On the Sampling Distribution of Entropy Function

Pan Fu-Charo
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/gradreports

Part of the Mathematics Commons

Recommended Citation
Fu-Charo, Pan, "On the Sampling Distribution of Entropy Function" (1979). All Graduate Plan B and other Reports. 1174.
https://digitalcommons.usu.edu/gradreports/1174

This Report is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Plan B and other Reports by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
ON THE SAMPLING DISTRIBUTION OF ENTROPY FUNCTION

by

Pan Fu-Charo

A report submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

in

Mathematics

Plan B

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1979
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>HARD AND FUZZY CLUSTERING</td>
<td>2</td>
</tr>
<tr>
<td>THEORY OF DISTRIBUTION OF ENTROPY</td>
<td>5</td>
</tr>
<tr>
<td>GENERATION OF RANDOM FUZZY C-PARTITIONS</td>
<td>9</td>
</tr>
<tr>
<td>SIMULATION OF THE SAMPLING DISTRIBUTION</td>
<td>13</td>
</tr>
<tr>
<td>STATISTICAL ANALYSIS OF MONTE CARLO RESULTS</td>
<td>16</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>22</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Simulation result of average entropy function $H_2(U)$</td>
<td>14</td>
</tr>
<tr>
<td>2.</td>
<td>Simulation result of average entropy function $H_3(U)$</td>
<td>15</td>
</tr>
<tr>
<td>3.</td>
<td>Chi-square ($\chi^2$) table with 6 degrees of freedom</td>
<td>17</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>The conjectured empirical distribution of $H_c(U)$</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>The histogram of empirical distribution of $H_c(U)$ for the case $n = 1$</td>
<td>20</td>
</tr>
</tbody>
</table>
INTRODUCTION

Since the introduction of Shannon's Entropy Function in a non-probabilistic setting [1], there has been an expectancy that it will provide a useful measure for the fuzzy cluster validity problem. We hope to demonstrate here (empirically) by Monte Carlo simulation that the entropy $H_c(U)$ of fuzzy c-partitions $(U)$ defined below has expectation

$$E(H_c(U)) = \sum_{k=2}^{c} \frac{1}{k}$$

and variance

$$\text{Var}(H_c(U)) = \frac{1}{n} \left( \sum_{k=2}^{c} \frac{1}{k^2} - \frac{c-1}{c+1} \left( \frac{\pi^2}{6} - 1 \right) \right).$$

Furthermore, for sufficiently large $n$, that $H_c$ is approximately normal.

In this report we propose a method which generates the required fuzzy matrix needed for the simulation. After we finish the simulation, we shall study the empirical distribution of $H_c(U)$ at $c=2$ and 3 using statistical hypothesis testing.
HARD AND FUZZY CLUSTERING

Let $X = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^s$ denote a finite data set; we call each $x_k \in \mathbb{R}^s$ a feature vector, and each of its components $x_{kj} \in \mathbb{R}$, $1 \leq j \leq s$, the $j$th feature or characteristic of $x_k$. Hard cluster analysis with respect to $X$ refers to the problem of choosing an integer $c$, $2 \leq c \leq n$, and a hard $c$-partition of $X$ the subsets (or clusters) of which group together elements in $X$ which are similar to one another according to some (mathematically) well defined clustering criterion. The variety of substructures and interrelationships existing in real data precludes the possibility of finding a universal criterion capable of identifying "optimal" partitionings of an arbitrary data set. It is precisely this fact that argues for the use of fuzzy sets in classification and clustering. A substantial literature concerning conventional clustering techniques is referenced in [2].

To describe fuzzy $c$-partitions, we first define fuzzy subsets. A function $u_i$ on $X$ with range $[0,1]$ is called a membership function, which characterizes a fuzzy subset [3] in $X$: $u_{ik} = u_i(x_k)$ is the grade of membership of $x_k$ in $u_i$. It is appropriate to emphasize here the $u_{ik}$ is not interpreted as the probability that $x_k$ comes from class $i$; rather,
it indicates how closely the features of $x_k$ agree with the characteristics distinguishing the $i$th subclass. Let us denote by

$$M_{fc} = \{u \in \mathbb{V}^n : u_{ik} \in [0,1] \forall i,k; \sum_{i=1}^C u_{ik} = 1 \forall k; \sum_{k=1}^n u_{ik} > 0 \forall i\}. \quad (1)$$

Where $\mathbb{V}$ is the vector space of all real $c \times n$ matrices. Each $u \in M_{fc}$ is called a fuzzy $c$-partition of $X$. The condition

$$\sum_{k=1}^n u_{ik} > 0$$

simply requires each fuzzy subset in the partition to be nonempty. The condition

$$\sum_{i=1}^C u_{ik} = 1$$

stipulates that each $x_k \in X$ must have total membership (in the data set $X$) of unity, but this membership can otherwise be distributed arbitrarily among the fuzzy subsets partitioning $X$. Several techniques concerning fuzzy clustering algorithms are contained in [4,5,6].

The principle difficulty in evaluating the results of clustering algorithm is inability to visualize the geometrical properties in high dimensional spaces. Therefore, to be able to properly interpret the result of a cluster-seeking procedure, we must resort to mathematical measures of cluster validity.
Many quantitative measures of clustering properties can be considered. For example, a distance table which shows the distance between "cluster centers" can be a very useful interpretation tool. A table which shows the average and the farthest distance from each cluster center is also useful, or the variances of a cluster domain about its mean can be used to infer the relative distribution of samples in the domain. With these techniques, properties of clusters can be described, if we assume the number of clusters can be described, if we assume the number of clusters is known. However, if we are exploring an essentially unknown set of data, the number of clusters is an unknown. In this situation, the techniques discussed above are insufficient to evaluate clustering results, so better measures are needed. In the next section we introduce two measures which seem useful for interpretation of cluster validity.
THEORY OF DISTRIBUTION OF ENTROPY

In this section, we introduce two measures of partition fuzziness which yield methods for evaluation of cluster validity.

(3.1) Definition [4] For $U \in M_{fc}$, let
\[ F_c(U) = \frac{\text{trace}(UU^T)}{n} = \frac{\|U\|^2}{n}. \]

$F_c$ defines the partition coefficient of $U$.

The use of $F_c$ in connection with clustering is exemplified in [4], where it is shown that $F_c$ satisfies

(a) $1/c \leq F_c \leq 1$
(b) $1/c = F_c(U) \iff U = [1/c]$ i.e., $u_{ik} = 1/c \forall i,k$
(c) $1 = F_c(U) \iff U \in M_c$ is hard.

We not define Average Fuzzy Entropy

(3.2) Definition [1] For $U \in M_{fc}$, and $a \in (1, \infty)$, let
\[ H_c(U) = \left( \frac{-\sum_{k=1}^{c} \sum_{i=1}^{n} u_{ik} \log_a u_{ik}}{n} \right). \]

$H_c$ is the average classification entropy in $U$, where

$u_{ik} \log_a u_{ik} = 0$ if $u_{ik} = 0$. 

The following two theorems show the qualitative equivalence of \( F_c \) and \( H_c \).

**Theorem 3.1** [1] \( U \in M_{fc} ; F_c \) as in (3.1)

(a) \( 0 \leq H_c(U) \leq \log_a c \)

(b) \( H_c(U) = \log_a c \iff U = [1/c] \) (i.e. \( u_{ik} = 1/c \))

\[ \forall i,k \iff F_c(U) = 1/c \]

(c) \( H_c(U) = 0 \iff U \in M_c \) is hard \( \iff F_c(U) = 1 \)

**Theorem 3.2** [1] \( \{ U \in M_{fc}, u_{ik} \in (0,1) \forall i,k, F_c \) as in (3.1), \( H_c \) as in (3.2), \( a \in (1,\infty), e = 2.718... \) \) then

\[ 1-F_c(U) < \left( \frac{H_c(U)}{\log_a e} \right) < \frac{1}{2}(c-F_c(U)) \]

take \( a=e \) then

\[ 1-F_c(U) < H_c(U) < \frac{1}{2}(c-F_c(U)) \]

These results suggest that minimization of \( H_c(U) \) (or equivalently, maximization of \( F_c(U) \)) may lead to an optimal choice of \( U \in M_{fc} \). Numerical experiments indicate that \( H_c \) may be somewhat more sensitive than \( F_c \). The question addressed below concerns the behavior of \( F_c \) and \( H_c \) for different \( c \) under certain assumptions about the distribution of \( U \).

In this report we place emphasis on empirical verification of the distribution of \( H_c(U) \) as reported in [7].
We will study the entropy function under the assumption that possible choices for one column of $U$ are uniformly distributed over a piece of a hyperplane in $\mathbb{R}^c$.

For $c = 2$ and $3$ the situation is this:

**Example 3.1** for $c = 2$ we are sampling from

$$\{x \mid x_1 + x_2 = 1\} \cap (\mathbb{R}^+)^2$$

Sample space for each column of $U \subseteq M_f^2$

**Example 3.2** for $c = 3$ we are sampling from

$$\{x \mid x_1 + x_2 + x_3 = 1\} \cap (\mathbb{R}^+)^3$$

Sample space for each column of $U \subseteq M_f^3$
For arbitrary $c$, i.e., $2 \leq c \leq n$, we are sampling from the hyperplane (3.3)

\[ S_C = \left\{ \{x \in R^c_i \mid \sum_{i=1}^{c} x_i = 1\} \right\} \]

One object of the present study is to verify empirically the following theorem.

**Theorem 3.3** [7] If $U \in M_{fc}$ and each column is uniform on $S_C$ then $H_C(U)$ has expectation

\[ E(H_C(U)) = \sum_{k=1}^{c} \frac{1}{k} \]

and variance

\[ \text{Var}(H_C(U)) = \frac{1}{n} \left( \sum_{k=2}^{c} \frac{1}{k^2} - \frac{c-1}{c+1} \left( \frac{\pi^2}{6} - 1 \right) \right) . \]

In the next section we will investigate the generation of random fuzzy c-partitions which simulate sampling from the set $S_C$ defined above.
GENERATION OF RANDOM FUZZY C-PARTITIONS

To do the simulation, we need random vectors from $S_c$. Toward this end, three tests were made on the uniform $[0,1]$ random number generator used: A sequence was considered to be "random" if it satisfactorily passed all three tests.

Chi-square test

The chi-square test is perhaps the best known of all statistical tests. In general, suppose that every observation can fall into one of $k$ categories. We take $n$ independent observations $x_1,x_2,\ldots,x_n$. Let $P_s$ be the hypothesized probability that each observation falls into the $s$th category and let $Y_s$ be the number of observations which actually do fall into category $s$. We form the statistic

$$\chi^2 = \sum_{s=1}^{k} \frac{(Y_s - nP_s)^2}{nP_s}$$ \hspace{1cm} (4.1)

Simplifying (4.1), we arrive at the formula

$$\chi^2 = \frac{1}{n} \sum_{s=1}^{k} \left( \frac{Y_s^2}{P_s} \right) - n$$ \hspace{1cm} (4.2)

which is approximately $\chi^2$ with $k-1$ degrees of freedom if the $\{P_s\}$ are known. Thus, a value of the statistic $\chi^2$ can be used with chi-square tables to decide whether to accept the sequence as "random" at a given level of significance.
The Kolmogorov-Smirnov test

If we are given n independent observations \( x_1, x_2, \ldots, x_n \) taken from some distribution specified by a continuous function \( F(x) \), we form the empirical cumulative distribution function

\[
F_n(x) = \frac{\text{number of } x_k \text{'s which are } \leq x}{n}
\]

To make this test, we form the statistic

\[
K_n = \sqrt{n} \left\{ \sup_{-\infty < x < \infty} \{ |F_n(x) - F(x)| \} \right\}.
\]

As in the chi-square test, we can look up the value \( K_n \) in a "percentile" table to determine whether to accept a sequence or not at a given level of significance.

The third set is called the Poker test (partition test).

Poker test

Given n groups of k successive numbers, we can count the number of k-tuples with s different values. A chi-square test is then made (with \( k-1 \) degrees of freedom) for the probability \( P_s \),

\[
P_s = \frac{d(d-1)(d-2)\ldots(d-s+1)}{d^k} \binom{k}{s},
\]

where \( 1 \leq d < \infty \) is the number of the observed samples in a group, and \( \binom{k}{s} \) is the stirling number defined in [8].
The randomness test was applied as follows: Let $k=10$ (equally divide $[0,1]$ into ten subintervals), take $d=12$ (we observe twelve random numbers in a group), $n=50$ (fifty outcomes form a chi-square test). Then we count the number of different subintervals the random numbers fall into (say $s$, then $Y_s = Y_s + 1$ in (4.2)). We continue this for 50 times, and then perform a chi-square test (4.2) with 9 degrees of freedom at a specified level of significance.

A sequence is considered "random" if it passes test 1 and 3 with chi-squared value $\chi^2$ at (4.2), and test 2 with $K_n$ value at (4.3), which fall into the 10 percent to 90 percent range of the percentile table.

In our simulation, it was found that results based on passed sequences were, for all practical purposes, seemingly identical with results based on untested sequences! The sample mean and variance of $H_c(U)$ were quite close in both cases. Therefore, the results reported below are based on sequences generated directly (without testing).

Next we consider how to generate fuzzy c-partition matrices $U$ satisfying (2.1) with columns which are uniform over the set $(S_c)$ at (4.5). Here we propose an algorithm based on the following conjecture [7].
Conjecture

Let \( r_k \) be uniform on \([0,1], k=1,2,\ldots,c-1\) \hspace{1cm} (4.4)

\[ u_1 = 1 - (r_1)^{c-1} \]

\[ u_2 = (1 - u_1) \left(1 - (r_2)^{c-2}\right) \]

\[ \vdots \]

\[ u_k = (1 - \sum_{j=1}^{k-1} u_j) \left(1 - (r_k)^{c-k}\right), \quad k=2,3,\ldots,c-1 \]

\[ \vdots \]

\[ u_c = 1 - \sum_{j=1}^{c-1} u_j \]

then \( U = (u_1,u_2,\ldots,u_c) \) is uniform on

\[ S_c = \{ U \in \mathbb{R}^c | \sum_{j=1}^{c} u_j = 1, \quad u_j \geq 0, \quad \forall j \} \] \hspace{1cm} (4.5)

In our algorithm, \( r_1,r_2,\ldots,r_{c-1} \) are drawn from the random number generator. Repeating this \( n \) times, we can generate a fuzzy c-partition matrix \( U \) for the simulation.
<table>
<thead>
<tr>
<th>n</th>
<th>Empirical mean</th>
<th>$\sigma$</th>
<th>Chi-square value $\chi^2$</th>
<th>Theoretical mean</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.4916</td>
<td>.19156</td>
<td>104.10</td>
<td>.5000</td>
<td>.18708</td>
</tr>
<tr>
<td>2</td>
<td>.5063</td>
<td>.14325</td>
<td>10.18</td>
<td>.5000</td>
<td>.13229</td>
</tr>
<tr>
<td>10</td>
<td>.4963</td>
<td>.05941</td>
<td>4.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.4991</td>
<td>.05432</td>
<td>3.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>.5026</td>
<td>.05632</td>
<td>3.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.5039</td>
<td>.05907</td>
<td>2.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>.4974</td>
<td>.03742</td>
<td>4.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.5018</td>
<td>.03627</td>
<td>4.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>.4986</td>
<td>.03939</td>
<td>3.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.5005</td>
<td>.03968</td>
<td>2.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.5007</td>
<td>.02908</td>
<td>1.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.4984</td>
<td>.02600</td>
<td>4.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.4994</td>
<td>.02673</td>
<td>3.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.4986</td>
<td>.02626</td>
<td>1.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.4992</td>
<td>.01848</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.4995</td>
<td>.01940</td>
<td>2.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.5006</td>
<td>.01856</td>
<td>3.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.4974</td>
<td>.01815</td>
<td>6.56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: We are sampling 100 times for each row.

$\sigma^2$ = sample variance of $H_2(U)$

$\sigma^2$ = theoretical variance of $H_2(U)$

$\chi^2$ = value of the chi-square test with 6 degrees of freedom under the hypothesis that the distribution of $H_2(U)$ comes from the normal distribution with theoretical mean and variance.
Table 2. Simulation result of average entropy function $H_3(U)$

<table>
<thead>
<tr>
<th>n</th>
<th>Empirical mean</th>
<th>$\sigma$</th>
<th>Chi-square value $\chi^2$</th>
<th>Theoretical mean</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.8301</td>
<td>.20461</td>
<td>15.08</td>
<td>.8333</td>
<td>.19646</td>
</tr>
<tr>
<td>2</td>
<td>.8392</td>
<td>.14410</td>
<td>11.86</td>
<td>.8333</td>
<td>.13892</td>
</tr>
<tr>
<td></td>
<td>.8316</td>
<td>.05820</td>
<td>3.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.8398</td>
<td>.06435</td>
<td>4.54</td>
<td>.8333</td>
<td>.06216</td>
</tr>
<tr>
<td></td>
<td>.8338</td>
<td>.06097</td>
<td>5.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.8386</td>
<td>.06344</td>
<td>4.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.8324</td>
<td>.03855</td>
<td>3.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>.8311</td>
<td>.03625</td>
<td>1.50</td>
<td>.8333</td>
<td>.03931</td>
</tr>
<tr>
<td></td>
<td>.8351</td>
<td>.03785</td>
<td>3.46</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.8355</td>
<td>.04206</td>
<td>4.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.8315</td>
<td>.02612</td>
<td>4.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>.8328</td>
<td>.03613</td>
<td>4.02</td>
<td>.8333</td>
<td>.02780</td>
</tr>
<tr>
<td></td>
<td>.8361</td>
<td>.02695</td>
<td>3.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.8320</td>
<td>.03183</td>
<td>3.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.8330</td>
<td>.01757</td>
<td>3.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>.8344</td>
<td>.02026</td>
<td>2.62</td>
<td>.8333</td>
<td>.01966</td>
</tr>
<tr>
<td></td>
<td>.8335</td>
<td>.01790</td>
<td>3.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.8347</td>
<td>.02136</td>
<td>6.26</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: We are sampling 100 times for each row.

$\bar{\sigma}^2$ = sample variance of $H_3(U)$

$\sigma^2$ = theoretical mean of $H_3(U)$

$\chi^2$ = value of the chi-square test with 6 degrees of freedom under the hypothesis that the distribution of $H_3(U)$ comes from the normal distribution with theoretical mean and variance.
STATISTICAL ANALYSIS OF MONTE CARLO RESULTS

As we scan Tables 1 and 2, note that for each result the empirical mean and variance of $H_c(U)$ are quite close to the theoretical mean and variance. This appears to confirm the validity of Theorem 3.3.

Now we will investigate the sampling distribution of $H_c(U)$. We will test the null hypothesis that the empirical distribution of $H_c(U)$ is normal, with mean and variance as in Theorem 3.3, because the central limit theorem guarantees that, in the limit the empirical distribution of the mean of any distribution function is normal. In Tables 1 and 2 we list the results of tests of goodness-of-fit (using the chi-square test), by placing each observation of $H_c(U)$ into one of the following categories:

\[
\begin{align*}
S_1 &= [0, u-1.067*\sigma] \\
S_2 &= [u-1.067*\sigma, u-.566*\sigma] \\
S_3 &= [u-.566*\sigma, u-.18*\sigma] \\
S_4 &= [u-.18*\sigma, u+.18*\sigma] \\
S_5 &= u+.18*\sigma, u+.566*\sigma] \\
S_6 &= [u+.566*\sigma, u+1.067*\sigma] \\
S_7 &= [u+1.067*\sigma, \log(\text{c})]
\end{align*}
\]

where $u = E(H_c(U))$

\[= \text{SQRT(Var}(H_c(U)))).\]
The probability that $H_C(U)$ falls into $X_i = \frac{1}{7}$ (i.e., $P_s$ at (4.2) = $\frac{1}{7}$, $\forall s$): thus, we test goodness-of-fit using the chi-squared test with 6 degrees of freedom.

For convenience we list the percentile table of the chi-square distribution with 6 degrees of freedom.

Table 3. Chi-square ($\chi^2$) table with 6 degrees of freedom

<table>
<thead>
<tr>
<th>$\chi^2$</th>
<th>.01</th>
<th>.025</th>
<th>.10</th>
<th>.20</th>
<th>.30</th>
<th>.50</th>
<th>.70</th>
<th>.80</th>
<th>.90</th>
<th>.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$</td>
<td>.87</td>
<td>1.24</td>
<td>1.64</td>
<td>3.07</td>
<td>3.83</td>
<td>5.35</td>
<td>7.23</td>
<td>8.56</td>
<td>12.6</td>
<td>14.4</td>
</tr>
</tbody>
</table>

With Table 3 we compare the sampling results for the chi-square values in Tables 1 and 2. We see that for the cases $n = 1$ and 2 the chi-square value is large, which means the empirical distribution of $H_C(U)$ definitely does not come from the normal distribution. However, as $n$ gets large $H_C(U)$ begins to agree with a normal distribution. Most of the chi-square values are reasonably small, which supports the hypothesis that the empirical distribution of $H_C(U)$ comes from the normal distribution with mean and variance (as in Theorem 3.3) at, say, the 70 percent level of significance.
If the above hypothesis is accepted, Figure 1 illustrates the conjectured behavior of the empirical distribution of $H_c(U)$ as a function of $c$.

Observe, however, that even for relatively large $n$, the test does not support this null hypothesis very strongly. There are two possible reasons for this: (1) The chi-square test is better for discrete distributions; (2) rate of convergence to the hypothesized normal is slow.

Based on the generated data, it seems clear that the sampling distribution of $H_c$ over single columns of $U$ is not normal, that is, the sampling distribution of Shannon's Entropy Function has mean and variance as given in Theorem 3.3, but is not normal. Histograms of the empirical distribution (Figure 2) for $c = 2,3,4,5,6,8$ support these observations.

The shape of the empirical distribution of $H_c$ suggests that it might be chi-squared with $c-1$ degrees of freedom. However, the variance of $H_c(U)$ decreases as $c$ increases (exclude $c=3$), whereas the variance of the chi-square distribution increases as $c$ increases. Although we didn't find a density function that the empirical distribution of $H_c(U)$ (at $n=1$) seems to agree with, the central limit theorem effectively eliminates the need for this knowledge. In actual cases, $n$ will be large enough (usually at least 50), so that the empirical distribution of the Average
Figure 1. The conjectured empirical distribution of $H_c(U)$. 

<table>
<thead>
<tr>
<th>$c$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{x}$</td>
<td>.50</td>
<td>.833</td>
<td>1.083</td>
<td>1.283</td>
<td>1.45</td>
<td>1.593</td>
<td>1.718</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>$\frac{.035}{n}$</td>
<td>$\frac{.0386}{n}$</td>
<td>$\frac{.0367}{n}$</td>
<td>$\frac{.0337}{n}$</td>
<td>$\frac{.0307}{n}$</td>
<td>$\frac{.0281}{n}$</td>
<td>$\frac{.0258}{n}$</td>
</tr>
</tbody>
</table>

$\bar{x}$ = mean of $H_c(U)$  
$\sigma^2$ = Variance of $H_c(U)$
Note: We sample 1000 times for each histogram, and for each interval, the interval is $1/20$ of $[0, \log(c)]$ for each $c$.

Figure 2. The histogram of empirical distribution of $H_c(U)$ for the case $n = 1$. 
Entropy function $H_c(U)$ over all columns of $U$ is satisfactorily close to the normal distribution with mean and variance as reported in Theorem 3.3 above.
REFERENCES


