Crystal production: integration OR summation?

(Written by Bernard L.W.L., QC Old Boy)

Situation: Assume workers are employed by a company to mine crystals in a project. Initially, there is only 1 worker on the 1st day. From the 2nd day onwards, a worker is added per day to involve in the mining project so as to increase the rate of production. If there is an increase of k in the rate of production in every subsequent day (in units of *crystals/worker/day*), (assume the workers' overall productivities do not diminish such that production rate would constantly be on the rise)
express the total no. of crystals produced from the 1st day to a certain day (t^{th} day)
in terms of suitable variables.

Approach 1:
Assume we plot the graph of no. of the crystals mined (y) against the no. of workers employed (x).
Let the number of crystals the first worker mined out on the 1st day be m, which in other words means the (daily) rate of production of this worker is m *crystals/worker/day*. Provided that there is an increase in production rate of k for the two workers on the 2nd day, let y_n be the number of crystals produced on the n^{th} day, then for the 2nd and 3rd days,

$\frac{y_2}{2} = m + k$, $\frac{y_3}{3} = \frac{y_2}{2} + k = (m+k) + k = m + 2k$

=> $y_2 = 2(m+k)$ => $y_3 = 3(m+2k)$

Similarly, for the n^{th} day,

$$\frac{y_n}{n} = m + (n-1)k$$

$$\Rightarrow y_n = n(m + (n-1)k)$$

Therefore, the total number of the crystals yield from the 1st day to t^{th} day is given by the summation (by rectangle rule)

$$\sum_{n=1}^{t} y_n = y_1 + y_2 + ... + y_t$$

$$= m + 2(m+k) + 3(m+2k) + ... + t(m+(t-1)k)$$

$$= (1+2+...+t)m + (1\cdot 2 + 2\cdot 3 + ... + (t-1)\cdot t)k ......(1)$$

$$= \frac{t(t+1)}{2}m + \sum_{n=1}^{t} n(n-1)k$$

$$=\frac{t(t+1)}{2}m+k\sum_{n=1}^{t}n^2-k\sum_{n=1}^{t}n$$
$$=\frac{t(t+1)}{2}m+\frac{t(t+1)(2t+1)}{6}k-\frac{t(t+1)}{2}k \qquad (\because \sum_{i=1}^{n}i^2=\frac{n(n+1)(2n+1)}{6}, n\in\mathbb{N})$$
$$=\frac{t(t+1)}{2}(m-k)+\frac{t(t+1)(2t+1)}{6}k......(2)$$

Alternatively,

From the above discussion, $y_n=n(m+(n-1)k)$,

Define the function $f(t)=t(m+(t-1)k)$ for the crystal mined on the t^{th} day.

$f(2)-f(1)=2(m+k)-m=m+2k$
$f(3)-f(2)=3(m+2k)-2(m+k)=m+4k$
Similarly, $\Delta f(t)=f(t)-f(t-1)=m+2(t-1)k$
Thus, the total crystal yield throughout the period required is given by:

$$\sum_{i=1}^{t}f(i)=f(1)+f(2)+...+f(t-1)+f(t)$$
$$=m+(m+m+2k)+(m+m+2k+m+4k)+...$$
$$+(m+m+2k+...+m+2((t-1)-1)k)$$
$$+(m+m+2k+...+m+2((t-1)-1)k+m+2(t-1)k)$$
$$=(1+2+3+...+t)m+(2+(2+4)+...+(2+4+...+2((t-1)-1)+2(t-1))k$$
$$=(1+2+3+...+t)m+2(1+(1+2)+...+(1+2+...+(t-2)+(t-1))k$$
$$=(1+2+3+...+t)m+2(\frac{(1)(1+1)}{2}+\frac{(2)(1+2)}{2}+...+\frac{(t-1)(1+(t-1))}{2})k$$
$$=(1+2+3+...+t)m+(1\cdot 2+2\cdot 3+...+(t-1)\cdot t)k=(1)$$

BUT here, since we are more eager to find $\sum f(i)$, the above steps may seem to be a lot more tedious and even unnecessary when compared to the former one.
The above calculations, however, are here to serve as a reference to verify (1) while **avoid applying the general formula** $y_n=n(m+(n-1)k)$ **directly.**
Indeed, below is an alternative way to simplify $(1\cdot 2+2\cdot 3+...+(t-1)\cdot t)k$.
Let $u(x)=(x-1)\cdot x$ and $v(x)=(x-1)\cdot x\cdot(x+1)$,*(the trick is done by letting* $u(x)$ *as the original function and* $v(x)$ *by attaching a subsequent factor to it)*
By considering $v(x-1)=(x-2)\cdot(x-1)\cdot x$,

$$v(x)-v(x-1)=(x-1)\cdot x\cdot(x+1)-(x-2)\cdot(x-1)\cdot x$$
$$=(x-1)\cdot x\cdot((x+1)-(x-2))$$
$$=3(x-1)\cdot x$$
$$=3u(x)$$

Taking summation on both sides,

$$\sum_{x=1}^{t}(v(x)-v(x-1))=3\sum_{x=1}^{t}u(x)$$

$$v(t)-v(0)=3\sum_{x=1}^{t}u(x)$$

$$=>\sum_{x=1}^{t}u(x)=\frac{1}{3}\left[(t-1)\cdot t\cdot(t+1)-0\right] \qquad \text{(Take } u(1)=v(1)=v(0)=0\text{)}$$

$$=\frac{t(t-1)(t+1)}{3}$$

(Readers may try to carry on the simplification in order to obtain the expression in (2).)

This method, despite being a bit more complicated, wouldavoid using the formula $\sum_{i=1}^{n}i^2=\frac{n(n+1)(2n+1)}{6}$ **.**

Approach 2:

For the crystal-worker graph plotted, the graph would be concaved upwards through the origin (no crystals at the start, $t=0$) because $\frac{d^2y_t}{dt^2}=2k>0$ (by 2nd derivative test) $(\because k>0)$.

This also agrees with the given fact that the production rate, or $\frac{dy_t}{dt}$, is increasing continuously (in the range $[0,\infty]$).

The total yield of crystal from the beginning till the t^{th} day is given by the definite integral $\int_0^t f(t)dt$ *(this integral takes into account the total yield as of the end of* t^{th} *day, OR the beginning of* $(t+1)^{th}$ *day).*

$$=\int_0^t t(m+(t-1)k)dt$$

$$=m\int_0^t tdt+k\int_0^t t(t-1)dt$$

$$=m\left[\frac{t^2}{2}\right]_0^t+k\left[\frac{t^3}{3}-\frac{t^2}{2}\right]_0^t$$

$$=\frac{t^2}{2}(m-k)+\frac{t^3}{3}k......(3)$$

Contradictions would arise because (2) and (3) are apparently not equal as long as $k \neq 0$,
So which one should be taken as the total crystal yield?

The reason why (2) differs from (3) is because of the different manners in which both methods rely upon in finding the sum of crystals throughout the t days.

Recall the definition of definite integral,
For a continuous function $f(x)$ in the region $[a,b]$ ($a \neq b$), if the area under the graph from the region is divided into n rectangles (similar to Fig. 1), let $\Delta x = \frac{b-a}{n}$,
then the definite integral

$$\int_a^b f(x)dx = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i)\Delta x$$

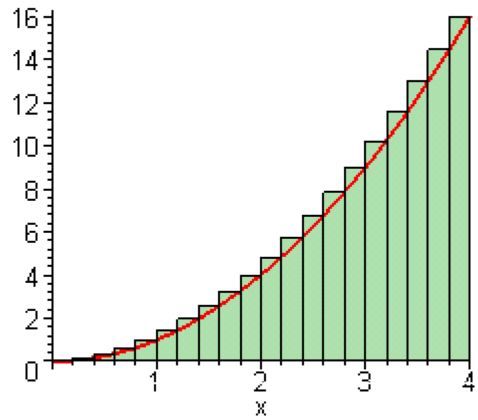

Fig. 1: dividing area below $f(x) = x^2$ into rectangles

For Approach 1, the evaluated sum is not exactly equal to the area under graph $f(t)$ from $x = 0$ to $x = t$, since the crystal yield is counted daily, we regard the sum as the total area of the t rectangles, where each length sub-interval $\Delta x = 1$,
Thus the sum in approach 1 can be interpreted as

$$\lim_{\Delta x \to 1} \sum_{i=1}^{t} f(x_i)\Delta x = \sum_{i=1}^{t} f(x_i) = f(x_1) + f(x_2) + ... + f(x_t) = y_1 + y_2 + ... + y_t = \sum_{n=1}^{t} y_n$$

But this is not the same as the definite integral $\int_0^t f(t)dt$ because Δx does not tend to be zero but is instead equal to 1.
As for Approach 2, the evaluated sum refers to the **exact** area under the graph from $x = 0$ to $x = t$ because we have taken $\Delta x \to 0$.
This explains why we arrive at different results when using approaches 1 and 2.
Here comes the more important question to ask, which approach comes with the correct crystal yield and why?

Approach 1 is the right one to adopt when we consider the case here where workers are the factor of production being used.
In fact, worker is a **discrete quantity** because workers must be counted in terms of integers (zero included), which is opposite to **continuous quantity** where the set covers all real numbers (such as decimals).
Because the sum we worked out in Approach 1 concerns the output value (yield of crystals) only when the independent variable (worker added on each day) is a positive integer (or natural numbers), thus it shows the exact crystal yield when it concerns the summation of y_n from $n = 1$ to $n = t$, where $n \in \mathbb{N}$ (the set of natural numbers also implies the discrete nature of workers).

Approach 2 would be feasible when the factor of production used is of continuous nature. Examples could be bacteria used in the fermentation of bread, as its amount would continuously increase from time to time, <u>**assuming that**</u> it follows the data given where the rate of production must change with <u>**the same increment**</u> when every day is passed.

From (2) in Approach 1, the crystal yield from the given period,
is found to be

$$\frac{t(t+1)}{2}(m-k)+\frac{t(t+1)(2t+1)}{6}k$$

$$=\frac{t^2+t}{2}(m-k)+\frac{2t^3+3t^2+t}{6}k$$

$$=\left(\frac{t^2}{2}+\frac{t}{2}\right)(m-k)+\left(\frac{t^3}{3}+\frac{t^2}{2}+\frac{t}{6}\right)k$$

$$=\left(\frac{t^2}{2}(m-k)+\frac{t^3}{3}k\right)+\frac{t}{2}(m-k)+\left(\frac{t^2}{2}+\frac{t}{6}\right)k$$

$$=(3)+\frac{t}{2}(m-k)+\left(\frac{t^2}{2}+\frac{t}{6}\right)k$$

Since $k, t > 0$, thus <u>(2) > (3)</u>.
However, some readers may cast a doubt on the case
when $m-k>0$, or $m>k$ is not true, then
(2) > (3) may not possibly stand.
Yet, for this case where the yield is on an increasing trend,
$m-k>0$ is<u>not</u> a must for the statement to stand.
(Readers may refer to Page 8 to study the particular case where it IS a must.)

For proving this fact, we can make use of the integral from the 1st day to $(t+1)^{th}$ day,

$$\int_0^{t+1} f(t)dt$$

$$= \int_0^{t+1} t(m+(t-1)k)dt$$

$$= m\left[\frac{t^2}{2}\right]_0^{t+1} + k\left[\frac{t^3}{3}-\frac{t^2}{2}\right]_0^{t+1}$$

$$= m\frac{(t+1)^2}{2} + k\left(\frac{(t+1)^3}{3}-\frac{(t+1)^2}{2}\right)$$

$$= m\frac{t^2+2t+1}{2} + k\frac{t^3+3t^2+3t+1}{3} - k\frac{t^2+2t+1}{2}$$

$$= \left(\frac{t^2}{2}(m-k)+\frac{t^3}{3}k\right) + \frac{2t+1}{2}m + \frac{6t^2-1}{6}k$$

$$= (3) + \frac{2t+1}{2}m + \frac{6t^2-1}{6}k$$

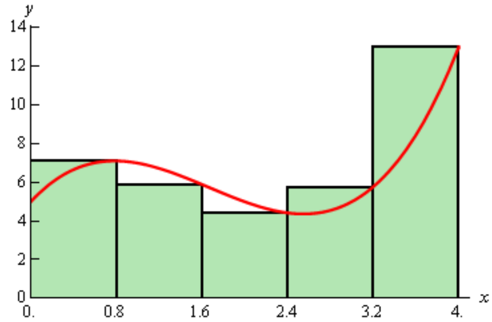

Fig. 2: Concaving upwards gives an overestimation while concaving downwards gives an underestimation.

Since $\int_0^{t+1} f(t)dt > \int_0^{t} f(t)dt$ for $f(t) > 0$,

thus $\frac{2t+1}{2}m + \frac{6t^2-1}{6}k > 0$ (considering $m,k>0$ and $t \geq 1$ ($\frac{6t^2-1}{6}k > 0$)).

Therefore(2) > (3) for all positive values of m and k .

The summation in Approach 1 is <u>overestimated</u>.

The absolute error of using rectangle rule is given by

$$|\text{Error}| = \frac{t}{2}(m-k) + \left(\frac{t^2}{2}+\frac{t}{6}\right)k \;=\; \underline{\underline{\frac{t}{2}m + \frac{t(3t-2)}{6}k}}$$

(2) and (3) together actually verify the fact that
when the graph of function is concaved upwards,
($\frac{d^2y}{dx^2} > 0$ within the region $[a,b]$ ($a,b \in \square$))
the area of rectangle(s) divided in this region exceeds
the area of portion under graph within this region
(or area of trapezoid when 'Trapezoidal Rule' is used)
And vice versa (See Fig. 2).

By using numerical methods such as trapezoidal rule,
Take the length sub-interval of x-axis, as $\Delta x = 1$, lower limit of integral (a)

and upper limit of integral (b) be 0 and t respectively,
Thus by estimating the integral of $f(t)$ with t sub-intervals,

$$\int_0^t f(t)dt$$
$$=\frac{\Delta x}{2}\left[f(0)+2f(1)+...+2f(t-1)+f(t)\right]$$
$$=\frac{(1)}{2}\left[0+2m+2(2(m+k))+2(3(m+2k))+...+2((t-1)(m+((t-1)-1)k)+t(m+(t-1)k)\right]$$
$$=(1+2+3+...+(t-1))m+(1\cdot 2+2\cdot 3+...+(t-2)\cdot(t-1))k+\frac{t(m+(t-1)k)}{2}$$
$$=\frac{(t-1)(1+(t-1))}{2}m+\left(\sum_{n=1}^{t}n(n-1)k-t\cdot(t-1)k\right)+\frac{t}{2}m+\frac{t(t-1)}{2}k$$
$$=\frac{t(t-1)}{2}m+\frac{t(t+1)(2t+1)}{6}k-\frac{t(t+1)}{2}k-\frac{t(t-1)}{2}k+\frac{t}{2}m \qquad \textbf{(From Approach 1)}$$
$$=\frac{t^2}{2}m+\frac{2t^3k-3t^2k+tk}{6}...(4)$$

In fact, (2) can also be expressed as:

$$\frac{t(t+1)}{2}(m-k)+\frac{t(t+1)(2t+1)}{6}k$$
$$=\frac{t^2}{2}m+\frac{t}{2}m-\frac{t(t+1)}{2}k+\frac{t(t+1)(2t+1)}{6}k$$
$$=\frac{t^2}{2}m+\frac{2t^3k-2tk+3mt}{6}$$

$$\because -3t^2k+tk+2tk=-3t(t-1)k<3mt(\because t>1)(m,k>0)$$
$$=>-3t^2k+tk<3mt-2tk$$
$$=>\frac{t^2}{2}m+\frac{2t^3k-3t^2k+tk}{6}<\frac{t^2}{2}m+\frac{2t^3k-2tk+3mt}{6}$$
$$\therefore (3)>(4).$$

Since both (3) and (4) are overestimates of $\int_0^t f(t)dt$,

$\therefore$ **(4) is therefore a better estimate** of the integral.
Thus, using **trapezoidal rule would give better approximations** than using rectangle rule when we consider dividing the area under the graph into the**same** number of sub-intervals.

For a graph of function to be concaved downwards $(f''(x) < 0)$,
there would be <u>two</u> possible cases for the scenario:

(1) The graph decreases from a positive value at an increasing rate;

(2) The graph increases from an initial value (zero/+ve) at an decreasing rate.

As it is more reasonable for the real-life case to have an increasing crystal yield on increasing the number of workers employed,
we would adopt the latter case (assumed with zero initial value for convenience).

The situation for such a case is more or less the same as that of the previous one, only that $f(t) = t(m-(t-1)k)$ instead of $f(t) = t(m+(t-1)k)$.

$(f''(t) = -2k < 0)$

$\because$ For the decreasing trend to become possible,

$t(m-(t-1)k) > 0$

$t > 0$ or $m-(t-1)k > 0$

$m > (t-1)k \geq k \ (t > 1, t \in \square)$

$\Rightarrow m > k$

<u>Unlike</u> the previous case (where $m > k$ is <u>not</u> necessary), $m > k$ **<u>must be</u>** true here such that the trend can happen.

Thus the total crystal yield required

$$= \int_0^t t(m-(t-1)k)dt$$

$$= m\int_0^t tdt - k\int_0^t t(t-1)dt$$

$$= m\left[\frac{t^2}{2}\right]_0^t - k\left[\frac{t^3}{3} - \frac{t^2}{2}\right]_0^t$$

$$= \frac{t^2}{2}(m+k) - \frac{t^3}{3}k$$

(Here, readers may try to evaluate the integrals by using rectangle rule and verify the result being an underestimated one, as well as using trapezoidal rule to, once again, show that it effectively minimizes the error when compared with many of the other numerical methods.)

The above results, no matter derived from rectangle rule or trapezoidal rule, are <u>NOT</u> absolutely accurate in finding the exact value of the definite integral $\int_0^t f(t)dt$.

Only when there are more sub-intervals divided in-between the upper and lower limits of integral, the approximation will get closer and closer to the exact value ($\Delta x \approx 0$).

In the following, we will try to consider dividing every single day into s **equal** sub-intervals such that $s\Delta x = 1, \Delta x = \frac{1}{s}$, and then obtain the most accurate value by taking the limit when s tends to be positive infinity ($s \to +\infty$).

Yet, within one day, the addition of every single worker in each sub-interval only leads to an **increase of only a part of** k such that after one day, the production rate would increase by k as stated in the situation.

In other words, every worker would help increase the production rate by: $\frac{k}{s}$.

$\therefore$ When the first worker is present, production rate = m

When the second worker is added, production rate = $m + \frac{k}{s}$

When the third worker is added, production rate = $m + \frac{k}{s} + \frac{k}{s} = m + \frac{2k}{s}$

For the s^{th} worker being added, production rate = $m + \frac{(s-1)k}{s}$

Also, for a random x^{th} worker being added, production rate = $m + \frac{(x-1)k}{s}$

$\therefore$ The production rate would become $m + k$ when the $(s+1)^{th}$ worker is added.

At the same time, the 1st day since the start is passed.

$\because$ We want to find the sum of all the crystals produced by the workers from the 1st day to the t^{th} day,

The last worker (the first worker on the t^{th} day) is the $(s(t-1)+1)^{th}$ one out of all.

$\therefore$ At this point, the production rate would end up as: $m + (t-1)k$.

Since the increase of s workers in this case is equal to an increase in productivity due to an increase of 1 worker in the original case,

Thus, on the addition of the second worker, productivity would equal to $1 + \frac{1}{s} = \frac{s+1}{s}$ of 1 worker in the original case.

On the addition of the third worker, productivity would equal to $\frac{s+1}{s} + \frac{1}{s} = \frac{s+2}{s}$ of 1 worker in the original case.

Similarly, for a random x^{th} worker being added, productivity would equal to $\frac{s+(x-1)}{s}$ of 1 worker in the original case.

Therefore, substitute $x=\frac{s+(x-1)}{s}$ into $f(x)=x(m+(x-1)k)$,

$$=> f(x)=\left(\frac{s+x-1}{s}\right)\left(m+\frac{x-1}{s}k\right).$$

The definite integral we want to find would be: $\int_{0}^{s(t-1)+1} f(x)dx$ (I).

Using the definition of definite integral,

$$I=\int_{0}^{s(t-1)+1} f(x)dx$$

$$=\lim_{s\to+\infty}\sum_{i=1}^{s(t-1)+1} f(x_i)\Delta x$$

$$\because \Delta x=\frac{1}{s},\text{ so } x_i=\frac{i}{s}.$$

$$\therefore I=\lim_{s\to+\infty}\sum_{i=1}^{s(t-1)+1}\left(\frac{s+\left(\frac{i}{s}\right)-1}{s}\right)\left(m+\frac{\left(\frac{i}{s}\right)-1}{s}k\right)\left(\frac{1}{s}\right)$$

$$=\lim_{s\to+\infty}\sum_{i=1}^{s(t-1)+1}\left(\frac{s^2+i-s}{s^3}\right)\left(m+\frac{(i-s)k}{s^2}\right)$$

$$=\lim_{s\to+\infty}\sum_{i=1}^{s(t-1)+1}\left(\frac{1}{s}+\frac{i}{s^3}-\frac{1}{s^2}\right)\left(m+\frac{ki-sk}{s^2}\right)$$

$$=\lim_{s\to+\infty}\sum_{i=1}^{s(t-1)+1}\left(\frac{m}{s}+\frac{ki-sk}{s^3}+\frac{mi}{s^3}+\frac{ki^2-ski}{s^5}-\frac{m}{s^2}-\frac{ki-sk}{s^4}\right)$$

$$=\lim_{s\to+\infty}\sum_{i=1}^{s(t-1)+1}\left(\frac{m}{s}+\frac{ki}{s^3}-\frac{k}{s^2}+\frac{mi}{s^3}+\frac{ki^2}{s^5}-\frac{ki}{s^4}-\frac{m}{s^2}-\frac{ki}{s^4}+\frac{k}{s^3}\right)$$

$$= \lim_{s\to+\infty} \sum_{i=1}^{s(t-1)+1} \left(\left(\frac{m}{s^3}\right)i + \left(\frac{k}{s^3} - \frac{2k}{s^4}\right)i + \left(\frac{k}{s^5}\right)i^2 + \left(\frac{m}{s} - \frac{k}{s^2} - \frac{m}{s^2} + \frac{k}{s^3}\right) \right)$$

$$= \lim_{s\to+\infty} \left(\begin{array}{l} \left(\frac{m}{s^3}\right)\left(\frac{(st-s+1)(st-s+2)}{2}\right) + \left(\frac{(ks-2k)}{s^4}\right)\left(\frac{(st-s+1)(st-s+2)}{2}\right) \\ + \left(\frac{k}{s^5}\right)\left(\frac{(st-s+1)(st-s+2)(2st-2s+1)}{6}\right) \end{array} \right)$$

$$= \ldots$$

$$= \frac{m}{2}t^2 - \frac{k}{2}t^2 + \frac{k}{3}t^3$$

$$= \frac{t^2}{2}(m-k) + \frac{t^3}{3}k$$, **which is equal to (3).**

$\therefore$ On dividing the sub-intervals within the upper and lower limits to an extent such that $\Delta x \to 0$, or $s \to +\infty$, it is possible to obtain almost the exact value of the definite integral which we would want to find.

Throughout the calculations, we would also need to apply the definition of definite integral so as to produce the closest result to the one obtained by doing integration.

References:

1. Definition of definite integral – http://en.wikibooks.org/wiki/Calculus/Definite_integral
2. Principle of Trapezoidal rule – http://en.wikipedia.org/wiki/Trapezoidal_rule

---