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

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

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:

$$\sum_{1 \leq p \leq N \text{ prime}} \frac{1}{p}.$$

---

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

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

---

**Ex**: 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;`

```pascal
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.

```pascal
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 $r$?

**Assume**: 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 $E[A_n]$. 

$\Omega=\{key=r[1], key=r[2] \& key \neq r[1], \ldots, key=r[n] \& key \neq r[1], 1 \leq i < n\}$

$$E[A_n] = \sum_{1 \leq k \leq n} k \cdot \Pr\{A_n = k\}$$

Assume: Uniform distribution, i.e., \( \Pr\{A_n = i\} = \frac{1}{n} \) if \( 1 \leq i \leq n \) then \( \frac{1}{n} \) else 0

$$E[A_n] = \sum_{1 \leq k \leq n} k \cdot \frac{1}{n} = \frac{1}{n} \sum_{1 \leq k \leq n} k \quad \text{why?}$$

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

\[ ca_0 + ca_1 = c(a_0 + a_1) \]

Let \( T_n = \sum_{1 \leq k \leq n} k \quad E[A_n] = \frac{1}{n} \cdot T_n \)

\( \text{How do we solve} \quad T_n = \sum_{1 \leq k \leq n} k? \)

Evaluate the first few terms & look for a pattern.

<table>
<thead>
<tr>
<th>( n )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_n )</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>10</td>
<td>15</td>
<td>21</td>
<td>28</td>
<td>36</td>
</tr>
</tbody>
</table>

No obvious pattern but they're called *triangular numbers* because of \( T_4 = \ldots \ldots \)

\[ \ldots \]

**Solution 0**: Look it up.

Sloane, *Handbook of Integer Sequences*

**Solution 1**: Guess & prove by induction.

**Theorem**: \( \sum_{1 \leq k \leq n} k = \frac{n(n+1)}{2} \)

**Proof**: Basis: 0 = 0

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

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

**Solution 2**: When Gauss was 9 yrs. old:

\( T_n = 1 + 2 + 3 + \ldots + (n-1) + n \)
Solution 3: Approximate $\sum$ by $\int$.

If $f$ is a continuous increasing function, and $a$ & $b$ are integers, then
\[
\int_{a-1}^{b} f(x) \, dx \leq \sum_{k=a}^{b} f(k) \leq \int_{a}^{b+1} f(x) \, dx
\]

If $f$ is a continuous decreasing function, and $a$ & $b$ are integers, then
\[
\int_{a}^{b+1} f(x) \, dx \leq \sum_{k=a}^{b} f(k) \leq \int_{a-1}^{b} f(x) \, dx
\]

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

What is $VA_n$? Remember that
\[
VA_n = E[A_n^2] - E[A_n]^2 = E[A_n^2] - \frac{(n+1)^2}{4}
\]
\[ \mathbb{E}[A_n^2] = \sum_{1 \leq k \leq n} k^2 \Pr(A_n = k) = \sum_{1 \leq k \leq n} \frac{k^2}{n} = \frac{1}{n} \sum_{1 \leq k \leq n} k^2 \]

How do we solve \( S_n = \sum_{0 \leq k \leq n} k^2 \)?

Evaluate the first few terms & look for a pattern.

\[
\begin{array}{cccccccccc}
 n & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
n^2 & 0 & 1 & 4 & 9 & 16 & 25 & 36 & 49 & 64 \\
S_n & 0 & 1 & 5 & 14 & 30 & 55 & 91 & 140 & 204 \\
\end{array}
\]

Solution 1: Approximate by integrals

\[ S_n = \sum_{0 \leq k \leq n} k^2 \approx \int_0^n x^2 \, dx = \left[ \frac{x^3}{3} \right]_0^n = \frac{n^3}{3} \]

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.

\[
S_{n+1} = \sum_{0 \leq k \leq n+1} k^2 = 0 + \sum_{1 \leq k \leq n+1} k^2 = (n+1)^2 + \sum_{0 \leq k \leq n} k^2
\]

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

How do we transform \( \sum_{1 \leq k \leq n+1} k^2 \Rightarrow S_n = \sum_{0 \leq k \leq n} k^2 \)?

**Commutative Law**: Let \( K \) be any finite set of integers. For any permutation \( \pi \) 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

\[
\sum_{1 \leq k \leq n+1} k^2 \Rightarrow \sum_{1 \leq k+1 \leq n+1} (k+1)^2 = \sum_{0 \leq k \leq n} (k+1)^2 = \sum_{0 \leq k \leq n} (k^2 + 2k + 1)
\]

**Associative Law**: Let \( K \) be any finite set of integers,

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

\[ (a_0 + b_0) + (a_1 + b_1) = (a_0 + a_1) + (b_0 + b_1) \]

\[
\sum_{1 \leq k \leq n+1} k^2 = \sum_{0 \leq k \leq n} k^2 + \sum_{0 \leq k \leq n} 2k + \sum_{0 \leq k \leq n} 1 = S_n + 2T_n + 1 = S_n + n^2 + 2n + 1
\]

From above we had \( S_n + 2T_n + n + 1 = n^2 + 2n + 1 + S \Rightarrow 2T_n = n^2 + n \Rightarrow T_n = \frac{n(n+1)}{2} \)
We got nowhere (on $S_n$), but note that using perturbation on $S_n = \sum_{0 \leq k \leq n} k^2$ yielded answer for $\sum_{0 \leq k \leq n} k$.

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

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

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

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

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

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

Setting sides equal$\Rightarrow$

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

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

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

Return to $T_n = \sum_{1 \leq k \leq n} k$. In general, an arithmetic progression is

$$T = \sum_{0 \leq k \leq n} (a + bk)$$

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

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

By associative law, adding equalities yields

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

By distributive law,

$$2T = (2a + bn) \sum_{0 \leq k \leq n} 1 = (2a + bn)(n+1)$$

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

Ex: 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) $\zeta$

Instances of Zipf's Law ___:
- 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  

\[ \mathbb{E}[A_n] = \sum_{1 \leq k \leq n} k \cdot \Pr\{A_n = k\} = \sum_{1 \leq k \leq n} k \frac{\zeta}{k} = \zeta \sum_{1 \leq k \leq n} 1 = n \zeta \]

What is \( \zeta \)? From laws of probability, \[ \sum_{1 \leq k \leq n} \frac{\zeta}{k} = 1 \Rightarrow \zeta = \sum_{1 \leq k \leq n} \frac{1}{k} \]

What is \[ \sum_{1 \leq k \leq n} \frac{1}{k} \] ? Common enough to warrant a name: \( H_n \), the \( n \)th Harmonic number (the \( k \)th harmonic produced by a string instrument has wavelength \( \frac{1}{k} \) fundamental tone)  

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

\[ \ln n \leq H_n \leq 1 + \ln n \]

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \ln n + \gamma )</th>
<th>( H_n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2.28333333333333</td>
<td>2.1866529124341003</td>
</tr>
<tr>
<td>20</td>
<td>3.597739657143682</td>
<td>3.572947273553991</td>
</tr>
<tr>
<td>50</td>
<td>4.499205338329423</td>
<td>4.489238005428146</td>
</tr>
</tbody>
</table>

**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 addressing) Hashing  

Given \( \text{dataarray} = \text{array [0..m-1]} \) of \( \text{datarecord} \)  

\( \text{datarecord} = \text{record} ...  
  k : \text{typekey} \)  
  ... end;  

\( \text{key} : \text{typekey} \)  

\( \text{function} \ h(\text{key} : \text{typekey}; i : 0..m-1) : 0..m-1 \)  

\( \text{var} \ r, \text{dataarray} \)  

Assume that for any \( \text{key} \) and any \( i \), we can distinguish between (mutually exclusive):  
- \( r[i] \) empty  
- \( r[i].k = \text{key} \)  
- \( r[i].k <> \text{key} \)  

For Random Hashing, assume a sequence of independent probes distributed uniformly over [0..m-1]
\text{Digit}: [\text{typekey}] \rightarrow \ [0..m-1] \text{ of hashing functions, } i=1,2,\ldots

\textbf{function} \ search (key :\text{typekey}; \var r :\text{dataarray}) : \text{integer} ;
\var i : \text{integer};
\begin{align*}
&\text{begin} \ i := 1; \\
&\quad \text{while } (\text{not empty}(r[h_i(key)]) \ \text{and} \ (r[h_i(key)].k <> key)) \ \text{do} \\
&\quad \quad \ i := i + 1; \\
&\quad \text{if} \ r[h_i(key)].k = key \ \text{then} \ \text{search} := i \\
&\quad \ \text{else} \ \text{search} := -1
\end{align*}
\text{end}

\text{Let r.v. } n - \text{# keys in } r
\begin{align*}
A_n \text{-# executions of while in successful search} \\
A'_n \text{-# executions of while in unsuccessful search}
\end{align*}

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

\(Pr\{A'_n > j\} = \alpha^j \cdot (1 - \alpha)\) (Test for \(\alpha = 0?\ \alpha = 1?)\)

\(E[A'_n] = \sum_{0 \leq j} j^* Pr[A'_n = j] = \sum_{0 \leq j} j^* \alpha^{j-1} (1 - \alpha) = (1 - \alpha) \sum_{0 \leq j} j^* \alpha^{j-1}\)

\textbf{Digression} \ : \text{What is } \sum_{0 \leq k} \alpha^k ?

\textbf{Attempt 1} \ : \text{Let } S_n = \sum_{0 \leq k \leq n} \alpha^k = 1 + \alpha + \alpha^2 + \ldots + \alpha^n
\begin{align*}
\alpha \quad S_n &= \sum_{0 \leq k \leq n} \alpha^{k+1} = \alpha + \alpha^2 + \ldots + \alpha^n + 1 \\
S_n - \alpha \quad S_n &= 1 - \alpha \ n + 1 \\
S_n &= \frac{1 - \alpha^{n+1}}{1 - \alpha} \text{ (for } \alpha \neq 1) \\
\text{As } n \rightarrow \infty, \ S_n \rightarrow \frac{1}{1 - \alpha} \text{ for } |\alpha| < 1.
\end{align*}

\text{Test with} \quad \alpha = \frac{1}{2}, \quad S_n = 1 + \frac{1}{2} + \frac{1}{4} + \ldots \rightarrow 2

\text{Test with} \quad \sum_{0 \leq k} \alpha^* \alpha^k, \ S_n \rightarrow \frac{1}{1 - \alpha}. \text{ Let } \alpha = \frac{9}{10}, \alpha = \frac{1}{10}

\text{As } n \rightarrow \infty S_n \rightarrow \frac{9}{10} + \frac{9}{10^2} + \frac{9}{10^3} + \ldots = \frac{9}{1 - \frac{9}{10}} = 1

\text{\textbf{Attempt 2} \ : (Perturbation technique)}}
\[ S_{n+1} = \sum_{0 \leq k \leq n+1} \alpha^k = S_n + \alpha^{n+1} \text{ and...} \]

\[ S_{n+1} = 1 + \sum_{1 \leq k \leq n+1} \alpha^k = 1 + \sum_{0 \leq k \leq n} \alpha^k = 1 + \alpha \sum_{0 \leq k \leq n} \alpha^k = 1 + \alpha S_n \]

\[ S_n + \alpha^{n+1} = 1 + \alpha S_n \implies S_n = \frac{1 - \alpha^{n+1}}{1 - \alpha} \]

Return from **Digression** :- \[ S_n = \sum_{0 \leq k \leq n} k \alpha^k \]

Try the perturbation technique directly:

\[ S_{n+1} = \sum_{0 \leq k \leq n+1} k \alpha^k = S_n + (n+1) \alpha^{n+1} \]

\[ S_{n+1} = 0 + \sum_{1 \leq k \leq n+1} k \alpha^k = \sum_{1 \leq k \leq n+1} (k+1) \alpha^{k+1} = \sum_{0 \leq k \leq n} (k+1) \alpha^{k+1} \]

\[ = \alpha S_n + \sum_{0 \leq k \leq n} \alpha^{k+1} = \alpha S_n + \frac{\alpha(1 - \alpha^{n+1})}{1 - \alpha} \]

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

\[ \ldots \]

\[ S_n = \alpha - (n+1) \alpha^{n+1} + n \alpha^{n+2} \quad \text{for } \alpha \neq 1 \]

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

Returning to hashing we had

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

\[ E[A_n] \text{ in H.W.#4-91-Pr.3} \]

**Products** :- Finite product \( a_1 a_2 \ldots a_n \) written \( \prod_{1 \leq k \leq n} a_k \). If \( n=0 \), product defined to be 1.

\[ \prod_{P(k)} a_k = e^{\sum_{1 \leq k \leq n} \ln a_k} \]

**Bounding Summations** :- \( \sum_{1 \leq k \leq n} k^4 \leq \sum_{1 \leq k \leq n} n^4 = n^5 \)

If \( a_{\max} = \max_{1 \leq k \leq n} a_k \), then \( \sum_{1 \leq k \leq n} k^4 \leq na_{\max} \)

Apply this to \( \sum_{0 \leq k \leq \infty} k^3 = 0 + \frac{1}{3} + \frac{2}{9} + \frac{3}{27} + \ldots \) with \( a_{\max} = 1/3 \),
yielding \[ \sum_{1 \leq k \leq \infty} \frac{k}{3^k} \leq \frac{\infty}{3}. \] Note that for \( k \geq 1, \)
\[
\frac{a_{k+1}}{a_k} = \frac{1}{3} \left( \frac{k+1}{k} \right) \leq \frac{2}{3}.
\]
Note: If \( r \) \( \exists \) \( a_{k+1} \leq r \) for all \( k \geq 0 \), then
\[
\sum a_k \leq \sum a_0 r^k = a_0 \sum r^k = a_0 \frac{1}{1-r} \quad (\text{likewise for finite sums})
\]
So \[
\sum_{0 \leq k \leq \infty} \frac{k}{3^k} \leq \sum_{0 \leq k \leq \infty} \frac{1}{3} \left( \frac{2}{3} \right)^k = \frac{1}{3} \left( \frac{1}{1-\frac{2}{3}} \right) = 1.
\]

---

\( \infty \)-summations:

\[
S = 1 + \frac{1}{2} + \frac{1}{4} + \ldots
\]
\[
2S = 2 + \frac{1}{2} + \frac{1}{4} + \ldots = 2 + S
\]
\[
S = 2 \quad (\text{also follows from geometric distribution})
\]
\[
T = 1 + 2 + 4 + 8 + \ldots
\]
\[
2T = 2 + 4 + 8 + \ldots = T - 1
\]
\[
T = -1
\]
Consider \( \sum_{k \in K} a_k \), where \( K \) could be \( \infty \).

Assume: All \( a_k \geq 0 \). If \( \exists \ A \) such that \( \forall \) finite \( F \subset K \), \( \sum_{k \in F} a_k \leq A \) 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 \( \sum_{k \in F} a_k \geq A \).) If \( K = \{0,1,2,\ldots\} \), \( \sum_{k \in K} a_k = \lim_{n \to \infty} \left( \sum_{0 \leq k \leq n} a_k \right) \).

For example, \( \sum_{0 \leq k} x^k = \lim_{n \to \infty} \frac{1 - x^{n+1}}{1-x} = \begin{cases} \frac{1}{1-x} & \text{if } 0 \leq x < 1 \\ \infty & \text{otherwise} \end{cases} \)

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

If the series is \( \text{absolutely convergent} \) (that is, \( \sum |a_k| \) converges), then its terms can be added in any order.