$1+2+3+...+(n-1)+n$ ✎ What is the ...?
Actually $1+2+...+n$ suffices.
How about 1+2+41.7 ?
For $a_1+a_2+...+a_n$, Lagrange(1772) introduced $\sum$ notation.

$$\sum_{k=1}^{n} a_k \text{ or } \sum_{1\le k\le n} a_k .$$

In general we write $\sum_{P(k)} a_k$ to denote the sum of all a_k s.t.:

- k integer,
- $P(k)$ for predicate (property) P

Implied conjunction if several properties: $\sum_{\substack{1\le p\le N \\ p \text{ prime}}} \frac{1}{p}$

<u>Warning</u>: tangent:

- $1/p$ of numbers near N divisible by p
- $\Pr\{p|n\}=(1/p)$ for $n \approx N$

Above summation approximate number of distinct prime factors of N
Summation$\approx \ln \ln N + 0.261972128$

<u>Ex</u>: Consider *Sequential Search*

```
{Return index of first occurrence of key in r[1..n], else return -1 }
function search(key : typekey; var r : datarray) : integer;
  var i : integer;
  begin i :=1;
          while (i < n) and (key <> r[i]) do i := i + 1;  {note potential error}
          if r[i ]= n then search := i else search := -1
  end;   {search }
```

✎ How can we speed it up? Use sentinel.

```
   begin
    temp := r[n];
    r[n] := key ;
    i  := 1;
    while (key <> r [i]) do i :=i +1;
    if ((i < n) or (key = temp)) then search :=i  else search :=-1;
    r[n] := temp;
```

✎ How many times is condition of loop executed i.e., how many probes are made of r?

<u>Assume</u>: Search is successful, that is, $\exists i$ such that $r[i]=key$. Choose a r.v. which measures work done, i.e., let A_n the number of probes of r for an array of size n . We want $E[A_n]$.

$\Omega=\{key=r[1]$, $key=r[2]$ &$key \ne r[1]$,...,
$key=r[n]$ &$key \ne r[i]$ $1\le i<n$ }

$$E[A_n] = \sum_{1\le k\le n} k * \Pr\{A_n = k\}$$

<u>Assume</u>: Uniform distribution, i.e., $\Pr\{A_n=i\}=$**if** $1\le i\le n$ **then** $\frac{1}{n}$ **else** 0

$$\mathrm{E}[A_n] = \sum_{1 \le k \le n} k * \frac{1}{n} = \frac{1}{n} \sum_{1 \le k \le n} k$$ ✎ why?

Distributive Law: Let K be any finite set of integers, $\sum_{k \in K} c * a_k = c \sum_{k \in K} a_k$

$ca_0 + ca_1 = c(a_0 + a_1)$

Let $T_n = \sum_{1 \le k \le n} k$ $\mathrm{E}[A_n] = \frac{1}{n} * T_n$

✎ How do we solve $T_n = \sum_{1 \le k \le n} k$?

Evaluate the first few terms & look for a pattern.

n	1	2	3	4	5	6	7	8
T_n	1	3	6	10	15	21	28	36

No obvious pattern but they're called *triangular numbers* because of

T_4 = ...
. . .
. .
.

Solution 0: Look it up.

Sloane, *Handbook of Integer Sequences*
CRC Standard Mathematical Tables
Abramowitz & Stegun, *Handbook of Mathematical Functions*
http://www.research.att.com/~njas/sequences/index.html

Solution 1: Guess & prove by induction.

Theorem: $\sum_{1 \le k \le n} k = \frac{n(n+1)}{2}$

Proof: Basis: 0=0

IH: Assume $\sum_{1 \le k \le n} k = \frac{n(n+1)}{2}$

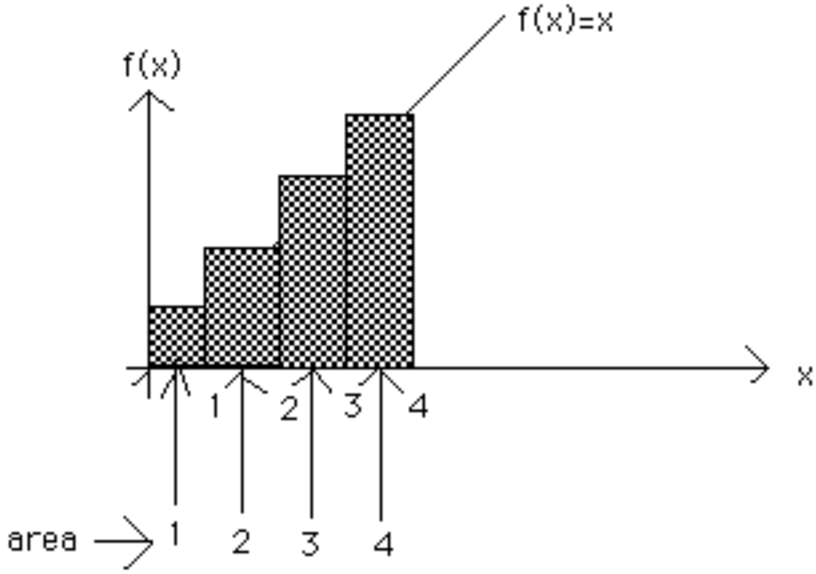
IS: $\sum_{1 \le k \le n+1} k = n+1+ \sum_{1 \le k \le n} k = n+1+\frac{n(n+1)}{2} = \frac{n^2+3n+2}{2} = \frac{(n+1)((n+1)+1)}{2}$

Solution 2: When Gauss was 9 yrs. old:

T_n = 1 + 2 + 3 +......+ $(n-1)$ + n
+T_n = n +$(n-1)$+$(n-2)$+......+ 2 + 1
$2T_n$ =$(n+1)$+$(n+1)$+$(n+1)$+.....+ $(n+1)$ + $(n+1)$ = $n(n+1)$

$$T_n = \frac{n(n+1)}{2}$$

Solution 3: Approximate $\sum$ by $\int$.



$$T_n = \sum_{1\le k\le n} k \approx \int_0^n x\,dx = \frac{x^2}{2}\Big]_0^n = \frac{n^2}{2}$$

If f is a continuous increasing function, and a & b are integers, then

$$\int_{a-1}^{b} f(x)dx \le \sum_{k=a}^{b} f(k) \le \int_{a}^{b+1} f(x)dx$$

If f is a continuous decreasing function, and a & b are integers, then

$$\int_{a}^{b+1} f(x)dx \le \sum_{k=a}^{b} f(k) \le \int_{a-1}^{b} f(x)dx$$

Returning to sequential search, E[A_n]= $\frac{1}{n}T_n = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2}$

What is VA_n? Remember that

$$VA_n = \mathrm{E}\left[A_n^2\right] - \mathrm{E}[A_n]^2 = \mathrm{E}\left[A_n^2\right] - \frac{(n+1)^2}{4}$$

$$\mathrm{E}\left[A_n^2\right] = \sum_{1\le k\le n} k^2 \Pr\{A_n = k\} = \sum_{1\le k\le n} k^2 \frac{1}{n} = \frac{1}{n}\sum_{1\le k\le n} k^2$$

✎ How do we solve $S_n = \sum_{0\le k\le n} k^2$?

Evaluate the first few terms & look for a pattern.

n	0	1	2	3	4	5	6	7	8
n^2	0	1	4	9	16	25	36	49	64
S_n	0	1	5	14	30	55	91	140	204

Solution 1: Approximate by integrals

$$S_n = \sum_{0\le k\le n} k^2 \approx \int_0^n x^2 dx = \frac{x^3}{3}\Big]_0^n = \frac{n^3}{3}$$

Solution 2: *Perturbation Method* :

-rewrite S_{n+1} by splitting off its first term

-rewrite S_{n+1} by splitting off its last term

-set these 2 expressions equal to each other and solve.

$$S_{n+1} = \sum_{0\le k\le n+1} k^2 \;=>\; 0 + \sum_{1\le k\le n+1} k^2 = (n+1)^2 + \sum_{0\le k\le n} k^2$$

We have two summations that are almost identical; if they were identical, they would cancel.

pHow do we transform $\sum_{1\le k\le n+1} k^2$ => $S_n = \sum_{0\le k\le n} k^2$?

Commutative Law: Let K be any finite set of integers. For any permutation π of the set of all integers, $\sum_{k\in K} a_k = \sum_{\pi(k)\in K} a_{\pi(k)}$

$$a_0+a_1=a_1+a_0$$

Choosing $\pi(k)=k+1$ (a permutation over the set of all integers) suggests replacing k by $k+1$ in $\sum_{1\le k\le n+1} k^2 \Rightarrow \sum_{1\le k+1\le n+1} (k+1)^2 = \sum_{0\le k\le n} (k+1)^2 = \sum_{0\le k\le n} (k^2+2k+1)$

Associative Law: Let K be any finite set of integers,

$$\sum_{k\in K} (a_k+b_k) = \sum_{k\in K} a_k + \sum_{k\in K} b_k$$

$(a_0+b_0) + (a_1+b_1) = (a_0+a_1) + (b_0+b_1)$

$$\sum_{1\le k\le n+1} k^2 = \sum_{0\le k\le n} k^2 + \sum_{0\le k\le n} 2k + \sum_{0\le k\le n} 1 = S_n + 2T_n + 1 = S_n + n^2 + 2n + 1$$

From above we had $S_n+2T_n+n+1=n^2+2n+1+S_n$ => $2T_n=n^2+n$ => $T_n = \frac{n(n+1)}{2}$

We got nowhere (on S_n), **but** note that using perturbation on $S_n = \sum_{0\le k\le n} k^2$ yielded answer for $\sum_{0\le k\le n} k$.

✐ Try perturbation method on $\sum_{0\le k\le n} k^3$ to solve $\sum_{0\le k\le n} k^2$?

$$0 + \sum_{1\le k\le n+1} k^3 = (n+1)^3 + \sum_{0\le k\le n} k^3$$

Lefthand side: Choosing $\pi(k)=k+1$ (a permutation over the set of all integers) suggests replacing k by $k+1$ in

$$\sum_{1\le k\le n+1} k^2 \Rightarrow \sum_{1\le k+1\le n+1} (k+1)^3 = \sum_{0\le k\le n} (k^3+3k^2+3k+1) = \sum_{0\le k\le n} k^3 + 3\sum_{0\le k\le n} k^2 + 3\sum_{0\le k\le n} k + (n+1)$$

$$= \sum_{0\le k\le n} k^3 + 3\sum_{0\le k\le n} k^2 + \frac{3n(n+1)}{2} + (n+1)$$

Righthand side: $n^3+3n^2+3n+1+\sum_{0\le k\le n} k^3$

Setting sides equal=>

$$3\sum_{0\le k\le n} k^2 + \frac{3n(n+1)}{2} + n + 1 = n^3+3n^2+3n+1 \Rightarrow \sum_{0\le k\le n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

Returning to original, $V(A_n)=E[A_n^2]-E[A_n]^2=E[A_n^2]-\frac{(n+1)^2}{4}$

$$=\frac{1}{n}\sum_{1\le k\le n}k^2-\frac{(n+1)^2}{4}=\frac{n^2-1}{12}$$

✎ Return to $T_n=\sum_{1\le k\le n}k$. In general, an *arithmetic progression* is $T=\sum_{0\le k\le n}(a+bk)$

Remember Gauss' trick of adding (pairwise) first & last terms,... By commutative law, replace $k \to \pi(k)=n-k$

$$T=\sum_{0\le n-k\le n}(a+b(n-k))=\sum_{0\le k\le n}(a+bn-bk)$$

By associative law, adding equalities yields

$$2T=\sum_{0\le k\le n}(a+bk+a+bn-bk)=\sum_{0\le k\le n}(2a+bn).$$

By distributive law, $2T=(2a+bn)\sum_{0\le k\le n}1=(2a+bn)(n+1)$

$T=(n+1)\frac{a+(a+bn)}{2}$ = # terms * average of first & last terms

<u>Ex</u>: Return to sequential search for another distribution. Assume r sorted by expected frequency of access & *Zipf's Distribution* $\Pr\{A_n=k\}\propto\frac{1}{k}$ or $\Pr\{A_n=k\}=\frac{\zeta}{k}$ for some (yet to be determined) ζ

<u>Instances of Zipf's Law</u>:

- populations of metropolitan areas
- word use frequency in language (averaged over body of speakers) (except Hebrew & an African language)
- word use frequency in a text for an author (used to identify authorship)
- frequency of occurrence of first digits

$$E[A_n]=\sum_{1\le k\le n}k*\Pr\{A_n=k\}=\sum_{1\le k\le n}k\frac{\zeta}{k}=\zeta\sum_{1\le k\le n}1=n\zeta$$

✎ What is ζ? From laws of probability, $\sum_{1\le k\le n}\frac{\zeta}{k}=1\Rightarrow\zeta=\frac{1}{\sum_{1\le k\le n}\frac{1}{k}}$

What is $\sum_{1\le k\le n}\frac{1}{k}$? Common enough to warrant a name: H_n, the $n^{\underline{th}}$ *Harmonic number*

(the $k^{\underline{th}}$ harmonic produced by a string instrument has wavelength $\frac{1}{k}$*fundamental tone)

$H_n\approx\ln n$. $\zeta=\frac{1}{H_n}\approx\frac{1}{\ln n}$. $E[A_n]\approx n\zeta\approx\frac{n}{\ln n}$, so search is $\approx\frac{1}{2}\ln n$ times faster.

$\ln n\le H_n\le 1+\ln n$. $\gamma=0.57721...$

$$H_n=\sum_{1\le k\le n}\frac{1}{k}=\ln n+\gamma+\frac{1}{2n}-\frac{1}{12n^2}+\frac{1}{120n^4}-\frac{1}{252n^6}+\ldots$$

where $\gamma=0.5772156649...$ (Euler's constant)

n	H_n	$\ln n+\gamma$
5	2.283333333333333	2.1866529124341003
20	3.597739657143682	3.572947273553991
50	4.499205338329423	4.489238005428146

Ex: For n /3 expected distance for head seek or element to move in random input sorting, see Pr. 4-H.W.#3-91

Ex: (Open addressng) Hashing

Given *dataarray* =**array** [0..m-1] **of** *datarecord*

datarecord =**record** ...

k :*typekey*

... **end**;

key :*typekey*

function h(*key* :*typekey*;i :integer) : 0..m-1

var r:*dataarray*

Assume that for any *key* and any i, we can distinguish between (mutually exclusive):

-$r[i]$ empty

-$r[i].k$ =*key*

-$r[i].k$ <>*key*

For *Random Hashing* , assume a sequence of independent probes distributed uniformly over [0..m-1]

h_i:[*typekey*]$\rightarrow$[0..m-1] of hashing functions, i=1,2,...

```
function search (key :typekey ;var r :dataarray ) : integer;
  var i :integer;
  begin i :=1;
        while ((not empty (r[h_i(key)]) and (r[h_i(key )].k <>key)) do
              i  := i  + 1;
        if r[h_i(key)].k=key  then search :=i
                              else search := -1
   end
```

Let r.v. n - # keys in r

A_n -# executions of **while** in successful search

A_n'-# executions of **while** in unsuccessful search

Assume: -For any *key*, i, Pr{ **not** *empty* ($r[h_i(key)].k$) }=$n/m=\alpha$

Pr{$A_n' > j$}=

Pr{ **not** *empty* ($r[h_1(key)].k$) } **&... &**Pr{ **not** *empty* ($r[h_j(key)].k$)}= α^j

(because of independence of probabilities)

Pr{$A_n' = j$}= $\alpha^{j-1}(1-\alpha)$ (Test for α=0? α=1?)

$$E[A_n'] = \sum_{0\le j} j * \Pr\{A_n' = j\} = \sum_{0\le j} j * \alpha^{j-1}(1-\alpha) = (1-\alpha)\sum_{0\le j} j * \alpha^{j-1}$$

Digression: What is $\sum_{0\le k} \alpha^k$?

Attempt 1: Let $S_n = \sum_{0\le k\le n} \alpha^k = 1 + \alpha + \alpha^2 + \ldots + \alpha^n$

$$\alpha S_n = \sum_{0\le k\le n} \alpha^{k+1} = \alpha + \alpha^2 + \ldots + \alpha^{n+1}$$

$$S_n - \alpha S_n = 1 - \alpha^{n+1}$$

$$S_n = \frac{1-\alpha^{n+1}}{1-\alpha} \quad \text{(for } \alpha \ne 1\text{)}$$

As $n \to \infty$, $S_n \to \frac{1}{1-\alpha}$ for $|\alpha|<1$.

Test with $\alpha = \frac{1}{2}$, $S_n = 1 + \frac{1}{2} + \frac{1}{4} + \ldots \to 2$

Test with $\sum_{0 \le k} \alpha * \alpha^k$, $S_n \to \frac{1}{1-\alpha}$. Let $\alpha = \frac{9}{10}, \alpha = \frac{1}{10}$

As $n \to \infty$, $S_n \to \frac{9}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \ldots =$

$$.9999\ldots = \frac{\frac{9}{10}}{1 - \frac{1}{10}} = 1$$

<u>Attempt 2</u>: (Perturbation technique)

$S_{n+1} = \sum_{0 \le k \le n+1} \alpha^k = S_n + \alpha^{n+1}$ and...

$$S_{n+1} = 1 + \sum_{1 \le k \le n+1} \alpha^k = 1 + \sum_{1 \le k+1 \le n+1} \alpha^{k+1} = 1 + \sum_{0 \le k \le n} \alpha^{k+1} = 1 + \alpha \sum_{0 \le k \le n} \alpha^k = 1 + \alpha S_n$$

$S_n + \alpha^{n+1} = 1 + \alpha S_n$ => $S_n = \frac{1-\alpha^{n+1}}{1-\alpha}$

Return from **<u>Digression</u>**:. $S_n = \sum_{0 \le k \le n} k\alpha^k$

Try the perturbation technique directly:

$$S_{n+1} = \sum_{0 \le k \le n+1} k * \alpha^k = S_n + (n+1)\alpha^{n+1}$$

$$S_{n+1} = 0 + \sum_{1 \le k \le n+1} k * \alpha^k = \sum_{1 \le k+1 \le n+1} (k+1) * \alpha^{k+1} = \sum_{0 \le k \le n} (k+1) * \alpha^{k+1}$$

$$= \alpha S_n + \sum_{0 \le k \le n} \alpha^{k+1} = \alpha S_n + \frac{\alpha(1-\alpha^{n+1})}{1-\alpha}$$

$$S_n + (n+1)\alpha^{n+1} = \alpha S_n + \frac{\alpha(1-\alpha^{n+1})}{1-\alpha}$$

....

$$S_n = \frac{\alpha - (n+1)\alpha^{n+1} + n\alpha^{n+2}}{(1-\alpha)^2} \quad \text{for } \alpha \ne 1$$

✎ What happens to S_n as $n \to \infty$? What happens to the terms $(n+1)\alpha^{n+1}$ and $n\alpha^{n+2}$? Do they grow or shrink? Look at the ratio between successive terms.

$\frac{(n+2)\alpha^{n+2}}{(n+1)\alpha^{n+1}} = \alpha \frac{n+2}{n+1} < 1$. So as $n \to \infty$, $S_n \to \frac{\alpha}{(1-\alpha)^2}$.

Returning to hashing we had

$$E[A_n'] = (1-\alpha)\sum_j j\alpha^{j-1} = \frac{1-\alpha}{\alpha}\sum_j j\alpha^j = \frac{1-\alpha}{\alpha}\frac{\alpha}{(1-\alpha)^2} = \frac{1}{1-\alpha}$$

$E[A_n]$ in H.W.#4-91-Pr.3

Products: Finite product $a_1 * a_2 * ... * a_n$ written $\prod_{1 \le k \le n} a_k$. If n=0, product defined to be 1.

$$\prod_{P(k)} a_k = e^{\sum_{P(k)} \ln a_k}$$

∞-summations:

$$S = 1 + \frac{1}{2} + \frac{1}{4} + \ldots$$

$$2S = 2 + 1 + \frac{1}{2} + \frac{1}{4} + \ldots = 2 + S$$

$S = 2$ (also follows from geometric distribution)

$T = 1 + 2 + 4 + 8 + ...$

$2T = 2 + 4 + 8 + ... = T - 1$

$T = -1$

Consider $\sum_{k \in K} a_k$, where K could be ∞.

Assume: All $a_k \ge 0$. If $\exists A$ such that $\forall$ finite $F \subset K$, $\sum_{k \in F} a_k \le A$ then we define $\sum_{k \in K} a_k$ to be the least such value A, else define $\sum_{k \in K} a_k = \infty$ (For any A, there is a finite F such that $\sum_{k \in F} a_k \ge A$.) If K={0,1,2,...}, $\sum_{k \in K} a_k = \lim_{n \to \infty} \left(\sum_{0 \le k \le n} a_k \right)$. For example,

$$\sum_{0 \le k} x^k = \lim_{n \to \infty} \frac{1 - x^{n+1}}{1 - x} = \begin{cases} \frac{1}{1-x} \text{ if } 0 \le x < 1 \\ \infty \text{ otherwise} \end{cases}$$

Remove above assumption: $\sum_{0 \le k} (-1)^k$ =1-1+1-1+1-...

=(1-1) + (1-1) + ... = 0

=1 - (1-1) - (1-1) -... = 1

If the series is *absolutely convergent* (that is, $\sum_{0 \le k} |a_k|$ converges), then its terms can be added in any order.