8-2005

A Limit Theorem in Cryptography.

Kevin Lynch
East Tennessee State University

A Limit Theorem in Cryptography

A thesis
presented to
the faculty of the Department of Mathematics
East Tennessee State University

In partial fulfillment of
the requirements for the degree
Master of Science in Mathematical Sciences

---

by
Kevin Roy Lynch
August 2005

Anant Godbole, Ph.D., Chair
Debra Knisley, Ph.D.
James Boland, Ph.D

Keywords: Cryptography, Stein-Chen Method, Poisson Approximation,
Cryptanalysis, Error Bound, Coupling Method, Even Word, Multiple of Three Word

ABSTRACT

A Limit Theorem in Cryptography

by

Kevin Roy Lynch

Cryptography is the study of encryptying and decrypting messages and deciphering encrypted messages when the code is unknown. We consider $\Lambda_\pi(\Delta x, \Delta y)$ which is a count of how many ways a permutation satisfies a certain property. According to Hawkes and O'Connor, the distribution of $\Lambda_\pi(\Delta x, \Delta y)$ tends to a Poisson distribution with parameter $\frac{1}{2}$ as $m \to \infty$ for all Δx,$\Delta y \in (\mathbf{Z}/q\mathbf{Z})^m$ minus 0. We give a proof of this theorem using the Stein-Chen method: As q^m approaches infinity, the distribution of $\Lambda_\pi(\Delta x, \Delta y)$ is approximately Poisson with parameter $\frac{1}{2}$. Error bounds for this approximation are provided.

DEDICATION

I dedicate this thesis to my aunt, Lisa Keel, who inspired me to pursue an advanced degree.

ACKNOWLEDGEMENTS

I am grateful to Dr. Anant Godbole, Dr. Debra Knisley, and Dr. James Boland for all their help during the preparation of this thesis. I would also like to thank my parents, Roy and Joan Lynch, for their unwavering encouragement during graduate school.

CONTENTS

LIST OF TABLES

1 INTRODUCTION

1.1 Introduction to Cryptography

'Cryptography is the study of encrypting and decrypting messages and deciphering encrypted messages when the code is unknown' [2]. Cryptography makes it possible to communicate in a covert manner:

> The basic objective of cryptography is to enable two people, usually referred to as Jennifer and James, to communicate over an insecure channel in such a way that an opponent, Sam, cannot understand what is being said. The information that Jennifer wants to send to James is called plaintext. Jennifer encrypts the plaintext, using a predetermined key, and sends the resulting ciphertext over a channel such as a telephone line. Sam, upon seeing the ciphertext in the channel by eavesdropping, cannot determine what the plaintext was, but James, who knows the encryption key, can decrypt the ciphertext and reconstruct the plaintext [3].

Let us consider the previous discussion in mathematical notation:

Definition 1.1.1

A *cryptosystem* is a five-tuple (P, C, Ω, E, D), where the following conditions are satisfied:

1. P is a finite set of possible plaintexts
2. C is a finite set of possible ciphertexts
3. Ω, the keyspace, is a finite set of possible keys

4. For each K in Ω, there is an encryption rule e_K in E and a corresponding decryption rule d_K in D. Each $e_K : P \to \mathrm{C}$ and $d_K : C \to P$ are functions such that $d_K(e_K(x)) = x$ for every plaintext element x in P [3].

Another definition of importance is that of cryptanalysis:

Definition 1.1.2

> *Cryptanalysis* is the process of attempting to compute the key K, given a string of ciphertext [3].

In the previous definition, each plaintext element was called x. Now, let us consider block-ciphers, which means that the x's are blocks of plaintext of a certain length.

Worth mentioning is a type of cryptanalysis which is used in the Digital Encryption attack:

> Differential Cryptanalysis is a so-called chosen plaintext attack to block ciphers. The general idea of differential cryptanalysis (for a spy who does not know the key K) is to look for pairs $\Delta x, \Delta y$ such that if the difference of plaintexts is Δx, then the difference of ciphertexts is Δy. If such a pair of $(\Delta x, \Delta y)$ occurs significantly more frequently than it should by pure randomness, then this allows one to extract information about the key K [1].

Now, consider the following key definition:

> Let $q \in N$, $q \geq 2$ and fix $\Delta x, \Delta y \in (\mathbf{Z}/q\mathbf{Z})^m$. Let π be a (uniformly distributed) random permutation of $(\mathbf{Z}/q\mathbf{Z})^m$ (which occurs due to a randomly chosen key K) and consider the random variable $\Lambda_\pi(\Delta x, \Delta y)$ giving the number of (unordered) pairs $(x, x') \subset (\mathbf{Z}/q\mathbf{Z})^m$ of plaintexts x, x' such that $x + x' = \Delta x$ and $y + y' = \Delta y$, where $y = \pi(x)$, $y' = \pi(x')$ are the corresponding ciphtertexts [1].

The notation $\Lambda_\pi(\Delta x, \Delta y)$ is a count of how many ways a permutation satisfies a certain critical property related to differential and linear cryptanalysis.

1.2 Definitions and Notation

The cardinality of the group $(\mathbf{Z}/q\mathbf{Z}, +)$ is symbolized by q and is the size of our alphabet. Thus, if $q = 26$ which is the cardinality of the set $\{0, 1, 2, ..., 25\}$, we have the correspondence between alphabetic characters and residues modulo 26: $A \leftrightarrow 0$, $B \leftrightarrow 1$, $C \leftrightarrow 2$, ..., $Z \leftrightarrow 25$.

The length of each word is represented by m. For example, the length of "CAT" is 3.

A word Δx is called *even* if it can be written as $x + x$. For example, "CAU" is an even word since it can be expressed as the sum of BAK and BAK. That is, CAU = BAK + BAK. Note that addition is done modulo q (26 in this case) and is performed componentwise.

A word Δx is called *odd* if it is not even.

1.3 Preliminary Results

The lemmas in this section are critical since the random variable $\Lambda_\pi(\Delta x, \Delta y)$ counts the number of unordered ways that a certain property holds. Key to that property is an enumeration of the ways in which a word can be decomposed into two or three parts.

Lemma 1.3.1

There are q^m ordered ways of expressing a word Δx of length m as $x + x' = \Delta x$.

Proof

First, let us prove that given any $\bar{x}$ and $\bar{a}$ in ($\mathbf{Z}$/q$\mathbf{Z}$, +), there exists a unique $\bar{y}$ such that $\bar{x} + \bar{y} = \bar{a}$. For each $\bar{x}$, there exists an element $\bar{x}^{-1}$ such that $\bar{x} + \bar{x}^{-1} = \bar{0}$. Hence, $\bar{x} + \bar{x}^{-1} + \bar{a} = \bar{a}$ and we see that $\bar{y} = \bar{x}^{-1} + \bar{a}$ is the unique element satisfying $\bar{x} + \bar{y} = \bar{a}$.

Now, for a word $\bar{a} = a_1 a_2 a_3 ... a_m$, there are q ways in which $x_1 + x_1' = a_1$, q ways in which $x_2 + x_2' = a_2$, q ways in which $x_3 + x_3' = a_3$, and finally there are q ways in which $x_m + x_m' = a_m$. Thus, by the multiplication principle, there are q^m ordered ways of expressing a word Δx of length m as $x + x' = \Delta x$. □

Lemma 1.3.2

If q is odd, then $x + x = \Delta x$ has only one solution. Note that Δx must be even if q is odd.

Proof

If x_1 is even we let $a_1 = \frac{x_1}{2}$. If x_1 is odd, $x_1 + q$ is even and $\frac{x_1+q}{2} + \frac{x_1+q}{2} = x_1 + q \equiv x_1 (mod q)$. So, we let $a_1 = \frac{x_1+q}{2}$. It may similarly be checked that this solution is unique.□

Lemma 1.3.3

The number of unordered ways of writing Δx as $x + x'$ is $\frac{q^m}{2}$, if Δx is odd and q is even. If q is odd, Δx cannot be odd since each word is even.

Proof

Any word Δx can be written in the form $\Delta x = a + b$ where a and b are m letter words. It is true that since Δx is odd, $a \neq b$. If we choose a, then b is automatically determined by modular arithmetic. Since b is determined by the choice of a and vice versa, if there are q^m ordered ways to choose a, then b is determined. By Theorem 1.3.1, there are q^m ordered ways of expressing $\Delta x = x + x'$. However, we must be concerned with double-counting. For example, CAT = COP + AME and CAT = AME + COP. So, if we count the two ways to express CAT: AME + COP and COP + AME, we are double-counting. To eliminate the problem of double-counting, we divide the number of ordered ways of expressing Δx, q^m, by 2. Therefore, the number of ordered ways of expressing $\Delta x = x + x'$ with Δx odd is $\frac{q^m}{2}$.□

Lemma 1.3.4

If q is even and Δx is even, the number of unordered ways of writing Δx as $x + x'$ is $\frac{q^m - 2^m}{2} + 2^m$.

Proof

First let us prove that the number of unordered ways of writing Δx as $x + x$ is 2^m if q is even.

One way of writing Δx as $x + x$ is using the fact that for every letter of Δx, $a = (\frac{a}{2} + \frac{a}{2}) mod q$. Another way of expressing an even letter is $a = (\frac{a}{2} + \frac{q}{2} + \frac{a}{2} + \frac{q}{2}) mod q = (a + q) mod q$. Now, consider $a = (a + 2q) mod q$. This would mean that we have written $a = \frac{a}{2} + q + \frac{a}{2} + q (mod q)$. But $\frac{a}{2} + q$ is not in our alphabet. Thus, there are only two ways of writing each letter of a word of length m . Hence, by the multiplication principle, there are a total of 2^m ways of writing Δx as $x + x$. If we subtract the number of ways of writing Δx as $x + x$, 2^m, from q^m ordered ways of expressing Δx as $x + x'$, we obtain $q^m - 2^m$. Well, $q^m - 2^m$ is the number of ordered ways of writing Δx as $x + x'$, $x \neq x'$ when Δx is even. However, if we divide the number of ordered ways of writing Δx as $x + x'$, q^m, by 2, we get $\frac{q^m - 2^m}{2}$. But $\frac{q^m - 2^m}{2}$ is not the number of unordered ways of writing Δx as $x + x'$. We need to add back the number of ways of writing Δx as $x + x$, 2^m. So now if we add 2^m to $\frac{q^m - 2^m}{2}$, we obtain $\frac{q^m - 2^m}{2} + 2^m$, the number of unordered ways of writing Δx as $x + x'$.□

Lemma 1.3.5

If q is odd, the total number of unordered ways of writing the even word, $\Delta x = x + x'$, is $\frac{q^m-1}{2} + 1 = \frac{q^m+1}{2}$.

Proof

If q is odd and Δx is even, there is exactly one way to express Δx as $x + x$ as seen in lemma 1.3.2. Now, $q^m - 1$ is equal to the number of ordered ways of writing Δx minus the number of ways of expressing an even word Δx in our odd alphabet as $x+x$. Since we have double-counted, we must divide q^m-1 by 2 to get the number of unordered ways of expressing an even word Δx when q is odd, but when $\Delta x \neq x+x$. Now, adding back this case, we get $\frac{q^m-1}{2} + 1$ as needed.□

Lemma 1.3.6

If $q \geq 2$, the number of ordered ways that we can write $\Delta x = a + b + c$ is q^{2m}.

Proof

Consider a word Δx of length m, $\Delta x = x_1 + x_2 + x_3$. Let x_1 and x_2 be arbitrary words. We know that for any word $x_1 = a_1a_2a_3 \ldots a_m$ of length m, there are q^m ways to write x_1 since there are q ways to choose a_1, q ways to choose a_2, and eventually q ways to choose a_m. So, by the multiplication principle, there are q^m ways to express a word x_1. Similarly, there are q^m ways of expressing x_2. But, now x_3 must be determined from the choices of x_1 and x_2. Now, let $\Delta x = c_1c_2 \ldots c_m$, $x_1 = a_1a_2a_3 \ldots a_m$, and $x_2 = b_1b_2 \ldots b_m$. Since $\Delta x = x_1+x_2+x_3$, we have by modular arithmetic, $x_3 = \Delta x - x_1 - x_2 = c_1c_2 \ldots c_m - a_1a_2 \ldots a_m - b_1b_2 \ldots b_m = d_1d_2 \ldots d_m$.

Hence, x_3 is determined by the choices of x_1 and x_2 and there is only one choice for x_3. Since there are q^m choices for x_1, q^m choices for x_2, and one choice for x_3,

by the multiplication principle, there are $q^m \cdot q^m \cdot 1 = q^{2m}$ ordered ways of writing $\Delta x = x_1 + x_2 + x_3$.□

Proposition 1.3.7

The approximate number of unordered ways of expressing $\Delta x = x_1 + x_2 + x_3$ is $\frac{q^{2m}}{6}$.

Proof

We know by Lemma 1.3.6, that the number of ordered ways of writing $\Delta x = x_1 + x_2 + x_3$ is q^{2m}. However, in the total number of ordered ways, q^{2m}, there is usually a six-fold duplication of words. For example, if $m = 3$, and the word is CAT, there are 3! or 6 ways of expressing CAT as the sum of ALE, CBC, and AON. That is,

$$\begin{aligned} \text{CAT} &= \text{ALE} + \text{CBC} + \text{AON} \\ &= \text{ALE} + \text{AON} + \text{CBC} \\ &= \text{CBC} + \text{ALE} + \text{AON} \\ &= \text{CBC} + \text{AON} + \text{ALE} \\ &= \text{AON} + \text{CBC} + \text{ALE} \\ &= \text{AON} + \text{ALE} + \text{CBC}. \end{aligned}$$

So, to avoid counting duplicate words, we divide by the total number of ordered ways of writing $\Delta x = x_1 + x_2 + x_3$ by 3!. Therefore, we have approximately $\frac{q^{2m}}{3!} = \frac{q^{2m}}{6}$ unordered ways of writing $\Delta x = x_1 + x_2 + x_3$, since "most" decompositions will be of the above form.□

Lemma 1.3.8

If $q \geq 2$ and Δx is not a multiple of three, then the exact number of unordered ways of writing $\Delta x = x_1 + x_2 + x_3$ is $\frac{q^{2m}-3q^m}{6} + q^m$. Note that q must be a multiple of three since if it is not, each word is a multiple of three.

Proof

Since Δx is not a multiple of three, it can be written as $x + y + z$ or $x + x + y$ but $x = y = z$ is impossible. It has been shown by Lemma 1.3.6, that there are q^{2m} ways of writing Δx as a sum of three parts. Consider the possibility $x + x + y$. Now, since there are q^m choices for each x, with y determined and thus $3q^m$ ordered choices. There are $3q^m$ ordered choices because for each Δx there are three possible representations for $\Delta x = x + x + y$: $x + x + y$, $y + x + x$, and $x + y + x$. To obtain the exact number of unordered ways of writing $\Delta x = x_1 + x_2 + x_3$, we first subtract the number of ordered ways of writing Δx as $x + x + y$ from the total number of ordered ways of writing $\Delta x = x_1 + x_2 + x_3$: $q^{2m} - 3q^m$. The expression $q^{2m} - 3q^m$, is the number of ordered ways of writing Δx as $x + y + z$ with x, y, and z distinct. If we divide $q^{2m} - 3q^m$ by 3!, the number of ways that $x + y + z$ can be expressed as a sum, we get the number of unordered ways. Thus, we obtain $\frac{q^{2m}-3q^m}{6}$ and since Δx is not a multiple of three, we do not consider $x + x + x$. Now we add q^m to $\frac{q^{2m}-3q^m}{6}$, the number of unordered representations of $x + x + y$. The expression $\frac{q^{2m}-3q^m}{6} + q^m$, gives the exact number of unordered ways to write $\Delta x = x_1 + x_2 + x_3$ when Δx not a multiple of three.□

Lemma 1.3.9

If $q \geq 2$, Δx is a multiple of three, and q is a multiple of three, then the exact number of unordered ways of expressing $\Delta x = x + y + z$ is $\frac{q^{2m}-3(q^m-3^m)-3^m}{6} + q^m$.

Proof

First, it is true that for every Δx which is a multiple of three, $\Delta x = a + a + a$, where $a = a_1 a_2 \ldots a_m$. If $\Delta x = x_1 x_2 \ldots x_m$, then $x_1 = 3a_1$, $x_2 = 3a_2$, $x_3 = 3a_3$, and finally $3a_m = x_m$. Therefore, if each component x_1,x_2, ..., x_m is a multiple of three, then Δx is a multiple of three. Now we need to show that $3a_1 = x_1$ has three solutions in $\mathbf{Z}/q\mathbf{Z}$. To show this, we need to prove that $a_1 = \frac{x_1}{3}$, $a_1 = \frac{x_1+q}{3}$, and $a_1 = \frac{x_1+2q}{3}$ are all between 0 and $q-1$. Since $q > 0$, $x_1 > 0$, x_1 is a multiple of three, and q is a multiple of three, it is a fact that $\frac{x_1+2q}{3}$ is divisible by three and therefore is greater than or equal to zero. Next, we need to show that $\frac{x_1+2q}{3} \leq q-1$. Consider $\frac{x_1+2q}{3} \leq q-1$. It is true that $\frac{x_1+2q}{3} \leq q-1 \Rightarrow x_1 + 2q \leq 3q-3 \Rightarrow x_1 \leq q-3$. but $x_1 \leq q-3 \leq q-1$ for $q > 0$, q a multiple of three. The case $\frac{x_1+q}{3}$ is proved similarly. Hence, $3a_1 = x_1$ has three solutions for a_1 if x_1 is a multiple of three. There are three ways of writing each letter of Δx if Δx is a multiple of three since there are three ways to write a_1, three ways to write a_2, and three ways to write a_m; by the multiplication principle, there are 3^m ways of writing Δx as $x + x + x$. First, we will subtract $q^m - 3^m$ from q^{2m}. Well, $q^m - 3^m$, is the difference between the number of representations of the form $x + x + y$ and the number of words that can be expressed as a multiple of three. If we multiply $q^{2m} - 3^m$ by three, we have the number of ordered ways of expressing $\Delta x = x + x + y$. From the term $q^{2m} - 3(q^{2m} - 3^m)$, we subtract the number of ways of expressing $\Delta x = x + x + x$, 3^m, to obtain the number

of ordered ways of expressing $\Delta x = x + y + z$. Dividing $q^{2m} - 3(q^2 - 3^m) - 3^m$ by 3! gives the unordered ways of expressing $\Delta x = x + y + z$. After the division, we have $\frac{q^{2m}-3(q^m-3^m)-3^m}{6}$. It is necessary to add q^m, the number of unordered ways of writing $\Delta x = x + x + y$ or $x + x + x$. Hence, we have $\frac{q^{2m}-3(q^m-3^m)-3^m}{6} + q^m$, the unordered number of ways of expressing $\Delta x = x + y + z$.□

Lemma 1.3.10

If $q \geq 2$, Δx is a multiple of three and $q \equiv 1(mod 3)$ or $q \equiv 2(mod 3)$, then the exact number of unordered ways of writing $\Delta x = x + y + z$ is $\frac{q^{2m}-3(q^m-1)-1}{6} + q^m$.

Proof

Consider $x_1 = 0, 3, 6, \ldots, q-1, q+2, q+5, \ldots, 2q-2, 2q+1, \ldots, 3q-3$. We see that these are all divisible by three and yield $a_1 = 0, 1, 2, \ldots, \frac{q-1}{3}, \frac{q+2}{3}, \ldots, \frac{2q-2}{3}, \ldots, q-1$. Also, $q+2, q+5, \ldots$ are congruent to $2, 5, \ldots, q-2$ and $2q+1, \ldots, 3q-3$ are congruent to $1, 4, 7, \ldots, q-3$. The case of $q \equiv 2(mod 3)$ is similar. First, we will subtract $3(q^m-1)$ from q^{2m}. Let us explain the previous sentence. We know that there are q^m unordered ways of writing Δx as $x + y + z$ or $x + x + x$. Since there is only one way to express $\Delta x = x + x + x$, the number of unordered ways of writing $\Delta x = x + x + y$ is $q^m - 1$. If we multiply $q^m - 1$ by three, we obtain the number of ordered ways of expressing $\Delta x = x + x + y$. From the term $q^{2m} - 3(q^m - 1)$, we subtract the number of ways of expressing $\Delta x = x + x + x$, one, to obtain the number of ordered ways of expressing $\Delta x = x+y+z$. Now, dividing $q^{2m}-3(q^m-1)-1$ by 3! gives the number of unordered ways of expressing $\Delta x = x + y + z$. After the division, we have $\frac{q^{2m}-3(q^m-1)-1}{6}$. It is necessary to add q^m, the number of unordered ways of writing Δx as $x + y + z$ or $x + x + x$. Hence, we have $\frac{q^{2m}-3(q^m-1)-1}{6} + q^m$, the exact number of unordered ways

of writing $\Delta x = x + y + z$, where Δx is a multiple of three.□

Let us consider an example that illustrates $q^{2m} - 3(q^m - 1) - 1$ is divisible by six. Suppose that $q \equiv 1(mod 3)$, $q = 5$ and $m = 3$. If $q = 5$ and $m = 3$, we have $\frac{5^6 - 3(5^3 - 1) - 1}{6} = 2542$. Now, suppose that $q \equiv 1(mod 3)$, $q = 4$ and $m = 3$. Now, we have $\frac{4^6 - 3(4^3 - 1) - 1}{6} = 651$. Thus, $q^{2m} - 3(q^m - 1) - 1$ is divisible by six.

1.4 Two Examples

In Section 1.1, we introduced the notation $\Lambda_\pi(\Delta x, \Delta y)$. We know that $\Lambda_\pi(\Delta x, \Delta y)$ is a count of how many times a permutation satisfies a certain property. In one situation, $\Lambda_\pi(\Delta x, \Delta y)$ may represent the number of times that we obtain "DOG" given the input "CAT" under the random permutation π. It may behoove us to consider an example to understand the meaning of $\Lambda_\pi(\Delta x, \Delta y)$. Suppose that $q = |\{0, 1\}| = 2$ and $m = 2$. Let us fix CA = 10 and DO = 11 there are $q^m = 2^2 = 4$ words. The four words are: 00, 01, 10, 11.

Now consider a permutation, symbolized by π of these four words:

$00 \rightarrow 00$

$01 \rightarrow 01$

$10 \rightarrow 11$

$11 \rightarrow 10$

There are 2 ways that we can obtain "CA" as a sum, $x + x'$:

$x_1 + x_1' = 10 + 00 = 10$, $x_2 + x_2' = 11 + 01 = 10$,

We are interested in the number of ways we can obtain "DO" from the images $\pi(x)$ and $\pi(x')$ of x and x'.

The sum of the images, $\pi(x) + \pi(x')$ are:

$\pi(x_1) + \pi(x_1') = 11 + 00 = 11 = DO \quad (1)$

$\pi(x_2) + \pi(x_2') = 10 + 01 = 11 = DO \quad (2)$

Notice that equations (1) and (2) both produce situations in which $\pi(x) + \pi(x')$ equals DO. Since $\Lambda_\pi(\Delta x, \Delta y)$ is the count of how many times the images of x and x' add to DO, $\Lambda_\pi(\Delta x, \Delta y) = 2$.

We know that for m letter words on an alphabet of size q, there are q^m possible words. For the total number of q^m words, there are $q^m!$ possible random permuations of those words. Let us define the random permutation in our preceding example to be π_1. There are $q^m = 2^2! = 4! = 24$ permutations of the four words. Hence in our notation, we have: π_1, π_2, ..., π_{24}.

By hand, I have determined all 24 possible permutations and computed $\Lambda_\pi(\Delta x, \Delta y)$ for each case. For this particular example, we have the following probability distribution of $\Lambda_\pi(\Delta x, \Delta y)$:

Table 1: Probability Distribution for $\Lambda_\pi(\Delta x, \Delta y)$

j	$P(\Lambda_\pi(\Delta x, \Delta y) = j)$
0	$\frac{2}{3}$
1	0
2	$\frac{1}{3}$

In our second example regarding $\Lambda_\pi(\Delta x, \Delta y)$, suppose that $q = |\{0, 1\}| = 2$ and $m = 3$. Let us fix "CAT" = 110 and "DOG" = 011 there are $q^m = 2^3 = 8$ words. The eight words are 111, 110, 101, 100, 011, 010, 001, and 000.

Now consider a permutation, π, of these eight words:

$111 \rightarrow 000$

$110 \rightarrow 101$

$101 \rightarrow 111$

$100 \rightarrow 100$

$011 \rightarrow 001$

$010 \rightarrow 010$

$001 \rightarrow 011$

$000 \rightarrow 110$

There are four ways that we can obtain "CAT" as a sum, $x + x'$:

$x_1 + x_1' = 111 + 001 = 110$, $x_2 + x_2' = 000 + 110 = 110$, $x_3 + x_3' = 100 + 010 = 110$, $x_4 + x_4' = 101 + 011 = 110$. This fact illustrates Lemma 1.3.3 since $q = 2$ is even and CAT $= 110$ is odd.

Again, we are interested in the number of ways we can obtain "DOG" from the images $\pi(x)$ and $\pi(x')$ of x and x'. The sum of the images, $\pi(x) + \pi(x')$ are:

$\pi(x_1) + \pi(x_1') = 000 + 011 = 011 = DOG$ (1)

$\pi(x_2) + \pi(x_2') = 110 + 101 = 011 = DOG$ (2)

$\pi(x_3) + \pi(x_3') = 100 + 010 = 110 \neq DOG$ (3)

$\pi(x_4) + \pi(x_4') = 111 + 001 = 110 \neq DOG$ (4)

Note that equations (1) and (2) produce situations in which $\pi(x) + \pi(x')$ equals DOG. Since $\Lambda_\pi(\Delta x, \Delta y)$ is the count of how many times the images of x and x' add to DOG, $\Lambda_\pi(\Delta x, \Delta y) = 2$.

Note that for any choice of q and m it is possible to determine a bound on $\Lambda_\pi(\Delta x, \Delta y)$ for each of the q^m random permutations. For instance, for $q = 5$, $m = 4$, Δx is HAWK, and Δy is PREY, since q is odd, HAWK is even and by Lemma 1.3.5, there are $\frac{5^4+1}{2} = 313$ ways to write HAWK as $x + x'$. Thus, $0 \leq \Lambda_\pi(\Delta x, \Delta y) \leq 313$.

Now, consider the following table which illustrates the computational complexity of $q^m!$:

Table 2: Values of q^m and $q^m!$ for Various Choices of q and m

q	m	q^m	$q^m!$
2	2	4	24
2	3	8	40320
3	2	9	362880
2	4	16	20922789888000

As you can see, even with small choices of q and m, the values of $q^m!$ grow very large. For $q = 2$ and $m = 8$, we have $2^8! = 256!$ permutations which is such a large number, a TI-83 cannot compute it. If $q = m = 5$, there are $5^5!$ permutations which even the supercomputer at Oak Ridge National Laboratory cannot calculate!

Only for very small values of q and m is it possible to calculate by hand all of the $q^m!$ permutations. Even with today's most powerful supercomputers, we cannot determine the number of permutations associated with large choices of q and m. However, with probability theory there is hope for finding a probability distribution for $\Lambda_\pi(\Delta x, \Delta y)$.

2 NEW RESULTS

2.1 The Stein-Chen Method for the Decomposition of Two Words

The following theorem is a two-fold improvement over the theorem proved by Daniel Neuenschwander [1]. First, we improve the existing theorem by providing a more concise proof. Second, we provide error bounds for the approximation of $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ by $P_o(\lambda(q,m))$. Also, Theorems 2a, 2b, and 3 are new results.

Theorem 1

If Δx, Δy are both odd and q is even as q^m approaches infinity, the distribution of $\Lambda_\pi(\Delta x, \Delta y)$ is approximately Poisson with parameter $\frac{1}{2}$. More specifically, the error in the approximation of $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ by $P_o(\lambda(q,m))$ is

$\sup_{A\subseteq Z^+} |P(\Lambda_\pi(\Delta x, \Delta y) \in A) - \sum_{k\in A} \frac{[\lambda(q,m)]^k}{k!} e^{-\lambda(q,m)}| \leq \frac{1}{q^m - 1} + \frac{1}{q^m}$, where $\lambda(q,m) \approx \frac{1}{2}$.

Proof

Define $\Lambda_\pi(\Delta x, \Delta y)$ as the number of episodes of $\frac{q^m}{2}$ so that $x + x' = \Delta x$, $y + y' = \Delta y$, $y = \pi(x)$, and $y' = \pi(x')$. Also, $\Delta x =$ "CAT" and $\Delta y =$ "DOG" are fixed words. Note that q can be any even integer and m is arbitrary and the monikers CAT and DOG are used just as running generic examples. The exact maximum possible number of episodes is equal to $\phi(q,m)$ and in this case $\phi(q,m) = \frac{q^m}{2}$. Define $\Lambda_\pi(\Delta x, \Delta y) = \sum_{j=1}^{\phi(q,m)} I_j$ where $I_j = 1$ if in the jth possibility $x + x' = \Delta x$ yields $y + y' = \Delta y$ and $I_j = 0$ if $x + x' = \Delta x$ does not yield $y + y' = \Delta y$. By the Stein-Chen method, the error in the approximation of $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ by $P_o(\lambda(q,m))$ is less than or equal to $P(I_1 = 1) + \sum_{j\neq 1} P(I_j \neq J_j)$, where I_j is as above and J_j is defined below. Now take fixed $j = 1 = (x, x') = (AME, COP)$ for example. Note that AME + COP

= CAT. By $j = 1 = (x, x')$, we mean $j = 1 = (x, x')$ is the first decomposition of Δx into $x + x'$, AME + COP, for example.

If $\pi(COP) = y$ and $\pi(AME) = y'$ so that $\pi(COP) + \pi(AME) = \Delta y = DOG$, then we are fine and we set $I_j = J_j$. Consider when $\pi(COP) + \pi(AME) = DAD + MOM = POQ \neq DOG$. Since the images of COP and AME do not add up to DOG, we must use the coupling method. Let us fix $\pi(COP) = DAD$. We know that $\pi(COP)+\pi(x') = DOG$, so it follows that $\pi(x') = AOD$ since $DAD+AOD = DOG$ and there exists a word KFC such that $\pi(KFC) = AOD$. So now, by the coupling method, we swap and set $\pi'(KFC) = MOM$ and $\pi'(AME) = AOD$ where π' is the altered permutation. Now, the only j that can change is $(x, x') = j$ so that $KFC + SUR = CAT$. Let us consider $P(I_j = 1, J_j = 0)$. In this case, $\pi(SUR) = DAD$, but now this is impossible since $\pi(COP) = DAD$. Thus, $P(I_j = 1, J_j = 0) = 0$.

Let us consider $P(I_j = 0, J_j = 1)$. We have $\pi'(KFC)+\pi'(SUR) = DOG$ and thus $MOM + \pi'(SUR) = DOG$ implies $\pi(SUR) = RAU$, since $MOM + RAU = DOG$. So, in any situation, we have $P(I_j = 0, J_j = 1)$ if the condition $\pi(SUR) = RAU$ is satisfied. The general case, without the use of specific names, is as follows. Let $\pi(x) = y$, $\pi(x') = z$, and $\pi(w) = y'$. By the definition of coupling, $\pi'(x) = y$, $\pi'(x') = y'$, $\pi'(w) = z$. In any situation, we have $P(I_j = 0, J_j = 1)$ if the condition $\pi(w') = z'$ is satisfied.

Let us consider $E(\Lambda_\pi(\Delta x, \Delta y))$.

$$\begin{aligned} \lambda(q, m) = E(\Lambda_\pi(\Delta x, \Delta y)) &= \phi(q, m)E(I_1) \\ &= \phi(q, m)P(I_1 = 1) \end{aligned}$$

$$
\begin{aligned}
&= \phi(q,m)\phi(q,m)\cdot 2\cdot \frac{(q^m-2)!}{q^m!} \\
&= [\phi(q,m)]^2\cdot 2\cdot \frac{(q^m-2)!}{q^m!} \\
&= (\frac{q^m}{2})^2\cdot \frac{2}{q^m(q^m-1)} \\
&= \frac{q^m}{2(q^m-1)} \\
&\approx \frac{1}{2}
\end{aligned}
$$

Recall that the error in the approximation of $\mathcal{L}(\Lambda_\pi(\Delta x,\Delta y))$ by $P_o(\lambda(q,m))$ is less than or equal to $P(I_1=1)+\sum_{j\neq 1}P(I_j\neq J_j)$. If $P(I_1=1)=\frac{\phi(q,m)\cdot 2}{q^m(q^m-1)}$, what is $\sum_{j\neq 1}P(I_j\neq J_j)$? Well, $\sum_{j\neq 1}P(I_j\neq J_j)=0$ unless $j=(KFC,SUR)$ and if $j=(KFC,SUR)$ a necessary condition is that $\pi(SUR)=RAU$. Now in any situation where we have $P(I_j=0,J_j=1)$ if $\pi(w')=z'$, there are now $(q^m-1)!$ total permutations of q^m words. Thus, $\sum_{j\neq 1}P(I_j\neq J_j)=\frac{(q^m-1)!}{q^m!}$ and $P(I_1=1)+\sum_{j\neq 1}P(I_j\neq J_j)=\frac{\phi(q,m)\cdot 2}{q^m(q^m-1)}+\frac{(q^m-1)!}{q^m!}$. By simplification, we obtain $P(I_1=1)+\sum_{j\neq 1}P(I_j\neq J_j)=\frac{1}{q^m-1}+\frac{1}{q^m}\to 0$ as $m\to\infty$ for every q.□

Theorem 2a

If Δx is even and Δy is odd with $q \geq 4$, q is even as q^m approaches infinity and the distribution of $\Lambda_\pi(\Delta x, \Delta y)$ is approximately Poisson with parameter $\frac{1}{2}$. More specifically, the error in the approximation of $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ by $P_o(\lambda(q,m))$ is $sup_{A \subseteq Z^+} |P(\Lambda_\pi(\Delta x, \Delta y) \in A) - \sum_{k \in A} \frac{[\lambda(q,m)]^k}{k!} e^{-\lambda(q,m)}| \leq \frac{1}{q^m - 1} + \frac{1}{q^m}$, where $\lambda(q,m) \approx \frac{1}{2}$.

Proof

Define $\Lambda_\pi(\Delta x, \Delta y)$ as $\Lambda_\pi(\Delta x, \Delta y) = \Lambda_1 + \Lambda_2$ where $\Lambda_1 = \sum_{j=1}^{\frac{q^m - 2^m}{2}} I_j$ and $\Lambda_2 = \sum_{k=1}^{2^k} I_k$. Note that $I_j = 1$ if in the jth possibility $x + x' = \Delta x$ yields $y + y' = \Delta y$ and $I_j = 0$ if $x + x' = \Delta x$ does not yield $y + y' = \Delta y$. Also, $I_k = 1$ if in the kth possibility $x + x = \Delta x$ yields $2\pi(x) = \Delta y$ and $I_k = 0$ if $x + x = \Delta x$ does not yield $2\pi(x) = \Delta y$. But, since it is impossible to have $\pi(x) + \pi(x) = \Delta y$ for Δy, q odd, it follows that $I_k = 0$. Therefore, $\Lambda_\pi(\Delta x, \Delta y) = \sum_{j=1}^{\frac{q^m - 2^m}{2}}$. By the Stein-Chen method, the error in the approximation of $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ by $P_o(\lambda(q,m))$ is less than or equal to $P(I_1 = 1) + \sum_{j \neq 1} P(I_j \neq J_j)$, where I_j is as above and J_j is defined below. Let us take fixed $j = 1 = (x, x') = (AME, COP)$ for example.

If $\pi(COP) = y$ and $\pi(AME) = y'$ so that $\pi(COP) + \pi(AME) = \Delta y = DOG$, then we are fine and we set $I_j = J_j$. Consider when $\pi(COP) + \pi(AME) = DAD + MOM = POQ \neq DOG$. Since the images of COP and AME do not add up to DOG, we must use the coupling method. Let us fix $\pi(COP) = DAD$. We know that $\pi(COP) + \pi(x') = DOG$, so it follows that $\pi(x') = AOD$ since $DAD + AOD = DOG$ and there exists a word KFC such that $\pi(KFC) = AOD$. So, by the coupling method, we swap and set $\pi'(KFC) = MOM$ and $\pi'(AME) = AOD$. The only j that can

change is $(x, x') = j$ so that $KFC + SUR = CAT$. Let us consider $P(I_j = 1, J_j = 0)$. Well, $\pi(SUR) = DAD$, but now this is impossible since $\pi(COP) = DAD$. Thus, $P(I_j = 1, J_j = 0) = 0$.

Let us consider $P(I_j = 0, J_j = 1)$. We have that $\pi'(KFC) + \pi'(SUR) = DOG$ and thus $MOM + \pi'(SUR) = DOG$ implies $\pi(SUR) = RAU$, since $MOM + RAU = DOG$. So, in any situation, we have $P(I_j = 0, J_j = 1)$ if the condition $\pi(SUR) = RAU$ is satisfied. Let us consider the general case, where $\pi(x) = y$, $\pi(x') = z$, and $\pi(w) = y'$. By the definition of coupling, $\pi'(x) = y$, $\pi'(x') = y'$, $\pi'(w) = z$. In any situation, we have $P(I_j = 0, J_j = 1)$ if the condition $\pi(w') = z'$ is satisfied.

Let us consider $E(\Lambda_\pi(\Delta x, \Delta y))$.

$$\begin{aligned}
\lambda(q, m) = E(\Lambda_\pi(\Delta x, \Delta y)) &= \phi(q, m) E(I_1) \\
&= (\frac{q^m - 2^m}{2}) E(I_1) \\
&= (\frac{q^m - 2^m}{2})(\frac{q^m}{2}) \frac{2(q^m - 2)!}{q^m!} \\
&= \frac{q^m - 2^m}{q^m - 1} \cdot \frac{1}{2} \\
&\approx \frac{1}{2}
\end{aligned}$$

provided that $q \geq 3$.

Recall that the error in the approximation of $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ by $P_o(\lambda(q,m))$ is less than or equal to $P(I_1 = 1) + \sum_{j \neq 1} P(I_j \neq J_j)$. Well, $P(I_1 = 1) = \frac{1}{q^m - 1}$, but what is $\sum_{j \neq 1} P(I_j \neq J_j)$? The $\sum_{j \neq 1} P(I_j \neq J_j) = 0$ unless $j = (KFC, SUR)$ and, if $j = (KFC, SUR)$, a necessary condition is that $\pi(SUR) = RAU$. In any situation, we have $P(I_j = 0, J_j = 1)$ if $\pi(w') = z'$ and there are now $(q^m - 1)!$ total permutations of q^m words. Thus, $\sum_{j \neq 1} P(I_j \neq J_j) = \frac{(q^m - 1)!}{q^m!}$ and
$P(I_1 = 1) + \sum_{j \neq 1} P(I_j \neq J_j) = \frac{1}{q^m - 1} + \frac{(q^m - 1)!}{q^m!}$. By simplification, we obtain $P(I_1 = 1) + \sum_{j \neq 1} P(I_j \neq J_j) = \frac{1}{q^m - 1} + \frac{1}{q^m} \to 0$ as $m \to \infty$ for every q.□

Theorem 2b

If Δx, Δy, are both even words and q is odd as q^m approaches infinity the distribution of $\Lambda_\pi(\Delta x, \Delta y)$is approximately Poisson with parameter $\frac{1}{2}$. More specifically, the total variation distance, d_{TV}, between $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ and $P_o(\lambda(q,m))$ is $d_{TV} \leq \frac{2}{q^m} + \frac{4}{q^m(q^m-1)}$, where $\lambda(q,m) = 1 \cdot \frac{1}{q^m} + \frac{q^m-1}{2} \cdot \frac{q^m-1}{2} \cdot 2 \cdot \frac{1}{q^m(q^m-1)} = \frac{1}{2}\frac{q^m+1}{q^m}$.

Proof

Define $\Lambda_\pi(\Delta x, \Delta y)$ as the number of episodes of $\frac{q^m-1}{2}$ so that $x + x' = \Delta x$, $y + y' = \Delta y$, $y = \pi(x)$, and $y' = \pi(x')$ or the one episode in which $x + x = \Delta x$, $y + y = \Delta y$, $y = \pi(x)$. The exact maximum possible number of episodes is equal to $\frac{q^m-1}{2} + 1$. Define $\Lambda_\pi(\Delta x, \Delta y) = \sum_{j=1}^{\frac{q^m-1}{2}} I_j + I_k$ where $I_j = 1$ if in the jth possibility $x + x' = \Delta x$ yields $y + y' = \Delta y$ with $y = \pi(x)$ and $y' = \pi(x')$ and $I_j = 0$ otherwise. Note that $I_k = 1$ if $x+x = \Delta x$, $y+y = \Delta y$, and $y = \pi(x)$. By the Stein-Chen method, the total variation distance, d_{TV}, between $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ and $P_o(\lambda(q,m))$ is less than or equal to $\frac{1}{\lambda(q,m)} \sum_{j=1}[P^2(I_j = 1) + P(I_j = 1) \sum_{i \neq j} P(I_j \neq J_i)]$, where I_j is as above and J_j is defined below. Take fixed $j = (x, x) = (MAE, MAE)$ for example, so that $\Delta x = YAI$. Let $\Delta y = WAS$ and $y = LAJ$ so $\Delta y = y + y$. If $\pi(MAE) = LAJ$, then we are fine and we set $I_j = J_j$. Consider when $\pi(MAE)+\pi(MAE) = WIZ+WIZ = SQY$. Since the images of MAE do not add up to WAS, we must use the coupling method. Now there exists MOP such that $\pi(MOP) = LAJ$. So, now by the coupling method, we swap and set $\pi'(MOP) = WIZ$ and $\pi'(MAE) = LAJ$. Now the only j that can change is $(x, x') = j$ so that $MOP + MMT = YAI$.

Let us consider $P(I_j = 1, J_j = 0)$. Given, $\pi(MMT) = ABC$, and $\pi(MOP) + \pi(MMT) = LAJ + ABC \neq WAS$. Thus, $I_j = 1$ is impossible and so $P(I_j =$

$1, J_j = 0) = 0$. Let us consider $P(I_j = 0, J_j = 1)$. We have that $\pi'(MOP) + \pi'(MMT) = WAS$ and thus $WIZ + \pi'(MMT) = WAS$ implies $\pi(MMT) = AST$ since $WIZ + AST = WAS$. So, in any situation, we have $P(I_j = 0, J_j = 1)$ if the condition $\pi(MMT) = AST$ is satisfied.

Now, let us consider the general case, where $\pi(x) = y$, $\pi(x') = z$, and $\pi(w) = y'$. By the definition of coupling, $\pi'(x) = y$, $\pi(x') = y'$, and $\pi'(w) = z$. In any situation, we have that $P(I_j = 0, J_j = 1)$ if the condition $\pi(w') = z'$ is satisfied.

There are other cases, for example, when $j = (x, x') = (AAA, YAI)$. If $I_j = 1$ we are fine, but let us consider the possibility that $I_j = 0$. Suppose that $\pi(AAA) = AAA$ and $\pi(YAI) = BBB$. Fix $\pi(AAA) = AAA$. There exists BBB such that $\pi(BBB) = YAI$. We swap and set $\pi'(BBB) = BBB$. The only j that can change is $j = (BBB, XZH)$ and $P(I_j = 0, J_j = 1)$ if $\pi(XZH) = VZR$. In any case, $\sum_{j \neq 1} P(I_j \neq J_j) = \frac{1}{q^m}$.

Let us consider $E(\Lambda_\pi(\Delta x, \Delta y))$.

$$\begin{aligned}
\lambda(q, m) = E(\Lambda_\pi(\Delta x, \Delta y)) &= \sum_{\frac{q^m - 1}{2}} P(I_j = 1) \\
&= \sum_{j=1} P(I_j = 1) + P(I_k = 1) \\
&= \frac{q^m - 1}{2} P(I_1 = 1) + P(I_k = 1) \\
&= \left(\frac{q^m - 1}{2}\right)^2 \cdot 2 \cdot \frac{(q^m - 2)!}{q^m!} + \frac{1}{q^m} \\
&= \frac{q^m - 1}{q^m} \cdot \frac{1}{2} + \frac{1}{q^m} \\
&\approx \frac{1}{2}
\end{aligned}$$

Consider the quantity

$\sum_{j=1}[P^2(I_j = 1) + P(I_j = 1)\sum_{i\neq j} P(I_i \neq J_i)](1)$. We know that (1) is equal to

$\sum_{j=1}^{\frac{q^m-1}{2}}[P^2(I_j = 1) + P(I_j = 1)\sum_{i\neq j} P(I_i \neq J_i)] + [P^2(I_k = 1) + P(I_k = 1)\sum \ldots](2)$.

Now, (2) equals

$\frac{q^m-1}{2}[\frac{(q^m-1)!^2}{(q^m)!} + \frac{(q^m-1)!}{q^m!}\frac{1}{q^m}] + \frac{1}{q^{2m}} + \frac{1}{q^m}\frac{1}{q^m}$ which equals

$\frac{q^m-1}{2}[\frac{1}{q^{2m}} + \frac{1}{q^{2m}}] + \frac{2}{q^{2m}} = \frac{q^m-1}{2}[\frac{2}{q^{2m}}] + \frac{2}{q^{2m}}(3)$.

The total variation distance, d_{TV}, between $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ and $P_o(\lambda(q,m))$ is symbolized $d_{TV}(\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y), P_o(\lambda(q,m))$. It is true that

$d_{TV}(\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y), P_o(\lambda(q,m)) \leq \frac{1}{\lambda(q,m)}\sum_{j=1}[P^2(I_j = 1) + P(I_j = 1)\sum_{i\neq j} P(I_i \neq J_i)](4)$. Now if we substitute the lower bound for $\lambda(q,m)$ and (3) into expression (4), we obtain

$[\frac{q^m-1}{2}\frac{2}{q^{2m}} + \frac{2}{q^{2m}}][\frac{1}{2}\frac{q^m-1}{q^m}]^{-1}$ (5). Thus, (5) simplifies to $\frac{\frac{q^m-1}{q^{2m}} + \frac{2}{q^{2m}}}{\frac{1}{2}\frac{q^m-1}{q^m}} = [\frac{q^m-1}{q^{2m}} + \frac{2}{q^{2m}}]\cdot 2\frac{q^m}{q^m-1} = \frac{2}{q^m} + \frac{4}{q^m(q^m-1)}$. Notice that $\frac{2}{q^m} + \frac{4}{q^m(q^m-1)} \to 0$ as $m \to \infty$ for every q.□

We could consider "Theorem 2c" where Δx is odd and both Δy and q are even. However, this case is handled as in Theorem 2b and is omitted.

Before we prove Theorem 3, it is necessary to prove two Lemmas.

Lemma 1

It is true that $d_{TV}(X,Y) \leq P(X \neq Y)$.

Proof

Recall $d_{TV}(X,Y) = sup_{A \subseteq \mathbf{Z}^+} |P(X \in A) - P(Y \in A)|$. Let A be any subset of $\mathbf{Z}^+$. Now, $P(X \in A) = P(X \in A \cap Y \in A) + P(X \in A, Y \notin A)) \leq P(Y \in A) + P(X \in A, Y \notin A)$. But, $X \in A$, $Y \notin A$ implies that $X \neq Y$. Thus, $P(X \in A, Y \notin A) \leq P(X \neq Y)$ and $P(X \in A) \leq P(Y \in A) + P(X \neq Y)$. Now, $P(X \in A) \leq P(Y \in A) + P(X \neq Y)$ implies that $P(X \in A) - P(Y \in A) \leq P(X \neq Y)$. Similarly, starting with Y, we have that $P(Y \in A) - P(X \in A) \leq P(X \neq Y)$, so $|P(X \in A) - P(Y \in A)| \leq P(X \neq Y)$ for every A. Hence, $sup_{A \subseteq \mathbf{Z}^+} |P(X \in A) - P(Y \in A)| \leq P(X \neq Y)$.□

Lemma 2 (Markov's Inequality)

If X is a discrete non-negative random variable, $P(X \geq a) \leq \frac{E(X)}{a}$. Specifically, if $a = 1$, we have $P(X \geq 1) \leq E(X)$.

Proof

Well, $E(X) = \sum_{x=0}^{\infty} xP(X = x) \geq \sum_{x=a}^{\infty} xP(X = x) \geq a \sum_{x=a}^{\infty} P(X = x) = aP(X \geq a)$. If $E(X) \geq aP(X \geq a)$ it follows that $\frac{E(X)}{a} \geq P(X \geq a)$. So, specifically if $a = 1$, we have that $E(X) \geq P(X \geq 1)$.□

Theorem 3

If $q \geq 6$, q is even and both Δx,Δy are even, as q^m approaches infinity, the distribution of $\Lambda_\pi(\Delta x, \Delta y)$ is approximately Poisson with parameter $\frac{1}{2}$. More specifically, the total variation distance, d_{TV} between $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ and $P_o(\lambda_1(q,m))$ is $d_{TV} \leq \frac{(q^m-2^m)}{q^m(q^m-1)} + \frac{2}{q^m} + (\frac{4}{q})^m$, where $\lambda_1(q,m) = \frac{(q^m-2^m)}{2}\frac{(q^m-2^m)}{2} \cdot 2 \cdot \frac{1}{q^m(q^m-1)} \approx \frac{1}{2}$.

Proof

Worth mentioning is the fact that an exact analysis, as done in Theorem 2, would be very complex. If we were to give an exact treatment as in Theorem 2, we would have $\Lambda_\pi(\Delta x, \Delta y) = \sum_{j=1}^{\frac{q^m-2^m}{2}} I_j + \sum_{i=1}^{2^m} I_j$. That is, we have $\Lambda_\pi(\Delta x, \Delta y) = \Lambda_1 + \Lambda_2$ where $\Lambda_1 = \sum_{j=1}^{\frac{q^m-2^m}{2}} I_j$ and $\Lambda_2 = \sum_{i=1}^{2^m} I_i$. We say that $I_j = 1$ if in the jth possibility $x_1 + x_2 = CAT$ yields $\pi(x_1) + \pi(x_2) = DOG$ and $I_j = 0$ otherwise. Also, $I_i = 1$ if $x + x = CAT$ yields $2\pi(x) = DOG$ and $I_j = 0$ otherwise. We are summing up to 2^m because there are 2^m unordered ways of writing $\Delta x = x + x$. For the case when Δx andΔy are even, we need to show that d_{TV} is small. To show that d_{TV} is small, we will use the triangle inequality. By the triangle inequality, we have that $d_{TV}(\Lambda, P_o(\lambda_1)) \leq d_{TV}(\Lambda_1, P_o(\lambda_1)) + d_{TV}(\Lambda_1, \Lambda)$. By Lemma 1, it is true that $d_{TV}(\Lambda, \Lambda_1) \leq P(\Lambda \neq \Lambda_1) = P(\Lambda_2 \geq 1)$. Now, by Lemma2, Markov's Inequality, we have that $P(\Lambda_2 \geq 1) \leq E(\Lambda_2)$.

Thus,

$$\begin{aligned} E(\Lambda_2) &= E(\sum_{1}^{2^m} I_j) \\ &= 2^m P(I_1 = 1) \\ &= 2^m \cdot 2^m \cdot \frac{1}{q^m} \\ &= \frac{4^m}{q^m} \\ &= (\frac{4}{q})^m \\ &= \lambda_2(q, m) \end{aligned}$$

It is a fact that $(\frac{4}{q})^m \to 0$ as $m \to \infty$ $\forall q \geq 5$. Recall that $d_{TV}(\Lambda, P_o(\lambda_1)) \leq d_{TV}(\Lambda_1, P_o(\lambda_1)) + d_{TV}(\Lambda_1, \Lambda)$. Now, $d_{TV}(\Lambda_1, P_o(\lambda_1)) + d_{TV}(\Lambda_1, \Lambda) \leq d(\Lambda_1, P_o(\lambda_1)) + E(\Lambda_2) \leq P(I_1 = 1) + \sum_{j=2} P(I_j \neq J_j) + E(\Lambda_2)$. We know that $P(I_1 = 1) = \frac{q^m - 2^m}{2} \cdot 2\frac{1}{(q^m)(q^m-1)}$. Now, we need to determine $\sum_{j=2} P(I_j \neq J_j)$. Before considering any cases, we need to make some declarations. Let the first fixed words BAT and RAT have the characteristic that BAT + RAT = CAT with $BAT = x_1$ and $RAT = x_2$. Note that we are no longer worrying about the modular arithmetic being "correct." Let COW be a generic word such that $COW + COW^c = DOG$. In addition, let DOE and OWL be words so that DOE + DOE = DOG and OWL + OWL = DOG. Consider $\pi(BAT) = COW^c$ where in COW^c the c represents the complement of COW i.e., the word which must be added to COW to obtain DOG. Also, $\pi(ELK) = COW$ and $\pi(RAT) = FOX$, and $\pi(PIG) = FOX^c$. Using the coupling method, we have $\pi'(BAT) = COW^c$, $\pi'(RAT) = COW$, and $\pi'(ELK) = FOX$. So, the only j that can change $j = (PIG, ELK)$ so that PIG + ELK = CAT. Let us clarify the preceding statement. By saying the only j that can change is $j = (PIG, ELK)$ we mean only

the sum of the images of PIG and ELK equal DOG. Therefore, $\sum_{j=2} P(I_j \neq J_j) = \frac{1}{q^m}$ and $d_{TV} \leq P(I_1 = 1) + \sum_{j=2} P(I_j \neq J_j) + E(\Lambda_2) \leq \frac{q^m - 2^m}{2} \cdot 2 \cdot \frac{1}{(q^m)(q^m-1)} + \frac{1}{q^m} + \frac{4^m}{q^m}$. Note that $\frac{q^m - 2^m}{2} \cdot 2 \cdot \frac{1}{(q^m)(q^m-1)} + \frac{1}{q^m} + \frac{4^m}{q^m} \to 0$ as $m \to \infty \ \forall q \geq 2$.

Let us consider case two where $\pi(BAT) = COW$ and $\pi(RAT) = DOE$ i.e., $x_1 = BAT$ and $x_2 = RAT$ are fixed words and $y_1 = COW$ is a word such that $COW + COW^c = DOG$. Also, $y_2 = DOE$ is a word such that DOE + DOE = DOG. We have $\pi(BAT) = COW$, $\pi(RAT) = DOE$, $\pi(ELK) = COW^c$. Applying the coupling method, we swap and set $\pi'(BAT) = COW$, $\pi'(RAT) = COW^c$, $\pi'(ELK) = DOE$, and $\pi(PIG) = OWL$. Now, $I_j = 1$ is impossible and $I_j = 0$ implies $J_j = 0$. The only j that can change is $j = (ELK, PIG)$. However, $P(I_j = 1) = 0$ and $I_j = 0$ implies that $J_j = 0$ and $P(I_j \neq J_j) = 0$ and $\sum_{j=2} P(I_j \neq J_j) = 0$. So, for d_{TV}, we have that $d_{TV} \leq P(I_1) = \frac{q^m - 2^m}{2} \cdot 2 \cdot \frac{1}{(q^m)(q^m-1)}$. In our last case to consider, $\pi(BAT) = DOE$, $\pi(ELK) = COW$, $\pi(RAT) = OWL$, and $\pi(FOX) = COW^c$. Let us now employ the coupling method. We swap and set $\pi'(BAT) = COW$, $\pi'(RAT) = COW^c$, and $\pi'(ELK) = DOE$, $\pi'(FOX) = OWL$. In this scenario, we have two $j's$ that can change. The first j is $j = (FOX, ELK)$, so that FOX + ELK = CAT i.e., if $ELK = FOX^c$ the second j is the same but we could have started with $\pi(ELK) = COW^c$ and $\pi(FOX) = COW$. Hence $\sum_{j \neq 1} P(I_j \neq J_j) = \frac{2}{q^m}$ and $d_{TV} \leq \frac{(q^m - 2^m)}{2} \cdot 2 \cdot \frac{1}{q^m(q^m-1)} + (\frac{4}{q})^m + \frac{2}{q^m}$. The largest of the three cases is the above one, leading to the result.

2.2 The Stein-Chen Method for the Decomposition of Three Words

We would like to prove the following:

If Δx and Δy are not multiples of three and $q \equiv 0(mod 3)$, as q^m approaches infinity the distribution of $\Lambda_\pi(\Delta x, \Delta y)$ is approximately Poisson.

Attempt at Proof

Let $\Delta x = EDC$ and $\Delta y = HAE$ be fixed words. Define $\Lambda_\pi(\Delta x, \Delta y)$ as $\Lambda_\pi(\Delta x, \Delta y) = \Lambda_1 + \Lambda_2$ where $\Lambda_1 = \sum_{j=1}^{\frac{q^m - 3q^m}{6}} I_j$ and $\Lambda_2 = \sum_{i=1}^{q^m} I_i$. Note that $I_j = 1$ if in the jth possibility $x + x' + x'' = \Delta x$ yields $y + y' + y'' = \Delta y$ and $I_j = 0$ if $x + x' + x'' = \Delta x$ does not yield $y + y' + y'' = \Delta y$. Also, $I_i = 1$ if in the ith possibility $x + x + y = \Delta x$ yields $\pi(x) + \pi(x) + \pi(y) = \Delta y$ and $I_i = 0$ if $x + x + y = \Delta x$ does not yield $\pi(x) + \pi(x) + \pi(y) = \Delta y$. Now, let us consider $\lambda(q, m)$, $P(I_j = 1)$, and $P(I_i = 1)$. Well, $\lambda(q, m) = (\frac{q^{2m} - 3q^m}{6})(\frac{q^{2m} - 3q^m}{6}) \cdot 6 \cdot \frac{(q^m - 3)!}{q^m!} + q^m \cdot q^m \cdot \frac{(q^m - 2)!}{q^m!} \approx \frac{q^m}{6} + 1 \approx \frac{q^m}{6}$. Also, $P(I_j = 1) = (\frac{q^{2m} - 3q^m}{6}) \cdot 6 \cdot \frac{(q^m - 3)!}{q^m!} \approx \frac{1}{q^m}$. Now, $P(I_i = 1) = \frac{(q^m)(q^m - 2)!}{q^m!} \approx \frac{1}{q^m}$. Let us now consider the approximate value of $\Lambda_\pi(\Delta x, \Delta y)$ where
$\Lambda_\pi(\Delta x, \Delta y) = \sum_{j=1}^{\frac{q^m - 3q^m}{6}} I_j + \sum_{i=1}^{q^m} I_i \approx \sum_{j=1}^{\frac{q^{2m}}{6}} I_j + \sum_{i=1}^{q^m} I_i$. The total variation distance, d_{TV}, between $\mathcal{L}(\Lambda_\pi(\Delta x, \Delta y))$ and $P_o(\lambda(q, m))$ is less than or equal to
$\frac{1}{\lambda(q,m)}[\sum_{k=1}^{\frac{q^{2m}}{6} + q^m} P^2(I_k = 1) + P(I_k = 1) \sum_{l \neq k} P(I_l \neq J_l)]$ which equals
$\frac{6}{q^m}[\frac{q^{2m}}{6}[\frac{1}{q^{2m}} + \frac{1}{q^m} \sum_{l \neq k} P(I_l \neq J_l)] + q^m[\frac{1}{q^{2m}} + \frac{1}{q^m} \sum_{l \neq k} P^*(I_l \neq J_l)]$ (1). Note that $\sum_{l \neq k} P(I_l \neq J_l)$ represents the sum of the probabilities with respect to

$\Delta x = x + y + z$, while $\sum_{l \neq k} P^*(I_l \neq J_l)$ represents the sum of the probabilities with respect to $\Delta x = x + x + y$. After simplifying (1), we obtain:

$\frac{1}{q^m} + \frac{6}{q^m} \cdot \frac{q^{2m}}{6} \cdot \frac{1}{q^m} \sum_{l \neq k} P(I_l \neq J_l) + \frac{6}{q^{2m}} + \frac{6}{q^m} \sum_{l \neq k} P^*(I_l \neq J_l)$ which equals

$\frac{1}{q^m} + \sum_{l \neq k} P(I_l \neq J_l) + \frac{6}{q^{2m}} + \frac{6}{q^m} \sum_{l \neq k} P^*(I_l \neq J_l)$. Let us consider $\sum_{l \neq k} P(I_l \neq J_l)$ and take fixed $l = (x, x', x'') = (x, y, z)$ so that $x + y + z = EDC$ for example. If $\pi(x) = u$, $\pi(y) = v$, and $\pi(z) = w$ so that $\pi(x) + \pi(y) + \pi(z) = \Delta y = HAE$ then we are fine and we set $I_l = J_l$. Consider when $\pi(x) + \pi(y) + \pi(z) = d + e + f \neq \Delta y$. Since the images of x, y, and z do not add up to Δy, we must use the coupling method. We fix $\pi(x) = d$, $\pi(y) = e$ and swap so that $\pi'(z) = HAE - d - e$ and $\pi'(z') = f$ where $\pi(z') = HAE - d - e$. Now, a class of l's that can change is (x, x', x'') so that $z' + a + b = \Delta x = EDC$. There are q^m possibilities for $z' + a + b$ Note that z' is fixed, $\pi(a)$ is arbitrary, and b is determined. It is true that $\pi(z') = HAE - d - e$ and $\pi(a)$ is an arbitrary word. If the condition $\pi(b) = d + e - \pi(a)$ is satisfied, we have $I_l = 1$ and thus $J_l = 0$ automatically. Therefore, $\sum P(I_l = 1, J_l = 0) = q^m \cdot \frac{1}{q^m} = 1$ which is not approximately zero as we require. Hence, $\sum P(I_l \neq J_l) \geq 1$ Thus, the Stein-Chen method of Poisson approximation is not applicable for the decomposition of three words.

Although the Stein-Chen method of Poisson approximation fails for the decomposition of three words, a commentary is needed. There are six cases to consider.

Case one

Δx, Δy are both multiples of three and $q \equiv 0(mod 3)$

Case two

Δx is a multiple of three, Δy is not a multiple of three and $q \equiv 0(mod 3)$

Case three

Δx is not a multiple of three, Δy is a multiple of three and $q \equiv 0(mod 3)$

Case four

Neither Δx nor Δy is a multiple of three and $q \equiv 0(mod 3)$. Note that this case was examined in the previous "theorem."

Case five

Δx, Δy are both multiples of three and $q \equiv 1(mod 3)$

Case six

Δx, Δy are both multiples of three and $q \equiv 2(mod 3)$

Let us calculate $\Lambda_\pi(\Delta x, \Delta y)$, $\lambda(q,m)$, and π_j, π_k, and π_l for each case.

For case one,

$\Lambda_\pi(\Delta x, \Delta y) = \sum_{i=1}^{\frac{q^{2m}-3(q^m-3^m)-3^m}{6}} I_i + \sum_{j=1}^{q^m-3^m} I_j + \sum_{k=1}^{3^m} I_k$ where $I_i = 1$ if in the ith possibility $x+x'+x'' = \Delta x$ yields $y+y'+y'' = \Delta y$ and $I_j = 1$ if in the jth possibility $x + x + x' = \Delta x$ yields $2\pi(x) + \pi(x') = \Delta y$ and $I_k = 1$ if in the kth possibility $x + x + x = \Delta x$ yields $3\pi(x) = \Delta y$

$\lambda(q,m) = (\frac{q^{2m}-3(q^m-3^m)-3^m}{6})(\frac{q^{2m}-3(q^m-3^m)-3^m}{6})6\frac{(q^m-3)!}{q^m!} + (q^m-3^m)(q^m-3^m)\frac{(q^m-2)!}{q^m!} + 3^m 3^m \frac{(q^m-1)!}{q^m!}$.

$\pi_j = (\frac{q^{2m}-3(q^m-3^m)-3^m}{6})6\frac{(q^m-3)!}{q^m!}$

$\pi_k = (q^m-3^m)\frac{(q^m-2)!}{q^m!}$

$\pi_l = 3^m\frac{(q^m-1)!}{q^m!}$.

For case two, we have

$\Lambda_\pi(\Delta x, \Delta y) = \sum_{n=1}^{\frac{q^{2m}-3(q^m-3^m)-3^m}{6}} I_n + \sum_{j=1}^{q^m-3^m} I_j$ where $I_n = 1$ if in the nth possibility $x + x' + x'' = \Delta x$ yields $y + y' + y'' = \Delta y$ and $I_j = 1$ if in the jth possibility $x + x + x' = \Delta x$ yields $2\pi(x) + \pi(x') = \Delta y$

$\lambda(q,m) = (\frac{q^{2m}-3(q^m-3^m)-3^m}{6})(\frac{q^{2m}-3q^m}{6}) \cdot 6 \cdot \frac{(q^m-3)!}{q^m!} + (q^m-3^m)(q^m)\frac{(q^m-2)!}{q^m!}$

$\pi_j = (\frac{q^{2m}-3q^m}{6}) \cdot 6 \cdot \frac{(q^m-3)!}{q^m!}$

$\pi_l = (q^m)\frac{(q^m-2)!}{q^m!}$

For case three, we have

$\Lambda_\pi(\Delta x, \Delta y) = \sum_{n=1}^{\frac{q^{2m}-3q^m}{6}} I_n + \sum_{p=1}^{q^m} I_p$ where $I_n = 1$ if in the nth possibility $x + x' + x'' = \Delta x$ yields $y + y' + y'' = \Delta y$ and $I_p = 1$ if in the pth possibility $x + x + x' = \Delta x$ yields $2\pi(x) + \pi(x') = \Delta y$

$\lambda(q,m) = (\frac{q^{2m}-3q^m}{6})(\frac{q^{2m}-3(q^m-3^m)-3^m}{6}) \cdot 6 \cdot \frac{(q^m-3)!}{q^m!} + q^m(q^m-3^m)\frac{(q^m-2)!}{q^m!}$

$\pi_j = (\frac{q^{2m}-3(q^m-3^m)-3^m}{6}) \cdot 6 \cdot \frac{(q^m-3)!}{q^m!}$

$\pi_k = (q^m - 3^m)\frac{(q^m-2)!}{q^m!}$.

Case four was discussed in the previous "theorem."

For case five, we have

$\Lambda_\pi(\Delta x, \Delta y) = \sum_{i=1}^{\frac{q^{2m}-3(q^m-3^m)-3^m}{6}} I_i + \sum_{j=1}^{q^m-3^m} I_j + \sum_{k=1}^{3^m} I_k$ where $I_i = 1$ if in the ith possibility $x+x'+x'' = \Delta x$ yields $y+y'+y'' = \Delta y$ and $I_j = 1$ if in the jth possibility $x + x + x' = \Delta x$ yields $2\pi(x) + \pi(x') = \Delta y$ and $I_k = 1$ if in the kth possibility $x + x + x = \Delta x$ yields $3\pi(x) = \Delta y$

$\lambda(q, m) = (\frac{q^{2m}-3(q^m-3^m)-3^m}{6})(\frac{q^{2m}-3(q^m-3^m)-3^m}{6})6\frac{(q^m-3)!}{q^m!} + (q^m - 3^m)(q^m - 3^m)\frac{(q^m-2)!}{q^m!} + 3^m 3^m \frac{(q^m-1)!}{q^m!}$.

$\pi_j = (\frac{q^{2m}-3(q^m-3^m)-3^m}{6})6\frac{(q^m-3)!}{q^m!}$

$\pi_k = (q^m - 3^m)\frac{(q^m-2)!}{q^m!}$

$\pi_l = 3^m \frac{(q^m-1)!}{q^m!}$.

For our last case, we have

$\Lambda_\pi(\Delta x, \Delta y) = \sum_{i=1}^{\frac{q^{2m}-3(q^m-3^m)-3^m}{6}} I_i + \sum_{j=1}^{q^m-3^m} I_j + \sum_{k=1}^{3^m} I_k$ where $I_i = 1$ if in the ith possibility $x+x'+x'' = \Delta x$ yields $y+y'+y'' = \Delta y$ and $I_j = 1$ if in the jth possibility $x + x + x' = \Delta x$ yields $2\pi(x) + \pi(x') = \Delta y$ and $I_k = 1$ if in the kth possibility $x + x + x = \Delta x$ yields $3\pi(x) = \Delta y$

$\lambda(q, m) = (\frac{q^{2m}-3(q^m-3^m)-3^m}{6})(\frac{q^{2m}-3(q^m-3^m)-3^m}{6})6\frac{(q^m-3)!}{q^m!} + (q^m - 3^m)(q^m - 3^m)\frac{(q^m-2)!}{q^m!} + 3^m 3^m \frac{(q^m-1)!}{q^m!}$.

$\pi_j = (\frac{q^{2m}-3(q^m-3^m)-3^m}{6})6\frac{(q^m-3)!}{q^m!}$

$\pi_k = (q^m - 3^m)\frac{(q^m-2)!}{q^m!}$

$\pi_l = 3^m \frac{(q^m-1)!}{q^m!}$.

3 CONCLUSION

While we have proved a large number of results showing $\Lambda_\pi(\Delta x, \Delta y)$ is approximately Poisson for $\Delta x = x + x'$ and $\Delta y = y + y'$, we have determined that $\Lambda_\pi(\Delta x, \Delta y)$ is not approximately Poisson for $\Delta x = x + y + z$. However, in the future, one may want to investigate if $\Lambda_\pi(\Delta x, \Delta y)$ is approximately Normal. There are three steps to consider in the Normal Approximation case:

Step 1: Calculate the Variance of $\Lambda_\pi(\Delta x, \Delta y)$, denoted $Var(\Lambda_\pi(\Delta x, \Delta y))$.

Step 2: Determine the dependence structure. That is, I_j is dependent on how many I_i's?

Step 3: Use Stein's Method for the Normal Distribution.

BIBLIOGRAPHY

[1] D. Neuenschwander, A Limit Theorem in Cryptology, *Limit Theorems in Probability and Statistics II*, **2** (1999), 437-442.

[2] F.S. Roberts, *Applied Combinatorics*, Prentice-Hall, Englewood Cliffs, 1984.

[3] D.R. Stinson, *Cryptography: Theory and Practice*, 2nd ed, Chapman and Hall/CRC, Boca Raton, 2002.

VITA

KEVIN R. LYNCH

Education

- Bachelor of Science Degree in Education, Tennessee Tech University, May 1998.
- Master of Science Degree in Mathematics, East Tennessee State University, August 2005.

Professional Experience

- Graduate Assistant/Teaching Assistant

  East Tennessee State University, Johnson City, TN, 2000-2002.