Introduction to Linear Algebra
- Gilbert Strang

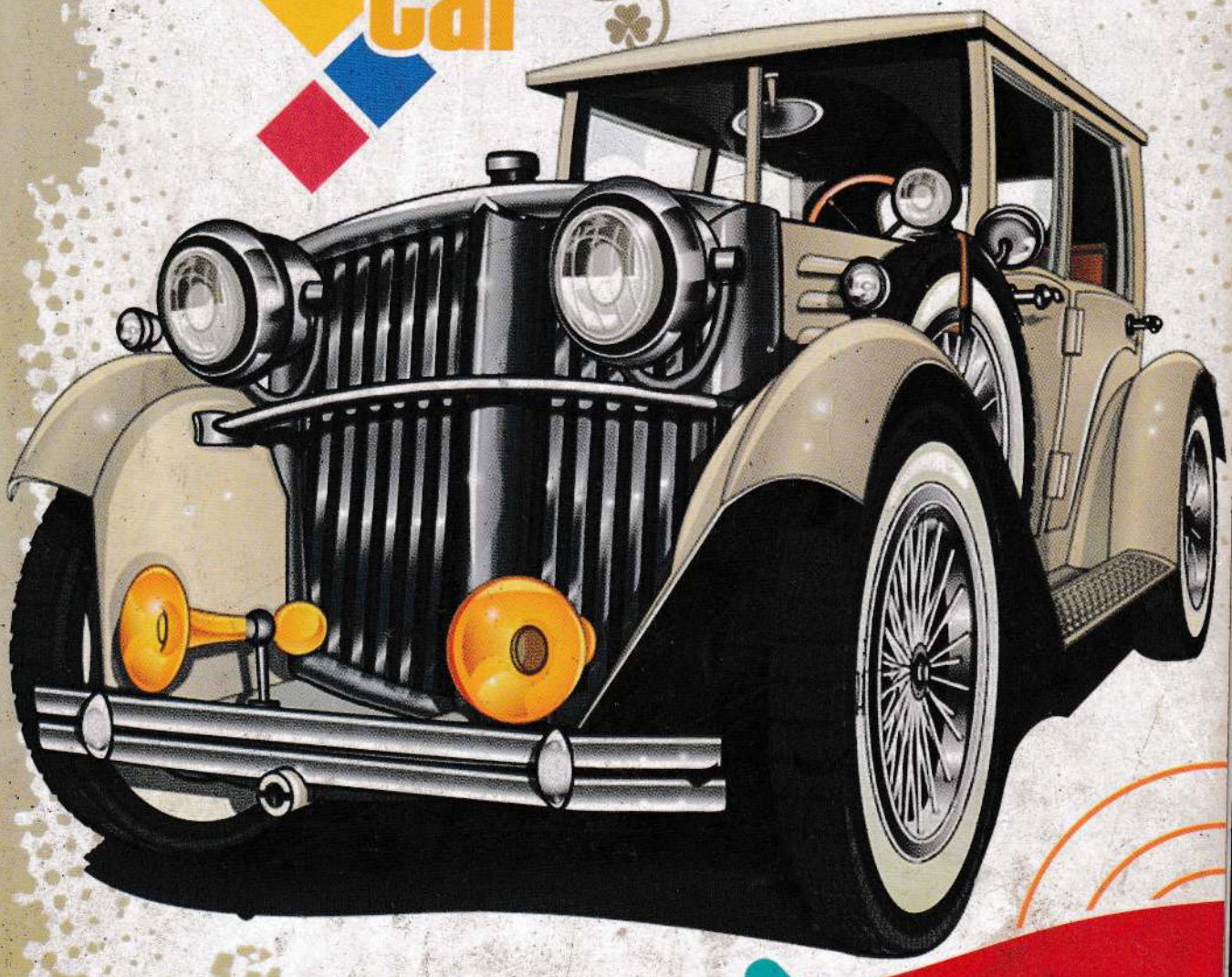
16

Linear Algebra for Cryptography

↓

Linear Algebra in Probability and Statistics

Vintage car



Note Book

soorajss1729@gmail.com +91-9400635788

INDEX

16

Name : SOORAJ S. Std. : Sec. : Roll No. :

School / College :

S. No.	Date	Title	Page No.	Teacher's Sign / Remarks
		INTRODUCTION TO LINEAR ALGEBRA - Gilbert Strang, MIT (5th Edition).		

□ Encryption with the Hill Cipher

The original cipher used the letters A to Z with $p = 26$. Hill chose an $n \times n$ encryption matrix E so that $\det(E)$ is not divisible by 2 or 13. Then the # $\det(E)$ has an inverse mod 26. and so does the matrix E. The inverse matrix $E^{-1} \equiv D \pmod{26}$ will be the decryption matrix that decodes the message.

Now,

convert each letter of the message into a # from 0 to 25.

Ignore spaces and divide the message into blocks $v_1, v_2, \ldots$, of size n. Then multiply each message block (mod p) by the encryption matrix E. The coded message is

$Ev_1, Ev_2, \ldots$

A code breaker will not know E or D.
And the block size 'n' is generally unknown
too.

Ex: **Hill Cipher**

A	B	C	D	...	X	Y	Z	.	?	_
0	1	2	3		23	24	25	26	27	28

MESSAGE : LINEAR_ALGEBRA

You'd want to scramble the order of the characters before assigning #'s.

We'll follow a Hill- 2-cipher

key matrix $A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$

L	I	N	E	A	R	_	A	L	G	E	B	R	A
11	8	13	4	0	17	28	0	11	6	4	1	17	0

Encoding the message,

$$\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} 11 \\ 8 \end{bmatrix} = \begin{bmatrix} 46 \\ 84 \end{bmatrix} = \begin{bmatrix} 17 \\ 26 \end{bmatrix} \pmod{29}$$

$$\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}\begin{bmatrix} 13 \\ 4 \end{bmatrix} = \begin{bmatrix} 38 \\ 72 \end{bmatrix} = \begin{bmatrix} 9 \\ 14 \end{bmatrix} \pmod{29}$$

This gives us the numerical message

46 84|38 72|51 85|56 112|40 74|11 21|34 68

We'll take all of the values modulo 29 in order to get #s that corresp. our starting alphabet.

17 26|9 14|22 27|27 25|11 16|11 21|5 10

This corresp. to the ciphertext message of:

R. JOW ? ? ZLQLV

We now has our encoded message.

$$\begin{bmatrix} 12 & 16 \\ 2 & 28 \end{bmatrix} \begin{bmatrix} 17 \\ 26 \end{bmatrix} = \begin{bmatrix} 620 \\ 762 \end{bmatrix}$$

$$\begin{bmatrix} 12 & 16 \\ 2 & 28 \end{bmatrix} \begin{bmatrix} 9 \\ 14 \end{bmatrix} = \begin{bmatrix} 332 \\ 410 \end{bmatrix}$$

This gives us:

620 762 | 332 410 | 696 800 | 724 754 | 388 470 | 468 610 | / 220 290

Taking all of the numbers module 29 gives us:

11 8 | 13 4 | 0 17 | 28 0 | 11 6 | 4 1 | / 17 0

LINEAR_ALGEBRA.

□ Finite Fields & Finite Vector Spaces

In algebra, a field $\mathbb{F}$ is a set of scalars that can be added & multiplied and inverted (except 0). From a field you build vectors $v = (f_1, f_2, \ldots, f_n)$. From linear combinations of vectors you build vector spaces. So linear algebra begins with a field $\mathbb{F}$.

* A group is a set G with a binary operation '$\cdot$' (a function from $G \times G$ into G) such that:

① The operation is associative

$$a \cdot (b \cdot c) = (a \cdot b) \cdot c \quad \forall\ a, b, c \in G$$

② There exists an identity element e

$$a \cdot e = e \cdot a = a \quad \forall\ a \in G$$

③ Each element a has an inverse element a^{-1}.

$$a \cdot a^{-1} = a^{-1} \cdot a = e$$

A group $(G, \cdot)$ is commutative (abelian) if $a \cdot b = b \cdot a \quad \forall\ a, b \in G$.

Ex:-

$(\mathbb{R}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R} - \{0\}, \times)$, $(\mathbb{Z}_n, +)$, $(\mathbb{Z}_p - \{0\}, \times)$, $(\mathbb{C}, +)$

A field is a set $\mathbb{F}$ with 2 binary operations $+$ and $\times$ such that:

① $(\mathbb{F}, +)$ is a commutative group with identity element 0.

ie., $\mathbb{F}$ is an Abelian group under $+$ with identity element 0.

② $(\mathbb{F} - \{0\}, \times)$ is a commutative group with identity element 1.

ie., The non-zero elements of $\mathbb{F}$ form an Abelian group under $\times$, with identity element 1.

③ The distributive law $a \times (b+c) = a \times b + a \times c$ holds $\forall\ a, b, c \in \mathbb{F}$

ie.,

$\times$ distributes over $+$

Finite fields

The finite field $\mathbb{F}_p$ contains only the numbers $0, 1, 2, \ldots, p-1$ where p is a prime #.

The field $\mathbb{F}_2$ has 2 members "0" and "1".

Addition table:

+	0	1
0	0	1
1	1	0

Multiplication table:

·	0	1
0	0	0
1	0	1

This is addition and multiplication "mod 2".

From this field $\mathbb{F}_2$ we can build vectors like $V=(0,0,1)$ and $W=(1,0,1)$. There are 3 components with 2 choices each: a total of $2^3 = 8$ different vectors in the vector space $(\mathbb{F}_2)^3$.

(a) The zero-dimensional subspace containing only $0=(0,0,0)$

(b) One-dimensional subspace containing 0 and a vector like V.

$V+V=0$

(c) 2D subspaces with a basis like V and W and 4 vectors $0, V, W, V+W$

(d) The full 3D subspace $(\mathbb{F}_2)^3$ with 8 vectors.

Bases for $(\mathbb{F}_2)^3$.

The standard Basis of $(\mathbb{F}_2)^3$ contains $(1,0,0)$ $(0,1,0)$ and $(0,0,1)$. These vectors are linearly independent and they span $(\mathbb{F}_2)^3$. Their 8 combinations with coefficients 0 & 1 fill all of $(\mathbb{F}_2)^3$.

Matrices that multiply $\vec{v} \in (\mathbb{F}_2)^3$.

1×3 or 2×3 or 3×3.

If 3×3, are they invertible?

Determinants can only be 0 (singular matrix) or 1 (invertible matrix).

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

2^9 possible matrices over $\mathbb{F}_2$.

Guess: most are singular

$$A\mathbf{x} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} \mathbf{x} = \begin{bmatrix} r_1 \cdot \mathbf{x} \\ r_2 \cdot \mathbf{x} \\ \vdots \\ r_m \cdot \mathbf{x} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \pmod p = 0 \pmod p$$

$$\Longrightarrow r_i \cdot \mathbf{x} = 0 \pmod p$$

$\therefore$ Every row of A is $\perp$ to every $\mathbf{x}$ in the nullspace $(\text{mod } p)$.

A basis for the usual $N(A)$ could include vectors that are $0 \pmod p$.

A field with $2^2 = 4$ members will not come from multiplication (mod 4), because 4 is not a prime.

Start with the numbers 0 and 1 in $\mathbb{F}_2$ and invent 2 more numbers a and $1+a$ provided they follow these 2 rules:

$a+a=0$

Beyond $p=2$, we have the fields $\mathbb{F}_p$ for all prime #s. They use addition & multiplication mod p. They are alphabets for codes. They provide the components for vectors $v=(f_1, f_2, \ldots, f_n)$ in the space $(\mathbb{F}_p)^n$. They provide the entries for matrices that multiply those vectors. Those fields $\mathbb{F}_p$ are the most frequently used finite fields.

10.7

1. If you multiply n whole numbers (even or odd) when is the answer odd? Translate into multiplication (mod 2): If you multiply 0's and 1's when is the answer 1?

Ans: Xing n whole #s gives an odd # only when all n #s are odd.

This translates to multiplication (mod 2).

Xing n 1's (or) 0's gives 1 only when all n numbers are 1.

2. If you add n whole # (even or odd) when is the sum of the #s odd? Translate into adding 0's and 1's (mod 2). When do they add to 1?

Ans: Adding n whole #s gives an odd # only when the n numbers include an odd # of odd numbers.
For addition of 1's and 0's (mod 2), the answer is odd when the # of 1's is odd.

10.7

1. If you multiply n whole numbers (even or odd) when is the answer odd? Translate into multiplication (mod 2): If you multiply 0's and 1's when is the answer 1?

Ans: Xing n whole #s gives an odd # only when all n #s are odd.

This translates to multiplication (mod 2).

Xing n 1's (or) 0's gives 1 only when all n numbers are 1.

2. If you add n whole # (even or odd) when is the sum of the #s odd? Translate into adding 0's and 1's (mod 2). When do they add to 1?

Ans: Adding n whole #s gives an odd # only when the n numbers include an odd # of odd numbers.
For addition of 1's and 0's (mod 2), the answer is odd when the # of 1's is odd.

3. (a) If $y_1 \equiv x_1$ and $y_2 \equiv x_2$, why is $y_1 + y_2 \equiv x_1 + x_2$? All are mod p.

Ans. $y_1 \equiv x_1 \pmod{p}$ & $y_2 \equiv x_2 \pmod{p}$

$y_1 - x_1 = q_1 p$ & $y_2 - x_2 = q_2 p$

$\Longrightarrow (y_1 + y_2) - (x_1 + x_2) = (q_1 + q_2) p = tp$

$(y_1 + y_2) \equiv (x_1 + x_2) \pmod{p}$

(b) Can you be sure that $x_1 + x_2$ is smaller than p?

Give an example where there is a smaller x with $y_1 + y_2 \equiv x \pmod{p}$

Ans: $5 \equiv 2 \pmod 3$ & $8 = 2 \pmod 3$

4. $p = 39$ is not prime. Find a # 'a' that has no inverse $z \pmod{39}$.
i.e., $az \equiv 1 \pmod{39}$ has no solution.
Then find a 2×2 matrix A that has no inverse matrix $Z \pmod{39}$.
i.e., $AZ = I \pmod{39}$ has no solution

Ans: ~~$az \equiv 1 \pmod{39}$ has a~~

$yz \equiv 1 \pmod p$ (or) $yz + tp = 1$ for some $t \in \mathbb{Z}$

If $\gcd(y, p) = 1$.

$az \equiv 1 \pmod{39}$ has a solution if $\gcd(a, 39) = 1$

Take $a = 8$, $\gcd(8, 39) = 1$

$$\begin{array}{r|l} 1 & 38 \,|\, 8 \\ & 1 \end{array}$$

$8 \equiv 8 \pmod{39}$

$8z = 8$

$$39 = 4(8) + 7 \quad \Big\} \quad 7 = 39 - 4(8)$$
$$8 = 1(7) + 1 \quad \Big\} \quad 1 = 8 - 1(7)$$

$$1 = 8 - 1(7) = 8 - 1\left[39 - 4(8)\right]$$

$$1 = 5(8) - 1(39)$$

$$5(8) - 1 = 1(39)$$

$$\Rightarrow 8^{-1} = 5 \equiv 5 \pmod{39}$$

If $a = 3$,

$$\gcd(3, 39) = 3 \neq 1$$

$$3z = 1 \pmod{39}$$

$$3 = 3 \pmod{39}$$

$$39 = 13(3)$$

No inverse exists

$AZ \equiv I \pmod{39}$

There is an inverse matrix mod p whenever the determinant of A is non-zero mod p, and p is a prime #.

6. Find a matrix that has independent columns in $\mathbb{R}^2$ but dependent columns (mod 5)

Ans: $A = \begin{bmatrix} 2 & 7 \\ 1 & 6 \end{bmatrix}$ is invertible since $\det(A) = 12 - 7 = 5 \neq 0$.

(mod 5) $A = \begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix}$

$A = \begin{bmatrix} 5 & 5 \\ 5 & 10 \end{bmatrix}$, $\det(A) = 50 - 25 = 25 \neq 0$

(mod 5) $A = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

7. What are all the 2×2 matrices of 0's and 1's that are invertible (mod 2)?

Ans: $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$

6 out of 16 possible 0-1 matrices are invertible.

8. Is the rowspace of 'A' still orthogonal to the nullspace in modular arithmetic (mod 11)? Are bases for those subspaces still bases (mod 11)?

Study & check previous section

Ans:

9. (Hill Cipher) Separate the message THISWHOLEBOOKISINCODE into blocks of 3 letters. Replace each letter by a # from 1 to 26 (normal order). Multiply each block by the 3×3 matrix L with 1's on and below the diagonal. What is the coded message in numbers and how would you decode it?

Ans:

	1	2	3	4	5	6	7	8	9	10	11	12	13
~~A~~	~~A~~	~~B~~	C	D	E	F	G	H	I	J	K	L	M

14	15	16	17	18	19	20	21	22	23	24	25	26
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Numbering the letters as they appear in the message.

THISWHOLEBOOKISINCODE

1 2 3|4 5 2|6 7 8|9 6 6|10 7 3|4 3 11|12 6 13 8

Multiply each block by $L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ to obtain Hill's cipher.

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ 9 \\ 11 \end{bmatrix}$$

Cipher:

1 3 6 | 4 9 (11) | 6 (13) (21) | 9 (15) (21) | (19) (13) (17) | 3 (14) (26) | 6 (19) (27)

If the cipher is ~~mod 10~~

mod 15

1 3 6 | 4 9 11 | 6 13 ~~21~~ | ~~9 15 21~~ | ~~3~~ 0 12 | 6 ~~4~~ 13

T I ~~O~~ S B T ~~O~~ D T ~~B~~ T L ~~I~~ ~~2~~ ~~C~~ ~~O~~ ~~W~~ D

Coded message

1	3	6	4	9	11	6	13	6	9	0	6	5	13	2
T	I	O	S	B	T	O	D	O	B	?	O	K	D	H
						3	14	11	6	4	12			
						I	~~S~~	N	O	S	C			

Coded message.

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 14 & 1 & 0 \\ 0 & 14 & 1 \end{bmatrix} \pmod{15}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 14 & 1 & 0 \\ 0 & 14 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 \\ 17 \\ 48 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \pmod{15}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 14 & 1 & 0 \\ 0 & 14 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 9 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 65 \\ 137 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 2 \end{bmatrix} \pmod{15}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 14 & 1 & 0 \\ 0 & 14 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 13 \\ 6 \end{bmatrix} = \begin{bmatrix} 6 \\ 97 \\ 188 \end{bmatrix} = \begin{bmatrix} 6 \\ 7 \\ 8 \end{bmatrix} \pmod{15}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 14 & 1 & 0 \\ 0 & 14 & 1 \end{bmatrix} \begin{bmatrix} 9 \\ 0 \\ 6 \end{bmatrix} = \begin{bmatrix} 9 \\ 126 \\ 6 \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 6 \end{bmatrix} \pmod{15}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 14 & 1 & 0 \\ 0 & 14 & 1 \end{bmatrix} \begin{bmatrix} 10 \\ 13 \\ 2 \end{bmatrix} = \begin{bmatrix} 10 \\ 153 \\ 184 \end{bmatrix} = \begin{bmatrix} 10 \\ 3 \\ 4 \end{bmatrix} \pmod{15}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 14 & 1 & 0 \\ 0 & 14 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 14 \\ 11 \end{bmatrix} = \begin{bmatrix} 3 \\ 56 \\ 207 \end{bmatrix} = \begin{bmatrix} 3 \\ 11 \\ 12 \end{bmatrix} \pmod{15}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 14 & 1 & 0 \\ 0 & 14 & 1 \end{bmatrix} \begin{bmatrix} 6 \\ 84 \\ 12 \end{bmatrix} = \begin{bmatrix} 3 \\ 56 \\ 207 \end{bmatrix} = \begin{bmatrix} 6 \\ 88 \\ 68 \end{bmatrix} = \begin{bmatrix} 6 \\ 13 \\ 8 \end{bmatrix} \pmod{15}$$

10. Suppose you know the original message (the plain text). Suppose, you also see the coded message. How could you start to discover the matrix in Hill's Cipher? For a long message do you expect success?

Ans:

First you have to discover the block size (= matrix size) and also the matrix L itself. Start with a guess for the block size. Then the plaintext and the coded cipher will tell you a series of matrix – vector products $Lx \equiv b$.

If the text is long enough (and the blocks are not too long) this is enough information to find L —

(or) to show that the block size must be wrong, when there is no L that gets all correct blocks $Lx \equiv b$

The extra difficulty is to find the value of p.

LINEAR ALGEBRA IN PROBABILITY AND STATISTICS

12

The sample mean starts with N samples $x_1, \ldots, x_N$ from a completed trial. Their mean is the avg. of the N observed samples:

Sample mean,

$$m = \mu = \frac{1}{N}(x_1 + x_2 + \cdots + x_N)$$

$$= \frac{1}{N}\sum_i x_i$$

The expected value of x starts with the probabilities $p_1, \ldots, p_n$ of $x_1, x_2, \ldots, x_n$

Expected value,

$$m = E[x] = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$$

$$= \sum_i p_i x_i$$

* $m = E[x]$ tells us what to expect,
$m = \mu$ tells us what we got.

* By taking many samples (large N), the sample results will come close to the probabilities. The "Law of Large Numbers" says that with probability 1, the sample mean will converge to its expected value $E[x]$ as the sample size N increases.

The variance σ^2 measures expected distance (squared) from the expected mean $E[X]$.

The sample variance S^2 measures actual distance (squared) from the sample mean.

Sample variance,

$$S^2 = \frac{1}{N-1}\left[(x_1-m)^2 + \cdots + (x_N-m)^2\right]$$

$$= \frac{1}{N-1}\sum_{i}(x_i-m)^2$$

$$\sum_i (x_i - m)^2 = \sum_i x_i^2 - N m^2$$

Now, start with probabilities P_i instead of samples. We find expected values instead of sample values.

Variance,

$$\sigma^2 = E\left[(x-m)^2\right] = P_1 (x_1 - m)^2 + \cdots\cdots + P_n (x_n - m)^2$$

$$= \sum_i P_i (x_i - m)^2$$

$m = E[x]$

Bessel's correction

Bessel's correction is the division of the sample variance by $(N-1)$ rather than N.

Ex:-

Suppose you have a bag with 3 cards in it. The cards are numbered 0, 2, 4.

Population of all $N=3$ cards in bag.

Population mean, $\mu = \frac{0+2+4}{3} = \frac{6}{3} = 2$

Population variance, $\sigma^2 = \frac{(0-2)^2 + (2-2)^2 + (4-2)^2}{3} = \frac{4+0+4}{3} = \frac{8}{3}$

There are 9 possible samples of 2 cards.

All possible samples of size $n=2$	sample average $\bar{x} = \frac{\sum x_i}{n}$	Sample variance $s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$
(0,0)	$\frac{0+0}{2} = 0$	$\frac{(0-0)^2+(0-0)^2}{1} = 0$
(0,2)	$\frac{0+2}{2} = 1$	$\frac{(0-1)^2+(2-1)^2}{1} = 2$
(0,4)	$\frac{0+4}{2} = 2$	$\frac{(0-2)^2+(4-2)^2}{1} = 8$
(2,0)	$\frac{2+0}{2} = 1$	$\frac{(2-1)^2+(0-1)^2}{1} = 2$
(2,2)	$\frac{2+2}{2} = 2$	$\frac{(2-2)^2+(2-2)^2}{1} = 0$
(2,4)	$\frac{2+4}{2} = 3$	$\frac{(2-3)^2+(4-3)^2}{1} = 2$
(4,0)	$\frac{4+0}{2} = 2$	$\frac{(4-2)^2+(0-2)^2}{1} = 8$
(4,2)	$\frac{4+2}{2} = 3$	$\frac{(4-3)^2+(2-3)^2}{1} = 2$
(4,4)	$\frac{4+4}{2} = 4$	$\frac{(4-4)^2+(4-4)^2}{1} = 0$

Average of all $\bar{x}$ sample averages $= \frac{0+1+2+1+2+3+2+3+4}{9} = \frac{18}{9} = 2$

$\Rightarrow$ Average of all $\bar{x} = \mu$

Average of all s^2 sample variances,

$$\frac{0+2+8+2+0+2+8+2+0}{9} = \frac{24}{9} = \frac{8}{3}$$

$\Longrightarrow$ Average of all $s^2 = \sigma^2$

List of all possible samples of size $n=2$	Sample average $\bar{x} = \frac{\sum x_i}{n}$	$\sum \frac{(x_i - \bar{x})^2}{n}$
(0,0)	$\frac{0+0}{2} = 0$	$\frac{(0-0)^2 + (0-0)^2}{2} = 0$
(0,2)	$\frac{0+2}{2} = 1$	$\frac{(0-1)^2 + (2-1)^2}{2} = 1$
(0,4)	$\frac{0+4}{2} = 2$	$\frac{(0-2)^2 + (4-2)^2}{2} = 4$
(2,0)	$\frac{2+0}{2} = 1$	$\frac{(2-1)^2 + (0-1)^2}{2} = 1$
(2,2)	$\frac{2+2}{2} = 2$	$\frac{(2-2)^2 + (2-2)^2}{2} = 0$
(2,4)	$\frac{2+4}{2} = 3$	$\frac{(2-3)^2 + (4-3)^2}{2} = 1$
(4,0)	$\frac{4+0}{2} = 2$	$\frac{(4-2)^2 + (0-2)^2}{2} = 4$
(4,2)	$\frac{4+2}{2} = 3$	$\frac{(4-3)^2 + (2-3)^2}{2} = 1$
(4,4)	$\frac{4+4}{2} = 4$	$\frac{(4-4)^2 + (4-4)^2}{2} = 0$

Average of all $\frac{\Sigma(x_i - \bar{x})^2}{n}$ for all samples,

$$\frac{0+1+4+1+0+1+4+1+0}{9} = \frac{12}{9} = \frac{4}{3} \neq \frac{8}{3} = \sigma^2$$

↳ what causes the bias ?

Ideally we could estimate the variance of the sample by subtracting each value from the population mean. However, since we don't know what the population mean is, we use the next best thing — the sample mean.

↑ This is where the bias comes in.

When you use the sample mean, you' are guaranteed that the mean lies somewhere within the range of your data points.
In fact, the mean of a sample minimizes the sum of squared deviations from the mean.

ie., the sum of deviations from the sample mean is always smaller than the sum of deviations from the population mean.

Proof of Bessel's correction.

"We would want the average of the sample variances for all possible samples to equal the population variance."

It seems like a logical property and a reasonable thing to happen. This is called "unbiased".

We want : avg. of (all possible sample variances) = Population variance.

$\Longrightarrow$ unbiased

Consider N independent & identically distributed (iid) random variables $x_1, x_2, \ldots, x_N$ and a sample mean $\bar{x}$

$$S^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

$\bar{x}$: sample mean

$$E[S^2] = E\left[\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right]$$

$$= E\left[\frac{1}{N}\sum_{i=1}^{N}\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)\right]$$

$$= E\left[\frac{1}{N}\sum_{i=1}^{N}x_i^2 - 2\bar{x}\cdot\frac{1}{N}\sum_{i=1}^{N}x_i + \frac{1}{N}\sum_{i=1}^{N}\bar{x}^2\right]$$

$$\sum_{i=1}^{N} x_i = N\bar{x}$$

$$E\left[\frac{1}{N}\sum_{i=1}^{N}x_i^2\right] = \frac{1}{N}E\left[\sum_{i=1}^{N}x_i^2\right] = \frac{1}{N}\sum_{i=1}^{N}E[x_i^2]$$

$$= E[x_i^2]$$

$$E[s^2] = E\left[\frac{1}{N}\sum_{i=1}^{N} x_i^2\right] - E[2\bar{x}^2] + E[\bar{x}^2]$$

$$= E[x_i^2] - E[\bar{x}^2]$$

For any random variable Y,

$$Var(Y) = E[Y^2] - E[Y]^2$$

$$\Longrightarrow E[Y^2] = Var[Y] + E[Y]^2$$

$$E[x_i^2] = Var[x_i] + E[x_i]^2$$

$$= \sigma^2 + \mu^2$$

$$E[\bar{x}^2] = Var[\bar{x}] + E[\bar{x}]^2$$

$$= \frac{\sigma^2}{N} + \mu^2$$

$$Var(\bar{x}) = Var\left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)$$

$$= \frac{1}{N^2} Var\left(\sum_{i=1}^{N} x_i\right)$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} Var(x_i)$$

$$= \frac{1}{N^2} \sum_{i=1}^{N} \sigma^2 = \frac{\sigma^2}{N}$$

$$E[s^2] = E[x_i^2] - E[\bar{x}^2]$$

$$= \left(\sigma^2 + \mu^2\right) - \left(\frac{\sigma^2}{N} + \mu^2\right)$$

$$= \sigma^2\left(1 - \frac{1}{N}\right) = \sigma^2\left(\frac{N-1}{N}\right)$$

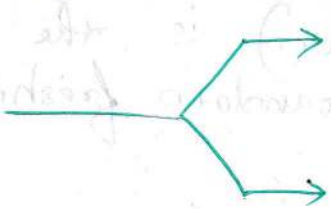
$$\Rightarrow E\left[\left(\frac{N}{N-1}\right)s^2\right] = E\left[\left(\frac{N}{N-1}\right)\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2\right]$$

$$= E\left[\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2\right] = \sigma^2$$

Continuous Probability Distributions

Upto now we have allowed for n possible outcomes $x_1, \ldots, x_n$. With ages 17, 18, 19 we only had $n = 3$.

If we measure age in days instead of years, there will be a 1000 possible ages (too many). Better to allow every # b/w 17 and 20 - a continuum of possible ages. Then the probabilities P_1, P_2, P_3 for ages x_1, x_2, x_3 have to move to a probability distribution $P(x)$ for a whole continuous range of ages $17 \leq x \leq 20$.



Uniform distribution

Suppose ages are uniformly distributed b/w 17 and 20. All ages b/w those numbers are "equally likely".

Of course,
any one exact age has no chance at all.
There is zero probability that you'll hit the exact number $x = 17.1$ or $x = 17 + \sqrt{2}$.

What you can truthfully provide (assuming our uniform distribution) is the chance $F(x)$ that a random freshman has age less than x:

The chance of age less than $x = 17$ is $F(17) = 0$

$\Rightarrow x \leq 17$ won't happen.

The chance of age less than $x = 20$ is $F(20) = 1$

$\rightarrow x \leq 20$ will happen.

The chance of age less than x is given by the cumulative probability $F(x)$,

$$F(x) = \frac{1}{3}\int_{17}^{x} dx = \frac{1}{3}\Big[x\Big]_{17}^{x} = \frac{1}{3}(x-17)$$

$$\int_{17}^{20} \frac{1}{3}\,dx = \frac{1}{3}\Big[x\Big]_{17}^{20} = 1$$

$F(17) = 0 \implies x \leq 17$ won't happen

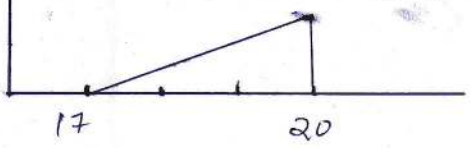
$F(20) = 1 \implies x \leq 20$ is sure.

B/w 17 & 20, the graph of the cumulative distribution $F(x)$ increases linearly for this uniform model.

Cumulative probability

$F(x)$: probability that a sample is below x

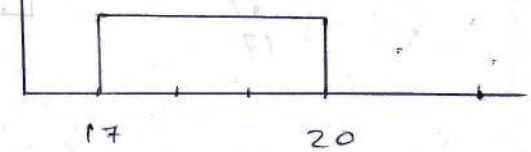
$$F(x) = \frac{1}{3}(x - 17)$$



probability density function (pdf)

– probability that a sample is near x,

$$p(x) = \frac{dF}{dx}$$



A cumulative probability refers to the probability that the value of a random variable falls within a specified range.
Frequently, cumulative probabilities refers to the probability that a random variable is less than or equal to a specified value.

Cumulative probability,

$$F(a) = \int_{-\infty}^{a} p(x)\, dx$$

$p(x)\,dx$: probability of a sample falling in b/w x and $x+dx$

$$p(x)\,dx \approx F(x+dx) - F(x)$$ [infinitesimally true]

$$\text{Probability of } a \le x \le b = \int_a^b p(x)\,dx = F(b) - F(a)$$

Mean & Variance of $p(x)$

Instead of adding $p_i x_i$ to get the mean (expected value), with a continuous distribution we integrate $x\,p(x)$:

Mean,

$$m = E[x] = \int x\,p(x)\,dx$$

$$m = E(x) = \frac{1}{3}\int_{x=17}^{20} x\,dx = \frac{1}{3}\left[\frac{x^2}{2}\right]_{17}^{20} = \frac{1}{6}\left[(20-17)(20+17)\right]$$

$$= \frac{1}{6} \times 3 \times 37 = 18.5$$

* For this uniform distribution, the mean m is halfway b/w 17 and 20.
The probability of a random value x below this halfway point, $m = 18.5$ is

$$F(18.5) = \frac{1}{2}.$$

$$F(m) = 1/2$$

The variance is the average squared distance to the mean. With N outcomes, σ^2 is the sum of $P_i(x_i - m)^2$.

For continuous random variable x, the sum changes to an integral

> Variance,
>
> $$\sigma^2 = E\left[(x-m)^2\right] = \int p(x)(x-m)^2 dx$$

When the ages are uniform b/w $17 \le x \le 20$, the integral can shift to $0 \le x \le 3$:

$$\sigma^2 = \int_{17}^{20} \frac{1}{3}(x-18.5)^2 dx = \int_0^3 \frac{1}{3}(x-18.5)^2 dx$$

$$= \frac{1}{9}\left[(x-18.5)^3\right]_0^3 = \frac{2}{9}\left[1.5\right]^3 = \frac{3}{4}$$

Uniform distribution for $0 \le x \le a$,

Probability density, $p(x) = \frac{1}{a}$

Cumulative probability, $F(x) = \frac{x}{a}$

Mean, $m = \frac{a}{2}$ (half way).

Variance, $\sigma^2 = \int_0^a \frac{1}{a}\left(x - \frac{a}{2}\right)^2 dx = \frac{a^2}{12}$

□ Normal distribution : Bell-shaped curve

(Gaussian distribution)

It is the most important of all probability density functions $p(x)$. The reason for its overwhelming importance comes from repeating an experiment and averaging the outcomes. The expts have their own distributions (like heads & tails). The avg. approaches a normal distribution.

Central limit theorem (informal).

* The avg. of N samples of "any" probability distribution approaches a normal distribution as $N \to \infty$.

O.M (22)
backside

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

For the "standard normal distribution", it is symmetric around $x=0$, so its mean value is $m=0$. It is chosen to have a standard variance $\sigma^2=1$. It is called $N(0,1)$.

Standard normal distribution,

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

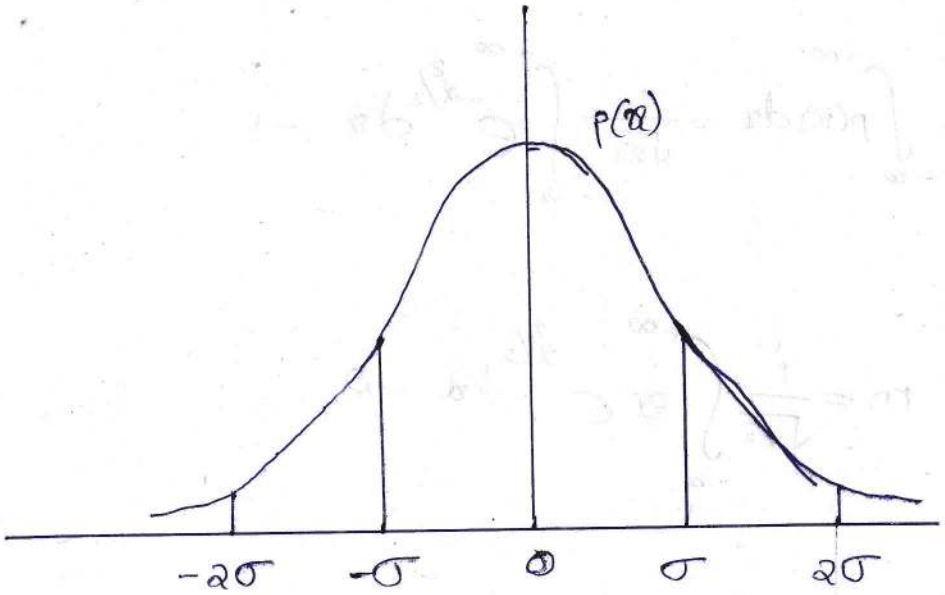
$$\int_{-\infty}^{+\infty} p(x)\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-x^2/2}\,dx = 1$$

$$m = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x\, e^{-x^2/2}\,dx = 0$$

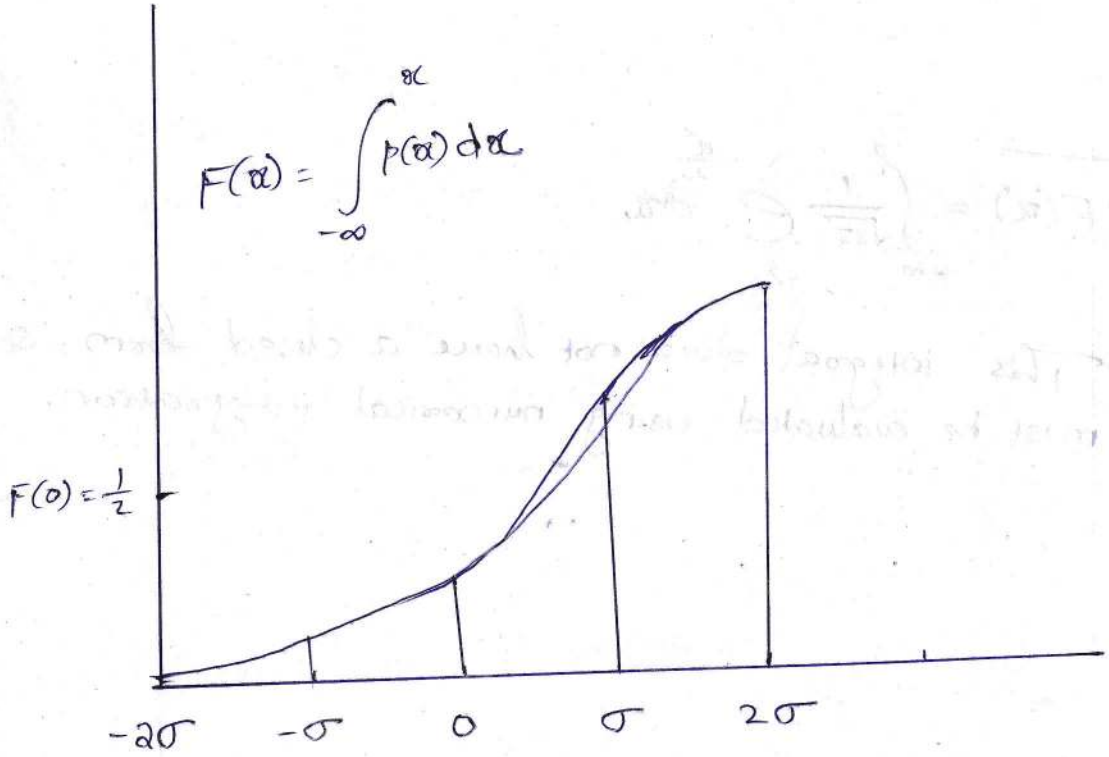
$$\sigma^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-0)^2 e^{-x^2/2}\,dx = 1$$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$$

This integral does not have a closed form, & must be evaluated using numerical integration.



$$F(x) = \int_{-\infty}^{x} p(x)\,dx$$



* Normal distribution $N(0, \sigma)$

The prob. that a random sample falls b/w $-\sigma$ & σ is,

$$F(\sigma) - F(-\sigma) = \int_{-\infty}^{\sigma} p(x)\,dx - \int_{-\infty}^{-\sigma} p(x)\,dx \approx \frac{2}{3} = 0.67$$

The probability that a random x lies b/w -2σ and 2σ is:

$$F(2\sigma) - F(-2\sigma) \approx 0.95$$

If you have an experimental result further than 2σ from the mean, it is fairly sure to be not accidental: chance = 0.05

Drug tests may look for a tighter confirmation, like probability 0.001.

Searching for the Higgs boson used a hyper-strict test of 5σ deviation from pure accident.

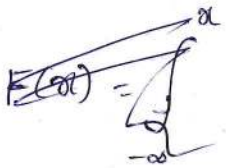
* The normal distribution with any mean m and standard deviation σ comes by shifting and stretching the standard $N(0,1)$.

Shift x to $x-m$

Stretch $x-m$ to $\frac{x-m}{\sigma}$

Gaussian density $p(x)$

Normal distribution $N(m,\sigma)$: $p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$



N Coin Flips & $N \to \infty$

Linearity

$$x_{new} = a x_{old} + b \implies m_{new} = a m_{old} + b$$

$$\sigma^2_{new} = a^2 \sigma^2_{old}$$

Fig. The probabilities $p = \left(\frac{1}{16}, \frac{4}{16}, \frac{6}{16}, \frac{4}{16}, \frac{1}{16}\right)$ for the # of heads in 4-flips.

These p_i approach a Gaussian distribution with variance $\sigma^2 = \frac{N}{4}$ centered at $m = \frac{N}{2}$.

To reach the standard Gaussian (mean = 0 & variance = 1) we shift and rescale these graphs.

If x is the # of heads in N flips - the avg. of N zero-one outcomes

— then x is shifted by its mean $m = \frac{N}{2}$ and rescaled by $\sigma = \frac{\sqrt{N}}{2}$ to produce the standard X:

Shifted & scaled

$$X = \frac{x - m}{\sigma} = \frac{x - \frac{N}{2}}{\sqrt{N}/2}$$

Subtracting m is "centering" or "detrending". The mean of X is zero.

Dividing by σ is "normalizing". The variance of X is 1.

At the center point $x = 0$, $e^{-x^2/2} = 1$.

The variance for N coin flips is $\sigma^2 = N/4$.

$\therefore$

The centre of the Bell-shaped curve has height, $\frac{1}{\sqrt{2\pi\sigma^2}} = \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sqrt{N/2}} = \sqrt{\frac{2}{N\pi}}$

Binomial theorem

The center probability $P_{N/2}$ for any even N.

The center probability ($\frac{N}{2}$ heads, $\frac{N}{2}$ tails) is:

$$P_{N/2} = {}^{N}C_{N/2}\, q^{N-\frac{N}{2}}\, p^{N/2}$$

$$= {}^{N}C_{N/2} \left(\frac{1}{2}\right)^{N/2} \left(\frac{1}{2}\right)^{N/2}$$

$$= \frac{1}{2^N} \frac{N!}{\left(\frac{N}{2}\right)! \left(\frac{N}{2}\right)!}$$

For large N,

Stirling's formula,

$$N! \approx \sqrt{2\pi N}\left(\frac{N}{e}\right)^N$$

Limit of coin-flip Center probability,

$$P_{N/2} = \frac{1}{2^N}\frac{N!}{(N/2)!\,(N/2)!} \approx \frac{1}{2^N}\frac{\sqrt{2\pi N}\,(N/e)^N}{\pi N\left(\frac{N}{2e}\right)^N}$$

$$= \frac{\sqrt{2}}{\sqrt{\pi N}} = \frac{1}{\sqrt{2\pi}\,\sigma}$$

$\sigma = \sqrt{N}/2$

Stirling's formula

$$N! \approx \sqrt{2\pi N}\left(\frac{N}{e}\right)^N$$

Proof : 1

OM(17)

$$\Gamma(x+1) = \int_0^\infty t^x e^{-t}\,dt$$

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\,dt$$

Integration by parts

$$\boxed{\Gamma(x+1) = x\,\Gamma(x)}$$

$$\Gamma(N+1) = N!$$

$$\Gamma(x+1) = \int_0^\infty t^x e^{-t} dt$$

÷ing by $\left(\frac{x}{e}\right)^x$,

$$\left(\frac{e}{x}\right)^x \Gamma(x+1) = \int_0^\infty \left(\frac{t}{x}\right)^x e^{-(t-x)} dt$$

Let $s = t - x$, so that

$$\left(\frac{e}{x}\right)^x \Gamma(x+1) = \int_{-x}^\infty \left(1+\frac{s}{x}\right)^x e^{-s} ds$$

$$= \int_{-x}^\infty f(s)\, ds$$

where, $f(s) = \left(1+\frac{s}{x}\right)^x e^{-s}$

$$\ln[f(s)] = x \ln\left(1+\frac{s}{x}\right) - s$$

Taylor series,

$$\ln(1+u) = u - \frac{u^2}{2} + \frac{u^3}{3} - \cdots$$

$\therefore$

$$\ln[f(s)] = x\left[\frac{s}{x} - \frac{1}{2}\left(\frac{s}{x}\right)^2 + \frac{1}{3}\left(\frac{s}{x}\right)^3 - \cdots\right] - s$$

$$= -\frac{1}{2}\frac{s^2}{x} + \frac{1}{3}\frac{s^3}{x^2} - \cdots$$

If x is sufficiently large,

$$\ln[f(s)] \approx \frac{-s^2}{2x}$$

$$\left(\frac{e}{x}\right)^x \Gamma(x+1) = \int_{-\infty}^{+\infty} e^{-s^2/2x}\, ds = \sqrt{2\pi x}$$

since, $\int_{-\infty}^{+\infty} e^{-at^2}\, dt = \sqrt{\pi/a}$

OM (22)
backside

$$x! = \Gamma(x+1) \approx \left(\frac{x}{e}\right)^x \sqrt{2\pi x}$$

$$\therefore \quad N! \approx \left(\frac{N}{e}\right)^N \sqrt{2\pi N}$$

Proof :2

Wallis' formula

$$\frac{\pi}{2} = \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdots\cdots\frac{2n}{2n-1}\cdot\frac{2n}{2n+1}\cdots\cdots$$

$$= \lim_{n\to\infty}\left[\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdots\cdots\frac{2n}{2n-1}\,\frac{2n}{2n+1}\right]$$

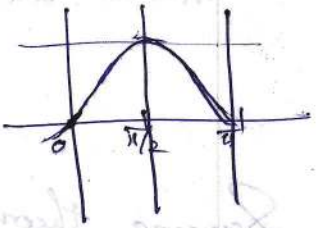
OM (17)

Proof for Wallis' formula

$$I_n = \int_0^{\pi/2} \sin^n x \, dx = \frac{n-1}{n} I_{n-2}$$

OM (17)

$$\frac{I_{2n+2}}{I_{2n}} = \frac{2n+2-1}{2n+2} = \frac{2n+1}{2n+2} = \frac{2+1/n}{2+2/n}$$



For $x \in [0, \pi/2]$,

$\sin x \leq 1 \Longrightarrow \sin^2 x \leq \sin x$

$\sin^2 x \leq \sin x \leq 1$

Xing by $\sin^{2n} x \geq 0$,

$$\sin^{2n+2} x \leq \sin^{2n+1} x \leq \sin^{2n} x$$

$$\int_0^{\pi/2} \sin^{2n+2} x \, dx \leq \int_0^{\pi/2} \sin^{2n+1} x \, dx \leq \int_0^{\pi/2} \sin^{2n} x \, dx$$

$$I_{2n+2} \leq I_{2n+1} \leq I_{2n}$$

$$\frac{I_{2n+2}}{I_{2n}} \leq \frac{I_{2n+1}}{I_{2n}} \leq 1$$

$$\frac{2+1/n}{2+2/n} \leq \frac{I_{2n+1}}{I_{2n}} \leq 1$$

$$\lim_{n\to\infty} \frac{2+1/n}{2+2/n} = \frac{2}{2} = 1 = \lim_{n\to\infty} 1$$

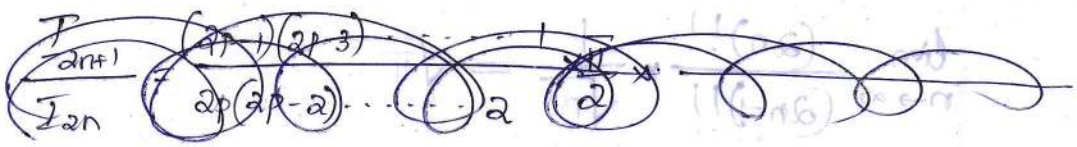
Squeeze theorem $\Longrightarrow$

$$\lim_{n\to\infty} \frac{I_{2n+1}}{I_{2n}} = 1$$

$$I_n = \int_0^{\pi/2} \sin^n x\,dx = \int_0^{\pi/2} \cos^n x\,dx$$

$$I_{2p} = \frac{(2p-1)(2p-3)\cdots\cdots 1}{2p(2p-2)\cdots\cdots 2} \cdot \frac{\pi}{2} = \frac{(2p-1)!!}{(2p)!!} \cdot \frac{\pi}{2}$$

$$I_{2p+1} = \frac{2p(2p-2)\cdots\cdots 2}{(2p+1)(2p-1)\cdots\cdots 1} = \frac{(2p)!!}{(2p+1)!!}$$



$$\frac{I_{2n+1}}{I_{2n}} = \frac{2n(2n-2)\cdots 2}{(2n+1)(2n-1)\cdots 1} \times \frac{2n(2n-2)\cdots 2}{(2n-1)(2n-3)\cdots 1} \times \left(\frac{2}{\pi}\right)$$

$$= \left(\frac{2}{\pi}\right) \times \frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdot\frac{6}{5}\cdot\frac{6}{7}\cdots\frac{2n}{2n-1}\cdot\frac{2n}{2n+1}$$

$$\lim_{n\to\infty} \frac{I_{2n+1}}{I_{2n}} = \lim_{n\to\infty} \left[\left(\frac{2}{\pi}\right)\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdots\frac{2n}{2n-1}\cdot\frac{2n}{2n+1}\right] = 1$$

$$\Longrightarrow \lim_{n\to\infty} \left[\frac{2}{1}\cdot\frac{2}{3}\cdot\frac{4}{3}\cdot\frac{4}{5}\cdots\frac{2n}{2n-1}\cdot\frac{2n}{2n+1}\right] = \frac{\pi}{2}$$

$$\lim_{n\to\infty} \left[\left(\frac{(2n)!!}{(2n-1)!!}\right)^2 \times \frac{1}{2n+1}\right] = \frac{\pi}{2}$$

$$\lim_{n\to\infty} \left[\frac{(2n)!!}{(2n-1)!!}\right]^2 \frac{1}{2n} = \frac{\pi}{2}$$

$$\lim_{n\to\infty} \frac{(2n)!!}{(2n-1)!!} \times \frac{1}{\sqrt{n}} = \sqrt{\pi}$$

$$\lim_{n\to\infty} \frac{(2n)!!}{(2n-1)!!} \times \frac{1}{\sqrt{n}} = \sqrt{\pi}$$

$$\lim_{n\to\infty} \frac{\left((2n)!!\right)^2}{(2n)!} \frac{1}{\sqrt{n}} = \sqrt{\pi}$$

$$\lim_{n\to\infty} \frac{2^{2n}(n!)^2}{(2n)!} \times \frac{1}{\sqrt{n}} = \sqrt{\pi}$$

Proof

$$\alpha = \lim_{n\to\infty} \frac{n!\, e^n}{n^n \sqrt{n}} = \lim_{n\to\infty} \frac{(2n)!\, e^{2n}}{(2n)^{2n}\sqrt{2n}}$$

Square the 1st expression & divide by the latter to get,

$$\alpha = \lim_{n\to\infty} \frac{(n!)^2 \cancel{e^{2n}}}{n^{2n}\, n} \times \frac{(2n)^{2n}\sqrt{2n}}{(2n)!\cancel{e^{2n}}}$$

$$= \lim_{n\to\infty} \frac{(n!)^2\, 2^{2n}\sqrt{2}}{(2n)!\sqrt{n}} \Longleftarrow \sqrt{2\pi} = \lim_{n\to\infty} \frac{n!\, e^n}{n^n\sqrt{n}}$$

$$\lim_{n\to\infty} \frac{n!\, e^n}{n^n \sqrt{2n}} = \sqrt{\pi}$$

$$n! \sim \sqrt{2n\pi}\left(\frac{n}{e}\right)^n$$

Proof :3

( )

$$\ln(n!) = \ln 1 + \ln 2 + \cdots + \ln n$$

$$= \sum_{k=1}^{n} \ln k$$

$$\approx \int_1^n (\ln x)\, dx$$

$$= \left[x \ln x - x\right]_1^n$$

$$= n \ln(n) - n + 1$$

$$k = \sqrt{2n\pi}\left(\frac{n}{e}\right)^n$$

$$\ln(k) = n\ln(n) - n + \frac{1}{2}\ln(2n\pi)$$

$$= \left(n+\frac{1}{2}\right)\ln(n) - n + \frac{1}{2}\ln(2\pi)$$

$$\frac{1}{2}\ln(2\pi) = 0.91868499$$

$n \to \infty$, $\quad n+\frac{1}{2} \to n$

$$\ln(n!) \approx n\ln(n) - n + 1 \approx \ln\left(\sqrt{2n\pi}\left(\frac{n}{e}\right)^n\right)$$