf{v})^{1/2} & \Sigma_{\mathbf{p}}(\mathbf{v})^{1/2}\Theta(\mathbf{v})\Sigma_{\mathbf{p}}(\mathbf{v})^{1/2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{4}\mathbf{I}_{K-1} & -\frac{1}{4}\mathbf{I}_{K-1} \\ -\frac{1}{4}\mathbf{I}_{K-1} & \frac{1}{4}\mathbf{I}_{K-1} \end{pmatrix}$$

Hence,

$$diag\{\mathbf{I}_{K-1}, \sqrt{\lambda}\mathbf{I}_{K-1}\} \left(\Sigma(\mathbf{v})_{-\lambda}^{1/2}\right)^{\tau} \nabla^2((A+B)(\mathbf{v}))\Sigma(\mathbf{v})_{-\lambda}^{1/2} diag\{\mathbf{I}_{K-1}, \sqrt{\lambda}\mathbf{I}_{K-1}\}$$

$$= \begin{pmatrix} \mathbf{I}_{K-1} & 0 \\ 0 & \sqrt{\lambda}\mathbf{I}_{K-1} \end{pmatrix} \begin{pmatrix} \frac{1}{4}\mathbf{I}_{K-1} & -\frac{1}{4}\mathbf{I}_{K-1} \\ -\frac{1}{4}\mathbf{I}_{K-1} & \frac{1}{4}\mathbf{I}_{K-1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_{K-1} & 0 \\ 0 & \sqrt{\lambda}\mathbf{I}_{K-1} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{4}\mathbf{I}_{K-1} & -\frac{\sqrt{\lambda}}{4}\mathbf{I}_{K-1} \\ -\frac{\sqrt{\lambda}}{4}\mathbf{I}_{K-1} & \frac{\lambda}{4}\mathbf{I}_{K-1} \end{pmatrix}$$

Therefore

$$(4.25) = \frac{1}{2}\left(\sqrt{N_{\mathbf{p}}}\Sigma(\mathbf{v})^{-1/2}(\hat{\mathbf{v}}-\mathbf{v})\right)^{\tau}\frac{1}{4}\begin{pmatrix} \mathbf{I}_{K-1} & -\sqrt{\lambda}\mathbf{I}_{K-1} \\ \\ -\sqrt{\lambda}\mathbf{I}_{K-1} & \lambda\mathbf{I}_{K-1} \end{pmatrix}\left(\sqrt{N_{\mathbf{p}}}\Sigma(\mathbf{v})^{-1/2}(\hat{\mathbf{v}}-\mathbf{v})\right)$$

$$=: \frac{1}{8}\left(\sqrt{N_{\mathbf{p}}}\Sigma(\mathbf{v})^{-1/2}(\hat{\mathbf{v}}-\mathbf{v})\right)^{\tau}\mathbb{V}\left(\sqrt{N_{\mathbf{p}}}\Sigma(\mathbf{v})^{-1/2}(\hat{\mathbf{v}}-\mathbf{v})\right)$$

which, using spectral decomposition, is equal to

$$\frac{1}{8}\left(\sqrt{N_{\mathbf{p}}}\Sigma(\mathbf{v})^{-1/2}(\hat{\mathbf{v}}-\mathbf{v})\right)^{\tau}\mathbb{Q}^{\tau}\mathbf{\Lambda}\mathbb{Q}\left(\sqrt{N_{\mathbf{p}}}\Sigma(\mathbf{v})^{-1/2}(\hat{\mathbf{v}}-\mathbf{v})\right)$$

where $\mathbf{\Lambda} = \text{diag}\{\zeta_1, \ldots, \zeta_{2K-2}\}$ with ζ_i being the eigenvalues of $\mathbb{V}$; and $\mathbb{Q}$ a $(2K-2)\times(2K-2)$ square matrix with columns that are the eigenvectors of $\mathbb{V}$ such that $\mathbb{Q}^{\tau}\mathbb{Q} = \mathbf{I}_{2K-2}$. By the Continuous Mapping Theorem, this converges in law to

$$\frac{1}{8}(\mathbb{Q}\mathbf{Z_2})^{\tau}\mathbf{\Lambda}(\mathbb{Q}\mathbf{Z_2}) =: \frac{1}{8}(\mathbb{W})^{\tau}\mathbf{\Lambda}(\mathbb{W}) = \frac{1}{8}\left(\sum_{i=1}^{2K-2}\zeta_i\mathbf{W}_i^2\right)$$

Note that since $\mathbb{Q}$ is a constant, we have

$$E(\mathbb{W}) = E(\mathbb{Q}\mathbf{Z_2}) = \mathbb{Q}E(\mathbf{Z_2}) = 0$$

and

$$Var(\mathbb{W}) = Var(\mathbb{Q}\mathbf{Z_2}) = \mathbb{Q}^{\tau}Var(\mathbf{Z_2})\mathbb{Q} = \mathbb{Q}^{\tau}\mathbf{I}_{2K-2}\mathbb{Q} = \mathbf{I}_{2K-2}$$

and so $\mathbb{W}$ also hast distribution standard multivariate normal. Hence for each i, $\mathbf{W}_i \sim N(0,1)$. Therefore we only need to find ζ_i, the eigenvalues of $\mathbb{V}$. This is done by solving the following equation:

$$0 = \det\{\mathbb{V} - \zeta \mathbf{I}_{2K-2}\} = \det \begin{pmatrix} (1-\zeta)\mathbf{I}_{K-1} & -\sqrt{\lambda}\mathbf{I}_{K-1} \\ \\ -\sqrt{\lambda}\mathbf{I}_{K-1} & (\lambda-\zeta)\,\mathbf{I}_{K-1} \end{pmatrix}$$

$$= \det\left\{(1-\zeta)\,(\lambda-\zeta)\,\mathbf{I}_{K-1} - \lambda \mathbf{I}_{K-1}\right\}$$

$$= \left((1-\zeta)\,(\lambda-\zeta) - \lambda\right)^{K-1} \det\left(\mathbf{I}_{K-1}\right)$$

Hence we have

$$0 = \left(\zeta(\zeta - (\lambda+1))\right)^{K-1}$$

which means that $\zeta = 0$ or $\zeta = 1+\lambda$ for $K-1$ times. Thus

$$\frac{1}{8}(\mathbb{Q}\mathbf{Z_2})^{\tau}\mathbf{\Lambda}(\mathbb{Q}\mathbf{Z_2}) = \frac{1}{8}\left(\sum_{i=1}^{2K-2} \zeta_i \mathbf{W}_i^2\right) \sim \frac{1}{8}\,(1+\lambda)\,\chi^2_{K-1}$$

□

Lemma 12. *When* $\mathbf{p} = \mathbf{q}$,

$$N_{\mathbf{p}}(\hat{A}_{JK_2} - \hat{A}_2^0) \xrightarrow{p} -\frac{1}{4}(1+\lambda)\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K} + \frac{1}{p_k}\right) - \sum_{m\neq n} \frac{p_n p_m}{p_K}\right)$$

where λ *is as in Condition* 1*. If* $\lambda = 1$*, this becomes*

$$-\frac{1}{2}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K} + \frac{1}{p_k}\right) - \sum_{m\neq n} \frac{p_n p_m}{p_K}\right)$$

Proof. Using Lemma 8,

$$N_{\mathbf{p}}(\hat{A}_{JK_2} - \hat{A}_2^0) = N_{\mathbf{p}}\left(\hat{A}_{JK_{2\mathbf{p}}} - \hat{A}_{2\mathbf{p}} + \hat{A}_{JK_{2\mathbf{q}}} - \hat{A}_{2\mathbf{q}}\right)$$

$$= -\frac{N_{\mathbf{p}}}{4(N_{\mathbf{p}}-1)}\sum_{k=1}^{K-1}(p_k + O(N^{-1/2}))(1 - p_k + O(N^{-1/2}))$$

$$\times\left(\frac{1}{p_K + O(N^{-1/2})} + \frac{1}{p_k + O(N^{-1/2})}\right)$$

$$+ \frac{N_{\mathbf{p}}}{4(N_{\mathbf{p}}-1)}\sum_{m\neq n}(p_n + O(N^{-1/2}))(p_m + O(N^{-1/2}))\left(\frac{1}{p_K + O(N^{-1/2})}\right)$$

$$- \frac{\lambda N_{\mathbf{q}}}{4(N_{\mathbf{q}}-1)}\sum_{k=1}^{K-1}(q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times\left(\frac{1}{q_K + O(N^{-1/2})} + \frac{1}{q_k + O(N^{-1/2})}\right)$$

$$+ \frac{\lambda N_{\mathbf{q}}}{4(N_{\mathbf{q}}-1)}\sum_{m\neq n}(q_n + O(N^{-1/2}))(q_m + O(N^{-1/2}))\left(\frac{1}{q_K + O(N^{-1/2})}\right)$$

$$\to -\frac{1}{4}\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K} + \frac{1}{p_k}\right) + \frac{1}{4}\sum_{m\neq n}\frac{p_n p_m}{p_K}$$

$$- \frac{\lambda}{4}\sum_{k=1}^{K-1} q_k(1-q_k)\left(\frac{1}{q_K} + \frac{1}{q_k}\right) + \frac{\lambda}{4}\sum_{m\neq n}\frac{q_n q_m}{q_K}$$

Since $\mathbf{p} = \mathbf{q}$, this is equivalent to

$$-\frac{1}{4}\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right)+\frac{1}{4}\sum_{m\neq n}\frac{p_n p_m}{p_K}$$

$$-\frac{\lambda}{4}\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right)+\frac{\lambda}{4}\sum_{m\neq n}\frac{p_n p_m}{p_K}$$

$$=-\frac{1}{4}(1+\lambda)\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right)-\sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

□

Lemma 13. *When* $\mathbf{p}=\mathbf{q}$,

$$N_{\mathbf{p}}(\hat{B}_{JK_2}-\hat{B}_2^0)\xrightarrow{p}\frac{1}{8}(1+\lambda)\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right)-\sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

where λ *is as in Condition* 1. *If* $\lambda=1$, *this becomes*

$$\frac{1}{4}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right)-\sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

Proof. Observe that

$$\hat{B}_{JK_2}=\hat{B}_{2\mathbf{p}}-\frac{N_{\mathbf{q}}-1}{N_{\mathbf{q}}}\sum_{j=1}^{N_{\mathbf{q}}}\left(\hat{B}_{2\mathbf{p}(-j)}-\hat{B}_{2\mathbf{p}}\right)$$

$$=\hat{B}_2^0-\frac{N_{\mathbf{p}}-1}{N_{\mathbf{p}}}\sum_{j=1}^{N_{\mathbf{p}}}\left(\hat{B}_2^{(-i)}-\hat{B}_2^0\right)-\frac{N_{\mathbf{q}}-1}{N_{\mathbf{q}}}\sum_{j=1}^{N_{\mathbf{q}}}\left(\hat{B}_{2\mathbf{p}(-j)}-\hat{B}_{2\mathbf{p}}\right)$$

Then using this and Lemma 10, we have

$$N_{\mathbf{p}}(\hat{B}_{JK_2}-\hat{B}_2^0)$$

$$\approx \frac{N_{\mathbf{p}}}{4(N_{\mathbf{p}}-1)} \sum_{k=1}^{K-1} (p_k + O(N^{-1/2}))(1 - p_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{N_{\mathbf{p}}}{4(N_{\mathbf{p}}-1)} \sum_{m \neq n} (p_n + O(N^{-1/2}))(p_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$$+ \frac{\lambda N_{\mathbf{q}}}{4(N_{\mathbf{q}}-1)} \sum_{k=1}^{K-1} (q_k + O(N^{-1/2}))(1 - q_k + O(N^{-1/2}))$$

$$\times \left(\frac{1}{p_K + q_K + O(N^{-1/2})} + \frac{1}{p_k + q_k + O(N^{-1/2})} \right)$$

$$- \frac{\lambda N_{\mathbf{q}}}{4(N_{\mathbf{q}}-1)} \sum_{m \neq n} (q_n + O(N^{-1/2}))(q_m + O(N^{-1/2})) \left(\frac{1}{p_K + q_K + O(N^{-1/2})} \right)$$

$$\xrightarrow{p} \frac{1}{4} \sum_{k=1}^{K-1} p_k(1 - p_k) \left(\frac{1}{p_K + q_K} + \frac{1}{p_k + q_k} \right) - \frac{1}{4} \sum_{m \neq n} p_n p_m \left(\frac{1}{p_K + q_K} \right)$$

$$+ \frac{\lambda}{4} \sum_{k=1}^{K-1} q_k(1 - q_k) \left(\frac{1}{p_K + q_K} + \frac{1}{p_k + q_k} \right) - \frac{\lambda}{4} \sum_{m \neq n} q_n q_m \left(\frac{1}{p_K + q_K} \right)$$

Since $\mathbf{p} = \mathbf{q}$, this is equivalent to

$$\frac{1}{8}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right) - \sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

$$+\frac{\lambda}{8}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right) - \sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

$$=\frac{1}{8}(1+\lambda)\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right) - \sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

□

The next Corollary follows directly from Lemmas 12 and 13.

Corollary 5. *When* $\mathbf{p} = \mathbf{q}$,

$$N_{\mathbf{p}}((\hat{A}_{JK_2} + \hat{B}_{JK_2}) - (\hat{A}_2^0 + \hat{B}_2^0))$$

$$\xrightarrow{p} -\frac{1}{8}(1+\lambda)\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right) - \sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

where λ *is as in Condition* 1*. If* $\lambda = 1$*, this becomes*

$$-\frac{1}{4}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K}+\frac{1}{p_k}\right) - \sum_{m\neq n}\frac{p_n p_m}{p_K}\right)$$

Using Slutsky's Theorem combined with Theorem 21 and Corollary 5, we obtain the following conclusion.

Theorem 22. *When* $\mathbf{p} = \mathbf{q}$,

$$N_{\mathbf{p}}(\hat{A}_{JK_2} + \hat{B}_{JK_2}) + \frac{1}{8}(1+\lambda)\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K} + \frac{1}{p_k}\right) - \sum_{m\neq n} \frac{p_n p_m}{p_K}\right)$$

$$\xrightarrow{L} \frac{1}{8}\,(1+\lambda)\,\chi^2_{K-1}$$

where λ *is as in Condition* 1*. If* $\lambda = 1$*, this becomes*

$$N_{\mathbf{p}}(\hat{A}_{JK_2} + \hat{B}_{JK_2}) + \frac{1}{4}\left(\sum_{k=1}^{K-1} p_k(1-p_k)\left(\frac{1}{p_K} + \frac{1}{p_k}\right) - \sum_{m\neq n} \frac{p_n p_m}{p_K}\right)$$

$$\xrightarrow{L} \frac{1}{4}\chi^2_{K-1}$$

Using the asymptotic distributions noted in Theorems 18 and 22, a hypothesis test of $H_0 : \mathbf{p} = \mathbf{q}$ can easily be derived.

5.1 One-Sample

For the one-sample situation, we have the test statistic

$$T_1 = 8N(\hat{A}_{JK_1} + \hat{B}_{JK_1}) + \left(\sum_{k=1}^{K-1} p_k(1-p_k) \left(\frac{1}{p_K} + \frac{1}{p_k} \right) - \sum_{m \neq n} \frac{p_n p_m}{p_K} \right) \tag{5.1}$$

where $\{p_1, \ldots, p_K\}$ is the known distribution we are testing against. T_1 is distributed χ^2_{K-1} under the null hypothesis. We reject when $T_1 > \chi^2_{K-1,\alpha}$.

When $\mathbf{p}$ and $\mathbf{q}$ are not equal, confidence intervals can be derived using the asymptotic standard normal approximations noted in Theorem 16. Therefore in the one-sample context, the $(1-\alpha)\%$ confidence interval for $A + B$ is

$$\hat{A}_{JK_1} + \hat{B}_{JK_1} \pm z_{\alpha/2} \sqrt{\frac{(a+b)^\tau(\hat{\mathbf{q}})\Sigma(\hat{\mathbf{q}})(a+b)(\hat{\mathbf{q}})}{N}}$$

5.2 Two-Sample

In the two-sample situation, we need to estimate the constant

$$\left(\sum_{k=1}^{K-1} p_k(1-p_k) \left(\frac{1}{p_K} + \frac{1}{p_k} \right) - \sum_{m \neq n} \frac{p_n p_m}{p_K} \right)$$

for the test statistic because we do not have a known distribution. Toward that end, let

$$\hat{r}_k = \frac{(X_k + Y_k) + I[(X_k + Y_k) = 0]}{N_{\mathbf{p}} + N_{\mathbf{q}}}$$

for $1 \leq k \leq K$, be the estimates of the probabilities of the mixed distribution between $\mathbf{p}$ and $\mathbf{q}$.

We use these estimates r_k for the test statistic

$$T_2 = \frac{8}{1+\lambda} N_{\mathbf{p}}(\hat{A}_{JK_2} + \hat{B}_{JK_2}) + \left(\sum_{k=1}^{K-1} \hat{r}_k(1-\hat{r}_k) \left(\frac{1}{\hat{r}_K} + \frac{1}{\hat{r}_k} \right) - \sum_{m \neq n} \frac{\hat{r}_n \hat{r}_m}{\hat{r}_K} \right) \tag{5.2}$$

Under the null hypothesis of $H_0 : \mathbf{p} = \mathbf{q}$, for all $1 \leq k \leq K$

$$\hat{r}_k \rightarrow p_k = q_k$$

which means that T_2 asymptotically distributed χ^2_{K-1}. If $\lambda = 1$, this becomes

$$T_2 = 4N_{\mathbf{p}}(\hat{A}_{JK_2} + \hat{B}_{JK_2}) + \left(\sum_{k=1}^{K-1} \hat{r}_k(1-\hat{r}_k) \left(\frac{1}{\hat{r}_K} + \frac{1}{\hat{r}_k} \right) - \sum_{m \neq n} \frac{\hat{r}_n \hat{r}_m}{\hat{r}_K} \right)$$

We reject when $T_2 > \chi^2_{K-1,\alpha}$.

When $\mathbf{p}$ and $\mathbf{q}$ are not equal, confidence intervals can be derived using the asymptotic standard normal approximations noted in Theorem 20. Thus, in the two-sample context, the $(1-\alpha)\%$ confidence interval for $A + B$ is

$$\hat{A}_{JK_2} + \hat{B}_{JK_2} \pm z_{\alpha/2} \sqrt{\frac{(a+b)^{\tau}(\hat{\mathbf{v}})\Sigma(\hat{\mathbf{v}})(a+b)(\hat{\mathbf{v}})}{N_{\mathbf{p}}}}$$

CHAPTER 6: IF K IS UNKNOWN

The situation which may arise is when the number of categories K is known to be finite, but the value itself is not known. The jackknife estimators presented here are not dependent on K being known, but for hypothesis testing it is necessary to determine the degrees of freedom for the critical value (χ^2_{K-1}). In general, estimating K with the observed number of categories is not very accurate. Some alternatives have been given in [24], and will be described briefly here so that they may be used in the hypothesis testing.

Let $K_{obs} = \sum_k I[Y_k > 0]$ and $M_r = \sum_k I[Y_k = r]$. The latest version of the estimator proposed by Chao is

$$\hat{K}_{Chao1a} = \begin{cases} K_{obs} + \left(\frac{N-1}{N}\right)\frac{M_1^2}{2M_2} & \text{if } M_2 > 0 \\ \\ K_{obs} + \left(\frac{N-1}{N}\right)\frac{M_1(M_1-1)}{2} & \text{if } M_2 = 0 \end{cases} \tag{6.1}$$

The paper [24] suggests three other estimators in Turing's perspective that will be given here as options to use when K is unknown. Let $\zeta_\nu = \sum_{k=1}^K p_k(1-p_k)^\nu$ for any integer ν. It can be verified that

$$Z_\nu = \sum_k \left[\hat{p}_k \prod_{j=1}^{\nu}\left(1 - \frac{Y_k - 1}{N - j}\right)\right]$$

is a uniformly minimum variance unbiased estimator (UMVUE) of ζ_ν for ν, $1 \le \nu \le N-1$. Let ν_N be such that

$$\nu_N = N - \max\{Y_k; k \geq 1\}$$

Then

$$K \approx K_{obs} + \frac{\zeta_{N-1}}{1 - \zeta_{\nu_N}/\zeta_{\nu_N-1}} \tag{6.2}$$

and that

It can be easily verified that $Z_{N-1} = M_1/N = T$, where T is Turing's formula. Replace ζ_{N-1} by $Z_{N-1} = T$, and $\zeta_{\nu_N}/\zeta_{\nu_N-1}$ by Z_{ν_N}/Z_{ν_N-1} into (6.2) to give the base estimator

$$\hat{K}_0 = K_{obs} + \frac{T}{1 - Z_{\nu_N}/Z_{\nu_N-1}} \tag{6.3}$$

The next estimator is a stretched version of the base estimator. Let $w_N \in (0, 1)$ be a user-chosen parameter, here demonstrated in the form

$$w_N = T^{\beta} \tag{6.4}$$

where T is Turing's formula. Then the stretched estimator is defined as

$$\hat{K}_1 = K_{obs} + \frac{T}{\left(1 - \dfrac{Z_{\nu_N}}{Z_{\nu_N-1}}\right)\left(1 - \dfrac{(1-w_N)\nu_N}{N}\right)} \tag{6.5}$$

According to [24], the stretched estimator has an improved performance over the base estimator when the distribution is not uniform, but it over-estimates K when there is uniformity. To adjust for this possibility, let

$$u_N = |(N-1)\ln(Z_1) - \ln(Z_{N-1})|$$

It can be shown that u_N is closer to 0 under a uniform distribution. Let

$$\beta^\flat = \min\{u_N, \beta\}$$

and

$$w_N^\flat = T^{\beta^\flat}$$

Then the suppressed estimator is defined as

$$\hat{K}_2 = K_{obs} + \frac{T}{\left(1 - \frac{Z_{\nu_N}}{Z_{\nu_N - 1}}\right)\left(1 - \frac{(1 - w_N^\flat)\nu_N}{N}\right)} \tag{6.6}$$

[24] states that $\hat{K}_0, \hat{K}_1$, and $\hat{K}_2$ are all consistent estimators for K. These estimators, along with Chao's estimator, which performs nearly identically to the base estimator K_0, will be used in the next chapter's simulations.

CHAPTER 7: SIMULATION STUDIES

The simulations are organized as follows. The scenarios considered are for $K = 30$ and $K = 100$, across three distributions: uniform, triangle, and power decay. There will be one section for each of these six distributions. In each section, first graphs will be shown of sample size N vs the average error for the plug-in estimator in red, and the average error for the jackknifed estimator proposed in this paper in blue. This is intended to exemplify the improved bias correction of the jackknife estimator.

Then, tables of the outcomes for different sample sizes, of testing the hypothesis $H_0 : \mathbf{p} = \mathbf{q}$ will be shown, which include both when the null hypothesis is true and when it is not. When the null hypothesis is true, the rates of rejection by sample size are given on the left side of the following tables. On the right side of the tables, the results are given for when $\mathbf{p} \neq \mathbf{q}$. T_1 and T_2 from (5.1) and (5.2), respectively, will be used as the test statistics for the jackknife estimator test. This is then compared with the corresponding hypothesis test that can be performed with the plug-in estimator. For the two-sample case, results for both the same sample size and different sample sizes will be given.

Additionally, results will be given for the possible scenario that K is unknown, using K_{obs}, $\hat{K}_{Chao1a}$, $\hat{K}_0$, $\hat{K}_1$, $\hat{K}_2$ from (6.1), (6.3), (6.5), (6.6) given in the previous chapter. Where necessary, the β value from (6.4) used here is 1/3.

7.1 Uniform Distribution: K=30

Suppose that $K = 30$ and that we have two equal uniform distributions, $\mathbf{p} = \mathbf{q} = \{1/30, \ldots 1/30\}$. The actual value of Jensen-Shannon Divergence in this case is obviously 0. The error tables are as follows.

Figure 7.1: **One-Sample**

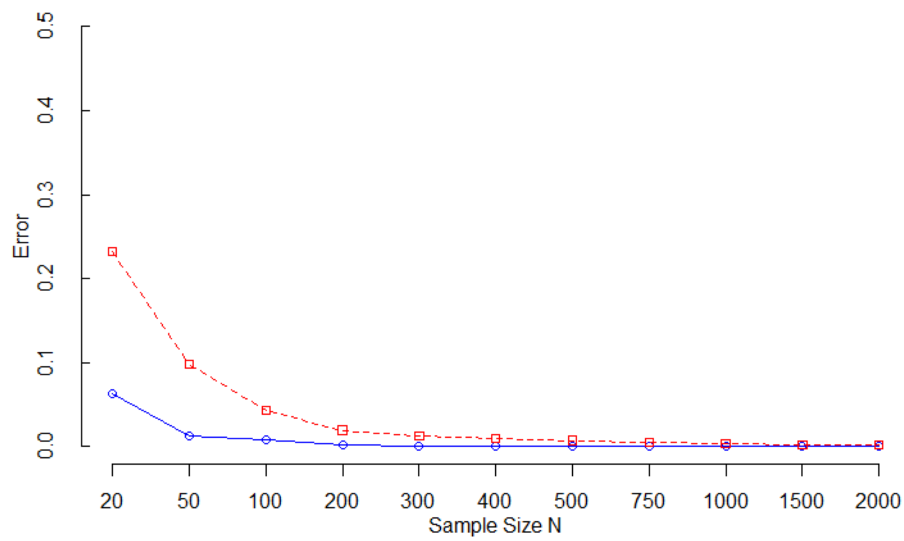
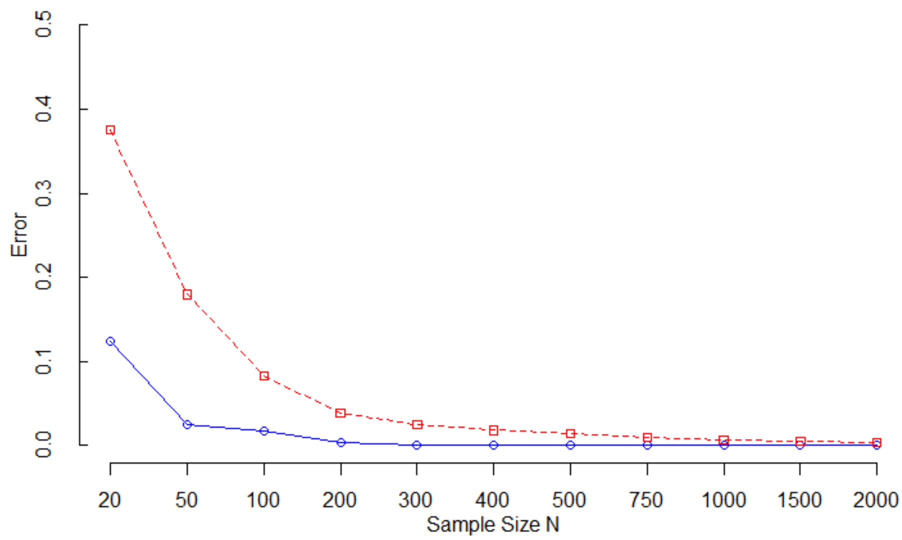


Figure 7.2: **Two-Sample**



Now suppose for **q**, that we subtract 1/200 from $\{q_1, \ldots, q_{15}\}$, and add 1/200 to $\{q_{16}, \ldots, q_{30}\}$. This adjusted **q** distribution juxtaposed on the uniform **p** looks something like this:

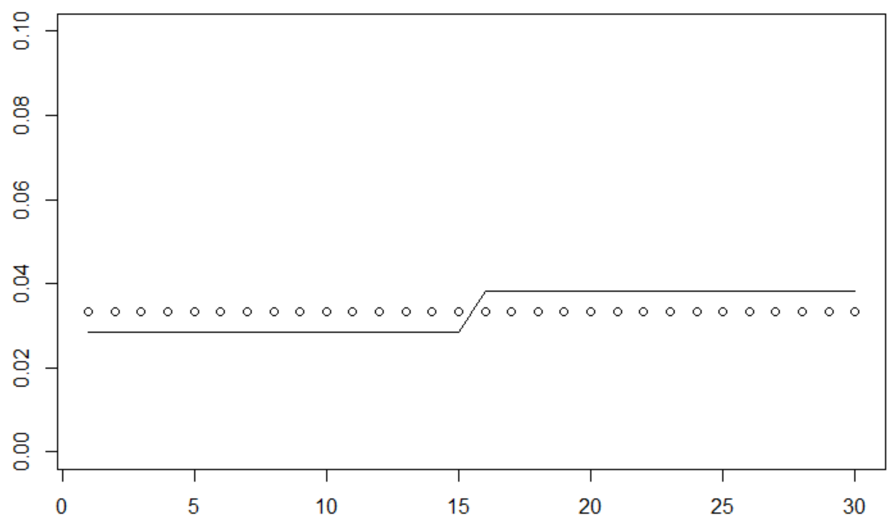


Figure 7.3

Here, between uniform **p** and this adjusted **q** given in Figure 7.3, the actual value of Jensen-Shannon Divergence is 0.002831143. For the alternative hypothesis when H_0 is false, **q** is given by Figure 7.3.

Figure 7.4: **One-Sample, *K* known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.318	0.123	0.398	0.133
100	0.218	0.048	0.299	0.094
200	0.104	0.039	0.228	0.105
300	0.075	0.043	0.264	0.177
400	0.073	0.042	0.33	0.267
500	0.07	0.055	0.429	0.352
750	0.055	0.043	0.601	0.564
1000	0.058	0.048	0.76	0.74
1500	0.041	0.039	0.956	0.953
2000	0.042	0.039	0.986	0.986

Figure 7.5: **One-Sample, $\boldsymbol{K}_{obs}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.582	0.273	0.627	0.304
100	0.248	0.073	0.337	0.12
200	0.102	0.048	0.219	0.098
300	0.076	0.041	0.288	0.2
400	0.065	0.051	0.344	0.269
500	0.063	0.044	0.44	0.383
750	0.061	0.048	0.652	0.61
1000	0.053	0.044	0.77	0.75
1500	0.049	0.044	0.942	0.937
2000	0.045	0.04	0.986	0.983

Figure 7.6: **One-Sample, $\hat{\boldsymbol{K}}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.349	0.149	0.412	0.193
100	0.198	0.064	0.266	0.095
200	0.118	0.046	0.23	0.104
300	0.077	0.037	0.282	0.181
400	0.062	0.039	0.333	0.267
500	0.066	0.05	0.413	0.362
750	0.055	0.035	0.598	0.564
1000	0.06	0.053	0.786	0.755
1500	0.053	0.047	0.952	0.948
2000	0.06	0.055	0.992	0.991

Figure 7.7: **One-Sample, $\hat{K}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.328	0.144	0.399	0.186
100	0.206	0.065	0.271	0.096
200	0.118	0.047	0.229	0.105
300	0.077	0.037	0.282	0.181
400	0.062	0.039	0.333	0.267
500	0.066	0.05	0.413	0.362
750	0.055	0.035	0.598	0.564
1000	0.06	0.053	0.786	0.755
1500	0.053	0.047	0.952	0.948
2000	0.06	0.055	0.992	0.991

Figure 7.8: **One-Sample, $\hat{K}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.242	0.11	0.292	0.132
100	0.159	0.056	0.215	0.081
200	0.099	0.041	0.208	0.095
300	0.077	0.037	0.282	0.181
400	0.062	0.039	0.333	0.267
500	0.066	0.05	0.413	0.362
750	0.055	0.035	0.598	0.564
1000	0.06	0.053	0.786	0.755
1500	0.053	0.047	0.952	0.948
2000	0.06	0.055	0.992	0.991

Figure 7.9: **One-Sample, $\hat{\boldsymbol{K}}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.275	0.117	0.332	0.144
100	0.167	0.057	0.223	0.082
200	0.099	0.041	0.208	0.095
300	0.077	0.037	0.282	0.181
400	0.062	0.039	0.333	0.267
500	0.066	0.05	0.413	0.362
750	0.055	0.035	0.598	0.564
1000	0.06	0.053	0.786	0.755
1500	0.053	0.047	0.952	0.948
2000	0.06	0.055	0.992	0.991

Figure 7.10: **Two-Sample, $\boldsymbol{K}$ known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.19	0.06	0.2	0.07
100	0.17	0.05	0.15	0.04
200	0.1	0.04	0.13	0.07
300	0.05	0.03	0.16	0.07
400	0.09	0.08	0.18	0.13
500	0.05	0.03	0.22	0.16
750	0.05	0.04	0.3	0.26
1000	0.07	0.07	0.41	0.4
1500	0.04	0.04	0.68	0.67
2000	0.05	0.04	0.8	0.79
2500	0.038	0.036	0.87	0.87
3000	0.071	0.069	0.96	0.96

Figure 7.11: **Two-Sample, K_{obs}**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.43	0.13	0.5	0.15
100	0.23	0.08	0.18	0.08
200	0.1	0.04	0.14	0.07
300	0.05	0.03	0.16	0.07
400	0.09	0.08	0.18	0.13
500	0.05	0.03	0.22	0.16
750	0.05	0.04	0.3	0.26
1000	0.07	0.07	0.41	0.4
1500	0.04	0.04	0.68	0.67
2000	0.05	0.04	0.8	0.79
2500	0.038	0.036	0.87	0.87
3000	0.071	0.069	0.96	0.96

Figure 7.12: **Two-Sample, $\hat{K}_{Chao1}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.23	0.106	0.287	0.128
100	0.157	0.035	0.175	0.038
200	0.078	0.033	0.141	0.074
300	0.059	0.036	0.122	0.081
400	0.06	0.039	0.173	0.132
500	0.068	0.053	0.205	0.173
750	0.059	0.047	0.289	0.269
1000	0.05	0.033	0.399	0.378
1500	0.041	0.039	0.62	0.61
2000	0.041	0.037	0.764	0.751
2500	0.038	0.036	0.876	0.869
3000	0.071	0.069	0.948	0.948

Figure 7.13: **Two-Sample, $\hat{\boldsymbol{K}}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.236	0.107	0.285	0.125
100	0.16	0.036	0.179	0.04
200	0.078	0.033	0.142	0.074
300	0.059	0.036	0.122	0.081
400	0.06	0.039	0.173	0.132
500	0.068	0.053	0.205	0.173
750	0.059	0.047	0.2389	0.269
1000	0.04	0.033	0.399	0.378
1500	0.041	0.039	0.62	0.61
2000	0.041	0.037	0.764	0.751
2500	0.038	0.036	0.876	0.869
3000	0.071	0.069	0.948	0.948

Figure 7.14: **Two-Sample, $\hat{\boldsymbol{K}}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.147	0.079	0.175	0.094
100	0.118	0.028	0.13	0.029
200	0.073	0.032	0.13	0.067
300	0.059	0.036	0.122	0.081
400	0.06	0.039	0.173	0.132
500	0.068	0.053	0.205	0.173
750	0.059	0.047	0.289	0.269
1000	0.04	0.033	0.399	0.378
1500	0.041	0.039	0.62	0.61
2000	0.041	0.037	0.764	0.751
2500	0.038	0.036	0.876	0.869
3000	0.071	0.069	0.948	0.948

Figure 7.15: **Two-Sample, $\hat{K}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.176	0.085	0.218	0.106
100	0.128	0.029	0.135	0.031
200	0.073	0.032	0.13	0.067
300	0.059	0.036	0.122	0.081
400	0.06	0.039	0.173	0.132
500	0.068	0.053	0.205	0.173
750	0.059	0.047	0.289	0.269
1000	0.04	0.033	0.399	0.378
1500	0.041	0.039	0.62	0.61
2000	0.041	0.037	0.764	0.751
2500	0.038	0.036	0.876	0.869
3000	0.071	0.069	0.948	0.948

Figure 7.16: **Two Sample Sizes, K known**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.065	0.202	0.088	0.228
50	150	0.217	0.064	0.243	0.09
100	300	0.135	0.044	0.199	0.06
200	600	0.072	0.029	0.151	0.086
300	900	0.062	0.039	0.202	0.153
400	1200	0.063	0.045	0.232	0.2
500	1500	0.048	0.033	0.296	0.259
750	2250	0.055	0.047	0.423	0.403
300	900	0.061	0.054	0.601	0.59
1500	4500	0.05	0.047	0.835	0.832
2000	6000	0.053	0.05	0.943	0.942
2500	7500	0.047	0.047	0.985	0.985
3000	9000	0.048	0.043	0.996	0.996

Figure 7.17: **Two Sample Sizes, K_{obs}**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.922	0.794	0.918	0.801
50	150	0.505	0.202	0.533	0.216
100	300	0.202	0.066	0.249	0.089
200	600	0.074	0.03	0.152	0.088
300	900	0.062	0.039	0.202	0.153
400	1200	0.063	0.045	0.232	0.2
500	1500	0.048	0.033	0.296	0.259
750	2250	0.055	0.047	0.423	0.403
300	900	0.061	0.054	0.601	0.59
1500	4500	0.05	0.047	0.835	0.832
2000	6000	0.053	0.05	0.943	0.942
2500	7500	0.047	0.047	0.985	0.985
3000	9000	0.048	0.043	0.996	0.996

Figure 7.18: **Two Sample Sizes, $\hat{K}_{Chao1}$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.283	0.341	0.291	0.329
50	150	0.26	0.114	0.3	0.135
100	300	0.15	0.059	0.202	0.075
200	600	0.071	0.03	0.152	0.088
300	900	0.062	0.039	0.202	0.153
400	1200	0.063	0.045	0.232	0.2
500	1500	0.048	0.033	0.296	0.259
750	2250	0.055	0.047	0.423	0.403
300	900	0.061	0.054	0.601	0.59
1500	4500	0.05	0.047	0.835	0.832
2000	6000	0.053	0.05	0.943	942
2500	7500	0.047	0.047	0.985	0.985
3000	9000	0.048	0.043	0.996	0.996

Figure 7.19: **Two Sample Sizes, $\hat{\boldsymbol{K}}_0$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.28	0.334	0.278	0.321
50	150	0.26	0.112	0.284	0.137
100	300	0.153	0.059	0.209	0.076
200	600	0.071	0.03	0.152	0.088
300	900	0.062	0.039	0.202	0.153
400	1200	0.063	0.045	0.232	0.2
500	1500	0.048	0.033	0.296	0.259
750	2250	0.055	0.047	0.423	0.403
300	900	0.061	0.054	0.601	0.59
1500	4500	0.05	0.0417	0.835	0.832
2000	6000	0.053	0.05	0.943	0.942
2500	7500	0.047	0.047	0.985	0.985
3000	9000	0.048	0.043	0.996	0.996

Figure 7.20: **Two Sample Sizes, $\hat{\boldsymbol{K}}_1$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.203	0.282	0.21	0.276
50	150	0.164	0.083	0.189	0.105
100	300	0.118	0.041	0.154	0.058
200	600	0.064	0.027	0.144	0.083
300	900	0.062	0.039	0.202	0.153
400	1200	0.063	0.045	0.232	0.2
500	1500	0.048	0.033	0.296	0.259
750	2250	0.055	0.047	0.423	0.403
300	900	0.061	0.054	0.601	0.59
1500	4500	0.05	0.047	0.835	0.832
2000	6000	0.053	0.05	0.943	0.942
2500	7500	0.047	0.047	0.985	0.985
3000	9000	0.048	0.043	0.996	0.996

Figure 7.21: **Two Sample Sizes, $\hat{\boldsymbol{K}}_2$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.244	0.306	0.248	0.298
50	150	0.202	0.095	0.231	0.115
100	300	0.128	0.047	0.165	0.059
200	600	0.064	0.027	0.144	0.083
300	900	0.062	0.039	0.202	0.153
400	1200	0.063	0.045	0.232	0.2
500	1500	0.048	0.033	0.296	0.259
750	2250	0.055	0.047	0.423	0.403
300	900	0.061	0.054	0.601	0.59
1500	4500	0.05	0.047	0.835	0.832
2000	6000	0.053	0.05	0.943	0.942
2500	7500	0.047	0.047	0.985	0.985
3000	9000	0.048	0.043	0.996	0.996

Clearly the jackknife estimator test converges to the size of the test $\alpha = 0.05$ more quickly than the plug-in estimator. And when the plug-in estimator test converges to $\alpha = 0.05$, the powers of the two tests are approximately equal.

7.2 Uniform Distribution: K=100

Next, suppose that $K = 100$ and we have two equal uniform distributions, $\mathbf{p} = \mathbf{q} = \{1/100, \ldots, 1/100\}$. Again we have the actual value of Jensen-Shannon Divergence at 0. The error tables are as follows, plug-in estimator in red and jackknife estimator in blue.

Figure 7.22: **One-Sample**

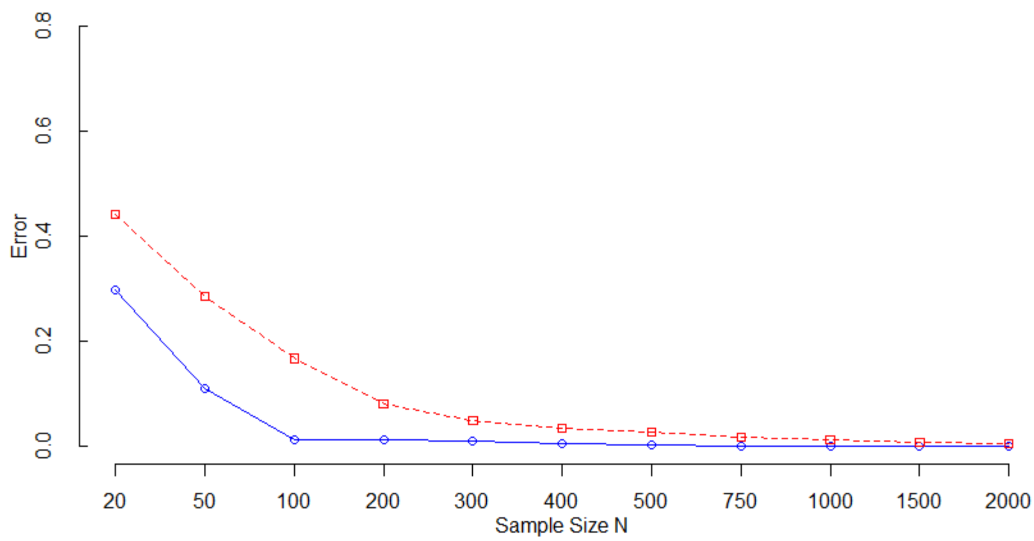
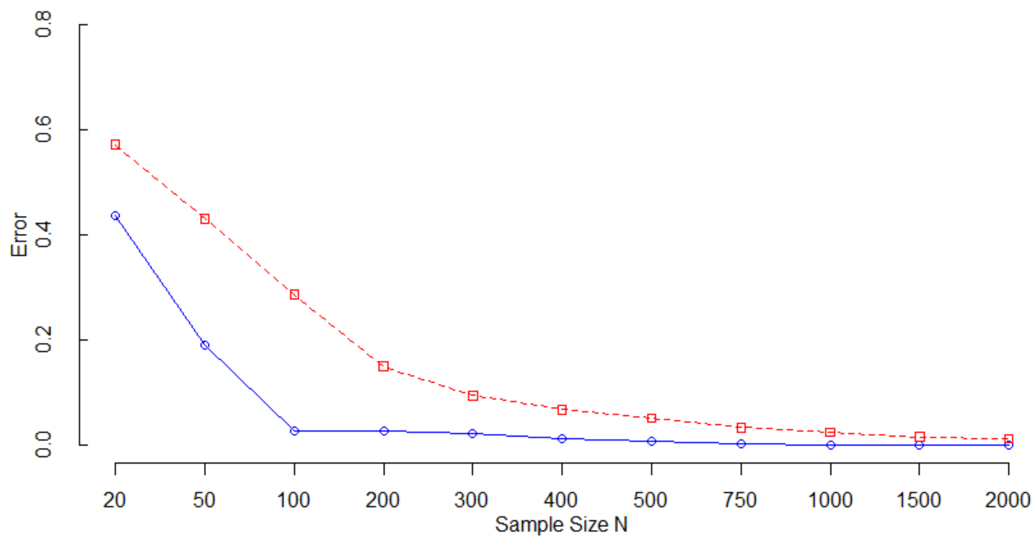


Figure 7.23: **Two-Sample**



Now suppose for $\mathbf{q}$, that we subtract 1/600 from $\{q_1, \ldots, q_{50}\}$, and add 1/600 to $\{q_{51}, \ldots, q_{100}\}$. This adjusted $\mathbf{q}$ distribution juxtaposed on the uniform $\mathbf{p}$ looks something like this:

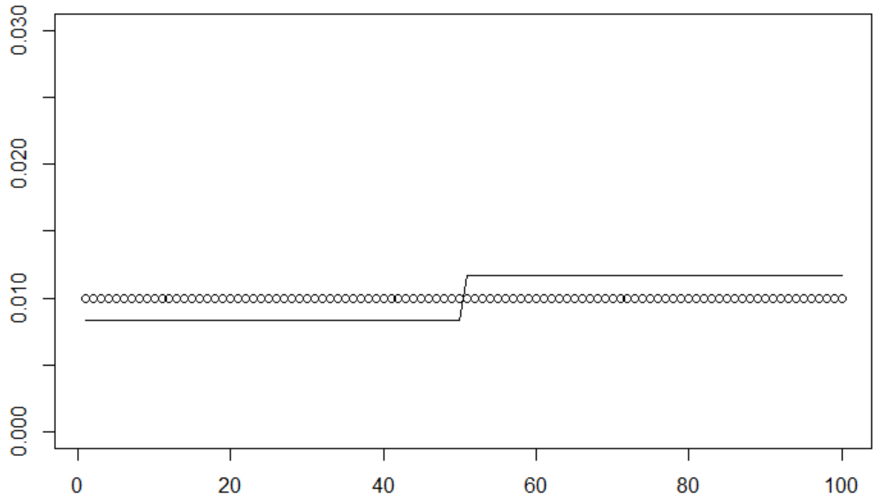


Figure 7.24

Here, between uniform **p** and this adjusted **q** given in Figure 7.24, the actual value of Jensen-Shannon Divergence is 0.003500705. For the alternative hypothesis when H_0 is false, **q** is given by 7.24.

Figure 7.25: **One-Sample, *K* known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.11	0.901	0.123	0.927
100	0.793	0.0252	0.861	0.363
200	0.635	0.041	0.738	0.071
300	0.408	0.013	0.609	0.055
400	0.275	0.012	0.562	0.072
500	0.204	0.019	0.498	0.089
750	0.121	0.026	0.553	0.25
1000	0.086	0.02	0.686	0.468
1500	0.086	0.042	0.868	0.778
2000	0.069	0.038	0.952	0.927

Figure 7.26: **One-Sample, $\boldsymbol{K}_{obs}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	1	1	1	1
100	1	0.812	0.999	0.874
200	0.857	0.148	0.915	0.192
300	0.518	0.039	0.695	0.106
400	0.327	0.022	0.603	0.097
500	0.231	0.025	0.518	0.106
750	0.124	0.027	0.553	0.254
1000	0.086	0.02	0.686	0.468
1500	0.086	0.042	0.868	0.778
2000	0.069	0.038	0.952	0.927

Figure 7.27: **One-Sample, $\hat{\boldsymbol{K}}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.388	0.609	0.396	0.644
100	0.617	0.336	0.678	0.407
200	0.585	0.081	0.674	0.106
300	0.402	0.022	0.572	0.074
400	0.264	0.015	0.538	0.079
500	0.203	0.022	0.482	0.093
750	0.124	0.024	0.55	0.247
1000	0.086	0.02	0.686	0.468
1500	0.086	0.042	0.868	0.778
2000	0.069	0.038	0.952	0.927

Figure 7.28: **One-Sample, $\hat{\boldsymbol{K}}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.395	0.612	0.407	0.649
100	0.612	0.337	0.675	0.406
200	0.584	0.08	0.677	0.107
300	0.405	0.022	0.577	0.073
400	0.263	0.016	0.542	0.079
500	0.205	0.022	0.483	0.093
750	0.124	0.024	0.55	0.247
1000	0.086	0.02	0.685	0.468
1500	0.086	0.042	0.868	0.778
2000	0.069	0.038	0.952	0.927

Figure 7.29: **One-Sample, $\hat{\boldsymbol{K}}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.297	0.591	0.308	0.623
100	0.411	0.224	0.491	0.299
200	0.354	0.038	0.43	0.063
300	0.234	0.013	0.405	0.033
400	0.179	0.007	0.414	0.053
500	0.143	0.017	0.376	0.062
750	0.105	0.019	0.508	0.214
1000	0.086	0.02	0.685	0.463
1500	0.086	0.042	0.868	0.778
2000	0.069	0.038	0.952	0.927

Figure 7.30: **One-Sample, $\hat{K}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.384	0.591	0.39	0.624
100	0.56	0.308	0.621	0.375
200	0.495	0.063	0.56	0.082
300	0.311	0.017	0.48	0.049
400	0.197	0.01	0.444	0.06
500	0.148	0.018	0.387	0.063
750	0.105	0.019	0.508	0.214
1000	0.086	0.02	0.685	0.463
1500	0.086	0.042	0.868	0.778
2000	0.069	0.038	0.952	0.927

Figure 7.31: **Two-Sample, K known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0	0.11	0	0.134
100	0.236	0.156	0.279	0.191
200	0.422	0.022	0.465	0.031
300	0.2269	0.01	0.387	0.015
400	0.163	0.009	0.316	0.017
500	0.156	0.014	0.24	0.029
750	0.079	0.016	0.277	0.095
1000	0.071	0.025	0.318	0.17
1500	0.077	0.045	0.471	0.361
2000	0.06	0.037	0.612	0.53
2500	0.052	0.036	0.746	0.703
3000	0.076	0.06	0.837	0.797
3500	0.053	0.045	0.895	0.876
4000	0.044	0.036	0.962	0.947

Figure 7.32: **Two-Sample, K_{obs}**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	1	0.966	1	0.975
100	0.995	0.766	0.994	0.787
200	0.752	0.097	0.801	0.116
300	0.418	0.025	0.531	0.036
400	0.215	0.015	0.383	0.034
500	0.172	0.021	0.266	0.035
750	0.079	0.017	0.281	0.097
1000	0.071	0.025	0.318	0.17
1500	0.077	0.045	0.471	0.361
2000	0.06	0.037	0.612	0.53
2500	0.052	0.036	0.746	0.703
3000	0.076	0.06	0.837	0.797
3500	0.053	0.045	0.895	0.876
4000	0.044	0.036	0.962	0.947

Figure 7.33: **Two-Sample, $\hat{K}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.111	0.225	0.127	0.263
100	0.313	0.235	0.359	0.278
200	0.417	0.039	0.443	0.044
300	0.271	0.013	0.376	0.021
400	0.164	0.013	0.32	0.022
500	0.156	0.017	0.236	0.027
750	0.076	0.015	0.275	0.094
1000	0.071	0.025	0.318	0.17
1500	0.077	0.045	0.471	0.361
2000	0.06	0.037	0.612	0.53
2500	0.052	0.036	0.746	0.703
3000	0.076	0.06	0.837	0.797
3500	0.053	0.045	0.895	0.876
4000	0.044	0.036	0.962	0.947

Figure 7.34: **Two-Sample, $\hat{K}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.108	0.222	0.121	0.257
100	0.307	0.229	0.353	0.275
200	0.413	0.038	0.447	0.045
300	0.272	0.013	0.375	0.022
400	0.163	0.013	0.319	0.023
500	0.156	0.016	0.239	0.027
750	0.076	0.016	0.275	0.095
1000	0.071	0.025	0.3181	0.17
1500	0.077	0.045	0.471	0.361
2000	0.06	0.037	0.912	0.53
2500	0.052	0.036	0.746	0.703
3000	0.076	0.06	0.837	0.797
3500	0.053	0.045	0.895	0.876
4000	0.044	0.036	0.962	0.947

Figure 7.35: **Two-Sample, $\hat{K}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.056	0.156	0.061	0.184
100	0.148	0.123	0.18	0.153
200	0.188	0.016	0.223	0.021
300	0.128	0.005	0.213	0.009
400	0.099	0.005	0.21	0.015
500	0.111	0.008	0.174	0.017
750	0.065	0.013	0.243	0.084
1000	0.071	0.024	0.313	0.167
1500	0.077	0.045	0.471	0.361
2000	0.06	0.037	0.612	0.53
2500	0.052	0.036	0.746	0.703
3000	0.076	0.06	0.837	0.797
3500	0.053	0.045	0.895	0.876
4000	0.044	0.036	0.962	0.947

Figure 7.36: **Two-Sample, $\hat{K}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.094	0.202	0.1	0.241
100	0.269	0.192	0.298	0.24
200	0.308	0.024	0.342	0.032
300	0.19	0.008	0.278	0.017
400	0.122	0.007	0.23	0.017
500	0.118	0.008	0.182	0.02
750	0.065	0.013	0.243	0.084
1000	0.071	0.024	0.313	0.167
1500	0.077	0.045	0.471	0.361
2000	0.06	0.037	0.612	0.53
2500	0.052	0.036	0.746	0.703
3000	0.076	0.06	0.837	0.797
3500	0.053	0.045	0.895	0.876
4000	0.044	0.036	0.962	0.947

Figure 7.37: **Two Sample Sizes, K known**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.018	0.442	0.022	0.503
100	300	0.448	0.14	0.506	0.197
200	600	0.406	0.033	0.508	0.058
300	900	0.264	0.014	0.389	0.032
400	1200	0.174	0.017	0.33	0.047
500	1500	0.144	0.016	0.315	0.066
750	2250	0.101	0.027	0.361	0.173
1000	3000	0.076	0.032	0.47	0.31
1500	4500	0.076	0.047	0.676	0.613
2000	6000	0.069	0.047	0.838	0.794
2500	7500	0.067	0.047	0.936	0.913
3000	9000	0.05	0.039	0.98	0.972
3500	10500	0.051	0.039	0.992	0.989
4000	12000	0.05	0.042	0.997	0.997

Figure 7.38: **Two Sample Sizes, $\boldsymbol{K}_{obs}$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	1	0.999	1	1
100	300	0.999	0.748	1	0.793
200	600	0.726	0.118	0.8	0.156
300	900	0.392	0.033	0.503	0.074
400	1200	0.225	0.037	0.387	0.062
500	1500	0.157	0.02	0.326	0.077
750	2250	0.101	0.028	0.362	0.174
1000	3000	0.076	0.032	0.47	0.31
1500	4500	0.076	0.047	0.676	0.613
2000	6000	0.069	0.047	0.838	0.794
2500	7500	0.067	0.047	0.936	0.913
3000	9000	0.05	0.039	0.98	0.972
3500	10500	0.051	0.039	0.992	0.989
4000	12000	0.05	0.042	0.997	0.997

Figure 7.39: **Two Sample Sizes, $\hat{\boldsymbol{K}}_{Chao1}$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.291	0.473	0.289	0.492
100	300	0.45	0.251	0.499	0.288
200	600	0.408	0.053	0.491	0.083
300	900	0.262	0.023	0.374	0.046
400	1200	0.18	0.025	0.322	0.053
500	1500	0.135	0.017	0.309	0.069
750	2250	0.097	0.028	0.36	0.172
1000	3000	0.046	0.032	0.47	0.31
1500	4500	0.046	0.047	0.676	0.613
2000	6000	0.069	0.047	0.838	0.794
2500	7500	0.067	0.047	0.936	0.913
3000	9000	0.05	0.039	0.98	0.972
3500	10500	0.051	0.039	0.992	0.989
4000	12000	0.05	0.042	0.997	0.997

Figure 7.40: **Two Sample Sizes,** $\hat{\boldsymbol{K}}_0$

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.281	0.48	0.284	0.497
100	300	0.447	0.25	0.498	0.282
200	600	0.407	0.053	0.487	0.084
300	900	0.259	0.023	0.375	0.046
400	1200	0.181	0.025	0.323	0.053
500	1500	0.136	0.017	0.31	0.069
750	2250	0.097	0.028	0.36	0.172
1000	3000	0.076	0.032	0.47	0.31
1500	4500	0.076	0.047	0.676	0.613
2000	6000	0.069	0.047	0.838	0.794
2500	7500	0.067	0.047	0.936	0.913
3000	9000	0.05	0.039	0.98	0.972
3500	10500	0.051	0.039	0.992	0.989
4000	12000	0.05	0.042	0.997	0.997

Figure 7.41: **Two Sample Sizes,** $\hat{\boldsymbol{K}}_1$

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.182	0.378	0.176	0.4
100	300	0.258	0.14	0.301	0.171
200	600	0.189	0.029	0.252	0.047
300	900	0.125	0.014	0.208	0.025
400	1200	0.113	0.01	0.217	0.037
500	1500	0.088	0.014	0.247	0.053
750	2250	0.083	0.025	0.336	0.153
1000	3000	0.076	0.032	0.47	0.31
1500	4500	0.0767	0.047	0.676	0.613
2000	6000	0.069	0.047	0.838	0.794
2500	7500	0.067	0.047	0.936	0.913
3000	9000	0.05	0.039	0.98	0.972
3500	10500	0.051	0.039	0.992	0.989
4000	12000	0.05	0.042	0.997	0.997

Figure 7.42: **Two Sample Sizes, $\hat{\boldsymbol{K}}_2$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.255	0.456	0.259	0.478
100	300	0.398	0.215	0.438	0.248
200	600	0.311	0.04	0.381	0.061
300	900	0.181	0.019	0.292	0.033
400	1200	0.131	0.013	0.255	0.042
500	1500	0.097	0.014	0.259	0.057
750	2250	0.083	0.025	0.336	0.153
1000	3000	0.076	0.032	0.47	0.31
1500	4500	0.076	0.047	0.676	0.613
2000	6000	0.069	0.047	0.838	0.794
2500	7500	0.067	0.047	0.936	0.913
3000	9000	0.05	0.039	0.98	0.972
3500	10500	0.051	0.039	0.992	0.989
4000	12000	0.05	0.042	0.997	0.997

7.3 Triangle Distribution: K=30

Next, suppose that $K = 30$ and we have two equal triangle distributions, $\mathbf{p} = \mathbf{q} = \{1/240, 2/240, \ldots, 15/240, 15/240, \ldots, 2/240, 1/240\}$. Again we have the actual value of Jensen-Shannon Divergence at 0. The error tables are as follows, plug-in estimator in red and jackknife estimator in blue.

Figure 7.43: **One-Sample**

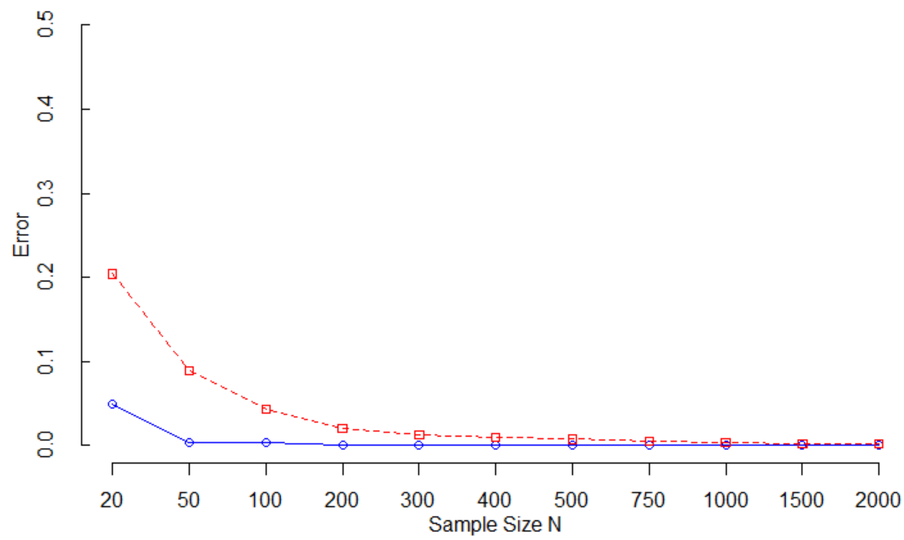
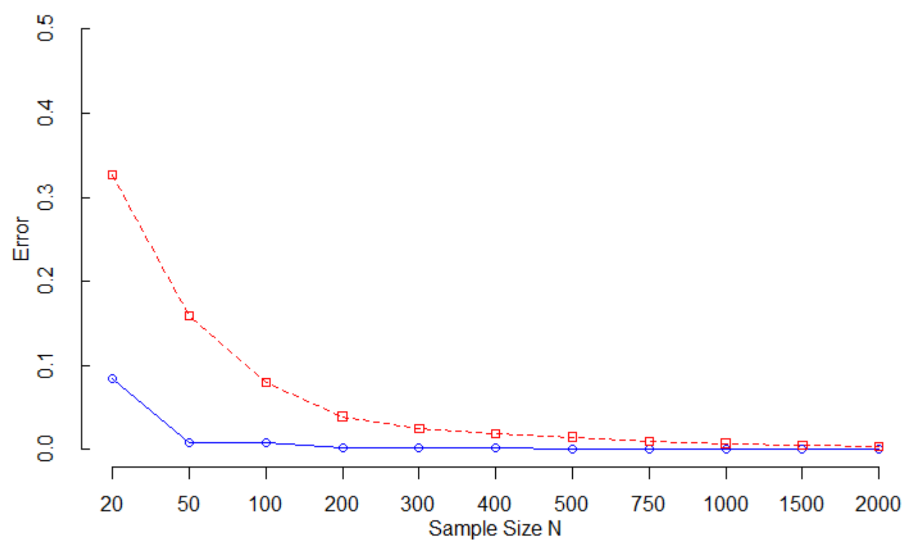


Figure 7.44: **Two-Sample**



Now suppose for $\mathbf{q}$, that we adjust $\mathbf{q}$ to be $\{1/240-1/1000, 2/240-2/1000, \ldots, 15/240-$

$15/1000, 15/240 + 15/1000, \ldots, 2/240 + 2/1000, 1/240 + 1/1000\}$. This adjusted $\mathbf{q}$ distribution juxtaposed on the original triangle $\mathbf{p}$ is demonstrated by the following:

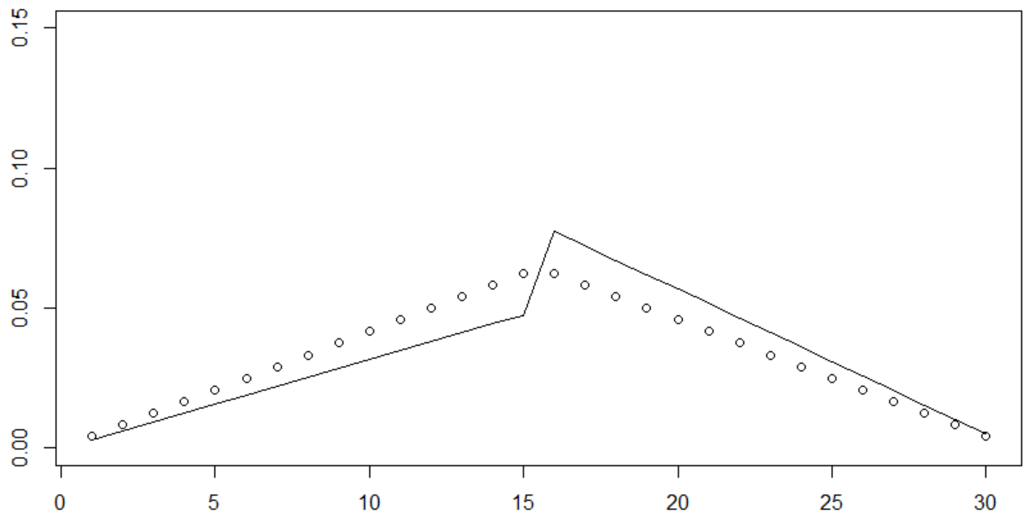


Figure 7.45

Here, the value of Jensen-Shannon divergence between these two distributions given in Figure 7.45 is 0.007324147. For the alternative hypothesis when H_0 is false, $\mathbf{q}$ is given by Figure 7.45.

Figure 7.46: **One-Sample, K known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.199	0.121	0.304	0.171
100	0.172	0.079	0.388	0.223
200	0.115	0.052	0.494	0.321
300	0.087	0.04	0.699	0.563
400	0.1	0.062	0.809	0.716
500	0.07	0.039	0.913	0.858
750	0.068	0.047	0.985	0.982
1000	0.058	0.039	1	1
1500	0.068	0.045	1	1
2000	0.059	0.052	1	1

Figure 7.47: **One-Sample, K_{obs}**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.63	0.372	0.711	0.448
100	0.325	0.178	0.58	0.345
200	0.161	0.084	0.549	0.382
300	0.114	0.06	0.727	0.591
400	0.118	0.076	0.819	0.73
500	0.078	0.048	0.915	0.864
750	0.071	0.05	0.985	0.982
1000	0.059	0.041	1	1
1500	0.069	0.046	1	1
2000	0.059	0.052	1	1

Figure 7.48: **One-Sample, $\hat{K}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.36	0.237	0.431	0.292
100	0.236	0.13	0.447	0.286
200	0.129	0.064	0.483	0.329
300	0.1	0.052	0.689	0.557
400	0.114	0.076	0.809	0.706
500	0.077	0.045	0.911	0.855
750	0.071	0.05	0.985	0.982
1000	0.058	0.041	1	1
1500	0.069	0.046	1	1
2000	0.059	0.052	1	1

Figure 7.49: **One-Sample,** $\hat{\boldsymbol{K}}_0$

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.366	0.244	0.44	0.3
100	0.248	0.138	0.465	0.295
200	0.133	0.067	0.491	0.335
300	0.1	0.052	0.691	0.555
400	0.112	0.075	0.808	0.706
500	0.075	0.044	0.91	0.853
750	0.07	0.05	0.985	0.979
1000	0.057	0.041	1	1
1500	0.069	0.046	1	1
2000	0.059	0.052	1	1

Figure 7.50: **One-Sample,** $\hat{\boldsymbol{K}}_1$

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.258	0.182	0.319	0.228
100	0.178	0.092	0.352	0.223
200	0.086	0.05	0.387	0.268
300	0.075	0.043	0.588	0.469
400	0.09	0.064	0.748	0.644
500	0.064	0.042	0.88	0.814
750	0.065	0.049	0.985	0.978
1000	0.057	0.041	1	1
1500	0.068	0.046	1	1
2000	0.059	0.052	1	1

Figure 7.51: **One-Sample, $\hat{K}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.283	0.197	0.344	0.244
100	0.181	0.093	0.354	0.225
200	0.086	0.05	0.387	0.268
300	0.075	0.043	0.588	0.469
400	0.09	0.064	0.748	0.644
500	0.064	0.042	0.88	0.814
750	0.065	0.049	0.985	0.978
1000	0.057	0.041	1	1
1500	0.068	0.046	1	1
2000	0.059	0.052	1	1

Figure 7.52: **Two-Sample, $\boldsymbol{K}$ known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.07	0.053	0.124	0.084
100	0.084	0.049	0.196	0.098
200	0.093	0.035	0.233	0.125
300	0.083	0.049	0.368	0.239
400	0.079	0.047	0.46	0.334
500	0.056	0.028	0.535	0.427
750	0.0854	0.034	0.784	0.717
1000	0.041	0.029	0.899	0.867
1500	0.061	0.052	0.99	0.987
2000	0.046	0.037	0.998	0.998

Figure 7.53: **Two-Sample, $\boldsymbol{K}_{obs}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.446	0.235	0.54	0.309
100	0.213	0.106	0.37	0.193
200	0.136	0.066	0.317	0.174
300	0.099	0.06	0.412	0.28
400	0.093	0.057	0.491	0.356
500	0.064	0.036	0.545	0.434
750	0.057	0.038	0.784	0.718
1000	0.044	0.033	0.899	0.867
1500	0.061	0.052	0.99	0.987
2000	0.046	0.037	0.998	0.998

Figure 7.54: **Two-Sample, $\hat{\boldsymbol{K}}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.195	0.116	0.269	0.184
100	0.124	0.075	0.249	0.133
200	0.099	0.052	0.264	0.149
300	0.084	0.053	0.368	0.258
400	0.089	0.055	0.463	0.342
500	0.062	0.035	0.534	0.423
750	0.057	0.037	0.784	0.717
1000	0.043	0.033	0.899	0.866
1500	0.061	0.052	0.99	0.987
2000	0.046	0.037	0.998	0.998

Figure 7.55: **Two-Sample, $\hat{K}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.198	0.126	0.274	0.188
100	0.131	0.079	0.263	0.143
200	0.105	0.052	0.266	0.151
300	0.085	0.054	0.372	0.26
400	0.089	0.054	0.46	0.339
500	0.061	0.035	0.532	0.421
750	0.057	0.036	0.78	0.714
1000	0.042	0.033	0.899	0.865
1500	0.061	0.052	0.99	0.987
2000	0.046	0.037	0.998	0.998

Figure 7.56: **Two-Sample, $\hat{K}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.123	0.078	0.178	0.128
100	0.089	0.052	0.184	0.104
200	0.066	0.035	0.188	0.101
300	0.066	0.038	0.294	0.194
400	0.061	0.042	0.409	0.303
500	0.052	0.032	0.482	0.392
750	0.054	0.035	0.768	0.704
1000	0.042	0.033	0.894	0.861
1500	0.061	0.052	0.99	0.987
2000	0.046	0.037	0.998	0.998

Figure 7.57: **Two-Sample, $\hat{K}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.141	0.094	0.198	0.142
100	0.092	0.052	0.187	0.106
200	0.066	0.035	0.188	0.101
300	0.066	0.038	0.294	0.194
400	0.061	0.042	0.409	0.303
500	0.052	0.032	0.482	0.392
750	0.054	0.035	0.768	0.704
1000	0.042	0.033	0.894	0.861
1500	0.061	0.052	0.99	0.987
2000	0.046	0.037	0.998	0.998

Figure 7.58: **Two Sample Sizes, K known**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.023	0.116	0.038	0.128
50	150	0.107	0.068	0.203	0.102
100	300	0.107	0.056	0.262	0.14
200	600	0.085	0.041	0.383	0.265
300	900	0.099	0.059	0.516	0.4
400	1200	0.083	0.044	0.659	0.553
500	1500	0.068	0.041	0.764	0.682
750	2250	0.053	0.031	0.928	0.907
300	900	0.078	0.064	0.988	0.983
1500	4500	0.058	0.053	1	1
2000	6000	0.065	0.057	1	1
2500	7500	0.055	0.049	1	1
3000	9000	0.06	0.058	1	1

Figure 7.59: **Two Sample Sizes, $\boldsymbol{K}_{obs}$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.866	0.74	0.864	0.746
50	150	0.48	0.276	0.602	0.381
100	300	0.267	0.153	0.44	0.262
200	600	0.149	0.073	0.463	0.313
300	900	0.11	0.079	0.554	0.432
400	1200	0.098	0.054	0.678	0.57
500	1500	0.081	0.05	0.772	0.692
750	2250	0.057	0.038	0.928	0.907
300	900	0.079	0.066	0.988	0.983
1500	4500	0.058	0.053	1	1
2000	6000	0.065	0.057	1	1
2500	7500	0.055	0.049	1	1
3000	9000	0.06	0.058	1	1

Figure 7.60: **Two Sample Sizes, $\hat{\boldsymbol{K}}_{Chao1a}$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.3	0.342	0.293	0.348
50	150	0.234	0.182	0.345	0.221
100	300	0.168	0.107	0.306	0.193
200	600	0.107	0.064	0.386	0.266
300	900	0.097	0.075	0.514	0.395
400	1200	0.092	0.054	0.65	0.549
500	1500	0.074	0.048	0.755	0.6866
750	2250	0.055	0.038	0.928	0.907
300	900	0.078	0.066	0.988	0.983
1500	4500	0.058	0.053	1	1
2000	6000	0.065	0.057	1	1
2500	7500	0.055	0.049	1	1
3000	9000	0.06	0.058	1	1

Figure 7.61: **Two Sample Sizes, $\hat{\boldsymbol{K}}_0$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.306	0.353	0.301	0.3538
50	150	0.249	0.184	0.36	0.23
100	300	0.174	0.115	0.327	0.199
200	600	0.108	0.065	0.393	0.275
300	900	0.099	0.074	0.518	0.395
400	1200	0.092	0.053	0.647	0.547
500	1500	0.072	0.047	0.753	0.686
750	2250	0.055	0.038	0.928	0.906
300	900	0.078	0.066	0.988	0.983
1500	4500	0.058	0.053	1	1
2000	6000	0.065	0.057	1	1
2500	7500	0.055	0.049	1	1
3000	9000	0.06	0.058	1	1

Figure 7.62: **Two Sample Sizes, $\hat{\boldsymbol{K}}_1$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.221	0.284	0.227	0.298
50	150	0.162	0.14	0.232	0.166
100	300	0.12	0.077	0.227	0.144
200	600	0.071	0.045	0.292	0.227
300	900	0.08	0.0631	0.417	0.337
400	1200	0.072	0.037	0.571	0.485
500	1500	0.062	0.041	0.709	0.646
750	2250	0.0551	0.037	0.924	0.903
300	900	0.075	0.063	0.988	0.982
1500	4500	0.058	0.053	1	1
2000	6000	0.065	0.057	1	1
2500	7500	0.055	0.049	1	1
3000	9000	0.06	0.058	1	1

Figure 7.63: **Two Sample Sizes, $\hat{\boldsymbol{K}}_2$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	60	0.265	0.33	0.264	0.318
50	150	0.186	0.15	0.258	0.186
100	300	0.123	0.078	0.228	0.147
200	600	0.071	0.045	0.292	0.227
300	900	0.08	0.061	0.417	0.337
400	1200	0.072	0.037	0.571	0.485
500	1500	0.062	0.041	0.709	0.646
750	2250	0.051	0.037	0.924	0.903
300	900	0.075	0.063	0.988	0.982
1500	4500	0.058	0.053	1	1
2000	6000	0.065	0.057	1	1
2500	7500	0.055	0.049	1	1
3000	9000	0.06	0.058	1	1

7.4 Triangle Distribution: K=100

Now, suppose that $K = 100$ and that we have two equal triangle distributions, $\mathbf{p} = \mathbf{q} = \{1/2550, 2/2550, \ldots, 50/2550, 50/2550, \ldots, 2/2550, 1/2550\}$. The actual value of Jensen-Shannon Divergence is 0. The error tables are as follows, plug-in estimator in red and jackknife estimator in blue.

Figure 7.64: **One-Sample**

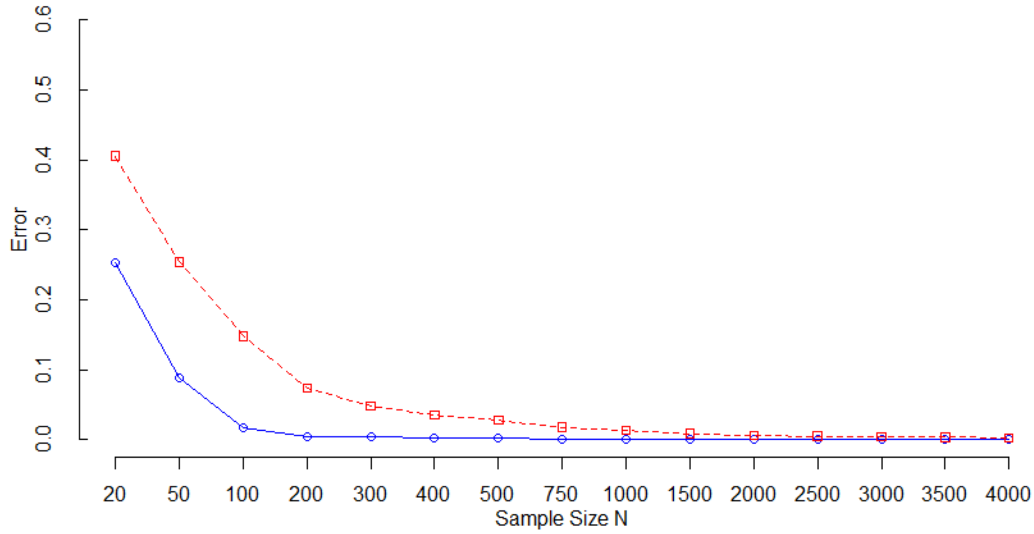
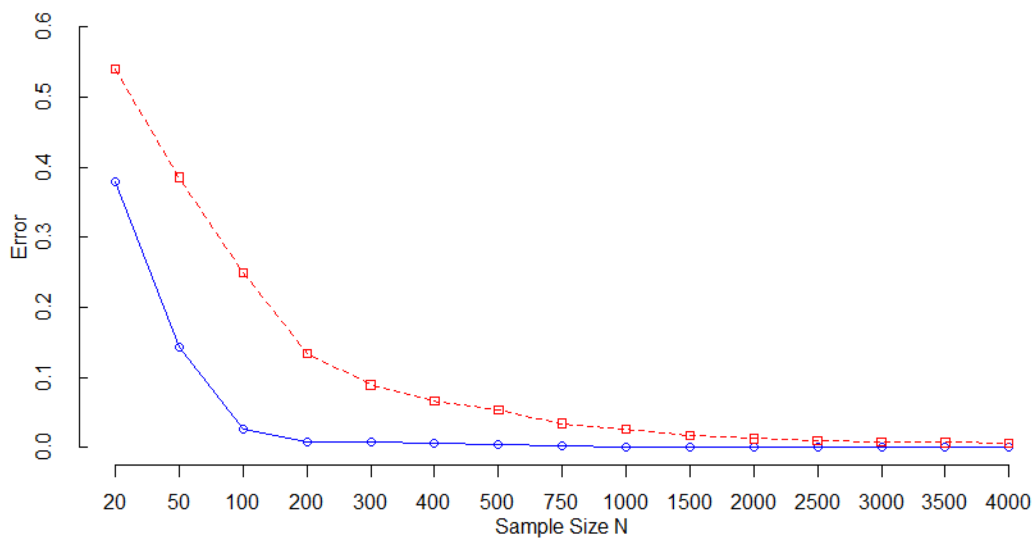


Figure 7.65: **Two-Sample**



Now suppose for $\mathbf{q}$, that we adjust $\mathbf{q}$ to be $\{1/2550-1/5000, 2/2550-2/5000, \ldots, 50/2550-$

$50/5000, 50/2550 + 50/5000, \ldots, 2/2550 + 2/5000, 1/2550 + 1/5000\}$. This adjusted $\mathbf{q}$ distribution juxtaposed on the original triangle $\mathbf{p}$ is demonstarted by the following:

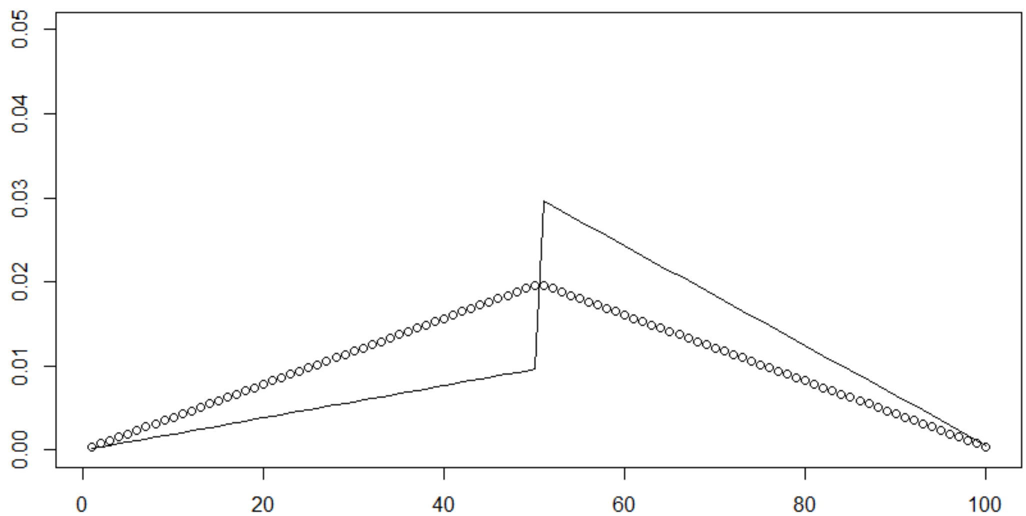


Figure 7.66

Here, the value of Jensen-Shannon divergence between these two distributions given in Figure 7.66 is 0.03531168. For the alternative hypothesis when H_0 is false, $\mathbf{q}$ is given by Figure 7.66.

Figure 7.67: **One-Sample, K known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0	0.998	0	0.995
50	0.001	0.775	0.005	0.81
100	0.336	0.296	0.515	0.457
200	0.395	0.102	0.726	0.265
300	0.355	0.068	0.792	0.302
400	0.279	0.049	0.852	0.381
500	0.231	0.045	0.87	0.474
750	0.178	0.04	0.972	0.788
1000	0.171	0.052	0.997	0.942
1500	0.116	0.035	1	0.998
2000	0.097	0.044	1	1
2500	0.083	0.039	1	1
3000	0.077	0.032	1	1

Figure 7.68: **One-Sample, K_{obs}**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	1	1	1	1
50	1	1	1	1
100	1	0.93	0.998	0.97
200	0.902	0.437	0.982	0.698
300	0.721	0.265	0.958	0.606
400	0.581	0.185	0.957	0.617
500	0.454	0.134	0.95	0.66
750	0.312	0.093	0.991	0.855
1000	0.254	0.091	0.997	0.961
1500	0.159	0.055	1	0.998
2000	0.12	0.058	1	1
2500	0.103	0.048	1	1
3000	0.091	0.042	1	1

Figure 7.69: **One-Sample, $\hat{\boldsymbol{K}}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.244	0.609	0.272	0.64
50	0.446	0.737	0.56	0.827
100	0.662	0.547	0.896	0.824
200	0.617	0.257	0.988	0.886
300	0.477	0.148	0.998	0.976
400	0.414	0.105	1	0.999
500	0.337	0.104	1	1
750	0.214	0.062	1	1
1000	0.201	0.063	1	1
1500	0.124	0.052	1	1
2000	0.116	0.056	1	1
2500	0.093	0.046	1	1
3000	0.082	0.037	1	1

Figure 7.70: **One-Sample, $\hat{\boldsymbol{K}}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.222	0.695	0.282	0.71
50	0.449	0.759	0.567	0.83
100	0.664	0.542	0.905	0.829
200	0.616	0.263	0.992	0.892
300	0.477	0.149	0.998	0.979
400	0.415	0.106	1	0.999
500	0.342	0.105	1	1
750	0.214	0.063	1	1
1000	0.204	0.062	1	1
1500	0.125	0.053	1	1
2000	0.116	0.057	1	1
2500	0.092	0.046	1	1
3000	0.083	0.036	1	1

Figure 7.71: **One-Sample, $\hat{K}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.127	0.695	0.175	0.71
50	0.319	0.662	0.435	0.759
100	0.438	0.383	0.74	0.688
200	0.368	0.131	0.948	0.76
300	0.256	0.079	0.988	0.898
400	0.185	0.058	1	0.983
500	0.161	0.051	1	0.994
750	0.098	0.028	1	1
1000	0.093	0.029	1	1
1500	0.058	0.034	1	1
2000	0.064	0.037	1	1
2500	0.056	0.025	1	1
3000	0.041	0.026	1	1

Figure 7.72: **One-Sample, $\hat{K}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.191	0.695	0.26	0.71
50	0.412	0.723	0.54	0.811
100	0.57	0.476	0.815	0.762
200	0.393	0.149	0.95	0.773
300	0.257	0.079	0.988	0.898
400	0.186	0.058	1	0.983
500	0.161	0.051	1	0.994
750	0.098	0.028	1	1
1000	0.093	0.029	1	1
1500	0.058	0.034	1	1
2000	0.064	0.037	1	1
2500	0.056	0.025	1	1
3000	0.041	0.026	1	1

Figure 7.73: **Two-Sample, K known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0	0	0	0
50	0	0.002	0	0.002
100	0.027	0.023	0.049	0.047
200	0.145	0.021	0.238	0.051
300	0.147	0.021	0.349	0.069
400	0.128	0.016	0.414	0.104
500	0.122	0.023	0.473	0.138
750	0.106	0.029	0.602	0.304
1000	0.108	0.034	0.758	0.499
1500	0.088	0.029	0.942	0.84
2000	0.064	0.031	0.989	0.965
2500	0.063	0.034	1	1
3000	0.05	0.028	1	1

Figure 7.74: **Two-Sample, K_{obs}**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	1	0	1	0
50	0.997	0.809	0.997	0.814
100	0.969	0.626	0.983	0.691
200	0.768	0.242	0.883	0.351
300	0.55	0.123	0.766	0.293
400	0.411	0.086	0.742	0.296
500	0.322	0.071	0.706	0.304
750	0.209	0.061	0.753	0.438
1000	0.185	0.065	0.835	0.604
1500	0.116	0.046	0.958	0.864
2000	0.087	0.041	0.993	0.97
2500	0.076	0.044	1	1
3000	0.066	0.032	1	1

Figure 7.75: **Two-Sample, $\hat{K}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.055	0	0.038	0
50	0.134	0.109	0.164	0.15
100	0.357	0.209	0.521	0.372
200	0.364	0.087	0.847	0.488
300	0.289	0.053	0.947	0.632
400	0.235	0.049	0.983	0.824
500	0.207	0.049	0.994	0.934
750	0.145	0.038	0.999	0.996
1000	0.116	0.038	1	1
1500	0.103	0.059	1	1
2000	0.076	0.04	1	1
2500	0.066	0.04	1	1
3000	0.059	0.027	1	1

Figure 7.76: **Two-Sample, $\hat{K}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.047	0	0.033	0
50	0.127	0.104	0.156	0.148
100	0.358	0.208	0.528	0.37
200	0.372	0.087	0.86	0.496
300	0.296	0.052	0.949	0.638
400	0.237	0.048	0.983	0.827
500	0.211	0.049	0.996	0.941
750	0.145	0.036	1	0.997
1000	0.118	0.039	1	1
1500	0.104	0.061	1	1
2000	0.076	0.04	1	1
2500	0.066	0.04	1	1
3000	0.057	0.027	1	1

Figure 7.77: **Two-Sample, $\hat{K}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.029	0	0.025	0
50	0.061	0.059	0.088	0.087
100	0.0176	0.096	0.286	0.208
200	0.0155	0.03	0.612	0.279
300	0.1	0.02	0.765	0.416
400	0.082	0.019	0.897	0.623
500	0.074	0.022	0.948	0.8
750	0.054	0.018	0.991	0.981
1000	0.046	0.02	0.996	0.991
1500	0.053	0.034	1	1
2000	0.036	0.022	1	1
2500	0.036	0.023	1	1
3000	0.031	0.013	1	1

Figure 7.78: **Two-Sample, $\hat{K}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.042	0	0.032	0
50	0.104	0.088	0.134	0.13
100	0.294	0.172	0.418	0.298
200	0.182	0.04	0.641	0.297
300	0.1	0.02	0.766	0.418
400	0.082	0.019	0.897	0.623
500	0.074	0.022	0.948	0.8
750	0.054	0.018	0.991	0.981
1000	0.046	0.02	0.996	0.991
1500	0.053	0.034	1	1
2000	0.036	0.022	1	1
2500	0.036	0.023	1	1
3000	0.031	0.013	1	1

Figure 7.79: **Two Sample Sizes, K known**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.001	0.086	0.012	0.158
100	300	0.095	0.07	0.522	0.322
200	600	0.191	0.039	0.912	0.59
300	900	0.22	0.044	0.988	0.875
400	1200	0.151	0.028	0.999	0.98
500	1500	0.142	0.038	1	0.998
750	2250	0.13	0.043	1	1
1000	3000	0.13	0.047	1	1
1500	4500	0.093	0.043	1	1
2000	6000	0.075	0.037	1	1
2500	7500	0.08	0.054	1	1
3000	9000	0.073	0.048	1	1

Figure 7.80: **Two Sample Sizes, K_{obs}**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	1	0.991	1	0.995
100	300	0.988	0.787	0.998	0.945
200	600	0.777	0.318	0.996	0.895
300	900	0.594	0.224	0.998	0.964
400	1200	0.465	0.133	1	0.995
500	1500	0.345	0.111	1	1
750	2250	0.245	0.099	1	1
1000	3000	0.176	0.089	1	1
1500	4500	0.128	0.062	1	1
2000	6000	0.104	0.047	1	1
2500	7500	0.099	0.06	1	1
3000	9000	0.086	0.053	1	1

Figure 7.81: **Two Sample Sizes, $\hat{K}_{Chao1a}$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.357	0.455	0.433	0.523
100	300	0.505	0.324	0.768	0.577
200	600	0.439	0.143	0.946	0.685
300	900	0.359	0.121	0.99	0.903
400	1200	0.272	0.089	0.999	0.973
500	1500	0.227	0.066	1	0.998
750	2250	0.168	0.067	1	1
1000	3000	0.146	0.067	1	1
1500	4500	0.095	0.054	1	1
2000	6000	0.083	0.04	1	1
2500	7500	0.086	0.051	1	1
3000	9000	0.076	0.05	1	1

Figure 7.82: **Two Sample Sizes, $\hat{K}_0$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.353	0.459	0.422	0.522
100	300	0.503	0.32	0.774	0.578
200	600	0.44	0.144	0.954	0.695
300	900	0.37	0.122	0.991	0.908
400	1200	0.282	0.089	0.999	0.976
500	1500	0.235	0.067	1	0.999
750	2250	0.168	0.067	1	1
1000	3000	0.148	0.066	1	1
1500	4500	0.097	0.054	1	1
2000	6000	0.083	0.042	1	1
2500	7500	0.086	0.05	1	1
3000	9000	0.076	0.051	1	1

Figure 7.83: **Two Sample Sizes, $\hat{\boldsymbol{K}}_1$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.232	0.342	0.2949	0.405
100	300	0.266	0.171	0.558	0.402
200	600	0.193	0.071	0.814	0.502
300	900	0.167	0.056	0.938	0.735
400	1200	0.105	0.026	0.979	0.905
500	1500	0.078	0.026	0.998	0.981
750	2250	0.071	0.027	0.996	0.994
1000	3000	0.065	0.03	1	1
1500	4500	0.047	0.027	1	1
2000	6000	0.037	0.021	1	1
2500	7500	0.043	0.03	1	1
3000	9000	0.055	0.04	1	1

Figure 7.84: **Two Sample Sizes, $\hat{\boldsymbol{K}}_2$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.316	0.422	0.387	0.488
100	300	0.401	0.252	0.687	0.495
200	600	0.217	0.086	0.828	0.522
300	900	0.168	0.057	0.938	0.735
400	1200	0.105	0.026	0.979	0.905
500	1500	0.078	0.026	0.998	0.981
750	2250	0.071	0.027	0.996	0.994
1000	3000	0.065	0.03	1	1
1500	4500	0.047	0.027	1	1
2000	6000	0.037	0.021	1	1
2500	7500	0.043	0.03	1	1
3000	9000	0.055	0.04	1	1

7.5 Power Decay Distribution: K=30

Next, suppose that $K = 30$ and we have two equal power decay distributions, $\mathbf{p} = \mathbf{q} = \{c_1/1^2, c_1/2^2, c_1/3^2, \ldots, c_1/30^2\}$, where c_1 is the adjusting constant to ensure the distribution sums to 1. Again we have the actual value of Jensen-Shannon Divergence at 0. The error tables are as follows, plug-in estimator in red and jackknife estimator in blue.

Figure 7.85: **One-Sample**

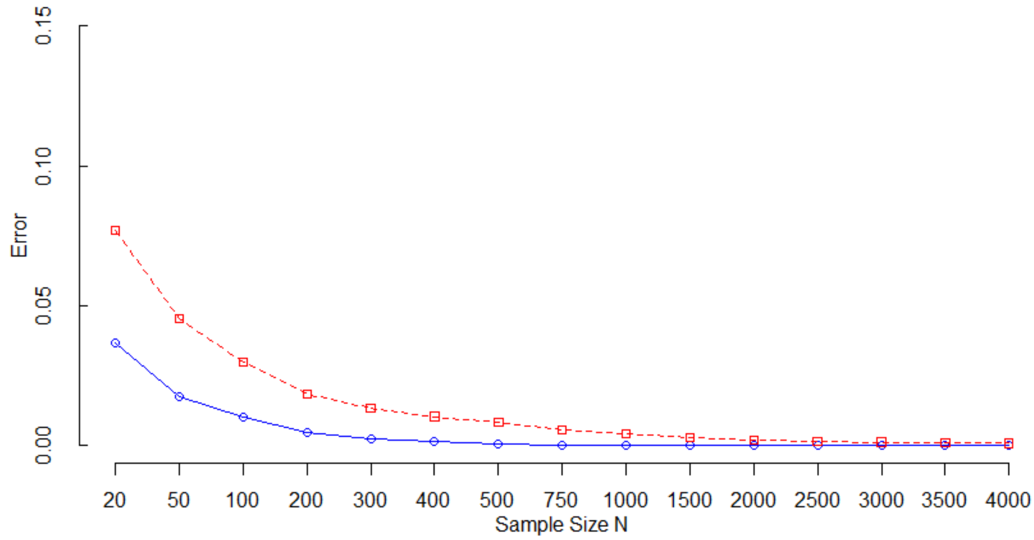
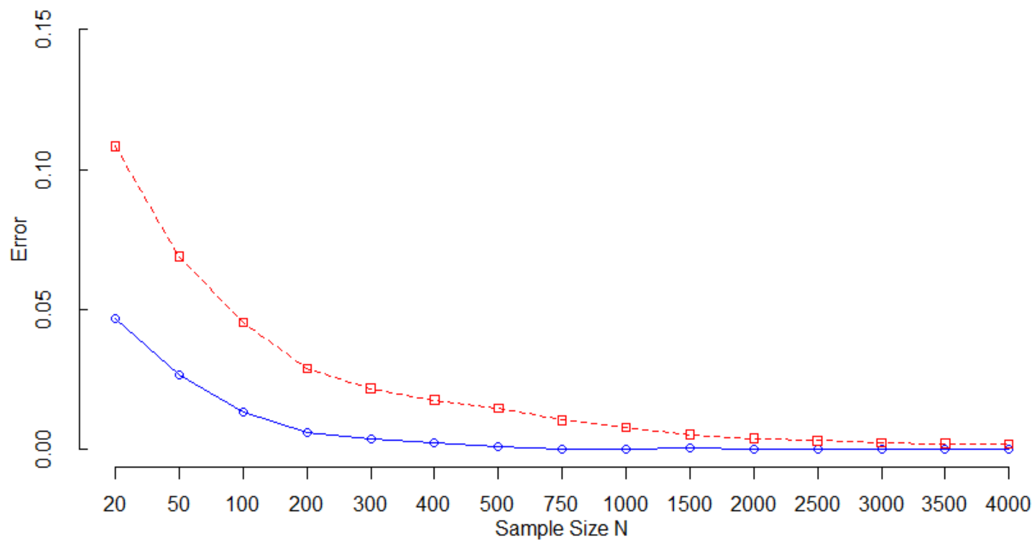


Figure 7.86: **Two-Sample**



Now suppose for **q**, that we adjust **p** to be $\{c_2/1^{2.2}, c_2/2^{2.2}, c_2/3^{2.2}, \ldots, c_2/30^{2.2}\}$, where c_2 is correspondingly adjusted to make the probabilities sum to 1. This adjusted **q** distribution juxtaposed on the original triangle **p** is demonstrated by the following:

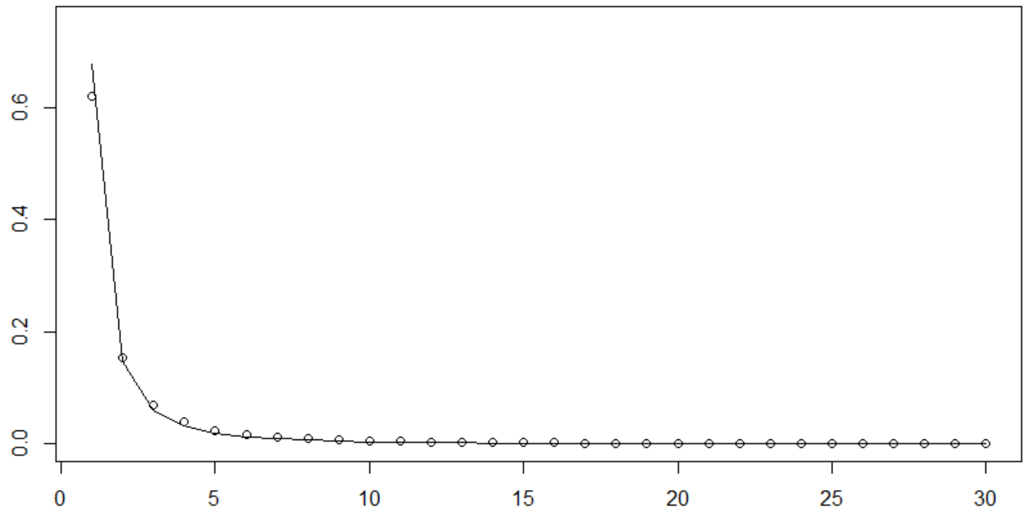


Figure 7.87

Here, the value of Jensen-Shannon divergence between these two distributions given in Figure 7.87 is 0.002538236. For the alternative hypothesis when H_0 is false, $\mathbf{q}$ is given by Figure 7.87.

Figure 7.88: **One-Sample, K known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0	0.147	0	0.122
100	0.005	0.22	0.002	0.227
200	0.029	0.212	0.022	0.295
300	0.063	0.217	0.098	0.355
400	0.1	0.172	0.175	0.388
500	0.117	0.157	0.29	0.444
750	0.178	0.128	0.545	0.544
1000	0.195	0.117	0.736	0.653
1500	0.182	0.091	0.927	0.863
2000	0.153	0.06	0.978	0.943

Figure 7.89: **One-Sample, K_{obs}**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.772	0.998	0.68	1
100	0.803	0.988	0.716	0.996
200	0.76	0.877	0.758	0.952
300	0.712	0.743	0.769	0.902
400	0.658	0.625	0.772	0.828
500	0.598	0.496	0.802	0.809
750	0.503	0.333	0.818	0.759
1000	0.426	0.255	0.885	0.798
1500	0.302	0.16	0.958	0.907
2000	0.223	0.104	0.982	0.949

Figure 7.90: **One-Sample, $\hat{K}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.339	0.823	0.243	0.865
100	0.305	0.705	0.292	0.714
200	0.298	0.515	0.293	0.623
300	0.328	0.43	0.335	0.56
400	0.3	0.366	0.358	0.486
500	0.293	0.308	0.423	0.511
750	0.258	0.21	0.531	0.527
1000	0.234	0.171	0.669	0.606
1500	0.21	0.123	0.886	0.796
2000	0.167	0.083	0.949	0.905

Figure 7.91: **One-Sample, $\hat{\boldsymbol{K}}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.457	0.951	0.346	0.967
100	0.508	0.904	0.425	0.917
200	0.483	0.715	0.486	0.835
300	0.491	0.603	0.526	0.763
400	0.463	0.493	0.552	0.681
500	0.412	0.389	0.608	0.677
750	0.365	0.266	0.693	0.66
1000	0.33	0.213	0.803	0.707
1500	0.255	0.14	0.937	0.868
2000	0.197	0.093	0.975	0.937

Figure 7.92: **One-Sample, $\hat{\boldsymbol{K}}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.41	0.927	0.287	0.944
100	0.43	0.87	0.363	0.891
200	0.415	0.662	0.398	0.78
300	0.424	0.544	0.447	0.709
400	0.403	0.432	0.488	0.633
500	0.358	0.353	0.546	0.637
750	0.323	0.238	0.641	0.612
1000	0.291	0.193	0.77	0.675
1500	0.239	0.13	0.924	0.846
2000	0.19	0.088	0.973	0.929

Figure 7.93: **One-Sample, $\hat{K}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.41	0.927	0.287	0.944
100	0.43	0.87	0.363	0.891
200	0.415	0.662	0.398	0.78
300	0.424	0.544	0.447	0.709
400	0.403	0.432	0.488	0.633
500	0.358	0.353	0.546	0.637
750	0.323	0.238	0.641	0.612
1000	0.291	0.193	0.77	0.675
1500	0.239	0.13	0.924	0.846
2000	0.19	0.088	0.973	0.929

Figure 7.94: **Two-Sample, K known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0	0	0	0
100	0.001	0.004	0	0.005
200	0.001	0.032	0.003	0.029
300	0.014	0.054	0.02	0.068
400	0.034	0.068	0.055	0.118
500	0.038	0.075	0.119	0.169
750	0.081	0.07	0.256	0.26
1000	0.111	0.072	0.404	0.338
1500	0.114	0.053	0.653	0.475
2000	0.091	0.045	0.787	0.601
2500	0.102	0.046	0.881	0.735
3000	0.081	0.034	0.937	0.837
3500	0.08	0.041	0.974	0.899
4000	0.083	0.035	0.978	0.947

Figure 7.95: **Two-Sample, K_{obs}**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.373	0.356	0.313	0.313
100	0.454	0.495	0.382	0.473
200	0.458	0.502	0.442	0.515
300	0.425	0.446	0.464	0.522
400	404	0.377	0.513	0.514
500	0.38	0.327	0.547	0.524
750	0.359	0.257	0.616	0.515
1000	0.293	0.198	0.641	0.521
1500	0.215	0.115	0.776	0.574
2000	0.157	0.073	0.83	0.652
2500	0.134	0.063	0.901	0.754
3000	0.108	0.051	0.942	0.849
3500	0.096	0.047	0.978	0.903
4000	0.092	0.046	0.978	0.947

Figure 7.96: **Two-Sample, $\hat{K}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.157	0.173	0.105	0.117
100	0.132	0.197	0.118	0.197
200	0.119	0.202	0.132	0.22
300	0.14	0.2	0.157	0.214
400	0.129	0.181	0.163	0.224
500	0.133	0.157	0.236	0.264
750	0.154	0.128	0.341	0.317
1000	0.155	0.119	0.43	0.351
1500	0.118	0.077	0.62	0.466
2000	0.111	0.055	0.756	0.568
2500	0.101	0.056	0.859	0.704
3000	0.086	0.047	0.927	0.817
3500	0.085	0.041	0.968	0.892
4000	0.086	0.041	0.977	0.947

Figure 7.97: **Two-Sample, $\hat{\boldsymbol{K}}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.198	0.226	0.142	0.168
100	0.208	0.333	0.194	0.278
200	0.217	0.335	0.231	0.339
300	0.231	0.322	0.256	0.367
400	0.213	0.26	0.296	0.35
500	0.217	0.213	0.36	0.394
750	0.233	0.18	0.453	0.406
1000	0.222	0.153	0.543	0.436
1500	0.162	0.089	0.704	0.516
2000	0.12	0.062	0.807	0.623
2500	0.118	0.057	0.886	0.74
3000	0.097	0.049	0.94	0.84
3500	0.091	0.044	0.976	0.9
4000	0.09	0.046	0.978	0.946

Figure 7.98: **Two-Sample, $\hat{\boldsymbol{K}}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.169	0.196	0.122	0.143
100	0.179	0.29	0.155	0.253
200	0.186	0.269	0.182	0.293
300	0.184	0.282	0.205	0.309
400	0.17	0.224	0.22	0.294
500	0.171	0.184	0.306	0.336
750	0.191	0.156	0.398	0.362
1000	0.191	0.143	0.505	0.403
1500	0.142	0.082	0.676	0.496
2000	0.117	0.059	0.786	0.599
2500	0.115	0.055	0.878	0.732
3000	0.09	0.049	0.938	0.833
3500	0.089	0.043	0.974	0.897
4000	0.089	0.044	0.978	0.943

Figure 7.99: **Two-Sample, $\hat{K}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.169	0.196	0.122	0.143
100	0.179	0.29	0.155	0.253
200	0.186	0.269	0.182	0.293
300	0.184	0.282	0.205	0.309
400	0.17	0.224	0.22	0.294
500	0.171	0.184	0.306	0.336
750	0.191	0.156	0.398	0.362
1000	0.191	0.143	0.505	0.403
1500	0.142	0.082	0.676	0.493
2000	0.117	0.059	0.786	0.599
2500	0.115	0.055	0.878	0.732
3000	0.09	0.049	0.938	0.833
3500	0.089	0.043	0.974	0.897
4000	0.089	0.044	0.978	0.943

Figure 7.100: **Two Sample Sizes, K known**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0	0.002	0	0
100	300	0.003	0.024	0	0.022
200	600	0.019	0.083	0.011	0.073
300	900	0.04	0.105	0.069	0.148
400	1200	0.072	0.118	0.12	0.184
500	1500	0.083	0.108	0.225	0.248
750	2250	0.101	0.09	0.414	0.363
300	900	0.1	0.055	0.575	0.47
1500	4500	0.117	0.068	0.808	0.668
2000	6000	0.116	0.066	0.918	0.825
2500	7500	0.098	0.049	0.973	0.941
3000	9000	0.09	0.042	0.989	0.973
3500	10500	0.083	0.036	0.997	0.997
4000	12000	0.08	0.043	0.998	0.997

Figure 7.101: **Two Sample Sizes, $\boldsymbol{K}_{obs}$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.307	0.72	0.545	0.646
100	300	0.317	0.707	0.589	0.703
200	600	0.62	0.678	0.633	0.688
300	900	0.587	0.568	0.665	0.667
400	1200	0.538	0.496	0.668	0.627
500	1500	0.494	0.416	0.686	0.627
750	2250	0.37	0.261	0.73	0.616
300	900	0.321	0.199	0.779	0.646
1500	4500	0.231	0.131	0.883	0.748
2000	6000	0.177	0.103	0.938	0.856
2500	7500	0.154	0.079	0.977	0.949
3000	9000	0.111	0.06	0.992	0.978
3500	10500	0.094	0.05	0.997	0.997
4000	12000	0.088	0.057	0.998	0.997

Figure 7.102: **Two Sample Sizes, $\hat{\boldsymbol{K}}_{Chao1a}$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.242	0.394	0.189	0.352
100	300	0.233	0.352	0.184	0.321
200	600	0.237	0.348	0.241	0.335
300	900	0.234	0.294	0.263	0.335
400	1200	0.219	0.281	0.278	0.338
500	1500	0.206	0.217	0.313	0.349
750	2250	0.173	0.165	0.436	0.398
300	900	0.167	0.113	0.558	0.467
1500	4500	0.139	0.095	0.741	0.632
2000	6000	0.125	0.078	0.879	0.789
2500	7500	0.121	0.063	0.954	0.92
3000	9000	0.089	0.05	0.985	0.97
3500	10500	0.083	0.046	0.997	0.994
4000	12000	0.078	0.053	0.997	0.996

Figure 7.103: **Two Sample Sizes, $\hat{K}_0$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.339	0.509	0.283	0.457
100	300	0.357	0.539	0.312	0.48
200	600	0.386	0.493	0.384	0.514
300	900	0.362	0.41	0.412	0.494
400	1200	0.338	0.369	0.447	0.48
500	1500	0.334	0.313	0.49	0.481
750	2250	0.257	0.204	0.579	0.513
300	900	0.237	0.145	0.684	0.568
1500	4500	0.182	0.107	0.827	0.694
2000	6000	0.142	0.091	0.913	0.83
2500	7500	0.139	0.075	0.972	0.932
3000	9000	0.104	0.057	0.989	0.977
3500	10500	0.09	0.048	0.997	0.996
4000	12000	0.086	0.057	0.998	0.997

Figure 7.104: **Two Sample Sizes, $\hat{K}_1$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.296	0.467	0.236	0.414
100	300	0.309	0.477	0.253	0.422
200	600	0.32	0.449	0.318	0.453
300	900	0.313	0.371	0.343	0.43
400	1200	0.283	0.327	0.372	0.412
500	1500	0.279	0.272	0.411	0.443
750	2250	0.214	0.182	0.527	0.468
300	900	0.199	0.133	0.637	0.53
1500	4500	0.17	0.1	0.804	0.67
2000	6000	0.138	0.089	0.907	0.82
2500	7500	0.127	0.068	0.971	0.931
3000	9000	0.1	0.053	0.988	0.973
3500	10500	0.087	0.047	0.997	0.996
4000	12000	0.084	0.055	0.998	0.996

Figure 7.105: **Two Sample Sizes, $\hat{\boldsymbol{K}}_2$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.296	0.467	0.236	0.414
100	300	0.309	0.477	0.253	0.422
200	600	0.32	0.449	0.318	0.453
300	900	0.313	0.371	0.343	0.43
400	1200	0.283	0.327	0.372	0.412
500	1500	0.279	0.272	0.411	0.443
750	2250	0.214	0.182	0.527	0.468
1000	3000	0.199	0.133	0.637	0.53
1500	4500	0.17	0.1	0.804	0.67
2000	6000	0.138	0.089	0.907	0.82
2500	7500	0.127	0.068	0.971	0.931
3000	9000	0.1	0.053	0.988	0.973
3500	10500	0.087	0.047	0.997	0.996
4000	12000	0.084	0.055	0.998	0.996

7.6 Power Decay Distribution: K=100

Next, suppose that $K = 100$ and we have two equal power decay distributions, $\mathbf{p} = \mathbf{q} = \{c_3/1^2, c_3/2^2, c_3/3^2, \ldots, c_3/100^2\}$, where c_3 is the adjusting constant to ensure the distribution sums to 1. Again we have the actual value of Jensen-Shannon Divergence at 0. The error tables are as follows, plug-in estimator in red and jackknife estimator in blue.

Figure 7.106: **One-Sample**

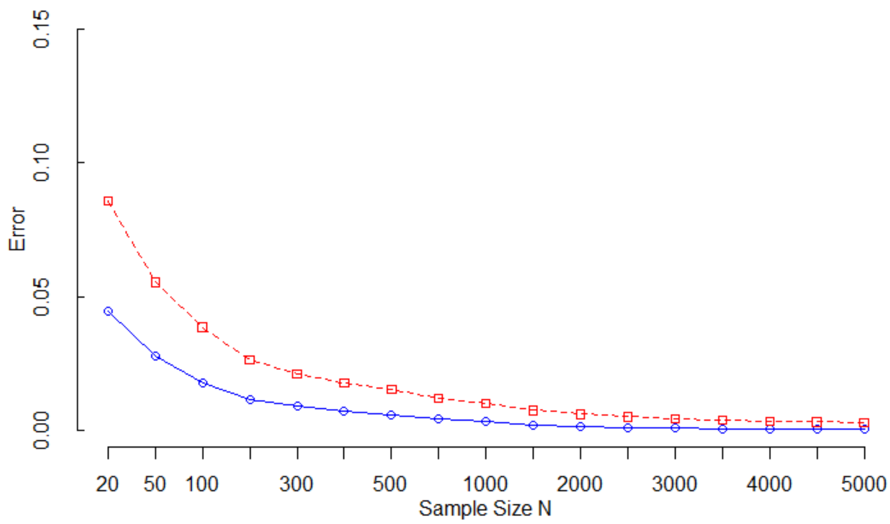
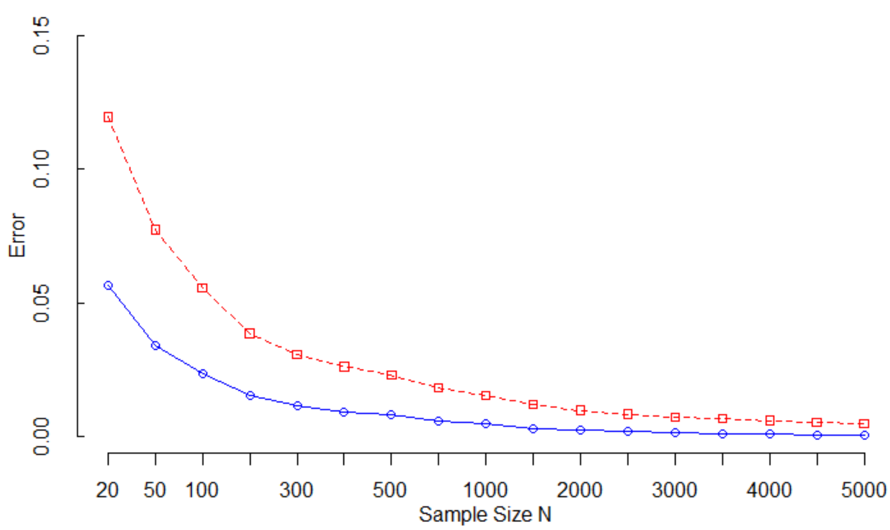


Figure 7.107: **Two-Sample**



Now suppose for $\mathbf{q}$, that we adjust $\mathbf{p}$ to be $\{c_4/1^{2.2}, c_4/2^{2.2}, c_4/3^{2.2}, \ldots, c_4/100^{2.2}\}$, where c_4 is correspondingly adjusted to make the probabilities sum to 1. This adjusted $\mathbf{q}$ distribution juxtaposed on the original triangle $\mathbf{p}$ is demonstrated by the following:

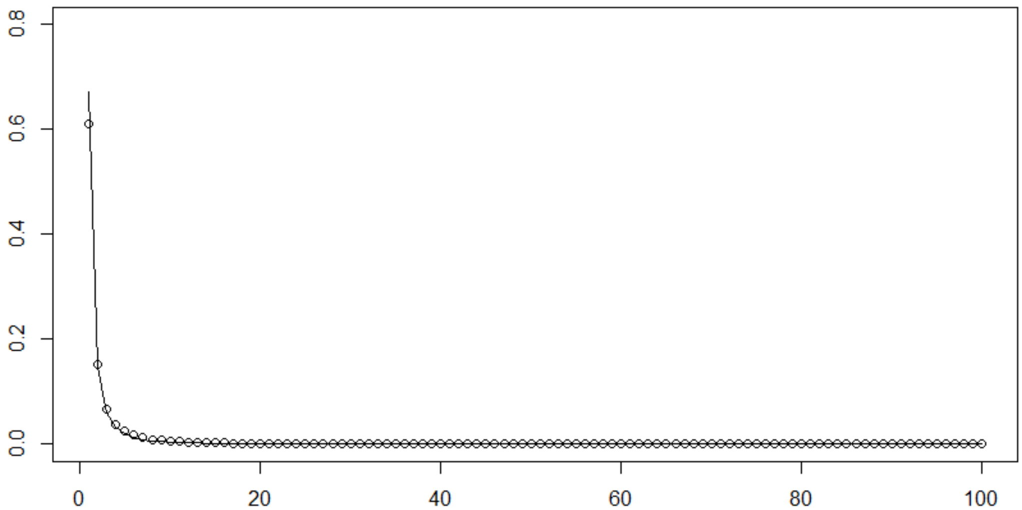


Figure 7.108

Here, between uniform $\mathbf{p}$ and this adjusted $\mathbf{q}$ given in Figure 7.108, the actual value of Jensen-Shannon Divergence is 0.00310155. For the alternative hypothesis when H_0 is false, $\mathbf{q}$ is given in Figure 7.108.

Figure 7.109: **One-Sample, K known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0	0.002	0	0.003
50	0	0.022	0	0.034
100	0	0.101	0	0.257
200	0	0.256	0	0.762
300	0	0.366	0	0.954
400	0	0.439	0.002	0.995
500	0	0.49	0.034	0.999
750	0	0.538	0.532	1
1000	0	0.554	0.955	1
1500	0.002	0.544	1	1
2000	0.007	0.503	1	1
2500	0.051	0.475	1	1
3000	0.089	0.417	1	1
3500	0.126	0.361	1	1
4000	0.15	0.306	1	1
4500	0.211	0.267	1	1
5000	0.259	0.273	1	1
5500	0.302	0.259	1	1
6000	0.285	0.194	1	1
6500	0.318	0.194	1	1
7000	0.321	0.139	1	1
7500	0.362	0.161	1	1
8000	0.381	0.161	1	1
8500	0.408	0.148	1	1
9000	0.384	0.132	1	1
9500	0.365	0.119	1	1
10000	0.37	0.1	1	1
10500	0.397	0.097	1	1
11000	0.382	0.104	1	1
11500	0.386	0.098	1	1
12000	0.37	0.073	1	1
12500	0.368	0.075	1	1
13000	0.329	0.085	1	1
13500	0.332	0.064	1	1
14000	0.364	0.093	1	1
14500	0.337	0.056	1	1
15000	0.34	0.066	1	1

Figure 7.110: **One-Sample, $\boldsymbol{K}_{obs}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.797	1	0.605	1
50	0.926	1	0.809	1
100	0.978	1	0.942	1
200	0.991	1	0.994	1
300	0.996	1	1	1
400	0.998	1	1	1
500	0.997	1	1	1
750	0.998	1	1	1
1000	0.998	1	1	1
1500	0.997	1	1	1
2000	0.999	1	1	1
2500	0.998	0.998	1	1
3000	0.991	0.987	1	1
3500	0.985	0.97	1	1
4000	0.985	0.934	1	1
4500	0.971	0.901	1	1
5000	0.971	0.873	1	1
5500	0.946	0.815	1	1
6000	0.94	0.725	1	1
6500	0.922	0.7	1	1
7000	0.913	0.636	1	1
7500	0.894	0.592	1	1
8000	0.888	0.558	1	1
8500	0.859	0.518	1	1
9000	0.858	0.431	1	1
9500	0.801	0.42	1	1
10000	0.796	0.375	1	1
10500	0.802	0.37	1	1
11000	0.778	0.343	1	1
11500	0.738	0.307	1	1
12000	0.709	0.296	1	1
12500	0.713	0.264	1	1
13000	0.659	0.221	1	1
13500	0.649	0.214	1	1
14000	0.652	0.23	1	1
14500	0.626	0.193	1	1
15000	0.608	0.185	1	1

Figure 7.111: **One-Sample, $\hat{K}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.388	1	0.236	1
50	0.368	1	0.242	1
100	0.341	0.994	0.329	0.997
200	0.4	0.962	0.484	0.989
300	0.396	0.942	0.616	0.977
400	0.409	0.924	0.71	0.969
500	0.398	0.905	0.808	0.978
750	0.449	0.912	0.924	0.985
1000	0.443	0.865	0.949	0.988
1500	0.491	0.825	0.99	0.995
2000	0.528	0.791	0.996	1
2500	0.52	0.732	0.998	0.999
3000	0.527	0.684	1	1
3500	0.506	0.609	1	1
4000	0.529	0.572	1	1
4500	0.51	0.497	1	1
5000	0.528	0.49	1	1
5500	0.526	0.454	1	1
6000	0.506	0.38	1	1
6500	0.515	0.355	1	1
7000	0.492	0.307	1	1
7500	0.494	0.299	1	1
8000	0.518	0.301	1	1
8500	0.509	0.279	1	1
9000	0.471	0.221	1	1
9500	0.461	0.219	1	1
10000	0.452	0.185	1	1
10500	0.462	0.189	1	1
11000	0.449	0.186	1	1
11500	0.446	0.173	1	1
12000	0.415	0.145	1	1
12500	0.423	0.141	1	1
13000	0.373	0.131	1	1
13500	0.378	0.109	1	1
14000	0.411	0.147	1	1
14500	0.378	0.105	1	1
15000	0.386	0.112	1	1

Figure 7.112: **One-Sample, $\hat{\boldsymbol{K}}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.41	1	0.268	1
50	0.555	1	0.408	1
100	0.626	1	0.61	1
200	0.77	1	0.882	1
300	0.8	1	0.966	1
400	0.813	1	0.994	1
500	0.849	1	0.999	1
750	0.881	1	1	1
1000	0.88	1	1	1
1500	0.886	0.99	1	1
2000	0.89	0.979	1	1
2500	0.887	0.96	1	1
3000	0.861	0.897	1	1
3500	0.842	0.847	1	1
4000	0.847	0.811	1	1
4500	0.821	0.746	1	1
5000	0.816	0.685	1	1
5500	0.792	0.623	1	1
6000	0.743	0.544	1	1
6500	0.737	0.513	1	1
7000	0.71	0.452	1	1
7500	0.716	0.419	1	1
8000	0.724	0.394	1	1
8500	0.692	0.369	1	1
9000	0.652	0.312	1	1
9500	0.61	0.287	1	1
10000	0.632	0.265	1	1
10500	0.611	0.247	1	1
11000	0.593	0.238	1	1
11500	0.584	0.218	1	1
12000	0.54	0.188	1	1
12500	0.554	0.186	1	1
13000	0.51	0.171	1	1
13500	0.492	0.141	1	1
14000	0.522	0.176	1	1
14500	0.49	0.138	1	1
15000	0.48	0.132	1	1

Figure 7.113: **One-Sample, $\hat{K}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.346	1	0.214	1
50	0.485	1	0.335	1
100	0.536	1	0.524	1
200	0.67	1	0.798	1
300	0.688	1	0.935	1
400	0.699	1	0.982	1
500	0.719	1	0.996	1
750	0.759	0.998	1	1
1000	0.76	0.997	1	1
1500	0.771	0.971	1	1
2000	0.783	0.948	1	1
2500	0.757	0.905	1	1
3000	0.746	0.838	1	1
3500	0.716	0.767	1	1
4000	0.721	0.714	1	1
4500	0.706	0.643	1	1
5000	0.686	0.571	1	1
5500	0.675	0.541	1	1
6000	0.622	0.456	1	1
6500	0.649	0.419	1	1
7000	0.615	0.355	1	1
7500	0.6	0.356	1	1
8000	0.605	0.331	1	1
8500	0.589	0.304	1	1
9000	0.537	0.258	1	1
9500	0.539	0.229	1	1
10000	0.52	0.213	1	1
10500	0.53	0.202	1	1
11000	0.498	0.204	1	1
11500	0.494	0.188	1	1
12000	0.475	0.146	1	1
12500	0.471	0.15	1	1
13000	0.418	0.137	1	1
13500	0.415	0.114	1	1
14000	0.454	0.151	1	1
14500	0.412	0.1	1	1
15000	0.416	0.113	1	1

Figure 7.114: **One-Sample, $\hat{\boldsymbol{K}}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
20	0.346	1	0.214	1
50	0.485	1	0.335	1
100	0.536	1	0.524	1
200	0.67	1	0.798	1
300	0.688	1	0.935	1
400	0.699	1	0.982	1
500	0.719	1	0.996	1
750	0.759	0.998	1	1
1000	0.76	0.997	1	1
1500	0.771	0.971	1	1
2000	0.783	0.948	1	1
2500	0.757	0.905	1	1
3000	0.746	0.838	1	1
3500	0.716	0.767	1	1
4000	0.721	0.714	1	1
4500	0.706	0.643	1	1
5000	0.686	0.571	1	1
5500	0.675	0.541	1	1
6000	0.622	0.456	1	1
6500	0.649	0.419	1	1
7000	0.615	0.355	1	1
7500	0.6	0.356	1	1
8000	0.605	0.331	1	1
8500	0.589	0.304	1	1
9000	0.537	0.258	1	1
9500	0.539	0.229	1	1
10000	0.52	0.213	1	1
10500	0.53	0.202	1	1
11000	0.498	0.204	1	1
11500	0.494	0.188	1	1
12000	0.475	0.146	1	1
12500	0.471	0.15	1	1
13000	0.418	0.137	1	1
13500	0.415	0.114	1	1
14000	0.454	0.151	1	1
14500	0.412	0.1	1	1
15000	0.416	0.113	1	1

Figure 7.115: **Two-Sample, K known**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0	0	0	0
100	0	0	0	0
200	0	0	0	0
300	0	0	0	0
400	0	0	0	0
500	0	0	0	0
750	0	0	0.002	0.038
1000	0	0	0.074	0.258
1500	0	0.003	0.774	0.889
2000	0	0.009	0.991	0.995
2500	0.002	0.019	1	1
3000	0.001	0.028	1	1
3500	0.006	0.032	1	1
4000	0.015	0.039	1	1
4500	0.021	0.046	1	1
5000	0.027	0.05	1	1
5500	0.044	0.058	1	1
6000	0.051	0.052	1	1
6500	0.068	0.061	1	1
7000	0.093	0.048	1	1
7500	0.101	0.069	1	1
8000	0.131	0.059	1	1
8500	0.125	0.063	1	1
9000	0.124	0.069	1	1
9500	0.152	0.055	1	1
10000	0.151	0.054	1	1
10500	0.168	0.056	1	1
11000	0.173	0.055	1	1
11500	0.182	0.045	1	1
12000	0.188	0.046	1	1
12500	0.187	0.051	1	1
13000	0.173	0.048	1	1
13500	0.19	0.046	1	1
14000	0.174	0.034	1	1
14500	0.177	0.032	1	1
15000	0.187	0.038	1	1

Figure 7.116: **Two-Sample, K_{obs}**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.456	0	0.319	0
100	0.567	0.006	0.497	0.002
200	0.686	0.325	0.761	0.333
300	0.728	0.591	0.913	0.764
400	0.76	0.683	0.952	0.909
500	0.772	0.729	0.979	0.966
750	0.831	0.818	1	0.997
1000	0.846	0.83	1	1
1500	0.861	0.821	1	1
2000	0.885	0.81	1	1
2500	0.879	0.793	1	1
3000	0.846	0.743	1	1
3500	0.849	0.699	1	1
4000	0.811	0.629	1	1
4500	0.818	0.626	1	1
5000	0.811	0.593	1	1
5500	0.808	0.562	1	1
6000	0.766	0.51	1	1
6500	0.754	0.496	1	1
7000	0.723	0.409	1	1
7500	0.711	0.415	1	1
8000	0.706	0.388	1	1
8500	0.686	0.358	1	1
9000	0.655	0.331	1	1
9500	0.638	0.294	1	1
10000	0.639	0.279	1	1
10500	0.602	0.248	1	1
11000	0.585	0.244	1	1
11500	0.553	0.222	1	1
12000	0.547	0.22	1	1
12500	0.533	0.193	1	1
13000	0.483	0.168	1	1
13500	0.484	0.164	1	1
14000	0.486	0.153	1	1
14500	0.445	0.137	1	1
15000	0.432	0.149	1	1

Figure 7.117: **Two-Sample, $\hat{K}_{Chao1a}$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.131	0	0.081	0
100	0.101	0.002	0.09	0
200	0.118	0.051	0.155	0.043
300	0.115	0.133	0.237	0.207
400	0.121	0.163	0.309	0.351
500	0.115	0.197	0.388	0.458
750	0.146	0.259	0.605	0.697
1000	0.139	0.266	0.725	0.814
1500	0.15	0.263	0.913	0.938
2000	0.182	0.277	0.969	0.978
2500	0.193	0.268	0.996	0.997
3000	0.194	0.247	0.997	0.998
3500	0.19	0.239	0.999	0.999
4000	0.21	0.213	0.999	1
4500	0.204	0.205	1	1
5000	0.234	0.211	1	1
5500	0.237	0.21	1	1
6000	0.238	0.168	1	1
6500	0.261	0.177	1	1
7000	0.217	0.142	1	1
7500	0.238	0.156	1	1
8000	0.272	0.163	1	1
8500	0.255	0.15	1	1
9000	0.234	0.138	1	1
9500	0.252	0.117	1	1
10000	0.26	0.116	1	1
10500	0.265	0.111	1	1
11000	0.274	0.102	1	1
11500	0.255	0.097	1	1
12000	0.259	0.084	1	1
12500	0.234	0.088	1	1
13000	0.217	0.083	1	1
13500	0.247	0.085	1	1
14000	0.231	0.07	1	1
14500	0.225	0.063	1	1
15000	0.222	0.077	1	1

Figure 7.118: **Two-Sample, $\hat{\boldsymbol{K}}_0$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.198	0	0.119	0
100	0.212	0.002	0.2	0
200	0.289	0.134	0.35	0.121
300	0.302	0.301	0.525	0.442
400	0.316	0.393	0.665	0.699
500	0.346	0.465	0.785	0.856
750	0.389	0.548	0.948	0.968
1000	0.39	0.553	0.994	0.995
1500	0.422	0.559	0.999	1
2000	0.476	0.546	1	1
2500	0.487	0.512	1	1
3000	0.465	0.465	1	1
3500	0.46	0.442	1	1
4000	0.45	0.389	1	1
4500	0.452	0.379	1	1
5000	0.465	0.363	1	1
5500	0.464	0.342	1	1
6000	0.446	0.303	1	1
6500	0.443	0.293	1	1
7000	0.414	0.243	1	1
7500	0.426	0.25	1	1
8000	0.434	0.255	1	1
8500	0.423	0.228	1	1
9000	0.392	0.196	1	1
9500	0.387	0.17	1	1
10000	0.408	0.161	1	1
10500	0.395	0.155	1	1
11000	0.387	0.142	1	1
11500	0.358	0.134	1	1
12000	0.372	0.128	1	1
12500	0.356	0.124	1	1
13000	0.315	0.112	1	1
13500	0.326	0.107	1	1
14000	0.327	0.091	1	1
14500	0.306	0.091	1	1
15000	0.297	0.099	1	1

Figure 7.119: **Two-Sample, $\hat{\boldsymbol{K}}_1$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.161	0	0.082	0
100	0.161	0.002	0.142	0
200	0.214	0.095	0.262	0.08
300	0.217	0.232	0.408	0.361
400	0.221	0.296	0.535	0.597
500	0.231	0.375	0.662	0.773
750	0.282	0.429	0.899	0.935
1000	0.272	0.438	0.967	0.989
1500	0.27	0.418	0.998	0.998
2000	0.305	0.42	1	1
2500	0.326	0.41	1	1
3000	0.316	0.356	1	1
3500	0.305	0.333	1	1
4000	0.297	0.29	1	1
4500	0.324	0.284	1	1
5000	0.324	0.277	1	1
5500	0.336	0.267	1	1
6000	0.314	0.216	1	1
6500	0.314	0.207	1	1
7000	0.283	0.181	1	1
7500	0.301	0.189	1	1
8000	0.337	0.194	1	1
8500	0.311	0.171	1	1
9000	0.302	0.147	1	1
9500	0.302	0.133	1	1
10000	0.308	0.134	1	1
10500	0.294	0.118	1	1
11000	0.297	0.112	1	1
11500	0.285	0.106	1	1
12000	0.296	0.097	1	1
12500	0.267	0.095	1	1
13000	0.241	0.092	1	1
13500	0.263	0.083	1	1
14000	0.256	0.069	1	1
14500	0.235	0.066	1	1
15000	0.238	0.078	1	1

Figure 7.120: **Two-Sample, $\hat{\boldsymbol{K}}_2$**

	H0 True		H0 False	
N	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	0.161	0	0.082	0
100	0.161	0.002	0.142	0
200	0.214	0.095	0.262	0.08
300	0.217	0.232	0.408	0.361
400	0.221	0.296	0.535	0.597
500	0.231	0.375	0.662	0.773
750	0.282	0.429	0.899	0.935
1000	0.272	0.438	0.967	0.989
1500	0.27	0.418	0.998	0.998
2000	0.305	0.42	1	1
2500	0.326	0.41	1	1
3000	0.316	0.356	1	1
3500	0.305	0.333	1	1
4000	0.297	0.29	1	1
4500	0.324	0.284	1	1
5000	0.324	0.277	1	1
5500	0.336	0.267	1	1
6000	0.314	0.216	1	1
6500	0.314	0.207	1	1
7000	0.283	0.181	1	1
7500	0.301	0.189	1	1
8000	0.337	0.194	1	1
8500	0.311	0.171	1	1
9000	0.302	0.147	1	1
9500	0.302	0.133	1	1
10000	0.308	0.134	1	1
10500	0.294	0.118	1	1
11000	0.297	0.112	1	1
11500	0.285	0.106	1	1
12000	0.296	0.097	1	1
12500	0.267	0.095	1	1
13000	0.241	0.092	1	1
13500	0.263	0.083	1	1
14000	0.256	0.069	1	1
14500	0.235	0.066	1	1
15000	0.238	0.078	1	1

Figure 7.121: **Two Sample Sizes, K known**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0	0	0	0
100	300	0	0	0	0
200	600	0	0	0	0
300	900	0	0	0	0
400	1200	0	0	0	0
500	1500	0	0	0	0
750	2250	0	0.001	0	0.004
1000	3000	0	0.004	0.001	0.026
1500	4500	0	0.032	0.025	0.165
2000	6000	0.002	0.068	0.174	0.412
2500	7500	0.01	0.101	0.511	0.676
3000	9000	0.025	0.113	0.789	0.849
3500	10500	0.058	0.122	0.91	0.923
4000	12000	0.065	0.128	0.982	0.962
4500	13500	0.079	0.113	0.992	0.982
5000	15000	0.099	0.112	0.999	0.997
5500	16500	0.137	0.115	1	0.999
6000	18000	0.155	0.123	1	1
6500	19500	0.15	0.095	1	1
7000	21000	0.193	0.097	1	1
7500	22500	0.199	0.097	1	1
8000	24000	0.187	0.083	1	1
8500	25500	0.19	0.071	1	1
9000	27000	0.19	0.061	1	1
9500	28500	0.214	0.081	1	1
10000	30000	0.211	0.063	1	1
10500	31500	0.228	0.07	1	1
11000	33000	0.213	0.074	1	1
11500	34500	0.206	0.059	1	1
12000	36000	0.223	0.068	1	1
12500	37500	0.213	0.057	1	1
13000	39000	0.217	0.061	1	1
13500	40500	0.21	0.048	1	1
14000	42000	0.208	0.05	1	1
14500	43500	0.192	0.048	1	1
15000	45000	0.206	0.042	1	1

Figure 7.122: **Two Sample Sizes, K_{obs}**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0	0	0.647	0.003
100	300	0	0	0.755	0.283
200	600	0	0	0.861	0.793
300	900	0	0	0.911	0.909
400	1200	0	0	0.95	0.942
500	1500	0	0	0.958	0.953
750	2250	0	0.001	0.981	0.979
1000	3000	0	0.004	0.996	0.992
1500	4500	0	0.032	0.999	0.997
2000	6000	0.002	0.068	1	0.999
2500	7500	0.01	0.101	1	0.999
3000	9000	0.025	0.113	1	0.999
3500	10500	0.962	0.89	1	1
4000	12000	0.937	0.818	1	1
4500	13500	0.93	0.78	1	1
5000	15000	0.884	0.71	1	1
5500	16500	0.892	0.713	1	1
6000	18000	0.866	0.626	1	1
6500	19500	0.852	0.591	1	1
7000	21000	0.836	0.543	1	1
7500	22500	0.79	0.494	1	1
8000	24000	0.766	0.465	1	1
8500	25500	0.747	0.405	1	1
9000	27000	0.686	0.352	1	1
9500	28500	0.703	0.342	1	1
10000	30000	0.67	0.319	1	1
10500	31500	0.655	0.3	1	1
11000	33000	0.642	0.271	1	1
11500	34500	0.624	0.264	1	1
12000	36000	0.602	0.246	1	1
12500	37500	0.553	0.213	1	1
13000	39000	0.512	0.209	1	1
13500	40500	0.524	0.183	1	1
14000	42000	0.524	0.196	1	1
14500	43500	0.491	0.152	1	1
15000	45000	0.459	0.165	1	1

Figure 7.123: **Two Sample Sizes, $\hat{K}_{Chao1a}$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.247	0.001	0.174	0
100	300	0.224	0.163	0.15	0.059
200	600	0.235	0.35	0.195	0.279
300	900	0.275	0.46	0.218	0.357
400	1200	0.261	0.479	0.236	0.415
500	1500	0.275	0.511	0.27	0.44
750	2250	0.318	0.527	0.374	0.531
1000	3000	0.306	0.524	0.387	0.544
1500	4500	0.353	0.537	0.572	0.685
2000	6000	0.37	0.511	0.704	0.767
2500	7500	0.386	0.482	0.795	0.841
3000	9000	0.387	0.463	0.886	0.896
3500	10500	0.369	0.397	0.926	0.921
4000	12000	0.337	0.364	0.951	0.943
4500	13500	0.374	0.35	0.969	0.955
5000	15000	0.358	0.314	0.988	0.981
5500	16500	0.391	0.314	0.994	0.988
6000	18000	0.361	0.268	1	0.999
6500	19500	0.364	0.258	0.998	0.998
7000	21000	0.377	0.247	0.999	0.998
7500	22500	0.35	0.217	1	0.999
8000	24000	0.36	0.196	0.999	0.997
8500	25500	0.315	0.168	1	1
9000	27000	0.297	0.157	1	1
9500	28500	0.319	0.154	1	1
10000	30000	0.296	0.146	1	1
10500	31500	0.317	0.148	1	1
11000	33000	0.311	0.121	1	1
11500	34500	0.306	0.124	1	1
12000	36000	0.306	0.125	1	1
12500	37500	0.278	0.111	1	1
13000	39000	0.275	0.096	1	1
13500	40500	0.27	0.091	1	1
14000	42000	0.27	0.104	1	1
14500	43500	0.249	0.085	1	1
15000	45000	0.254	0.089	1	1

Figure 7.124: **Two Sample Sizes, $\hat{K}_0$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.389	0.002	0.285	0
100	300	0.439	0.296	0.346	0.127
200	600	0.548	0.664	0.429	0.516
300	900	0.606	0.778	0.508	0.658
400	1200	0.64	0.816	0.616	0.785
500	1500	0.66	0.841	0.641	0.777
750	2250	0.702	0.852	0.798	0.88
1000	3000	0.724	0.847	0.839	0.896
1500	4500	0.725	0.812	0.936	0.947
2000	6000	0.732	0.797	0.978	0.974
2500	7500	0.738	0.762	0.996	0.99
3000	9000	0.724	0.719	0.996	0.989
3500	10500	0.71	0.656	0.999	0.994
4000	12000	0.689	0.606	1	0.997
4500	13500	0.674	0.558	1	0.998
5000	15000	0.623	0.491	1	1
5500	16500	0.657	0.512	1	0.999
6000	18000	0.612	0.434	1	1
6500	19500	0.607	0.385	1	1
7000	21000	0.591	0.366	1	1
7500	22500	0.562	0.32	1	1
8000	24000	0.538	0.314	1	1
8500	25500	0.508	0.242	1	1
9000	27000	0.439	0.222	1	1
9500	28500	0.483	0.225	1	1
10000	30000	0.463	0.204	1	1
10500	31500	0.447	0.193	1	1
11000	33000	0.444	0.171	1	1
11500	34500	0.434	0.162	1	1
12000	36000	0.418	0.16	1	1
12500	37500	0.382	0.145	1	1
13000	39000	0.359	0.129	1	1
13500	40500	0.371	0.123	1	1
14000	42000	0.346	0.134	1	1
14500	43500	0.334	0.106	1	1
15000	45000	0.321	0.114	1	1

Figure 7.125: **Two Sample Sizes, $\hat{K}_1$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.331	0.002	0.235	0
100	300	0.371	0.26	0.262	0.105
200	600	0.431	0.588	0.33	0.43
300	900	0.498	0.708	0.402	0.559
400	1200	0.515	0.738	0.463	0.684
500	1500	0.517	0.746	0.491	0.686
750	2250	0.559	0.771	0.641	0.813
1000	3000	0.576	0.771	0.683	0.828
1500	4500	0.572	0.728	0.848	0.904
2000	6000	0.584	0.706	0.934	0.945
2500	7500	0.596	0.675	0.981	0.974
3000	9000	0.572	0.614	0.991	0.983
3500	10500	0.55	0.556	0.995	0.987
4000	12000	0.529	0.509	0.999	0.995
4500	13500	0.526	0.465	0.999	0.995
5000	15000	0.474	0.394	1	0.999
5500	16500	0.533	0.4	1	0.998
6000	18000	0.482	0.341	1	0.999
6500	19500	0.47	0.311	1	1
7000	21000	0.465	0.308	1	1
7500	22500	0.44	0.259	1	1
8000	24000	0.428	0.237	1	1
8500	25500	0.385	0.197	1	1
9000	27000	0.358	0.165	1	1
9500	28500	0.368	0.172	1	1
10000	30000	0.365	0.166	1	1
10500	31500	0.651	0.162	1	1
11000	33000	0.346	0.134	1	1
11500	34500	0.353	0.134	1	1
12000	36000	0.344	0.127	1	1
12500	37500	0.31	0.113	1	1
13000	39000	0.301	0.102	1	1
13500	40500	0.304	0.095	1	1
14000	42000	0.285	0.106	1	1
14500	43500	0.266	0.087	1	1
15000	45000	0.269	0.09	1	1

Figure 7.126: **Two Sample Sizes, $\hat{K}_2$**

		H0 True		H0 False	
N1	N2	Plug-in rejection rate	Jackknife rejection rate	Plug-in rejection rate	Jackknife rejection rate
50	150	0.331	0.002	0.235	0
100	300	0.371	0.26	0.262	0.105
200	600	0.431	0.588	0.33	0.43
300	900	0.498	0.708	0.402	0.559
400	1200	0.515	0.738	0.463	0.684
500	1500	0.517	0.746	0.491	0.686
750	2250	0.559	0.771	0.641	0.813
1000	3000	0.576	0.771	0.683	0.828
1500	4500	0.572	0.728	0.848	0.904
2000	6000	0.584	0.706	0.934	0.945
2500	7500	0.596	0.675	0.981	0.974
3000	9000	0.572	0.614	0.991	0.983
3500	10500	0.55	0.556	0.995	0.987
4000	12000	0.529	0.509	0.999	0.995
4500	13500	0.526	0.465	0.999	0.995
5000	15000	0.474	0.394	1	0.999
5500	16500	0.533	0.4	1	0.998
6000	18000	0.482	0.341	1	0.999
6500	19500	0.47	0.311	1	1
7000	21000	0.465	0.308	1	1
7500	22500	0.44	0.259	1	1
8000	24000	0.428	0.237	1	1
8500	25500	0.385	0.197	1	1
9000	27000	0.358	0.165	1	1
9500	28500	0.368	0.172	1	1
10000	30000	0.365	0.166	1	1
10500	31500	0.351	0.162	1	1
11000	33000	0.346	0.134	1	1
11500	34500	0.353	0.134	1	1
12000	36000	0.344	0.127	1	1
12500	37500	0.31	0.113	1	1
13000	39000	0.301	0.102	1	1
13500	40500	0.304	0.095	1	1
14000	42000	0.285	0.106	1	1
14500	43500	0.266	0.087	1	1
15000	45000	0.269	0.09	1	1

CHAPTER 8: EXAMPLES WITH REAL DATA

8.1 ONE-SAMPLE

The demographics of the immigrants to the U.S. are dynamic, changing from year to year. A goodness of fit test of one time frame against an earlier time frame can be used to test whether or not the changes over time are statistically significant. Here, suppose we have U.S. immigration population data by race from the year 2011, and can obtain a sample from the year 2016 of size $N = 1000$. The population data from 2011 is as follows:

	2011	
Region	**Number**	**Percentage**
Americas	419,996	39.60%
East and Southeast Asia	252,594	23.82%
South Asia	123,625	11.66%
North Africa and West/Central Asia	92,239	8.70%
Europe	83,736	7.90%
Sub-Saharan Africa	83,400	7.86%
Australia and Oceania	4,962	0.47%
Total	1,060,552	100.00%

Figure 8.1

To conduct the hypothesis test, we assume that the year 2011 distribution proportions are the "known" distribution. Using this and the sample from 2016, we obtain $\hat{A}_{JK_1} + \hat{B}_{JK_1} = 0.003043507$. Since this is a one-sample situation, we use T_1 from (5.1), which yields

$$T_1 = 8N(\hat{A}_{JK_1} + \hat{B}_{JK_1}) + \left(\sum_{k=1}^{K-1} p_k(1-p_k) \left(\frac{1}{p_K} + \frac{1}{p_k} \right) - \sum_{m \neq n} \frac{p_n p_m}{p_K} \right)$$

$$= 30.34806$$

This is clearly greater than the critical value $\chi^2_{K-1,0.01} = 16.81189383$, with $K = 7$, and the p-value is 0.0000337494. Therefore we can say with 99% confidence that there is a statistically significant change in race demographics in the U.S. from the year 2011 to 2016.

The 2016 population data is eventually obtained, and is given in the following table:

	2011		2016	
Region	**Number**	**Percentage**	**Number**	**Percentage**
Americas	419,996	39.60%	506,852	42.90%
East and Southeast Asia	252,594	23.82%	242,541	20.53%
South Asia	123,625	11.66%	121,715	10.30%
North Africa and West/Central Asia	92,239	8.70%	121,041	10.25%
Europe	83,736	7.90%	93,556	7.92%
Sub-Saharan Africa	83,400	7.86%	90,167	7.63%
Australia and Oceania	4,962	0.47%	5,577	0.47%
Total	1,060,552	100.00%	1,181,449	100%

Figure 8.2

The true Jensen-Shannon Divergence between the two populations is 0.0014745343, and so clearly the test correctly rejected the null hypothesis.

8.2 TWO-SAMPLE

Every country in the world has its own unique partition of individuals which subscribe to particular religions (or lack thereof), which can be conceived of as a multinomial distribution. Estimating Jensen-Shannon Divergence could be applicable in this context, measuring the “difference” or “distance” between two of these distributions for two different countries. With this in mind, two samples of size $N_{\mathbf{p}} = N_{\mathbf{q}} = 500$ were obtained from the religious demographics of Australia and Canada during the year 2011. The possible categories of religion that the individuals sampled could choose from are:

Anglican	Spiritualist
Baha'i	United Church
Baptist	No Religion
Buddhist	Jehovah's Witnesses
Church of Christ	Presbyterian and Reformed
Hindu	Zoroastrianism
Islam	Catholic
Judaism	Brethren
Latter Day Saints	Aboriginal Religion
Lutheran	Seventh Day Adventist
Pentecostal	Orthodox Christian
Salvation Army	Other Christian
Sikh	Other Religions

Figure 8.3

To test whether the religious make-up of the two countries is indeed different, a hypothesis test is conducted using the two aforementioned samples, which yields $\hat{A}_{JK_2} + \hat{B}_{JK_2} = 0.04388825$. Using T_2 from (5.2), with $\lambda = N_{\mathbf{p}}/N_{\mathbf{q}} = 1$, and noting that $K = 26$, we have

$$T_2 = 4N_{\mathbf{p}}(\hat{A}_{JK_2} + \hat{B}_{JK_2}) + \left(\sum_{k=1}^{K-1} \hat{r}_k(1 - \hat{r}_k) \left(\frac{1}{\hat{r}_K} + \frac{1}{\hat{r}_k} \right) - \sum_{m \neq n} \frac{\hat{r}_n \hat{r}_m}{\hat{r}_K} \right)$$

$$= 112.7308$$

Comparing this to the critical value of $\chi^2_{K-1,0.01} = 44.31410490$, and noting that the p-value is 0, clearly results in a rejected hypothesis. Therefore, we can say with 99% confidence that the two populations of Australia and Canada have different distributions over types of religion.

The population data from which the samples came is displayed in the following table:

	Australia		Canada	
	Number	Percentage	Number	Percentage
Anglican	3,679,907	17.11%	1,631,845	4.97%
Baha'i	13,705	0.06%	18,945	0.06%
Baptist	352,499	1.64%	635,840	1.94%
Buddhist	528,979	2.46%	366,830	1.12%
Church of Christ	49,688	0.23%	15,815	0.05%
Hindu	275,536	1.28%	497,960	1.52%
Islam	476,291	2.21%	1,053,945	3.21%
Judaism	97,334	0.45%	329,500	1.00%
Latter Day Saints	59,770	0.28%	108,665	0.33%
Lutheran	251,932	1.17%	478,180	1.46%
Pentecostal	237,984	1.11%	478,705	1.46%
Salvation Army	60,162	0.28%	70,955	0.22%
Sikh	72,296	0.34%	454,960	1.38%
Spiritualist	11,553	0.05%	4,315	0.01%
United Church	1,065,794	4.96%	2,007,610	6.11%
No Religion	4,796,786	22.30%	7,762,200	23.63%
Jehovah's Witnesses	85,636	0.40%	137,780	0.42%
Presbyterian and Reformed	599,517	2.79%	495,350	1.51%
Zoroastrianism	2,541	0.01%	6,130	0.02%
Catholic	5,439,266	25.29%	12,809,535	38.99%
Brethren	21,732	0.10%	18,110	0.06%
Aboriginal Religion	7,362	0.03%	64,940	0.20%
Seventh Day Adventist	63,002	0.29%	66,940	0.20%
Orthodox Christian	604,333	2.81%	549,465	1.67%
Other Christian	568,853	2.64%	2,253,860	6.86%
Other Religions	2,085,261	9.70%	533,945	1.63%
Total	21,507,719	100.00%	32,852,325	100.00%

Figure 8.4

The true Jensen-Shannon Divergence for this population is 0.03423257. Therefore the test correctly rejected the null hypothesis.

APPENDIX A: ADDITIONAL PROOFS

Lemma 14. *Let* $\mathbf{v}$ *and* $\hat{\mathbf{v}}$ *be defined as in (2.7) and (2.8), respectively. Additionally, note that we can write*

$$A(\mathbf{v}) = \frac{1}{2}\left(\sum_{k=1}^{K-1} p_k \ln(p_k) + \left(1 - \sum_{k=1}^{K-1} p_k\right)\ln\left(1 - \sum_{k=1}^{K-1} p_k\right)\right)$$

$$+ \frac{1}{2}\left(\sum_{k=1}^{K-1} q_k \ln(q_k) + \left(1 - \sum_{k=1}^{K-1} q_k\right)\ln\left(1 - \sum_{k=1}^{K-1} q_k\right)\right)$$

and

$$B(\mathbf{v}) = -\sum_{k=1}^{K-1} \frac{p_k + q_k}{2} \ln\left(\frac{p_k + q_k}{2}\right)$$

$$- \frac{\left(1 - \sum_{k=1}^{K-1} p_k\right) + \left(1 - \sum_{k=1}^{K-1} q_k\right)}{2} \ln\left(\frac{\left(1 - \sum_{k=1}^{K-1} p_k\right) + \left(1 - \sum_{k=1}^{K-1} q_k\right)}{2}\right)$$

Then the first and second partial derivatives for each p_k *and* q_k *are*

$$\frac{\partial}{\partial p_k} A(\mathbf{v}) = \frac{1}{2} \ln\left(\frac{p_k}{p_K}\right) \tag{A.1}$$

$$\frac{\partial}{\partial q_k} A(\mathbf{v}) = \frac{1}{2} \ln\left(\frac{q_k}{q_K}\right) \tag{A.2}$$

$$\frac{\partial}{\partial p_k} B(\mathbf{v}) = \frac{\partial}{\partial q_k} B(\mathbf{v}) = -\frac{1}{2} \ln\left(\frac{p_k + q_k}{p_K + q_K}\right) \tag{A.3}$$

and

$$\frac{\partial^2}{\partial p_k^2} A(\mathbf{v}) = \frac{1}{2}\left(\frac{1}{p_k} + \frac{1}{p_K}\right) \tag{A.4}$$

$$\frac{\partial^2}{\partial q_k^2} A(\mathbf{v}) = \frac{1}{2}\left(\frac{1}{q_k} + \frac{1}{q_K}\right) \tag{A.5}$$

$$\frac{\partial^2}{\partial p_{k_i}\partial p_{k_j}} A(\mathbf{v}) = \frac{1}{2p_K} \tag{A.6}$$

$$\frac{\partial^2}{\partial q_{k_i}\partial q_{k_j}} A(\mathbf{v}) = \frac{1}{2q_K} \tag{A.7}$$

$$\frac{\partial^2}{p_k q_k} B(\mathbf{v}) = \frac{\partial^2}{\partial p_k^2} B(\mathbf{v}) = \frac{\partial^2}{\partial q_k^2} B(\mathbf{v}) = -\frac{1}{2}\left(\frac{1}{p_k + q_k} + \frac{1}{p_K + q_K}\right) \tag{A.8}$$

$$\frac{\partial^2}{\partial p_{k_i}\partial q_{k_j}} B(\mathbf{v}) = \frac{\partial^2}{\partial p_{k_i}\partial p_{k_j}} B(\mathbf{v}) = \frac{\partial^2}{\partial q_{k_i}\partial q_{k_j}} B(\mathbf{v}) = -\frac{1}{2(p_K + q_K)} \tag{A.9}$$

Proof. For each k, $1 \leq k \leq K-1$,

$$\begin{aligned} \frac{\partial}{\partial p_k} A(\mathbf{v}) &= \frac{1}{2}\left(1 + \ln(p_k) + \left(-1 - \ln\left(1 - \sum_{k=1}^{K-1} p_k\right)\right)\right) \\ &= \frac{1}{2}\left(\ln(p_k) - \ln(p_K)\right) \end{aligned}$$

and

$$\frac{\partial}{\partial p_k} B(\mathbf{v}) = -\frac{1}{2}\left(1 + \ln\left(\frac{p_k + q_k}{2}\right)\right) - \frac{1}{2}\left(1 + \ln\left(1 - \frac{\sum_{k=1}^{K-1} p_k + \sum_{k=1}^{K-1} q_k}{2}\right)\right)$$

$$= -\frac{1}{2}\left(\ln\left(\frac{p_k + q_k}{2}\right) - \ln\left(1 - \frac{\sum_{k=1}^{K-1} p_k + \sum_{k=1}^{K-1} q_k}{2}\right)\right)$$

$$= -\frac{1}{2}\left(\ln\left(\frac{p_k + q_k}{2}\right) - \ln\left(\frac{p_K + q_K}{2}\right)\right)$$

The partials with respect to q_k are obtained similarly by symmetry. The second derivatives follow immediately from the first derivatives.

□

Lemma 15.

$$E\left(\sum_{k=1}^{K-1}(\hat{p}_k - p_k)\right)^2 = \sum_{k=1}^{K-1}\frac{p_k(1-p_k)}{N_{\mathbf{p}}} - \sum_{j\neq k}\frac{p_j p_k}{N_{\mathbf{p}}}$$

Proof.

$$E\left(\sum_{k=1}^{K-1}(\hat{p}_k - p_k)\right)^2 = Var\left(\sum_{k=1}^{K-1}\hat{p}_k\right)$$

$$= \sum_{k=1}^{K-1} Var(\hat{p_k}) + \sum_{j\neq k} Cov(p_j, p_k)$$

$$= \sum_{k=1}^{K-1}\frac{p_k(1-p_k)}{N_{\mathbf{p}}} - \sum_{j\neq k}\frac{p_j p_k}{N_{\mathbf{p}}}$$

□

The following lemma comes from [9] and is used only for reference.

Lemma 16. *Let* $\mathbf{G}$ *and* $\mathbf{H}$ *be arbitrary nonsingular matrices with* $\mathbf{H}$ *having rank one, then*

$$(\mathbf{G}+\mathbf{H})^{-1} = \mathbf{G}^{-1} - \frac{1}{1+g}\mathbf{G}^{-1}\mathbf{H}\mathbf{G}^{-1} \tag{A.10}$$

where $g = tr\{\mathbf{H}\mathbf{G}^{-1}\}$.

Bibliography

[1] Antos, A. and Kontoyiannis, I. (2001). Convergence Properties of Functional Estimates for Discrete Distributions. *Random Structures and Algorithms*, 19, 163-193.

[2] Blyth, C.R. (1959). Note on Estimating Information. *Annals of Mathematical Statistics*, 30, 71-79.

[3] Cover, Thomas M. and Thomas, Joy A. (2006). Elements of Information Theory, Second Edition. *Wiley Series in Telecommunications and Signal Processing*, 19.

[4] Crooks, Gavin E. (2008). Inequalities between the Jenson-Shannon and Jeffreys divergences. *Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA*

[5] 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.

[6] Kullback, S. and Leibler, R.A. (1951). On Information and Sufficiency. *Annals of Mathematical Statistics*, 22(1), 79-86.

[7] Lin, Jianhua (1991). Divergence Measures Based on the Shannon Entropy. *IEEE Transactions on Information Theory*, VOL. 37, NO. I, 145-151.

[8] Miller, G.A. and Madow, W.G. (1954). On the Maximum-Likelihood Estimate of the Shannon-Wiener Measure of Information. Operational Applications Laboratory, Air Force, Cambridge Research Center, Air Research and Development Command, Report AFCRC-TR-54-75; Luce, R.D., Bush, R.R., Galanter, E., Eds.; Bolling Air Force Base: Washington, DC, USA.

[9] Miller, Kenneth S. (1981). On the Inverse of the Sum of Matrices. *Riverside Research Institute*. 67.

[10] Paninski, L. (2003). Estimation of Entropy and Mutual Information. *Neural Computation*. 15, 1191-1253.

[11] Pearson, K. (1900). On a Criterion that a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is such that it can be Reasonably Supposed to have Arisen from Random Sampling. *Philosophical Magazine*, Series 5, 50, 157-175. (Reprinted in 1948 in Karl Pearson's Early Statistical Papers, ed by E.S. Pearson, Cambridge: Cambridge University Press.)

[12] Pearson, K. (1922). On the Chi Square Test of Goodness of Fit. *Biometrika*, 9, 22-27.

[13] Schechtman, Edna and Wang, Suojin. (2002). Jackknifing Two-Sample Statistics.

[14] Schindelin, J.E. and Endres, D.M. (2003) A new metric for probability distributions. *IEEE Transactions on Information Theory*, 1858 - 1860.

[15] Shannon, C.E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, 27, 379-423 and 623-656.

[16] Vinh, N.X., Epps, J. and Bailey, J. (2010). Information Theoretic Measures for Clusterings Comparison: Variants, Properties, Normalization and Correction for Chance. *Journal of Machine Learning Research*, 11, 2837-2854.

[17] Yao, Y.Y. (2003). Information-Theoretic Measures for Knowledge Discovery and Data Mining. *Entropy Measures, Maximum Entropy Principle and Emerging Applications*, Karmeshu (ed.), Springer, 115-136.

[18] Zhang, Z. (2017). Statistical Implications of Turing's Formula. *John Wiley & Sons, Inc.*

[19] Zhang, Z. (2012). Entropy Estimation in Turing's Perspective. *Neural Computation*, 24(5), 1368-1389.

[20] Zhang, Z. (2013b). Asymptotic Normality of an Entropy Estimator with Exponentially Decaying Bias. *IEEE Transactions on Information Theory*, 59(1), 504-508.

[21] Zhang, Z. and Zhang, X. (2012). A Normal Law for the Plug-in Estimator of Entropy. *IEEE Transactions on Information Theory*, 58(5), 2745-2747.

[22] Zhang, Z. and Zheng, L. (2015). A Mutual Information Estimator with Exponentially Decaying Bias. *Statistical Applications in Genetics and Molecular Biology*, 14(3), 243-252.

[23] Zhang, Z. and Zhou, J. (2010). Re-Parameterization of Multinomial Distribution and Diversity Indices. *Journal of Statistical Planning and Inference*, 140(7), 1731-1738.

[24] Zhang, Zhiyi, Chen, Chen and Zhang, Jialin (2018). Estimation of population size in entropic perspective . *Communications in Statistics - Theory and Methods.*