Biometrika (1961), 48, 1 *and* 2, *p.* 1
Printed in Great Britain

Studies in the history of probability and statistics

XI. Daniel Bernoulli on maximum likelihood

BY M. G. KENDALL

Research Techniques Division, London School of Economics and Political Science

1. Almost as soon as the calculus of probabilities began to take a definite shape mathematicians were concerned with the use of probabilistic ideas in reconciling discrepant observations. James Bernoulli's *Ars Coniectandi* was published in 1713. Within 9 years we find Roger Cotes (1722), in a work on the estimation of errors in trigonometrical mensuration, discussing what would nowadays be described as an estimation problem in a plane. Let p, q, r, s be four different determinations of a point o, with weights P, Q, R, S which are inversely proportional to distance from o (*pondera reciproce proportionalia spatiis evagationum*). Put weights P at p, etc., and find their centre of gravity z. This, says Cotes, is the most probable site of o. (*Dico punctum z fore locum obiecti maxime probabilem, qui pro vero eius loco tutissime haberi potest.*) Cotes does not say why he thinks this is the most probable position or how he arrived at the rule.

2. According to Laplace this result of Cotes was not applied until Euler (1749) used it in some work on the irregularities in the motion of Saturn and Jupiter. Further attacks on the problem of a somewhat similar kind were employed by Mayer (1750) in a study of lunar libration and by Boscovich (1755) in measurements on the mean ellipticity of the earth. There was evidently a good deal of interest being taken in the combination of observations about the middle of the eighteenth century. The ideas, as was only natural, were often intuitive and sometimes obscurely expressed, but the fundamental questions seem to have been asked at quite an early stage. For example, Simpson (1757) refers to a current opinion that one good observation was as accurate as the arithmetic mean of a set, and although from that point onwards a series of writers argued for the arithmetic mean, Laplace (1774), in his first great memoir, was clearly aware that for some distributions of error there were better estimators such as the median.

3. Simpson (1756, 1757) was the first to introduce the concept of distribution of error and to consider continuous distributions. But like most of his contemporaries he regarded it as inevitable to impose two conditions: first, the distributions must be symmetrical; secondly, they must be finite in range. Lagrange reproduced Simpson's work without acknowledgement in a memoir published between 1770 and 1773, but Lagrange's contributions are more of analytical than of probabilistic interest.

4. Daniel Bernoulli was born in 1700 and lived to be 82. Throughout his productive life he made contributions to the theory of probability and although his mathematical methods are not now of much importance, the originality of his thinking on such matters as moral expectation entitles him to a permanent place among the founders of the subject. In particular, the memoir on maximum likelihood reproduced in the following pages is astonishingly in advance of its time. The author was 78 when it was published and it appears that he excogitated the basic ideas for himself without reference to previous writings. The memoir may, in actual fact, have been written rather earlier. Laplace's

article of 1774 refers to *manuscripts* of Bernoulli and Lagrange which he had heard of but not seen. An announcement of their existence, says Laplace sublimely, *reawakened* his interest in the subject. Laplace was 25 at the time.

5. I am much indebted to my colleague Mr C. G. Allen for the translations of the articles by Bernoulli and Euler which follow. They are, I felt, of sufficient interest to justify the publication of an English version, especially Bernoulli's. The reasoning is so clear that I can leave Daniel to tell his own story, but perhaps I may direct attention to two points:

(*a*) Influenced by the belief that an error distribution must have a finite range, Bernoulli runs into trouble with the parameter determining that range. He assumes a semi-circular distribution and lays down the peculiar condition that any distribution must be abrupt at its terminals. Once this is done, however, his formulation of maximum likelihood is clear and explicit and he derives what would nowadays be called the ML equations by differentiating the likelihood of the sample.

(*b*) In § 16 he is right on the verge of a principle of minimal variance. In comparing two methods of estimation he points out that one (the ML method) gives samples which are closer to the true value than the other.

6. The commentary by Euler seems to me of less value. He points out, correctly in my opinion, that the ML principle is arbitrary in the sense that there is no logical reason to believe that observations come from a generating system which gives them the greatest probability. (Bernoulli admits that his reasoning on this point is metaphysical, but at least he does reason about it.) Euler then goes on to propound principles which seem to me to be much more open to doubt than the one he is trying to replace. His examples at the end, in which he has to manoeuvre his error-range to avoid imaginary solutions, ends rather lamely with the conclusion that it doesn't matter much anyway. However, it is always of interest to read what a great mind has to offer on a subject. Nor should we forget, perhaps, that at the time of publication Euler himself was 71 and had been blind for 10 years.

REFERENCES

BOSCOVICH, R. G. (1755). (In Maire, C. and Boscovich.) *De litteraria expeditione per Pontificiam ditionem ad dimetiendos duos meridiani gradus*. Romae.

COTES, R. (1722). Aestimatio errorum in mixta mathesi, per variationes partium trianguli plani et spherici. *Opera Miscellanea*, Cantabrigiae.

EULER, L. (1749). *Pièce qui a remporté le prix de l'Académie Royale des Sciences en* 1748, *sur les inégalités du mouvement de Saturne et de Jupiter*. Paris.

LAGRANGE, J-L. (1770–3). Mémoire sur l'utilité de la méthode de prendre le milieu des résultats de plusieurs observations etc. *Miscellanea Taurinensia*, 5, 167.

LAPLACE, P. S. (1774). Déterminer le milieu que l'on doit prendre entre trois observations données d'un même phénomène. *Mém. Acad. Paris (par divers savants)*, 4, 634.

MAYER, T. (1750). Abhandlung über die Umwälzung des Mondes um seine Axe. *Kosmographische Nachrichten und Sammlung*.

SIMPSON, T. (1756). A letter...on the advantage of taking the mean of a number of observations in practical astronomy. *Phil. Trans.* **44**, 82.

SIMPSON, T. (1757). An attempt to show the advantage arising by taking the mean of a number of observations in practical astronomy. *Miscellaneous Tracts*, London.

The most probable choice between several discrepant observations and the formation therefrom of the most likely induction

By DANIEL BERNOULLI†

Translated by C. G. Allen

British Library of Political and Economic Science, London School of Economics and Political Science

1. Astronomers as a class are men of the most scrupulous sagacity; it is to them therefore that I choose to propound those doubts that I have sometimes entertained about the universally accepted rule for handling several slightly discrepant observations of the same event. By this rule the observations are added together and the sum divided by the number of observations; the quotient is then accepted as the true value of the required quantity, until better and more certain information is obtained. In this way, if the several observations can be considered as having, as it were, the same weight, the centre of gravity is accepted as the true position of the objects under investigation. This rule agrees with that used in the theory of probability when all errors of observation are considered equally likely.

2. But is it right to hold that the several observations are of the same weight or moment, or equally prone to any and every error? Are errors of some degrees as easy to make as others of as many minutes? Is there everywhere the same probability? Such an assertion would be quite absurd, which is undoubtedly the reason why astronomers prefer to reject completely observations which they judge to be too wide of the truth, while retaining the rest and, indeed, assigning to them the same reliability. This practice makes it more than clear that they are far from assigning the same validity to each of the observations they have made, for they reject some in their entirety, while in the case of others they not only retain them all but, moreover, treat them alike. I see no way of drawing a dividing line between those that are to be utterly rejected and those that are to be wholly retained; it may even happen that the rejected observation is the one that would have supplied the best correction to the others. Nevertheless, I do not condemn in every case the principle of rejecting one or other of the observations, indeed I approve it, whenever in the course of observation an accident occurs which in itself raises an immediate scruple in the mind of the observer, before he has considered the event and compared it with the other observations. If there is no such reason for dissatisfaction I think each and every observation should be admitted whatever its quality, as long as the observer is conscious that he has taken every care.

3. Let us compare the observer with an archer aiming his arrows at a set mark with all the care that he can muster. Let his mark be a continuous vertical line so that only deviations in a horizontal direction are taken into account; let the line be supposed to be drawn in the middle of a vertical plane erected perpendicular to the axis of vision, and let the whole of the plane on either side be divided into narrow vertical bands of equal width. Now if the arrow be loosed several times, and for each shot the point of impact be examined and its distance from the vertical mark noted on a sheet, though the outcome cannot in the least be exactly predicted, yet there are many assumptions that can reasonably be made and

† This memoir and the following commentary by Euler appeared in Latin in the memoirs of the Academy of St Petersburg, *Acta Acad. Petrop.* (1777), pp. 3–33. A photostat copy has been deposited in the library of the Royal Statistical Society.

which can be useful to our inquiry, provided all the errors are such as may easily be in one direction as the other, and their outcome is quite uncertain, being decided only as it were by unavoidable chance. In astronomy, likewise, anything which admits of correction *a priori* is not reckoned as an error. When all those corrections have been made which theory enjoins, any further correction which is necessary in order to reconcile the several slightly discrepant observations which differ slightly from each other is a matter solely for the theory of probability. What in particular happens in the course of observation, *ex hypothesi* we scarcely know, but this very ignorance will be the refuge to which we are forced to flee when we take our stand on what is not truest but most likely, not certain but most probable (*non verissimum sed verisimillimum, non certum sed probabilissimum*), as the theory of probability teaches. Whether that is always and everywhere identical with the usually accepted arithmetical mean may reasonably be doubted.

4. Errors, which are unavoidable in observation, may indeed affect individual observations; nevertheless, any given observation has its own rights and could not be impugned if it were the only one that had been made. Any observation must therefore be in itself sound and good, and no-one ought to assign any other value than that ascertained thereby; but since they are mutually contradictory, a value has to be assigned to the whole complex of observations without touching the parts. In this way a definite error is attributed to the individual observations; but I think that of all the innumerable ways of dealing with errors of observation one should choose the one that has the highest degree of probability for the complex of observations as a whole.

The rule which I here propound will be accepted by all, provided that the degree of probability in respect of a given observation can be defined in terms of a point which is assumed to be true. I freely admit that this last condition has not been definitely met; at the same time I am convinced that all things are not equally uncertain and that better results can be got than can be expected from the commonly accepted rule. Let us see if certain assumptions should not properly be made in this argument which contribute something to a higher probability. I will begin the examination with some general considerations.

5. If the archer whom I mentioned in § 3 makes innumerable shots, all with the utmost possible care, the arrows will strike sometimes the first band next to the mark, sometimes the second, sometimes the third and so on, and this is to be understood equally of either side whether left or right. Now is it not self-evident that the hits must be assumed to be thicker and more numerous on any given band the nearer this is to the mark? If all the places on the vertical plane, whatever their distance from the mark, were equally liable to be hit, the most skilful shot would have no advantage over a blind man. That, however, is the tacit assertion of those who use the common rule in estimating the value of various discrepant observations, when they treat them all indiscriminately. In this way, therefore, the degree of probability of any given deviation could be determined to some extent *a posteriori*, since there is no doubt that, for a large number of shots, the probability is proportional to the number of shots which hit a band situated at a given distance from the mark.

Moreover, there is no doubt that the greatest deviation has its limits which are never exceeded and which indeed are narrowed by the experience and skill of the observer. Beyond these limits all probability is zero; from the limits towards the mark in the centre the probability increases and will be greatest at the mark itself.

6. The foregoing give some idea of a scale of probabilities for all deviations, such as each observer should form for himself. It will not be absolutely exact, but it will suit the nature

of the inquiry well enough. The mark set up is, as it were, the centre of forces to which the observers are drawn; but these efforts are opposed by innumerable imperfections and other tiny hidden obstacles which may produce in the observations small chance errors. Some of these will be in the same direction and will be cumulative, others will cancel out, according as the observer is more or less lucky. From this it may be understood that there is some relation between the errors which occur and the actual true position of the centre of forces; for another position of the mark the outcome of chance would be estimated differently. So we arrive at the particular problem of determining the most probable position of the mark from a knowledge of the positions of some of the hits. It follows from what we have adduced that one should think above all of a scale (*scala*) between the various distances from the centre of forces and the corresponding probabilities. Vague as is the determination of this scale, it seems to be subject to various axioms which we have only to satisfy to be in a better case than if we suppose every deviation, whatever its magnitude, to occur with equal ease and therefore to have equal probability. Let us suppose a straight line in which there are disposed various points, which indicate of course the results of different observations. Let there be marked on this line some intermediate point which is taken as the true position to be determined. Let perpendiculars expressing the probability appropriate to a given point be erected. If now a curve is drawn through the ends of the several perpendiculars this will be the scale of the probabilities of which we are speaking.

7. If this is accepted, I think the following assumptions about the scale of probabilities can hardly be denied.

(*a*) Inasmuch as deviations from the true intermediate point are equally easy in both directions, the scale will have two perfectly similar and equal branches.

(*b*) Observations will certainly be more numerous and indeed more probable near to the centre of forces; at the same time they will be less numerous in proportion to their distance from that centre. The scale therefore on both sides approaches the straight line on which we supposed the observed points to be placed.

(*c*) The degree of probability will be greatest in the middle where we suppose the centre of forces to be located, and the tangent to the scale for this point will be parallel to the aforesaid straight line.

(*d*) If it is true, as I suppose, that even the least-favoured observations have their limits, best fixed by the observer himself, it follows that the scale, if correctly arranged, will meet the line of the observations at the limits themselves. For at both extremes all probability vanishes and a greater error is impossible.

(*e*) Finally, the maximum deviations on either side are reckoned to be a sort of boundary between what can happen and what cannot. The last part, therefore, of the scale, on either side, should approach steeply the line on which the observations are sited, and the tangents at the extreme points will be almost perpendicular to that line. The scale itself will thus indicate that it is scarcely possible to pass beyond the supposed limits. Not that this condition should be applied in all its rigour if, that is, one does not fix the limits of error over-dogmatically.

8. If we now construct a semi-ellipse of any parameter on the line representing the whole field of possible deviations as its axis, this will certainly satisfy the foregoing conditions quite well. The parameter of the ellipse is arbitrary, since we are concerned only with the proportion between the probabilities of any given deviation. However elongated or compressed the ellipse may be, provided it is constructed on the same axis, it will perform the

same function; which shows that we have no reason to be anxious about an accurate description of the scale. In fact we can even use a circle, not because it is proved to be the true scale by mathematical reasoning, but because it is nearer the truth than an infinite straight line parallel to the axis, which supposes that the several observations are of equal weight and probability, however distant from the true position. This circular scale also lends itself best to numerical calculations; meanwhile it is worth observing in advance that both hypotheses come to the same whenever the several observations are considered to be infinitely small. They also agree if the radius of the auxiliary circle is supposed to be infinitely large, as if no limits were set to the deviations. Thus if the deviation of an observation from the true position is thought of as the sine of a circular arc, the probability of that observation will be the cosine of the same arc. Let the auxiliary semicircle, which I have just described, be called *the controlling semicircle* (*moderator*). Where the centre of this semicircle is located, the true position, which fits the observations best, is to be fixed. Admittedly our hypothesis is, to some extent, precarious, but it is certainly to be preferred to the common one, and will not be hazardous to those who understand it, since the result that they will arrive at will always have a higher probability than if they had adhered to the common method. When by the nature of the case a certain decision cannot be reached, there is no other course than to prefer the more probable to the less probable.

9. I will illustrate this line of argument by a trivial example. The particular problem is the reconciliation of discrepant observations; it is therefore a question of difference of observations. Now if a dice-thrower makes three throws with one die so that the second exceeds the first by one and the third exceeds the second by two, the throws may arise in three ways, viz. 1, 2, 4 or 2, 3, 5 or 3, 4, 6. None of these throws is to be preferred to the other two, for each is in itself equally probable. If you prefer the one in the middle, viz. 2, 3, 5, the preference is illogical. The same sort of thing happens if you choose to consider observations which, so far as you are concerned, are accidental, whether they are astronomical or of some other kind, as equally probable. Now suppose the thrower produces the same result by throwing a pair of dice three times. There will then be eight different ways in which he would obtain this result, viz. 2, 3, 5; 3, 4, 6; 4, 5, 7; 5, 6, 8; 6, 7, 9; 7, 8, 10; 8, 9, 11 and 9, 10, 12. But they are far from being all equally probable. It is well known that the respective probabilities are proportional to the numbers 8, 30, 72, 100, 120, 80, 40 and 12. From this known scale I have better right to conclude that the fifth set has happened than that any other has, because it has the highest probability; and so the three throws of a pair of dice will have been 6, 7 and 9. No-one, however, will deny that the first set 2, 3 and 5 might possibly have happened, even though it has only a fifteenth part of the probability corresponding to the fifth set. Forced to choose, I simply choose what is most probable. Although this example does not quite square with our argument, it makes clear what contribution the investigation of probabilities can make to the determination of cases. Now I will come more to grips with the actual problem.

10. First of all, I would have every observer ponder thoroughly in his own mind and judge what is the greatest error which he is morally certain (though he should call down the wrath of heaven) he will never exceed however often he repeats the observation. He must be his own judge of his dexterity and not err on the side of severity or indulgence. Not that it matters very much whether the judgement he passes in this matter is fitting or somewhat flighty. Then let him make the radius of the *controlling circle* equal to the aforementioned greatest error; let this radius be r and hence the width of the whole doubtful field $= 2r$.

If you desire a rule on this matter common to all observers, I recommend you to suit your judgement to the actual observations that you have made: if you double the distance between the two extreme observations, you can use it, I think, safely enough as the diameter of the controlling circle, or, what comes to the same thing, if you make the radius equal to the difference between the two extreme observations. Indeed, it will be sufficient to increase this difference by half to form the diameter of the circle if several observations have been made; my own practice is to double it for three or four observations, and to increase it by half for more. Lest this uncertainty offend any one, it is as well to note that if we were to make our controlling semicircle infinite we should then coincide with the generally accepted rule of the arithmetical mean; but if we were to diminish the circle as much as possible without contradiction, we should obtain the mean between the two extreme observations, which as a rule for several observations I have found to be less often wrong than I thought before I investigated the matter.

11. After all these preliminaries it remains to determine the position of the controlling circle, since it is at the centre of this circle that the several observations should be deemed to be, as it were, concentrated. The aforesaid position is deduced from the fact that the whole complex of observations would occur more easily, and therefore more probably, for this location than for any other position of the circle. We shall have the true degree of probability for the whole complex of observations if we note the probability corresponding to the several observations that have been carried out and multiply all the probabilities by each other, just as we did in § 9. Then the product of the multiplication is to be differentiated and the differential put $= 0$. In this way we shall obtain an equation whose root will give the distance of the centre from any given point.

Put the radius of the controlling circle $= r$; the smallest observation $= A$; the second $A+a$; the third $A+b$; the fourth $A+c$, and so on; the distance of the centre of the controlling semicircle from the smallest observation $= x$, so that $A+x$ will denote the quantity which is most probably to be assumed on the basis of all the observations. By our hypothesis the probability for the first observation alone is to be expressed by $\sqrt{\{r^2-x^2\}}$; for the second observation by $\sqrt{\{r^2-(x-a)^2\}}$; for the third by $\sqrt{\{r^2-(x-b)^2\}}$; for the fourth by $\sqrt{\{r^2-(x-c)^2\}}$ and so on. Then I would have the several probabilities multiplied together according to the rules of the theory of probability, which gives

$$\sqrt{\{r^2-x^2\}} \times \sqrt{\{r^2-(x-a)^2\}} \times \sqrt{\{r^2-(x-b)^2\}} \times \sqrt{\{r^2-(x-c)^2\}} \times \ldots.$$

Finally, if the differential of this product is put $= 0$, the equation, by virtue of our hypotheses, gives the required value x as having the highest probability. As, however, the aforesaid quantity is to be brought to its maximum value, it is obvious that its square will simultaneously be brought to the same state. So we can use, for ease of calculation, a formula which is composed entirely of rational terms, viz.

$$(r^2-x^2) \times \{r^2-(x-a)^2\} \times \{r^2-(x-b)^2\} \times \{r^2-(x-c)^2\} \times \ldots$$

and the differential is once more put $= 0$. For the rest, as many factors are to be taken as there were observations.

12. If a single observation was made, we must accept the observation as true. Now this is shown by our hypothesis. If only the first factor r^2-x^2 is taken, we shall have $-2x\,dx = 0$ or $x = 0$ and consequently $A+x = A$. So in this case our hypothesis agrees with the common one.

If two observations have been made, A and $A+a$, two factors are to be taken, namely

$$\{r^2-x^2\}\times\{r^2-(x-a)^2\} \quad \text{or} \quad r^4-2r^2x^2+x^4+2ar^2x-a^2r^2+2ax^3\times a^2x^2,$$

the differential of which

$$= -4r^2x\,dx+4x^3dx+2ar^2dx-6ax^2dx+2a^2x\,dx = 0 \quad \text{or} \quad 2x^3-3ax^2-2r^2x+a^2x+ar^2 = 0.$$

The only useful root which this equation gives is $x = \frac{1}{2}a$, and $A+x = A+\frac{1}{2}a$. This also is the teaching of the common hypothesis. This agreement holds whatever be the radius of the controlling circle, a fact which shows clearly enough, in the case of several observations, that the size of our controlling circle in an enterprise of this sort need not be strictly exact, and one should not expect it to be. What is awkward—and I do not conceal it—is that for several observations a very long calculation is required, and so I hardly dare propose more than general discussions of these cases. Let me at least expound the theory of three observations, which is of the highest importance.

13. When we have three observations to deal with, viz. A; $A+a$ and $A+b$, we shall have three factors

$$\{r^2-x^2\}\times\{r^2-(x-a)^2\}\times\{r^2-(x-b)^2\},$$

for which we have to find the maximum value. If now these factors are actually multiplied together we shall obtain

$$\begin{gathered} r^6+2ar^4x-3r^4x^2-4ar^2x^3+3r^2x^4+2ax^5-x^6 \\ -a^2r^4-2ab^2r^2x+2b^2r^2x^2+2ab^2x^3-b^2x^4+2bx^5 \\ -b^2r^4+2br^4x-a^2b^2x^2-4br^2x^3-4abx^4 \\ +a^2b^2r^2-2a^2br^2x+4abr^2x^2+2a^2bx^3-a^2x^4 \\ +2a^2r^2x^2. \end{gathered}$$

If this expression is differentiated, and then after division by dx is put $= 0$ to obtain the maximum value, the following general equation for any three observations whatsoever will result

$$\begin{gathered} 2ar^4-6r^4x-12ar^2x^2+12r^2x^3+10ax^4-6x^5 \\ -2ab^2r^2+4b^2r^2x+6ab^2x^2-4b^2x^3+10bx^4 \\ +2br^4-2a^2b^2x-12br^2x^2-16abx^3 \\ -2a^2br^2+8abr^2x+6a^2bx^2-4a^2x^3 \\ +4a^2r^2x = 0. \end{gathered}$$

The root of this equation, which is indeed of the fifth degree and consists of twenty terms, gives the distance of the centre of the controlling circle from the first observation, and the quantity $A+x$ gives the value which is most probably to be deduced from the three observations which have been made.

14. Unless the force of our fundamental arguments has been most attentively weighed there will be few perhaps who will see any relation whatever between the enormous equation and what seems to be a very simple question; for the common answer is $x = \frac{1}{3}(a+b)$. Nevertheless, our equation corresponds well enough to notions which crop up elsewhere, some of which I will now expound.

(a) If the radius of the controlling circle is supposed to be infinite compared with a and b, all terms are to be rejected except those in which r rises to the highest power, in which case our equation is reduced to this very simple one $2ar^4+2br^4-6r^4x=0$ or $x=\frac{1}{3}(a+b)$. So the common rule is contained in our equation. If, however, our definition set out in § 10 is considered, it will be obvious how unfitting is the hypothesis of an infinite radius and how manifestly some more suitable one could be substituted for it.

(*b*) If we put $b=2a$, it is obvious that $x=a$ whatever value is given to the radius r, and that too will be common to both theories. Let us see therefore what our equation shows for this case. Substituting for b the equation becomes

$$\begin{gathered}6ar^4-6r^4x-36ar^2x^2+12r^2x^3+30ax^4-6x^5\\-12a^3r^2+36a^2r^2x+36a^3x^2-52a^2x^3\\-8a^4x=0.\end{gathered}$$

Now this equation, whatever be the value of r, is satisfied by $x=a$, which the nature of the case demands.

(*c*) If $b=-a$, x must equal 0 whatever be the value of r. This too is beautifully shown by our equation, which now becomes

$$\begin{gathered}-6r^4x+12r^2x^3-6x^5\\-2a^4x+8a^2x^3=0.\end{gathered}$$

A glance will show that the useful root is $x=0$.

15. This and other similar corollaries sufficiently confirm the real connexion of our fundamental arguments with the question under discussion, however enormous the equation we have found may seem in so simple an inquiry. I proceed to examples in which the radius of the controlling circle is neither infinite nor indifferent, which is where practically all cases belong. In these examples our new theory always produces a different result from the common one; and the more the intermediate observation approaches either extreme, the greater the difference. It is on the discussion of these cases that the matter hinges, so we must have recourse to purely numerical examples.

Example 1. Let us assume three observations

$$A;\quad A+0{\cdot}2\quad\text{and}\quad A+1,$$

so that

$$a=0{\cdot}2\quad\text{and}\quad b=1$$

and let the value to be assumed as most likely from these three observations be $A+x$. The common rule gives $x=0{\cdot}4$. Let us see the new one which to my mind is more probable, and let us put $r=1$ (cf. § 10). The following purely numerical equation results

$$1{\cdot}92-0{\cdot}32x-12{\cdot}96x^2+4{\cdot}64x^3+12x^4-6x^5=0.$$

the solution of which is approximately $x=0{\cdot}4427$, which exceeds the commonly accepted value by more than a tenth. This marked excess is due to the fact that the middle observation is much nearer to the first than to the third. From this it is easily deduced that the excess will be changed to a defect if the middle observation is nearer to the third than to the first, and that the nearer the middle observation is to the mean between the two extreme observations, the smaller will be this defect. To test this conjecture I retain the other values and change only the middle observation, as follows.

Example 2. Let a now $=0{\cdot}56$, and as before $r=b=1$. By the commonly accepted rule we shall have $x=0{\cdot}52$. Let us see what happens with ours. The equation of § 13 gives the following numerical equation

$$1{\cdot}3728+3{\cdot}1072x-13{\cdot}4784x^2-2{\cdot}2144x^3+15{\cdot}6x^4-6x^5=0$$

which is approximately satisfied by $x = 0{\cdot}5128$. In accordance with our principles, the value of x is less than the arithmetical mean which is usually accepted, but the difference between the two is now quite small, viz. 0·0072, exactly as I had anticipated would be the case. Hence it can also be seen that the greatest difference between the two estimates occurs when it so happens that two observations exactly coincide and only the third diverges. There are two cases, viz. when $a = 0$ and when $a = b$. I will expound the result in each case.

Example 3. Put $a = 0$, leaving the remaining denominations unaltered. Dividing by $2b - 2x$ we have the following numerical equation

$$1 - 6x^2 - 2x^3 + 3x^4 = 0,$$

which is approximately satisfied by $x = 0{\cdot}3977$, whereas the value of x obtained from the common rule is $x = 0{\cdot}3333$. The former exceeds the latter by 0·0644. If, however, we put $a = b$ and divide by $2x$, the following equation results

$$4 - 6x - 6x^2 - 10x^3 - 3x^4 = 0.$$

This is approximately satisfied by $x = 0{\cdot}6022$, while the common value is 0·6666. So the difference between the two is once more 0·0644, but this time our new value is less than the common one, whereas previously it was greater. It is clear from this that our method takes better aim at a certain intermediate point than does the common method. Evidence of this sort does much to commend the method that I propose, and I will go a little more closely into this consideration, if so be that an *argumentum ad hominem* may be accepted in a matter which does not admit of mathematical demonstration.

16. If we combine the two cases in example 3, and suppose that six observations have been made, viz. $A, A, A+b$ and $A+b, A+b, A$, it is obvious that three observations support the value A and the same number the value $A+b$. We see by § 12 that in this case both methods give the required mean value as $A + \frac{1}{2}b$, or for example 3, $A + 0{\cdot}5$; or, omitting the constant quantity A, simply 0·5. This value, derived from the six observations combined, will not be doubted by anyone. Now let us divide these six observations into two other triads, namely $A, A, A+1$ and $A+1, A+1, A$. In this case, rejecting once more the quantity A, the commonly accepted rule gives for the first triad $0{\cdot}\dot{3}$ and for the second $0{\cdot}\dot{6}$, both differing, the first by defect and the second by excess, by $0{\cdot}1\dot{6}$ from the mean 0·5. So for either triad of observations taken separately the common theory involves an error of $0{\cdot}1\dot{6}$, while ours involves an error of 0·1022, which is notably smaller. A great deal more evidence of this kind could be adduced to give further support to our fundamental argument; but I am afraid I should appear immoderate if I went on extending something which cannot be settled with certainty and absolute perfection. We have no higher aim than to be able to distinguish what is more probable from what is less.

17. Such further perfection as we may reasonably expect will consist in a stricter and more accurate determination of the controlling scale and its width. I will add a few further comments on this topic. It is obvious from the foregoing considerations that our estimates are not so very different from the commonly accepted rule: so it is a question of a certain correction which this rule appears to allow. This correction is provided by the actual divergences of the observations from the required true point, since they can be so arranged, for any given width of the controlling scale, as to make the most probable fit with this point. But for my part I can see no way of strictly determining the width of the aforesaid scale except that which I mentioned in § 10. If an observer, through undue mistrust of his own powers, enlarges the

dimensions of the controlling semicircle excessively, it will not give all the help it might, but what it gives will be more certain; if on the other hand he contracts the scale unduly, other things being equal he will arrive at a correction which is a little greater and somewhat less probable. Prudence seems to be as necessary here as sharp-sightedness. Should you wish to use the observations that have actually been made as a basis for an *a posteriori* estimate of the width of the controlling scale to be applied, it will be prudent to weigh in your own mind whether one should consider the observations to have turned out luckily or not. The more you assign to good luck, the less you can attribute to the skill applied in observing, and the larger accordingly will be the controlling circle which you will apply. In § 13 I assumed $r = b$; in other words the radius of the controlling circle equalled the distance between the two extreme observations. I admit, however, on better reflexion that this size of radius seems to me to argue somewhat excessive confidence; it would be safer certainly in future to put $r = \frac{3}{2}b$ or even $r = 2b$. If so, the correction would come out notably smaller but all the more certain and trustworthy.

18. If there is any validity in our principles, though they are metaphysical rather than mathematical, we may justly conclude therefrom that one should seldom if ever reject an observation, and never without the utmost circumspection. I have already given my opinion on this subject in § 2. The whole complex of observations is simply a chance event modified and confined within certain limits by the skill of the observer. It may well happen, though very rarely, that of three observations two are miraculously identical, while the third by ill luck is very wide of the other two. But if this happens to me and I am certain that I have not unduly contracted the limits of maximum possible error or shown undue confidence in my skill, I should not hesitate to refer the examination of the whole case to our principles and form my estimate from them. Only the observer must give the same attention to each of the observations. I should like them all treated equally.

19. The only remaining caution refers to the controlling scale which I have applied. We have taken a semicircle as answering sufficiently the conditions set out in § 7 and at the same time most suited to the calculations that have to be carried out. Meanwhile it is worthy of note that there are other infinite curves which undoubtedly lead to the same equation as I set out in § 13. In § 11 we made the probabilities, for a circular scale, proportional respectively to the perpendiculars

$$\sqrt{(r^2-x^2)}; \quad \sqrt{\{r^2-(x-a)^2\}}; \quad \sqrt{\{r^2-(x-b)^2\}}.$$

Now if instead of a semicircle we suppose a parabola (*arcum parabolicum*) constructed on the line $2r$, with its axis passing perpendicularly through the middle of this line, then keeping the same notation, we shall have perpendiculars, or the corresponding probabilities expressed by them,

$$\frac{\rho}{r^2}(r^2-x^2), \quad \frac{\rho}{r^2}\{r^2-(a-x)^2\}, \quad \frac{\rho}{r^2}\{r^2-(b-x)^2\}, \quad \text{etc.,}$$

where the new letter ρ denotes the longest perpendicular at the abscissa $x = 0$. Now since the factor ρ/r^2 is common to all the terms we can simply substitute unity for this factor when we have brought the product of all the several probabilities to a maximum. It follows from this that the parameter of the parabola is always arbitrary. I also pointed out in the aforementioned § 11 that if this product has been brought to its maximum all its powers will at the same time be maximized or minimized. It is obvious from this that both scales, the parabolic and the circular, lead to the same required value of x. Furthermore, it is evident that innumerable other scales fulfil the same function; they will all have this property, that from their peak they approach in either direction the line $2r$, on which the several observations are necessarily supposed to lie, and intersect it. Therefore all scales of this sort achieve our aim, and we need not be too pedantic in this matter, since we are content to strive for something better if not for the best.

20. Finally, as regards the awkward, not to say monstrous, form of our fundamental equation set out in § 13, we can mend the awkwardness somewhat; for I express the useful root as approximately

$$x = \frac{a+b}{3} + \frac{2a^3 - 3a^2b - 3ab^2 + 2b^3}{27r^2}.$$

The first term is none other than the common arithmetical mean for three observations, the second indicates approximately the further correction required by our principles. This root indeed will agree all the more accurately with the equation of § 13, the greater is assumed to be the width of the controlling scale indicated by $2r$. Far be it from us, however, to increase the value of the letter r unnecessarily merely to make calculation easier, for every useless increase takes away a little from the amount of our correction. Nor would it be less dangerous to attribute too much to one's powers of observation and so shorten the radius r unjustifiably. 'There are fixed bounds, outside of which justice cannot exist': cf. § 10. Our principles themselves show that it is impossible for r to be less than $\frac{1}{2}b$, since this involves the manifest contradiction of positing as impossible something which is supposed to have actually happened. I have not concealed, however, the somewhat free assumptions that have been made in the course of our argument; but I should not have thought that all our methods of judging the observations that have been made ought to be rejected on that account. Of this at least I am convinced, that *the common rule for three observations gives somewhat too small a result when* $a < \frac{1}{2}b$ *and too large a result if* $a > \frac{1}{2}b$, *and cannot ever be applied with greater certainty than when the intermediate observation is approximately equidistant from the two extremes*. Secondly, I think it probable that our equation in § 13 gives a safer and better determination of the position to be selected, provided the radius of the controlling circle is not rashly diminished beyond the limits which the powers of the observer permit: cf. § 17. The question that I have dealt with is properly this: given three or more shots of an arrow marked on a straight line, to determine the most probable position of the point at which the archer was aiming. But any and every observer who understands these things will form for himself criteria which will answer his purpose, according to the nature of the material (*argumento*) which he has to hand, provided he makes cautious use of the rules derived from the theory of combinations.

Recapitulation. By its very nature our problem is indeterminate, inasmuch as it depends on the practice, experience, and skill of the observer, on the precision of the instruments, on the keenness of the senses, in short on countless circumstances which may be more or less favourable. Account will be taken of all these things in assuming the width of the field of possible deviations; on this subject I have given my opinion, with all circumspection. Secondly, one has to examine the casual working of chance in favour of any given deviation (lit. the working of the casual chance which favours any deviation), since it is advantageous if any given deviation is assigned the probability which from the nature of the case fits it.† To be sure, this scale of probabilities remains in its turn uncertain and undetermined, should

† Reading *cuiuis aberrationi* for the *cuiuis aberratione* of the text.

an accurate one be desired, but displays, nevertheless, by the very nature of the case, several properties; and if these are satisfied, it may be considered to be sufficiently known, as I learned from several experiments. So a method comes to light of expressing in accordance with the proven precepts of the theory of probability the absolute probability appropriate to any given system of observations for any assumed location of that system. It only remains then to select that location of the system in question which enjoys the highest probability. It certainly seems extraordinary to me that the algebraic equation defining this location, which is so far-fetched and rises to the fifth degree for only three observations, which is expressed in a very large number of terms and is deduced from principles never used before, nevertheless, from whatever point it is examined, gives rise to nothing which is in the least displeasing, still less leads to any absurd result. The upshot of the calculations in any example is little different from that which is indicated by the common method, provided one does not recklessly jump at the precepts which I have laid down. Where the comparison of three given observations shows that the middle one is approximately equidistant from the extremes, we shall adhere without hesitation to the common rule; but if the two intervals are notably unequal, I think it is better to have recourse to our theory, provided one follows the precepts I have set out and exercises the greatest prudence in fixing just bounds to the field of possible deviations. All this I should wish to have weighed in the balance of metaphysics rather than mathematics. Those who are most shocked by our principles will have nothing further to contradict if only they make the field of possible deviations as large as possible.

Observations on the foregoing dissertation of Bernoulli

By L. Euler

1. The question which our distinguished friend Bernoulli handles here is one of no little moment, namely, how an unknown quantity should be derived from several observations which vary slightly from each other. To make the nature of the question easier to discern clearly, let us suppose that the elevation of the pole star at some place or other has to be discovered and that the observations made to this end have the following different values:

$$\Pi + a, \quad \Pi + b, \quad \Pi + c, \quad \Pi + d, \quad \text{etc.},$$

where the letters a, b, c, d, etc., are taken to be expressed in seconds. From these the true elevation of the pole at this place, $\Pi + x$, is to be deduced. Generally this quantity x is obtained by taking the arithmetic mean of all the quantities a, b, c, d, etc. Hence if the number of observations $= n$, $x = (a+b+c+d+\text{etc.})/n$.

2. In this rule it is obviously assumed that all observations are of the same degree of goodness. For if some were more exact than others, account ought to be taken of this distinction in the computation. Now although there is no apparent reason in the circumstances why one of these observations should be accorded a greater value than the rest, nevertheless, the learned author observes that these observations ought to be awarded a higher degree of goodness the nearer they approach to the truth, just as that class of observations which is thought to depart too far from the truth is usually completely rejected. The whole business therefore amounts to this: to show how the degree of goodness appropriate to the several observations is to be estimated.

3. According to the view of the distinguished author, it will be convenient to consider the deviation of each observation from the truth as already known. This will be $x-a$ for the first observation, $x-b$ for the second, $x-c$ for the third, etc., but the defect of each observation should be estimated not so much from these differences as from their squares, since the defect itself is to be reckoned as the same whether the observation errs by excess or defect. Hence if some observation agrees perfectly with the truth, its defect will be zero. If therefore the degree of goodness of this observation is indicated by r^2, it is obvious that the degree of goodness of the first observation must be indicated by $r^2-(x-a)^2$, that of the second by $r^2-(x-b)^2$, of the third by $r^2-(x-c)^2$ and so on, the value of r being such that for an observation which is to be all but rejected the degree of goodness vanishes. If we assume that this happens in the observation which gives $\Pi+u$, then since the degree of goodness of this would be $r^2-(x-u)^2$, it must be laid down in all cases that $r^2=(x-u)^2$.

4. Having established these conclusions concerning the degree of goodness of each observation, the distinguished author appeals to the following principle, for which indeed he gives no reason: that the product of all the formulae expressing the degrees of goodness of the several observations should be allotted a maximum value. On this principle therefore he bids one differentiate this product and equate the differential with nought, since this equation will then give the true value of x. This he illustrates with some examples based on sets of three observations, deriving therefrom values of x which seem to be quite in conformity with the truth.

5. This principle for only three observations led to an equation of the fifth degree, whose root x had to be found; and anyone who wished to apply the principle to four observations would arrive at an equation of the seventh degree. Five observations would lead to one of the ninth degree and so on. It is thus abundantly evident that this method cannot possibly be used where there are several observations, and this is in fact candidly conceded by the distinguished author, who presents the whole dissertation as a purely metaphysical speculation.

6. As, however, the distinguished author has not supported this principle of the maximum by any proof, he will not take it amiss if I propound certain doubts about it. If we assume that among the observations in question there is one that should be almost rejected, whose degree of goodness would accordingly be as small as possible, it is evident that the product of all the formulae mentioned would in fact be reduced to nothing, so that it could not possibly be considered as a maximum, no matter how great it might be, were that observation omitted. Now the principles of the theory of probability make it abundantly clear that the value of the unknown quantity x should come out the same whether an observation such as this, which has no goodness at all, is introduced into the calculation or totally rejected.

7. I do not think that it is necessary in this question to have recourse to the principle of the maximum, since the undoubted precepts of the theory of probability are quite sufficient to resolve all questions of this kind. If the first observation, which gave $\Pi+a$, is assigned the amount or degree of goodness (*pretium seu gradum bonitatis*) α, the second β, the third γ, the unknown quantity x is given by the rules of this theory thus:

$$x=\frac{a\alpha+b\beta+c\gamma+d\delta+\text{etc.}}{\alpha+\beta+\gamma+\delta+\text{etc.}}.$$

Hence $$\alpha(x-a)+\beta(x-b)+\gamma(x-c)+\delta(x-d)+\text{etc.}=0.$$

Now it is clear that if all the grades of goodness were equal and the number of observations were n, we should have $x = (a+b+c+d+\text{etc.})/n$, as required by the common rule. From which it follows that different values may emerge for the unknown quantity x to the extent that the degrees of goodness differ.

8. Since therefore, as the distinguished author himself states, the grades of goodness indicated by the letters α, β, γ, δ are

$$\alpha = r^2-(x-a)^2, \quad \beta = r^2-(x-b)^2,$$
$$\gamma = r^2-(x-c)^2, \quad \delta = r^2-(x-d)^2, \quad \text{etc.},$$

the equation we have found becomes

$$r^2(x-a)+r^2(x-b)+r^2(x-c)$$
$$-(x-a)^3-(x-b)^3-(x-c)^3\,\text{etc.} = 0.$$

Hence if the number of observations $= n$ and we put for brevity's sake

$$a+b+c+d+\text{etc.} = A,$$
$$a^2+b^2+c^2+d^2+\text{etc.} = B,$$
$$a^3+b^3+c^3+d^3+\text{etc.} = C,$$

that equation is reduced to the following fairly simple form

$$nr^2x-Ar^2-nx^3+3Ax^2-3Bx+C = 0.$$

Thus we arrive at a cubic equation, from which the unknown x can easily be found, whatever the number of observations n.

9. If we regard the quantity r as infinite, which is the case when all the observations are assigned the same degree of goodness, then we may neglect all the other terms and directly deduce from this equation the following

$$x = \frac{A}{n} = \frac{a+b+c+d+\text{etc.}}{n}$$

just as is required by the rule which is commonly adopted. If we designate this value by the letter p, and substitute Π for $\Pi+p$ in the observations themselves, we shall have to diminish the several numbers a, b, c, d, etc., by the same quantity p, and thus the sum of them all, for which we put A, will equal 0. To avoid, however, the introduction of new letters into the calculation at this point we can from the beginning so constitute the quantity Π that if the values of the several observations are given as $\Pi+a$, $\Pi+b$, $\Pi+c$, $\Pi+d$, etc., the sum of the letters $a+b+c+d+\text{etc.} = 0$. Then to discover the quantity x we shall have the following much simpler equation

$$nx^3-nr^2x+3Bx-C = 0,$$

from which would follow, if r were infinite, $x = 0$. It is clear from this that if this equation has several real roots, the smallest should be taken as x, so that the required true value will be $\Pi+x$.

10. This same question can, however, be referred even to a quadratic equation by introducing the sort of observation which after weighing all the circumstances we decide should be totally rejected. Let such an observation be $\Pi+u$, and since *ex hypothesi* its degree of

goodness $r^2-(x-u)^2 = 0$, $r^2 = (x-u)^2$. The introduction of this value in the last equation that we found produces the following form

$$2nux^2-nu^2x+3Bx-C = 0.$$

It will be convenient to regard the term $-nu^2x$ in this equation as the greatest, so that the equation can be expressed as follows

$$x(nu^2-3B-2nux) = -C$$

from which follows
$$x = \frac{-C}{nu^2-3B-2nux},$$

where by substituting for x the value just obtained we get the following continued fraction

$$x = \frac{-C}{nu^2-3B+}\;\frac{2nuC}{nu^2-3B+}\;\frac{2nuC}{nu^2-3B+}, \quad \text{etc.},$$

a form which will soon give the true value of x itself.

11. Since the distinguished author has founded his solution on the principle of the maximum, it will not now be difficult to produce an analytical formula of this sort which, when made equal to its maximum, yields the true value of x. Let us use for this purpose the form first discovered

$$r^2(x-a)+r^2(x-b)+r^2(x-c)+\text{etc.}$$
$$-(x-a)^3-(x-b)^3-(x-c)^3-\text{etc.} = 0,$$

which may be regarded as the differential of some formula which is to be raised to its maximum. This formula itself will emerge, if this expression is put in the form of a differential and integrated. Multiplying by $4dx$ and integrating we obtain

$$2r^2(x-a)^2+2r^2(x-b)^2+2r^2(x-c)^2+\text{etc.}$$
$$-(x-a)^4-(x-b)^4-(x-c)^4-\text{etc.}+\text{constant}.$$

If we assume $-nr^4$ as the constant, there being n observations, by change of sign the following formula results

$$\{r^2-(x-a)^2\}^2+\{r^2-(x-b)^2\}^2+\{r^2-(x-c)^2\}^2+\text{etc.}$$

12. In place therefore of the formula which our distinguished friend Bernoulli thought should be made equal to its maximum we have now arrived at another formula very well suited to the nature of the question, which when brought to its maximum gives the true value of x, since this formula is obtained by adding together the squares of all the degrees of goodness.

13. To furnish an example of our method, let us consider the observations by which the longitude of the observatory of St Petersburg is deduced from the difference between the meridians of the observatories of Paris and St Petersburg. These are reported as follows:

I	1° 51′ 50″	IV	1° 51′ 50″
II	1° 51′ 52″	V	1° 51′ 50″
III	1° 51′ 39″	VI	1° 51′ 50″

Taking the arithmetic mean of these in the usual way we obtain 1° 51′ 48½″.

14. Now let us apply our formulae to this case, taking $\Pi = 1°\,51'\,48\tfrac{1}{2}''$. The values of our six letters a, b, c, d, e, f will be

$$a = 1\tfrac{1}{2},\quad b = 3\tfrac{1}{2},\quad c = -9\tfrac{1}{2},\quad d = 1\tfrac{1}{2},\quad e = 1\tfrac{1}{2},\quad f = 1\tfrac{1}{2}.$$

Their sum $A = 0$; the sum of the squares B is found to be $\frac{1}{2}(223)$; the sum of the cubes $= -801$. Hence our equation for $n = 6$ will be

$$\begin{aligned} 12ux^2 - 6u^2x + 801 \\ + 334\tfrac{1}{2}x = 0. \end{aligned}$$

15. Now let us *define* the number u from a case which the author of the observations thinks should be rejected, such as $1°\,52'\,20''$, which gives $u = 31\frac{1}{2}$. Let us suppose that $u = 30$, making our quadratic equation

$$360x^2 - 5065\tfrac{1}{2}x + 801 = 0$$

instead of which we may write in round figures

$$36x^2 = 500x - 80.$$

From this
$$x = \frac{250 \pm \sqrt{59,620}}{36},$$

that is either
$$x = \frac{250+244}{36} = 14 \quad \text{or} \quad x = \frac{250-244}{36} = \frac{1}{6}.$$

The latter value only can be considered, and might have been obtained immediately by neglecting the first term in the equation: the value of x would then have been $\frac{50}{8}$ or approximately $\frac{1}{6}$. The required difference of the meridians will therefore be $1°\,51'\,48\frac{2}{3}''$.

16. Again, suppose that observation had been rejected which gave $1°\,51'\,0''$: we then have $u = -48\frac{1}{2}$. Let us take $u = -48$, giving the equation

$$-576x^2 - 13,489\tfrac{1}{2}x + 801 = 0.$$

Neglecting the first term we obtain $x = \frac{8}{135} = \frac{1}{17}$. Now since this observation would have deserved to be rejected, if u had been in the neighbourhood of -300, hence, carrying out the calculation as before, x would have come out as about $\frac{1}{6}$. It is clear from this that in this case we could have been content with the common rule, since not even a second's difference is involved.

17. Since, however, the third of these observations differs so much from the others, it will perhaps be convenient to set the limit not far from it. If we were to do this for the case $1°\,51'\,33\frac{1}{2}''$, $u = -15''$, our equation would accordingly be

$$-180x^2 - 1000x + 80 = 0,$$

the smaller root of which equation will be $\frac{12}{18} = \frac{2}{3}$. Hence the difference of the meridians would have come out as $1°\,51'\,49\frac{1}{6}''$. It is once more clear from this case that no notable error is to be feared, unless we make a quite monstrous mistake in assuming a value for u. In this matter it will suffice to note that nu^2 must always be much larger than $3B$.

18. In particular this method deserves to be applied to those observations from which the learned Lexell not long ago determined the parallax of the sun. From these we take, purely by way of example, the following four conclusions drawn from the observations, namely (I) 8·52; (II) 8·43; (III) 8·86; (IV) 8·28. Taking the arithmetic mean of these we get 8·52. If therefore we put $\Pi = 8{\cdot}52$ the values of the four letters a, b, c, d can be fixed as follows:

$$a = 1, \quad b = 9, \quad c = -34, \quad d = +24$$

so that the sum comes out as $A = 0$.† All these numbers of course denote hundredths of a second. The sum of the squares $B = 1814$, the sum of the cubes $C = -24,750$.

† According to his original usage all these signs should be reversed. Presumably a has been rounded up to unity to make the sum of deviations zero.

19. If we now assume as the term where the degree of goodness vanishes $u = 40$, our equation emerges as

$$320x^2 - 948x + 24{,}750 = 0.$$

From this the value of x itself comes out as imaginary. Let us accordingly assume $u = 50$; the equation will then become

$$400x^2 - 10{,}000x + 24{,}750$$
$$+ 5442x = 0$$

and we still arrive at an imaginary result. If, however, we take $u = 60$ the smaller value of x will be $3\frac{5}{12}$, which might seem to be too large. If we admit it the parallax of the sun would be 8·555. But let us note that larger values of u give smaller values for x. Since the application of this method is so vague, we may well doubt whether in this fashion we can arrive any closer to the truth, and perhaps it will suffice to have learnt at any rate, whether the value of x will come out positive or negative.

20. In this case, to be sure, we have seen that the value of x is certainly positive, since we have found a negative number for C. Hence we may profitably observe in general that whenever C comes out positive, x becomes negative, while if C is negative the value of x will be positive. In either case it must of necessity be so small that the result will hardly differ from the common rule. This at any rate can be added, that the larger the number C, the greater must necessarily be the value of x. For if the sum of the cubes C actually vanished, then x would always $= 0$, whatever value is accepted for u, just as the common rule requires.

21. Thus, notwithstanding the uncertainty produced by the number u, it seems that something reasonably probable can be laid down even if we cannot reach certainty, if we pay attention to the following points. First, it is certain that whenever the sum of the cubes $C = 0$, x will always $= 0$. Secondly, the larger the quantity C, the larger will be the value of x itself, with the opposite sign. Thirdly, it is clear enough that the quantity nu^2 must be very much greater than the quantity $3B$. In view of this we can lay it down with reasonable probability that $x = -C/\lambda nB$, where the number λ, it is true, is left to our judgement. However, it will meet all cases and depart hardly at all from the truth, if we put $\lambda = 2$ or at most $\lambda = 3$. The resulting difference will usually be so unimportant that we hardly need consider it. For the case where the greatest error is to be feared would undoubtedly be that in which several observations, i in number, agree entirely in each giving the value a, while the one remaining observation gives $-ia$, so that the sum of them all $A = 0$. The sum of the squares $B = ia^2 + i^2a^2 = i(i+1)a^2$; the sum of the cubes $= ia^3 - i^3a^3 = -i(i^2-1)a^3$. For $n = i+1$ our formula gives

$$x = +\frac{i(i^2-1)a}{\lambda i(i+1)^2} = \frac{(i-1)a}{\lambda(i+1)}.$$

If therefore i is very large and we take $\lambda = 2$ the result is $x = \frac{1}{2}a$. In the earlier example where $n = b$, $B = 11\frac{1}{2}$ and $C = -801$, $x = +801/(12 \times 111\frac{1}{2}) = \frac{3}{5}$ approximately. In the second, where $n = 4$, $B = 1814$, $C = -24{,}750$, $x = 24{,}750/(8 \times 1814) = \frac{8}{5}$ approximately. These values do not appear to involve anything absurd.

If, however, anyone thinks that it would be more reasonable to take $\lambda = 3$, I hardly think the difference is worth arguing about, since the very nature of the observations does not admit of a greater degree of precision.