THE UNDER-OVER-UNDER THEOREM

FRANCIS SCHEID, Boston University

An experimenter has made three measurements of a quantity Q, at time t_1, t_2, and t_3. Conforming to currently accepted ritual of experimentation he plots three points. What they look like is shown in Figure 1.

From discussions with expert theorists in his field our man knows that the points should fall on a straight line. The relationship between Q and t should be linear. The formula $Q=Mt+B$ is indicated. Unfortunately, theory does not say *which* straight line. It does not predict the values of M and B. Indeed the purpose of the present experiment was the obtaining of this missing information. Many experiments are similarly inspired. One of Albert Einstein's accomplishments as a young man (before relativity) was the determination of an unknown coefficient in a theoretically derived formula for Brownian motion, the random motion of suspended particles in a supporting medium (dust in quiet air for example). So, theory often leaves little problems of this sort in its wake.

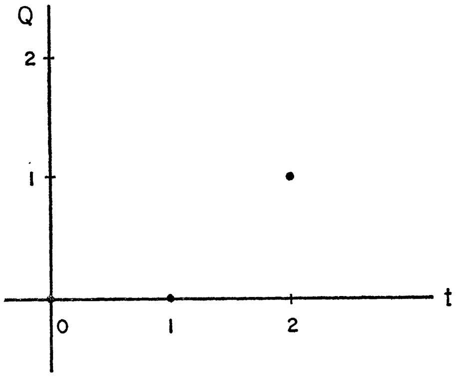

FIG. 1

Apparently, however, experiment also allows room for opinion. The three points in the diagram do not fall on a straight line. Taking his data home with him for the evening our man, in a moment of weakness, wishes he had made only two measurements. Then realizing full well that such thoughts are idle he decides to make the best of what he has, applies a transparent ruler to the diagram, and after a few moments of indecision draws his choice. It passes under, over, and under the three points in turn (Fig. 2). From his diagram he estimates

$$M = \text{rise/run} = .5,\ B = -.2$$

and instantly wonders just how good these values are. Has he made the best choice of straight line?

He is not by any means the first human being to face this question. Indeed, various "best" solutions have been suggested over the passing years. What many believe to be the best of the best appears first in the work of the nineteenth-century mathematician Čebyšev (alias Tchebysheff, alias Tchebycheff, alias Chebyshev). Chebyshev argues as follows. Choose any line whatever. It will pass over or under the three points by amounts h_1, h_2, h_3 to be called errors (positive for over, negative for under). A direct hit simply makes one error zero. Let H be the largest error size, the worst miss, for the line chosen. For every line there is such an H. Then, of all possible straight lines, the best according to Chebyshev is the one whose H, the biggest miss, is smallest. The maximum error size must be minimized. It is a MINMAX problem.

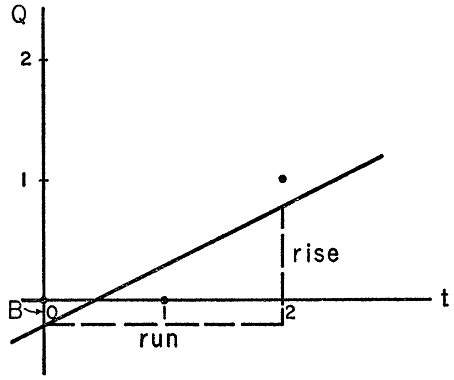

FIG. 2

Now let's take the analytic view. The equation $y=Mt+B$ represents a straight line. With M and B unspecified it could be almost any line. We are to determine M and B by applying the Chebyshev idea of bestness to the experimental data. The values of Y may then be thought of as corrections of the data values Q. Let the data be quite generally

$$\begin{matrix} t_1 & t_2 & t_3 \\ Q_1 & Q_2 & Q_3 \end{matrix}$$

so that the errors become

$$h_1 = y_1 - Q_1, \qquad h_2 = y_2 - Q_2, \qquad h_3 = y_3 - Q_3, \tag{1}$$

with

$$y_1 = Mt_1 + B, \qquad y_2 = Mt_2 + B, \qquad y_3 = Mt_3 + B. \tag{2}$$

A few lines of algebra will now convince the deepest skeptic that

$$(t_3 - t_2)y_1 - (t_3 - t_1)y_2 + (t_2 - t_1)y_3 = 0.$$

(Just substitute y_1, y_2, y_3 from above.) For brevity this is rewritten as

$$\beta_1 y_1 - \beta_2 y_2 + \beta_3 y_3 = 0, \tag{3}$$

where $\beta_1 = t_3 - t_2$, $\beta_2 = t_3 - t_1$, $\beta_3 = t_2 - t_1$. It may be assumed that $t_1 < t_2 < t_3$ so that all three β's are positive.

We now raise the question "Is there a line for which

$$h_1 = h, \qquad h_2 = -h, \qquad h_3 = h,$$

making the three errors of *equal size* and *alternating sign?*" If h is positive this line goes over-under-over the three points, while if h is negative the line goes under-over-under.

This question may seem thoroughly irrelevant and immaterial. But in fashionable courtroom language "it will be connected up." Tradition in mathematical journalism frowns on following the winding paths taken by our mathematical ancestors to their own discoveries. To continue the metaphor, it is considered wasteful of time to re-enact the crime. Our question happens to be a short cut, though this is perhaps not at once obvious.

If a line having the suggested property does exist it is easy to deduce what h must be. For then

$$y_1 = Q_1 + h, \qquad y_2 = Q_2 - h, \qquad y_3 = Q_3 + h,$$

and substituting into (3),

$$\beta_1(h + Q_1) - \beta_2(-h + Q_2) + \beta_3(h + Q_3) = 0,$$

leading quickly to

$$h = -\frac{\beta_1 Q_1 - \beta_2 Q_2 + \beta_3 Q_3}{\beta_1 + \beta_2 + \beta_3}, \tag{4}$$

so that only one h is possible. Our line must therefore pass through the three points

$$\begin{array}{ccc} t_1 & t_2 & t_3 \\ Q_1 + h & Q_2 - h & Q_3 + h \end{array}$$

with h given by (4).

But this is slightly severe. It is common knowledge that passing through just two specified points is all that one should ask of a straight line. It narrows the field of competition to exactly one candidate. That there is a line through the above *three* points is therefore newsworthy, if true. That it is true may be discovered by comparing the slopes of P_1P_2 and P_2P_3 (P_1, P_2, and P_3 being the three points, taken left to right). It requires a little persistence, but using (4) and the fact that $\beta_1 + \beta_3 = \beta_2$ the equality of these slopes emerges and settles the issue beyond further doubt.

The answer to our question is thus, "Yes, there is a line, exactly one, which makes the errors h_1, h_2, h_3 of equal size and alternating sign."

Now comes the connecting up. The line just found will be called the Chebyshev line, for it is the best line in Chebyshev's sense. The proof is easy. As indicated, h, $-h$, h are the errors for the Chebyshev line. Let h_1, h_2, h_3 be the errors for *any other line*. Then using (4), (1) and (3) in succession

$$
\begin{aligned}
h &= -\frac{\beta_1(y_1 - h_1) - \beta_2(y_2 - h_2) + \beta_3(y_3 - h_3)}{\beta_1 + \beta_2 + \beta_3} \\
&= -\frac{(\beta_1 y_1 - \beta_2 y_2 + \beta_3 y_3) - (\beta_1 h_1 - \beta_2 h_2 + \beta_3 h_3)}{\beta_1 + \beta_2 + \beta_3} \qquad (5)\\
&= \frac{\beta_1 h_1 - \beta_2 h_2 + \beta_3 h_3}{\beta_1 + \beta_2 + \beta_3}.
\end{aligned}
$$

But if H is, as suggested before, the maximum of $|h_1|$, $|h_2|$, $|h_3|$, then because all β's are positive the right side of (5) is certainly increased if h_1, h_2, h_3 are replaced by H, $-H$, H respectively. This means

$$
(6) \qquad |h| \leqq \frac{\beta_1 H + \beta_2 H + \beta_3 H}{\beta_1 + \beta_2 + \beta_3} = H.
$$

Thus the maximum error size, $|h|$, of the Chebyshev line is no larger than the maximum, H, for any other line. This proves that $|h|$ is the required minmax.

By example, if the three points of our original diagram are as they seem to be

t:	0	1	2
Q:	0	0	1

then $\beta_1 = 2-1=1$, $\beta_2=2-0=2$, $\beta_3=1-0=1$, and by (4),

$$
h = -\frac{(1)(0) - (2)(0) + (1)(1)}{1 + 2 + 1} = -\tfrac{1}{4}.
$$

The Chebyshev line thus passes through

t:	0	1	2
y:	$-\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{4}$

and so has the equation $y = \frac{1}{2}t - \frac{1}{4}$ as may easily be verified. This line passes under, over and under the three points missing each by the same amount, and yields the predictions $M = .5$, $B = -.25$, which are certainly very close to the results our experimenter obtained by eye. As a matter of fact the average person faced with the same problem will draw by eye a line fairly close to the Chebyshev line.

Can there be two best lines? In other words, can some other line have the same maximum error as the Chebyshev line? In this case $H = |h|$, and the *equality* sign holds in (6). At this point a few moments of quiet thought are more useful than a thousand words, and should eventually bring conviction that the substitution which carries (5) into (6) cannot preserve equality unless h_1, h_2, h_3 are of size $|h|$ and of alternating sign. But it is these features which led us to the three points through which the Chebyshev line passes. Surely there are not *two* straight lines through these three points. The uniqueness of the best line is a fact.

Let's suppose next that our experimenter is not entirely satisfied with his first effort. Returning to his apparatus next morning he discovers that he has forgotten to shut it off. This brings mixed emotions, since it is an expensive affair and he has not been authorized to let it go on so long. However, it would be criminal to waste such a golden opportunity, and accordingly he obtains two additional readings before his conscience forces him to shut down. Adding these two readings to yesterday's plot he has the picture (Fig. 3) before him. (We may assume that his readings are from a counter, so that all have integer values.) It is now plain that his old line is too high. It passes over both new points, missing the first by quite a bit. Having no reason to suppose this par-

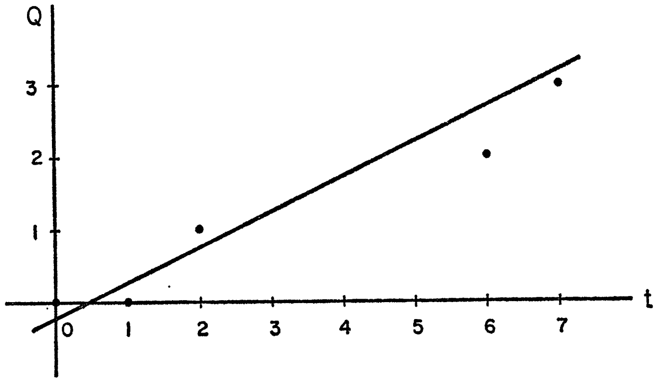

FIG. 3

ticular reading any less worthy than the other four, he once again applies the transparent ruler and draws a new line. Figure 4 shows what the picture now looks like. From the new line he makes the estimates

$$M = \text{rise/run} = 2.75/7 = .39,\ B = -.1,$$

which differ, of course, from his previous estimates but which he suspects are somewhat closer to the fact.

And what does Chebyshev have to say about this more complicated problem? Which line is really the best line? Chebyshev's answer is the same as before. It is the line for which the maximum error size H is the smallest. Becoming analytic once more let the data be rather generally

$$t_1 \cdots t_n$$
$$Q_1 \cdots Q_n.$$

In our example we take the values suggested by the diagram

t:	0	1	2	6	7
Q:	0	0	1	2	3.

For any straight line $y=Mt+B$ the ordinates y_i and errors h_i are again given by equations just like (1) and (2), there simply being more of them. Again H is the maximum of the quantities $|h_i|$ and the best line for the given points, according to Chebyshev, is the line whose H is the smallest. Now, how can this best line be found? The following *exchange method* provides a very effective answer to this question. It proceeds in four basic steps, and does yield the best line as we shall see.

STEP ONE. *Choose any three of the data points.* (*A set of three data points will be called a triple. This step simply selects an initial triple. It will be changed in Step four.*) *Proceed to Step two.*

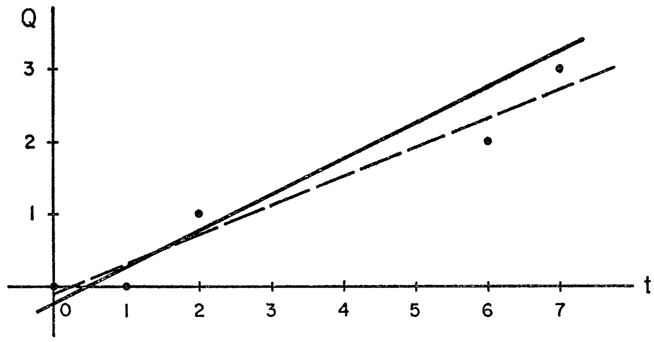

FIG. 4

STEP TWO. *Find the Chebyshev line for this triple as illustrated above. The value h for this line will, of course, be calculated in the process. Proceed to Step three.*

In illustration of these first two steps let us choose the initial triple

t:	0	1	2
Q:	0	0	1

consisting of the first three of the five points. This is, of course, the triple occurring in our original example. The Chebyshev line for this triple was found to be $y=\frac{1}{2}t-\frac{1}{4}$ with $h=-\frac{1}{4}$. This line passes under, over and under the first three of our five points with equal error sizes. But as noted earlier its errors at the remaining two points suggest a new choice of line. The exchange method makes this new choice as follows.

STEP THREE. *Calculate the errors at all remaining points for the Chebyshev line just found. Call the largest of these h_i values (in absolute value) H. If $|h|=H$ the search is over. The Chebyshev line for the chosen triple is also the best line for the entire point set.* (*We shall prove this in a moment.*) *If $|h|<H$ proceed to Step four.*

In our example we find

$$h_1 = -\tfrac{1}{4} = h, \quad h_2 = \tfrac{1}{4} = -h, \quad h_3 = -\tfrac{1}{4} = h, \quad h_4 = \tfrac{3}{4}, \quad h_5 = \tfrac{1}{4}.$$

Thus $H=\frac{3}{4}$ and since this exceeds $|h|$ we are not finished and proceed as instructed to

STEP FOUR. *This is the exchange step. Choose a new triple as follows. Add to the old triple a data point at which the greatest error size H occurs. Then discard one of the former points, in such a way that the remaining three have errors of alternating sign. (A moment's practice will show that this is always possible.) Then return, with the new triple, to Step two.*

In the example we must now include the fourth point and eliminate the first. On the new triple

$$\begin{array}{lccc} t: & 1 & 2 & 6 \\ Q: & 0 & 1 & 2 \end{array}$$

the errors h_2, h_3, h_4 found a moment ago do have the required alternation of sign. With this new triple we return as instructed to Step two:

Again we are to find a Chebyshev line. The computation follows the pattern of our first effort:

$$\beta_1 = 6 - 2 = 4, \quad \beta_2 = 6 - 1 = 5, \quad \beta_3 = 2 - 1 = 1,$$

$$h = -\frac{(4)(0) - (5)(1) + (1)(2)}{4 + 5 + 1} = \frac{3}{10}.$$

The Chebyshev line must pass through

$$\begin{array}{lccc} t: & 1 & 2 & 6 \\ y: & \frac{3}{10} & \frac{7}{10} & \frac{23}{10} \end{array}$$

and so has the equation

$$y = \tfrac{2}{5}t - \tfrac{1}{10}. \tag{7}$$

This again completes Step two. Repeating Step three as instructed we find

$$h_1 = -\tfrac{1}{10}, \quad h_2 = \tfrac{3}{10} = h, \quad h_3 = -\tfrac{3}{10} = -h, \quad h_4 = \tfrac{3}{10} = h, \quad h_5 = -\tfrac{3}{10}.$$

This time $H=\frac{3}{10}$ so that $|h| = H$ and the job is done. The Chebyshev line (7) on the triple at $t=(1, 2, 6)$ is the best line for the set of five points with which we began. Its maximum error, $H=\frac{3}{10}$, is the smallest possible for any line. This will now be proved in the general case. Let us stop, however, to record the latest predictions of M and B, namely,

$$M = .4, \quad B = -.1,$$

which should certainly tend to boost sales of transparent rulers.

There are two open questions. The exchange method can come to a halt only in Step three, and then only if no error exceeds $|h|$. How do we know this condition is ever satisfied? (Conceivably we could be computing forever.) That is the first question. The second is, of course, assuming the condition *is* satisfied at some stage, why is the last Chebyshev line also the best line in the Chebyshev sense for the whole point set?

To answer the first question, recall that after any particular exchange the old Chebyshev line has errors of size $|h|$, $|h|$, H on the new triple. Also recall that $|h| < H$ (or we would have stopped) and that these errors alternate in sign. The Chebyshev line for this new triple is then found. Call its errors on the triple h^*, $-h^*$, h^*. Returning to (5) now, with the old Chebyshev line playing the role of "any other line," we have

$$h^* = \frac{\beta_1 h_1 - \beta_2 h_2 + \beta_3 h_3}{\beta_1 + \beta_2 + \beta_3},$$

where h_1, h_2, h_3 are the numbers h, h, H with alternating sign. Because of this alternation of sign all three terms in the numerator of this fraction *have the same sign*, so that

$$|h^*| = \frac{\beta_1 |h| + \beta_2 |h| + \beta_3 H}{\beta_1 + \beta_2 + \beta_3}$$

if we assume that the error H is at the third point, just to be specific. (It really makes no difference in which position it goes.) In any event, $|h^*| > |h|$ because $|h| < H$. The new Chebyshev line has a *greater* error size on its triple than the old one had (on its own triple). This result now gives excellent service. If it comes as a surprise, look at it this way. The old line in our example gave excellent service ($h = \frac{1}{4}$) on its own triple, but poor service ($H = \frac{3}{4}$) elsewhere. The new line gave good service ($h = \frac{3}{10}$) on its own triple, and just as good service ($H = h$) on the other points also.

We can now prove that the exchange method must come to a stop sometime. For, there are only so many triples. And *no triple is ever chosen twice*, since as just proved the h values increase steadily. At some stage the condition $|h| = H$ will be satisfied.

The second question is almost as easily answered as the first. For the last Chebyshev line, and the h value of its triple, the maximum error size on the whole point set is $H = |h|$, (or we would have proceeded by another exchange to still another triple). If $h_1, \cdots, h_n$ are the errors for any other line, then by (6) $|h| \leqq \max |h_i|$, where h_i is restricted to the three points of the last triple. But then certainly $|h| \leqq \max |h_i|$, where h_i is unrestricted, for including the remaining points can only make the right side even bigger. Thus the maximum error, $H = |h|$, of this final Chebyshev line is the smallest maximum error of all. It is the minmax. The best line in the Chebyshev sense is one which has errors of equal size and alternating sign on a properly chosen triple. This is what we have light-heartedly designated as the under-over-under theorem. Taking one moment more, let us again show that the best line is unique. If a second line had maximum error (on the entire point set) equal to that of the final Chebyshev line, which is $|h|$, then *on the final triple* its maximum error would certainly be no greater. But it could also be no less, for, as we saw earlier, no line can outdo the Chebyshev line on its own triple. So on the final triple both lines would have the same maximum error. Our earlier proof of uniqueness now applies. The two lines must be the same.

As a final example we may suppose that our laboratory scientist has received a new research grant, which restores his fiscal respectability and permits him to repeat his experiment on a grander scale. The results are the following experimenter's delight.

t:	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Q:	0	1	1	2	1	3	2	2	3	5	3	4	5	4	5	
t:	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Q:	6	6	5	7	6	8	7	7	8	7	9	11	10	12	11	13

With this wealth of expensive information he abandons his transparent ruler and consults higher authority on best-fitting straight lines. The search for the correct triple and corresponding Chebyshev line is on. There are 31 points to choose among. First graders may soon be learning that the number of possible triples in this accumulation is $C(31, 3)$ or $31!/(28!\,3!) = 4495$. Finding the correct triple in this haystack appears at first glance to be a task well left to the arithmetical wizardry of automatic computing machines. Fortunately the exchange method does have the repetitious nature so easy to explain to (program for) such machines. And fortunately too the method wastes very little time on inconsequential triples. It hastens to its target like a hungry lion. The IBM 650 at Boston University took less than five minutes and only three exchanges, starting from the horrible initial triple at $t = (0, 1, 2)$, to produce the Chebyshev line (shown in Fig. 5).

$$y = .38095t - .28571$$

on the triple at $t = (9, 24, 30)$. The value of h for this line is 1.85714 and the maximum error outside the triple is only 1.61905 so that truly this is the best line in Chebyshev's sense. It yields the predictions

$$M = .38095, \qquad B = -\ .28571$$

for the long-sought coefficients. These may be compared with the cruder values found earlier, and presumably will serve the purpose for the present. If the forward march of science leads to the suspicion that these predictions are grossly in error, then perhaps a better experimental method can be devised, or hopefully more funds will become available. But for the given data the consensus of current mathematical literates is that the Chebyshev line is the best.

The exchange method has rooted out the correct triple so brilliantly in this example that the use of high-speed machines seems more a luxury than a necessity. It is interesting to watch the progress of the computation through the points included in the successive triples and the corresponding h and H values.

Triple at $t=$	h	H
(0, 1, 2)	.25000	5.25000
(0, 1, 24)	.35417	−3.89583
(1, 24, 30)	−1.75862	−2.44828
(9, 24, 30)	−1.85714	−1.85714

Note that in this case no unwanted point is ever brought into the triple. Three points are needed, three exchanges suffice. Note also the steady increase of $|h|$, as forecast by theory.

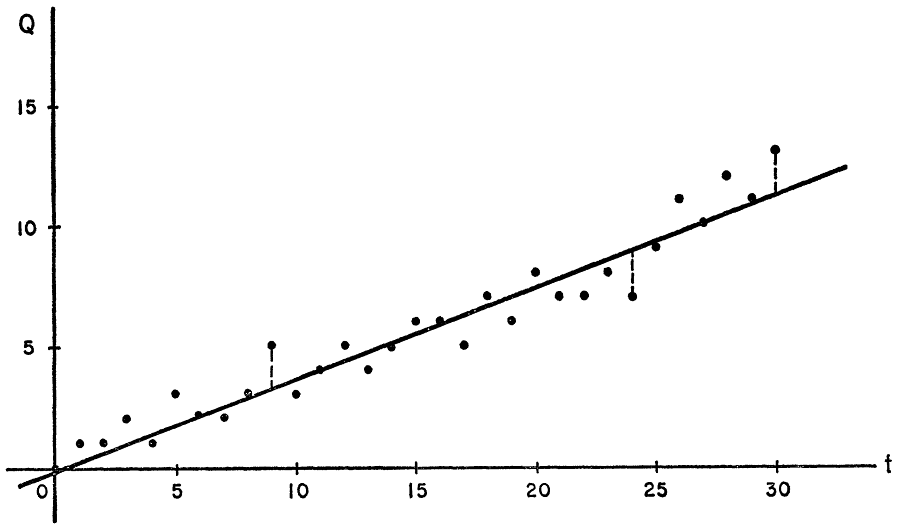

FIG. 5

Though the machine is really unnecessary for the above calculations, the existence of modern computers has led during the past decade to a rapid spread of the Chebyshev method to more complicated problems of approximation. Not only may straight lines be fitted by the minmax criterion, but a variety of more sophisticated curves as well. For these the search may be much more laborious (perhaps for a correct octuple). For various reasons such problems of approximation are fundamental in modern computation, and it is easy to believe that the machine becomes man's indispensable colleague.

PROBLEM. *Show that the best line of our last example (shown in Fig. 5) makes direct hits at $t=6$ and at $t=27$. There are no other direct hits.*