The Genetic Code and the I Ching

Jan Krämer

Theoretical Computer Science Dept. of Computer Science
RWTH Aachen University

Helping Donald Knuth Seminar 2005

Outline

Outline

I Ching, the Bible of Changes

I Ching

- One of the five classics of confucian wisdom
- Contains 64 chapters
- Each chapter is marked by an hexagram (for example ䷡)

I Ching, the Bible of Changes

King Wen's ordering

- The original ordering of the chapters in the *I Ching*
- This implies a certain ordering of the hexagrams
- (The *I Ching* has been ascribed to King Wen c. 1100 B.C.)

Example

1: ䷀, 2: ䷁, 3: ䷂, 4: ䷃, . . ., 63: ䷾, 64: ䷿

DNA and Proteins

DNA for Computer Scientists

- DNA sequence strings over 4 letter alphabet $\{T, C, A, G\}$ (nucleotides)
- Tripplet of nucleotides form a codon
- Each codon maps to an amino acid $\{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y\}$
- Proteins are strings of amino acids

Mapping DNA to proteins

Algorithm

- Chop DNA sequence into codons (x, y, z)
- Map $T \rightarrow 1, C \rightarrow 2, A \rightarrow 3, G \rightarrow 4$
- Lookup amino acid in matrices below
  $(x, y, z) \rightarrow$ matrix x, row y, collum z
- Keep going until you hit seperator '–'

Amino Acid Lookup Matrix

$$\begin{pmatrix} F & S & Y & C \\ F & S & Y & C \\ L & S & - & - \\ L & S & - & W \end{pmatrix} \begin{pmatrix} L & P & H & R \\ L & P & H & R \\ L & P & Q & R \\ L & P & Q & R \end{pmatrix} \begin{pmatrix} I & T & N & S \\ I & T & N & S \\ I & T & K & R \\ M & T & K & R \end{pmatrix} \begin{pmatrix} V & A & D & G \\ V & A & D & G \\ V & A & E & G \\ V & A & E & G \end{pmatrix}$$

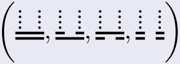
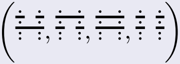
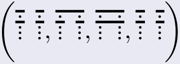
Mapping Hexagrams to Amino Acids

Question

Show that there is a simple way to map each codon into a hexagram of the I Ching, with the property that the 21 possible outcomes $\{A, C, D, \ldots, W, Y, -\}$ correspond to 21 consecutive hexagrams of the King Wen ordering.

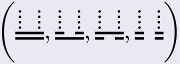
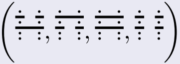
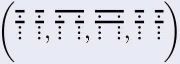
Mapping Hexagrams to Amino Acids

Solution

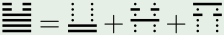
- Represent (T, C, A, G) by
  -  in first nucleotide
  -  in second nucleotide
  -  in third nucleotide

Mapping Hexagrams to Amino Acids

Solution

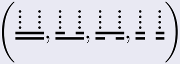
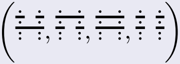
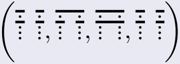
- Represent (T, C, A, G) by
  -  in first nucleotide
  -  in second nucleotide
  -  in third nucleotide

Example

- 
-  $\Leftrightarrow$ codon *TTC*
-  $\Leftrightarrow$ amino acid *F*

Mapping Hexagrams to Amino Acids

Solution

- Represent (T, C, A, G) by
  -  in first nucleotide
  -  in second nucleotide
  -  in third nucleotide
- Hexagrams 34 - 54 mapped to 21 possible amino acids

Q&A Session 1

Q&A

Any Questions?

Outline

Question 7.2.1.7-2(b)

Question

Is that a sensational discovery?

Question 7.2.1.7-2(b)

Question

Is that a sensational discovery?

Knuth's Answer

No

Approach

Note

- The mapping between hexagramms and amino acids, can be reduced to a serialization of the amino acid matrices.
- The number of the hexagram is the position of the mapped amino acid in the string

Approach

Note

- The mapping between hexagramms and amino acids, can be reduced to a serialization of the amino acid matrices.
- The number of the hexagram is the position of the mapped amino acid in the string

Question to Solve

How many permutation of the aminoacid matrices exists that contain a run of 21 distinct consecutive elements.

Number of Total Possible Permutations

Knuth States

$$\binom{64}{6,6,6,4,4,4,4,4,3,3,2,2,2,2,2,2,2,2,2,2,1,1} \approx 2.3 \times 10^{69}$$

Where Does That Formula Come From?

Multiset: $\{n_1 * x1, \ldots, n_r * x_r\}$, $n_1 + \ldots + n_r = n$

$\binom{n}{n_1,\ldots,n_r}$: multinomial coefficient

You draw the positions for:

- The k_1 elements of type x_1
- $\ldots$
- The k_r elements of type x_r

$$\binom{n}{n_1,\ldots,n_r} = \binom{n}{n_1}\binom{n-n_1}{n_2}\ldots\binom{n-n_1-\ldots-n_{r-1}}{n_r} = \frac{n!}{n_1!\ldots n_r!}$$

Number of Total Possible Permutations

Knuth States

$$\binom{64}{6,6,6,4,4,4,4,4,3,3,2,2,2,2,2,2,2,2,2,1,1} \approx 2.3 \times 10^{69}$$

Exact Solution

$$\binom{64}{6,6,6,4,4,4,4,4,3,3,2,2,2,2,2,2,2,2,2,1,1} =$$

$$2316278295874198383354290667060208126761409719904245070233600000000000 \approx 2.3 \times 10^{69}$$

Multiset Permutations with Run of r Distinct Elements

Explanations

Principle of inclusion and exclusion:
Permutations with 21 consecutive elements $-$ multiples

Definitions

$$
\begin{aligned}
\textbf{Multiset} &= \{(n_1+1)*x_1,(n_2+1)*x_2,\ldots,(n_r+1)*x_r\} \\
n &= n_1+n_2+\ldots+n_r
\end{aligned}
$$

Multiset Permutations with Run of r Distinct Elements

$$r! -$$

Explanations

Number of permutations of the 21 consecutive elements

Definitions

$$\begin{aligned} \textbf{Multiset} &= \{(n_1+1)*x_1, (n_2+1)*x_2, \ldots, (n_r+1)*x_r\} \\ n &= n_1+n_2+\ldots+n_r \end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$\binom{n}{n1, \ldots, n_r} r! -$$

Explanations

Number of permutations of the rest of the elements

Definitions

$$\begin{aligned} \textbf{Multiset} &= \{(n_1+1)*x_1, (n_2+1)*x_2, \ldots, (n_r+1)*x_r\} \\ n &= n_1+n_2+\ldots+n_r \end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! -$$

Explanations

Number of places where the consecutive elements can be inserted

Definitions

$$\begin{aligned} \textbf{Multiset} &= \{(n_1+1)*x_1,(n_2+1)*x_2,\ldots,(n_r+1)*x_r\} \\ n &= n_1+n_2+\ldots+n_r \end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! -$$

Where Do the Multiples Come From?

Select x,y,v,w,α so that $|xvy| = n$, $w = \sigma(v)$ and $|v\alpha| = r$ distinct elements.
$xv\alpha wy$ can be created by:

- αw in xvy, or
- $v\alpha$ in xwy

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! -$$

A Special Case Makes Our Life Easier

$xv\alpha wy$ of length $n+r$ has been multiply inserted
$w=\sigma(v)$ but both $v\alpha$ and αw must contain all possible elements exactly once
$\Rightarrow$ α must contain all elements that only occur once
Since we have such an element (W) in the sequence of amino acids
$\Rightarrow$ This case is indeed the only of multiple insertions we must consider

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! - \sum_{k=1}^{r}$$

Explanations

Vary over $|v|$ in $xv\alpha wy$

Definitions

$$\begin{aligned}\textbf{Multiset} &= \{(n_1+1)*x_1, (n_2+1)*x_2, \ldots, (n_r+1)*x_r\} \\ n &= n_1+n_2+\ldots+n_r \\ n_r &= 0\end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! - \sum_{k=1}^{r} \qquad \binom{n-k}{n_1 \quad ,\ldots,n_r}$$

Explanations

Generate number of possible words xy for $xv\alpha wy$ with $|v| = k$

Definitions

$$\begin{aligned}
\textbf{Multiset} &= \{(n_1+1)*x_1, (n_2+1)*x_2, \ldots, (n_r+1)*x_r\} \\
n &= n_1 + n_2 + \ldots + n_r \\
n_r &= 0
\end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! - \sum_{k=1}^{r} \sum_{\substack{0\le d_1,\ldots,d_r\le 1\\ d_1+\ldots+d_r=k}} \binom{n-k}{n_1-d_1,\ldots,n_r-d_r}$$

Explanations

Generate number of possible words xy for $xv\alpha wy$ with $|v| = k$
but take into account the possible elements that can make up v

Definitions

$$\begin{aligned} \textbf{Multiset} &= \{(n_1+1)*x_1, (n_2+1)*x_2, \ldots, (n_r+1)*x_r\} \\ n &= n_1+n_2+\ldots+n_r \\ n_r &= 0 \end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! - \sum_{k=1}^{r} \sum_{\substack{0\le d_1,\ldots,d_r\le 1\\ d_1+\ldots+d_r=k}} \binom{n-k}{n_1-d_1,\ldots,n_r-d_r}$$

Explanations

This defines not only the number of xys but the specific elements contained in v, w and α

Definitions

$$\begin{aligned} \textbf{Multiset} &= \{(n_1+1)*x_1,(n_2+1)*x_2,\ldots,(n_r+1)*x_r\} \\ n &= n_1+n_2+\ldots+n_r \\ n_r &= 0 \end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! - \sum_{k=1}^{r}(n+1-k) \qquad \sum_{\substack{0\le d_1,\ldots,d_r\le 1\\ d_1+\ldots+d_r=k}}\binom{n-k}{n_1-d_1,\ldots,n_r-d_r}$$

Explanations

Number of places where we can split xy

Definitions

$$\begin{aligned}\textbf{Multiset} &= \{(n_1+1)*x_1,(n_2+1)*x_2,\ldots,(n_r+1)*x_r\}\\ n &= n_1+n_2+\ldots+n_r\\ n_r &= 0\end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! - \sum_{k=1}^{r}(n+1-k) \qquad a_k \sum_{\substack{0\le d_1,\ldots,d_r\le 1\\ d_1+\ldots+d_r=k}}\binom{n-k}{n_1-d_1,\ldots,n_r-d_r}$$

Explanations

Number of possible permutation for v
(2134) is not an interesting permutation for v
as that would have been covered in $k=3 \Rightarrow |v=3|$

Definitions

$$\begin{aligned}
\textbf{Multiset} &= \{(n_1+1)*x_1,(n_2+1)*x_2,\ldots,(n_r+1)*x_r\}\\
n &= n_1+n_2+\ldots+n_r\\
n_r &= 0\\
a_k &= \text{number of indecomposable permutations with } k \text{ elements}
\end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! - \sum_{k=1}^{r}(n+1-k){\color{red}k!} \qquad a_k \sum_{\substack{0\le d_1,\ldots,d_r\le 1\\ d_1+\ldots+d_r=k}} \binom{n-k}{n_1-d_1,\ldots,n_r-d_r}$$

Explanations

Number of permutations of v we can insert as w

Definitions

$$\begin{aligned}
\textbf{Multiset} &= \{(n_1+1)*x_1,(n_2+1)*x_2,\ldots,(n_r+1)*x_r\}\\
n &= n_1+n_2+\ldots+n_r\\
n_r &= 0\\
a_k &= \text{number of indecomposable permutations with } k \text{ elements}
\end{aligned}$$

Multiset Permutations with Run of r Distinct Elements

$$(n+1)\binom{n}{n1,\ldots,n_r}r! - \sum_{k=1}^{r}(n+1-k)k!(r-k)!a_k \sum_{\substack{0\leq d_1,\ldots,d_r\leq 1\\ d_1+\ldots+d_r=k}} \binom{n-k}{n_1-d_1,\ldots,n_r-d_r}$$

Explanations

Number of permutations of α we can insert

Definitions

$$\begin{aligned}
\textbf{Multiset} &= \{(n_1+1)*x_1,(n_2+1)*x_2,\ldots,(n_r+1)*x_r\}\\
n &= n_1+n_2+\ldots+n_r\\
n_r &= 0\\
a_k &= \text{number of indecomposable permutations with } k \text{ elements}
\end{aligned}$$

Using the Formula

Knuth States

$$402880402175789790003993681964551328451668718750185553920000000 \approx 2.4 \times 10^{63}$$

Using the Formula

Knuth States

$$240288040217578979000399368196455132845166871875018555392000000 \approx 2.4 \times 10^{63}$$

Result of Verification

$$240288040217578979000399368196455132845166871875018555392000000 \approx 2.4 \times 10^{63}$$

Bringing it Together

It would seem that only one of every million permutations contains a run of 21 distinctive elements.

Bringing it Together

It would seem that only one of every million permutations contains a run of 21 distinctive elements.

But

There are $4!^3\binom{6}{2,2,2} = 1244160$ (Verified) ways to represent codons as in part (a) and effectively change the ordering of the hexagrams.

Bringing it Together

It would seem that only one of every million permutations contains a run of 21 distinctive elements.

But

There are $4!^3\binom{6}{2,2,2} = 1244160$ (Verified) ways to represent codons as in part (a) and effectively change the ordering of the hexagrams.

Thus

The one in a million chance cannot be seen as a proof that the authors of the I Ching must have forseen the Genetic code.

Empirical Test

Knuth States

About 31% of all permutations turn out to have a suitable codon mapping.

Empirical Test

Knuth States

About 31% of all permutations turn out to have a suitable codon mapping.

Method of Verification

- I developed a C++ application that
    - Randomly generates permutation of hexagrams
- Searches for codon mappings

Empirical Test

Knuth States

About 31% of all permutations turn out to have a suitable codon mapping.

Result of Verification

After 1100 tested permutations, 347 were found to have mappings

Empirical Test

Knuth States

About 31% of all permutations turn out to have a suitable codon mapping.

Result of Verification

After 1100 tested permutations, 347 were found to have mappings
$\Rightarrow$ 31%

Q&A Session 2

Q&A

Any Questions?

Outline

My Recomendation

My Recomendation

Knuths provided answer is correct, albeit extremely slow to retrace due to the long, underdocumented formula.

Final Recomendation

Let the Group Decide

What do you think?