Stephen M. Stigler

THE HISTORY OF STATISTICS

The Measurement of Uncertainty before 1900

THE BELKNAP PRESS OF HARVARD UNIVERSITY PRESS
Cambridge, Massachusetts, and London, England

1. *Least Squares and the Combination of Observations*

Adrien Marie Legendre (1752–1833)

THE METHOD of least squares was the dominant theme — the leitmotif — of nineteenth-century mathematical statistics. In several respects it was to statistics what the calculus had been to mathematics a century earlier. "Proofs" of the method gave direction to the development of statistical theory, handbooks explaining its use guided the application of the higher methods, and disputes on the priority of its discovery signaled the intellectual community's recognition of the method's value. Like the calculus of mathematics, this "calculus of observations" did not spring into existence without antecedents, and the exploration of its subtleties and potential took over a century. Throughout much of this time statistical methods were commonly referred to as "the combination of observations." This phrase captures a key ingredient of the method of least squares and describes a concept whose evolution paced the method's development. The method itself first appeared in print in 1805.

Legendre in 1805

In March of 1805 political Europe focused its attention uneasily on France. The 1801 Peace of Amiens was crumbling, and preparations were under way for a new round of war, one that would begin that autumn with the battle of Trafalgar and the opening of the Napoleonic campaigns with victories at Ulm and Austerlitz. Scientific Europe also looked to France; there the discipline was intellectual, not martial, and was subsequently more sure and longer lived than that of the new emperor of the French.

In March of 1805 Laplace celebrated his fifty-sixth birthday, prepared the fourth volume of his *Traité de mécanique céleste* for press, and, perhaps, began to return his thoughts to the completion of a book on probability he had first contemplated more than twenty years before. Also in March of 1805 another French mathematical scientist, Adrien Marie Legendre, sent to press the final pages of a lengthy memoir that contained the first publication of what is even today the most widely used nontrivial technique of mathematical statistics, the method of least squares.

Legendre (born 18 September 1752, died 10 January 1833) was a mathematician of great breadth and originality. He was three years Laplace's junior and succeeded Laplace successively as professor of mathematics at the Ecole Militaire and the Ecole Normale. Legendre's best-known mathematical work was on elliptic integrals (he pioneered this area forty years before Abel and Jacobi), number theory (he discovered the law of quadratic reciprocity), and geometry (his *Eléments de géométrie* was among the most successful of such texts of the nineteenth century). In addition, he wrote important memoirs on the theory of gravitational attraction. He was a member of two French commissions, one that in 1787 geodetically joined the observatories at Paris and Greenwich and one that in 1795 measured the meridian arc from Barcelona to Dunkirk, the arc upon which the length of the meter was based. It is at the nexus of these latter works in theoretical and practical astronomy and geodesy that the method of least squares appeared.

In 1805 (the appendix we shall discuss is dated 6 March 1805) Legendre published the work by which he is chiefly known in the history of statistics, *Nouvelles méthodes pour la détermination des orbites des comètes.* At eighty pages this work made a slim book, but it gained a fifty-five-page supplement (and a reprinted title page) in January of 1806, and a second eighty-page supplement in August of 1820. The appendix presenting the method of least squares occupies nine of the first eighty pages; it is entitled "Sur la méthode des moindres quarrés." For stark clarity of exposition the presentation is unsurpassed; it must be counted as one of the clearest and most elegant introductions of a new statistical method in the history of statistics. In fact, statisticians in the succeeding century and three-quarters have found so little to improve upon that, but for his use of $\int$ instead of Σ to signify summation, the explanation of the method could almost be from an elementary text of the present day. Legendre began with a clear statement of his objective:

On the Method of Least Squares

In most investigations where the object is to deduce the most accurate possible results from observational measurements, we are led to a system of equations of the form

$$E = a + bx + cy + fz + \&c.,$$

in which, a, b, c, f, &c. are known coefficients, varying from one equation to the other, and x, y, z, &c. are unknown quantities, to be determined by the condition that each value of E is reduced either to zero, or to a very small quantity. (Legendre, 1805, p. 72)

We might write this today as

$$E_i = a_i + b_i x + c_i y + f_i z + \ . \ . \ .$$

or

$$a_i = -b_i x - c_i y - f_i z - \ . \ . \ . + E_i,$$

but we would probably join Legendre in calling the E's "errors." When the number of equations equaled the number of unknowns, Legendre saw no problem. But when there were more equations than unknowns, it became impossible to choose values for the unknowns that would eliminate all the errors. He noted that there was an element of arbitrariness in any way of, as he put it, "distributing the errors among the equations," but that did not stop him from dramatically proposing a single best solution:

Of all the principles that can be proposed for this purpose, I think there is none more general, more exact, or easier to apply, than that which we have used in this work; it consists of making the sum of the squares of the errors a *minimum*. By this method, a kind of equilibrium is established among the errors which, since it prevents the extremes from dominating, is appropriate for revealing the state of the system which most nearly approaches the truth. (Legendre, 1805, pp. 72–73)

Minimize the sum of the squares of the errors! How simple—but was it practical? Legendre wasted no time in writing down the equations he derived by differentiating the sum of squared errors $(a + bx + cy + fz +$

&c.)^2 + $(a' + b'x + c'y + f'z + \&c.)^2 + (a'' + b''x + c''y + f''z + \&c.)^2$ + &c. with respect to $x, y, \ldots$, namely,

$$0 = \int ab + x \int b^2 + y \int bc + z \int bf + \&c.$$

$$0 = \int ac + x \int bc + y \int c^2 + z \int fc + \&c.$$

$$0 = \int af + x \int bf + y \int cf + z \int f^2 + \&c.$$

> where by $\int ab$ we understand the sum of the similar products $ab + a'b' + a''b''$ + &c.; and by $\int b^2$ the sum of the squares of the coefficients of x, that is, $b^2 + b'^2 + b''^2$ + &c., and so on. (Legendre, 1805, p. 73)

Lest there be any doubt as to how to form these, the "normal equations" (a name introduced later by Gauss), Legendre restated the rule in italicized words:

> In general, *to form the equation of the* minimum *with respect to one of the unknowns, it is necessary to multiply all the terms of each equation by the coefficient of the unknown in that equation, taken with its proper sign, and then to find the sum of all these products.* (Legendre, 1805, p. 73)

The equations, as many as there were unknowns, could then be solved by "ordinary methods."

The practicality of Legendre's principle was thus self-evident; all it required was a few simple multiplications and additions and the willingness to solve a set of linear equations. Because the latter step would be required in any case, even if the number of observations equaled the number of unknowns, the formation of the "equations of the minimum" was all that Legendre added to the calculations. It was a small computational price to pay for such an appealingly evenhanded resolution to a difficult problem. In case more convincing was needed, Legendre added these points:

(1) If a perfect fit were possible, his method would find it.

(2) If it were subsequently decided to discard an equation (say, if its "error" were too large), it would be a simple matter to revise the equations by subtracting the appropriate terms.

(3) The arithmetic mean was a special case of the method, found when there is a single unknown with constant coefficient $b = b' = \ldots = 1$.

(4) Likewise, finding the center of gravity of several equal masses in space was a special case. By analogy with this last case Legendre closed his introduction of the method with these words, "We see, therefore, that the method of least squares reveals, in a manner of speaking, the center around which the results of observations arrange themselves, so that the deviations from that center are as small as possible" (Legendre, 1805, p. 75).

Legendre followed this explanation with a worked example using data from the 1795 survey of the French meridian arc, an example involving three unknowns in five equations. The clarity of the exposition no doubt contributed to the fact that the method met with almost immediate success. Before the year 1805 was over it had appeared in another book, Puissant's *Traité de géodésie* (Puissant, 1805, pp. 137–141; Puissant did use Σ for summation); and in August of the following year it was presented to a German audience by von Lindenau in von Zach's astronomical journal, *Monatliche Correspondenz* (Lindenau, 1806, p. 138–139). An unintended consequence of Legendre's publication was a protracted priority dispute with Carl Friedrich Gauss, who claimed in 1809 that he had been using the method since 1795 (Chapter 4).

Ten years after Legendre's 1805 appendix, the method of least squares was a standard tool in astronomy and geodesy in France, Italy, and Prussia. By 1825 the same was true in England.[1] The rapid geographic diffusion of the method and its quick acceptance in these two fields, almost to the exclusion of other methods, is a success story that has few parallels in the history of scientific method. It does, however, raise a number of important questions, including these: Did the introduction of the method create a historical discontinuity in the development of statistics, or was it a natural, if inspired, outgrowth of previous approaches to similar problems? What were the characteristics of the problems faced by eighteenth-century astronomers and geodesists that led to the method's introduction and easy acceptance? In what follows I shall attempt to answer these questions by examining several key works in these fields. In particular, I shall argue that least squares was but the last link in a chain of development that began about 1748, and that by the late 1780s methods were widely known and used that were, for practical purposes, adequate for the problems faced. I shall show how the deceptively simple concept that there was a potential gain to be achieved through the combination of observational data gathered under differing circumstances proved to be a major stumbling block in early work and how the development of these methods required the combination of extensive empirical experience and mathematical or mechanical insight.

1. The earliest English translation was by George Harvey (1822), "On the Method of Minimum Squares." Before the mid 1820s, it was common for Legendre's name for his method, "Moindres quarrés," to be translated into English as "minimum squares" or "small squares" rather than as "least squares."

Cotes's Rule

By the middle of the eighteenth century at least one statistical technique was in frequent use in astronomy and navigation: the taking of a simple arithmetic mean among a small collection of measurements made under essentially the same conditions and usually by the same observer (Plackett, 1958). But it is worth emphasizing that, widespread as the practice was, it was followed only in a narrowly conceived set of problems. Astronomers averaged measurements they considered to be equivalent, observations they felt were of equal intrinsic accuracy because the measurements had been made by the same observer, at the same time, in the same place, with the same instrument, and so forth. Exceptions, instances in which measurements not considered to be of equivalent accuracy were combined, were rare before 1750.

One possible exception is a rule found in a work of Roger Cotes published in 1722 (published posthumously, for Cotes died in 1716).

> Let p be the place of some object defined by observation, q, r, s the places of the same object from subsequent observations. Let there also be weights P, Q, R, S reciprocally proportional to the displacements which may arise from the errors in the single observations, and which are given from the given limits of error; and the weights P, Q, R, S are conceived as being placed at p, q, r, s, and their centre of gravity Z is found: I say the point Z is the most probable place of the object, and may be most safely had for its true place. (Cotes, 1722, p. 22; based on the translation by Gowing, 1983, p. 107)

Cotes's rule can be (and has been) read as recommending a weighted mean, or even as an early appearance of the method of least squares (De Morgan, 1833–1844, "Least Squares"). However, it has about it a vagueness that could only be cleared up by one or more accompanying examples, examples Cotes did not provide. To understand the genesis of the method of least squares, we must look not just at what investigators say they are doing (and how the statement might be most charitably interpreted in the light of later developments) but also at what was actually done. Cotes's rule had little or no influence on Cotes's immediate posterity. In the literature of the theory of errors its earliest citation seems to be that by Laplace (1812, p. 346; 1814, p. 188).

Tobias Mayer and the Libration of the Moon

The development of the method of least squares was closely associated with three of the major scientific problems of the eighteenth century: (1) to determine and represent mathematically the motions of the moon; (2) to account for an apparently secular (that is, nonperiodic) inequality that had been observed in the motions of the planets Jupiter and Saturn; and (3) to determine the shape or figure of the earth. These problems all involved astronomical observations and the theory of gravitational attraction, and they all presented intellectual challenges that engaged the attention of many of the ablest mathematical scientists of the period.

It seems most appropriate to begin with two works: "Recherches sur la question des inégalités du mouvement de Saturne et de Jupiter," written by Leonhard Euler and published in 1749; and "Abhandlung über die Umwalzung des Monds um seine Axe und die scheinbare Bewegung der Mondsflecten," written by Tobias Mayer and published in 1750. Although these works were not the first to consider their respective subjects (the inequalities of the motions of Jupiter and Saturn and the libration of the moon), they were among the best of the early treatments of these subjects. Because they were widely read, they greatly influenced later workers and, from a statistical point of view, form a unique and dramatic contrast in the handling of observational evidence. Together they tell a story of statistical success (by a major astronomer, Mayer) and statistical failure (by a leading mathematician, Euler). They show why the discovery of the method of least squares was not possible in the intellectual climate of 1750, and they highlight the conceptual barriers that had to be crossed before this climate became sufficiently tropical to support the later advances of Legendre, Gauss, and Laplace.

We shall consider Mayer's statistical success first. Notwithstanding the principal "monthly" regularity in the motion of the moon about the earth, its detailed motion is extraordinarily complex. In the eighteenth century the problem of accurately accounting for these minor perturbations in the moon's movements, either by a mathematical formula or by an empirically determined table describing future lunar positions, was of great scientific, commercial, and even military significance. Its scientific importance lay in the general desire to show that Newtonian gravitational theory can account for the movements of our nearest celestial neighbor (within the always decreasing limits of observational error) if allowance is made for the attraction of other bodies (such as the sun), for periodic changes in the earth's and the moon's orbits, and for the departures from sphericity of the shapes of the earth and moon.[2] But it was the potential commercial and military usefulness of a successful accounting of the moon (as an aid to navigation) that was primarily responsible for the widespread attention the problem received. Over the previous nineteen centuries, from Hipparchus and Ptolemy to Newton and Flamsteed, the linked development of theoretical and practical astronomy had played the key role in freeing ship's navigators from a dependence upon land sightings as a way of deter-

2. As eloquent testimony to the difficulty of the problem, we have Newton reportedly telling Halley that lunar theory "made his head ache and kept him awake so often that he would think of it no more" (Berry, 1898, p. 240).

mining the ship's position. The developments of better nautical instruments—including the sextant in 1731—and a more accurate understanding of astronomical theory, increasingly enabled navigators to map their ships' courses across previously trackless seas. By 1700 it had become possible to determine a ship's latitude at sea with relative precision by the fixed stars—simply by measuring the angular elevation of the celestial pole above the horizon. The determination of longitude, however, was not so simple. Indeed, in 1714 England established the "commissioners for the discovery of longitude at sea," a group that by 1815 had disbursed £101,000 in prizes and grants to achieve its goal. The two most promising methods of ascertaining longitude at sea were the development of an accurate clock (so that Greenwich time could be maintained on shipboard and longitude determined by the comparison of the fixed stars' positions and Greenwich time) and the creation of lunar tables that permitted the determination of Greenwich time (and thus of longitude) by comparison of the moon's position and the fixed stars.

Johann Tobias Mayer (1723–1762) had already made a name for himself as a cartographer and practical astronomer by the time he undertook a study of the moon in 1747 (a study that eventually led to the preparation of lunar tables that were to earn his widow £3,000 from the British commissioners in 1765). The specific work of Mayer that most influenced statistical practice was his study, published in 1750, of the librations of the moon.

The popular notion that the moon always presents the same face to the earth is not literally true. The moon in fact is subject to "libration": The face viewed from earth varies, so that over an extended period of time about 60 percent of the moon's surface is visible from earth. Two sources of this libration were known to Galileo: the apparent diurnal libration due to the earth's rotation and a libration in latitude due principally to the fact that the moon's axis of rotation is not perpendicular to the earth's orbital plane about the sun. By the time of Mayer's work it was known that the earth's location was at a focus, not the center, of the moon's elliptical orbit. Thus the moon's rotation at a uniform speed produced a third type of libration, one of longitude.

Over the period from April 1748 to March 1749, Mayer made numerous observations of the positions of several prominent lunar features; and in his 1750 memoir he showed how these data could be used to determine various characteristics of the moon's orbit (Figure 1.1). His method of handling the data was novel, and it is well worth considering this method in detail, both for the light it sheds on his pioneering, if limited, understanding of the problem and because his approach was widely circulated in the major contemporary treatise on astronomy, having signal influence upon later work.

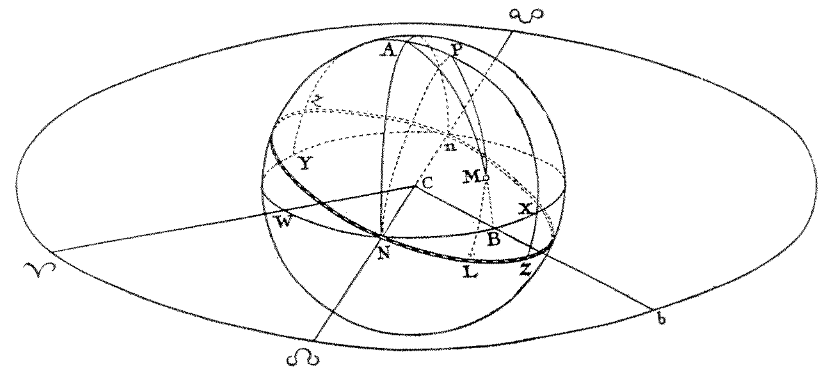

Figure 1.1. Tobias Mayer's original drawing of the moon. (From Mayer, 1750, table VI.)

Mayer's method for the resolution of inconsistent observational equations can be discerned in his discussion of the position of the crater Manilius. Figure 1.2 represents the moon, which Mayer considered as a sphere. The great circle QNL represents the moon's true equator, and P is the moon's pole with respect to this equator, one end of its axis of revolution. The great circle DNB is that circumference (or apparent equator) of the moon that is seen from earth as parallel to the plane of the ecliptic, the plane of the earth's orbit about the sun, and A is the pole of the moon with respect to DNB, its apparent pole as viewed by an earthbound astronomer oriented by the ecliptic. The point γ (the point on the circle DNB in the direction from the moon's center C toward the equinox) was taken as a reference point. The circle DNB and the pole A will vary with time, as a result of the libration of the moon, but they form the natural system of coordinates at a given time. The equator QNL and the pole P are fixed but not observable from earth. Mayer's aim was to determine the relationship between these coordinate systems and thus accurately determine QNL and P. He accomplished this by making repeated observations of the crater Manilius. Now, in Figure 1.2, M is the position of Manilius, and PL and AB are meridian quadrants through M with respect to the two polar coordinate systems. Mayer was able to observe the position of M on several occasions with respect to the constantly changing coordinate system determined by DNB and A; that is, subject to observational error he could at a given time measure the arcs $AM \equiv h$ and $\gamma B \equiv g$.

To determine the relationship between the coordinate systems, Mayer sought to find the fixed, but unknown, arc length $AP \equiv \alpha$, the true

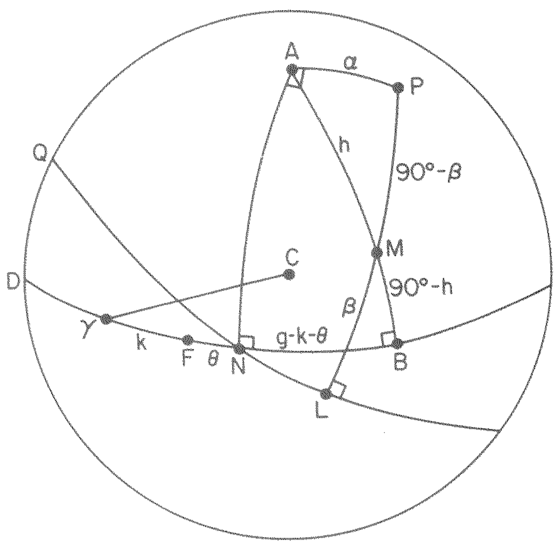

Figure 1.2. The moon. M, The crater Manilius; NL, the moon's equator; P, the moon's equatorial pole; NB, the circumference parallel to the plane of the ecliptic; A, the pole of NB; Cγ, the direction of the equinox from the moon's center C; F, the node of the moon's orbit and the plane of the ecliptic. See text for more details.

latitude of Manilius $\beta = \mathrm{ML}$, and the distance θ between the unknown node or point of intersection of the two circles (N) and the known point of intersection F of the plane of the orbit of the moon and the circle DNB. He let $k = \gamma\mathrm{F}$ be the observed longitude of F. Then g, h, and k were observable and varied from observation to observation as a result of the motion of the moon (and observational error); and α, θ, and β were fixed and unknown, to be determined from the observations. Because NAP forms a right angle, a basic identity of spherical trigonometry implies that these quantities are related nonlinearly by the equation.

$$\text{(A)} \qquad \sin\beta = \cos\alpha\cos h + \sin\alpha\sin h\sin(g - k - \theta).$$

Now, Mayer knew that both α and θ were small (in the neighborhood of 2 or 3 degrees), and he proceeded, via several trigonometric identities, to derive an almost linear approximation to this equation under the supposition that $\cos\alpha$, $\cos\theta$, and $\cos(\beta - 90° + h)$ were approximately 1.0, and

$\sin(\beta - 90° + h) \cong \beta - 90° + h$. He was led[3] to the equation

$$\text{(C)} \qquad \beta - (90° - h) = \alpha\sin(g - k) - \alpha\sin\theta\cos(g - k),$$

which provided at least an approximate relationship between the observations and the unknowns.

At this point Mayer was in sight of his goal. He only needed to take observations for three different days, solve the resulting three linear equations for β, α, and $\alpha\sin\theta$ (and then solve for θ), and he would be done. His problem, however, was that he suffered an embarrassment of riches—he had twenty-seven days' observations of Manilius. The resulting twenty-seven equations are given in Table 1.1.

The form of Mayer's problem is almost the same as that of Legendre; Legendre might have written E for the discrepancy (due to the linear approximation and observational error) between the two sides; he would then have had

$$\mathrm{E} = (90° - h) - \beta + \alpha\sin(g - k) - \alpha\sin\theta\cos(g - k).$$

In Mayer's form the equations came to be called the *equations of condition* because they expressed a condition or relationship that would hold if no errors were present. The modern tendency would be to write, say, $(h - 90°) = -\beta + \alpha\sin(g - k) - \alpha\sin\theta\cos(g - k) + \mathrm{E}$, treating $h - 90°$ as the dependent variable and $-\beta$, α, and $-\alpha\sin\theta$ as the parameters in a linear regression model.

How did Mayer address his overdetermined system of equations? His approach was a simple and straightforward one, so simple and straightforward that a twentieth-century reader might arrive at the very mistaken opinion that the procedure was not remarkable at all. Mayer divided his equations into three groups of nine equations each, added each of the three groups separately, and solved the resulting three linear equations for α, β, and $\alpha\sin\theta$ (and then solved for θ). His choice of which equations belonged in which groups was based upon the coefficients of α and $\alpha\sin\theta$. The first group consisted of the nine equations with the largest positive values for the coefficient of α, namely, equations 1, 2, 3, 6, 9, 10, 11, 12, and 27. The second group were those with the nine largest negative values for this coefficient: equations 8, 18, 19, 21, 22, 23, 24, 25, and 26. The remaining nine equations formed the third group, which he described as having the largest values for the coefficient of $\alpha\sin\theta$.

3. Mayer started with $\sin(g - k - \theta) = \sin(g - k)\cos\theta - \sin\theta\cos(g - k) \cong \sin(g - k) - \sin\theta\cos(g - k)$. Then, setting $\cos\alpha = 1$ in equation (A), he had equation (B) $\sin\beta - \cos h = \sin\alpha\sin h\sin(g - k) - \sin\alpha\sin h\sin\theta\cos(g - k)$; letting $x = \beta - (90° - h)$, a small quantity, he had $\sin\beta = \cos(h - x) = \cos h\cos x + \sin h\sin x \cong \cos h + x\sin h$, which with (B) gives $x\sin h \cong \sin\alpha\sin h\sin(g - k) - \sin\alpha\sin h\sin\theta\cos(g - k)$, and setting $\alpha \cong \sin\alpha$ and dividing through by $\sin h$ gives (C).

Table 1.1. Mayer's twenty-seven equations of condition, derived from observations of the crater Manilius from 11 April 1748 through 4 March 1749.

Eq. no.	Equation	Group
1	$\beta - 13°10' = +0.8836\alpha - 0.4682\alpha \sin\theta$	I
2	$\beta - 13°8' = +0.9996\alpha - 0.0282\alpha \sin\theta$	I
3	$\beta - 13°12' = +0.9899\alpha + 0.1421\alpha \sin\theta$	I
4	$\beta - 14°15' = +0.2221\alpha + 0.9750\alpha \sin\theta$	III
5	$\beta - 14°42' = +0.0006\alpha + 1.0000\alpha \sin\theta$	III
6	$\beta - 13°1' = +0.9308\alpha - 0.3654\alpha \sin\theta$	I
7	$\beta - 14°31' = +0.0602\alpha + 0.9982\alpha \sin\theta$	III
8	$\beta - 14°57' = -0.1570\alpha + 0.9876\alpha \sin\theta$	II
9	$\beta - 13°5' = +0.9097\alpha - 0.4152\alpha \sin\theta$	I
10	$\beta - 13°2' = +1.0000\alpha + 0.0055\alpha \sin\theta$	I
11	$\beta - 13°12' = +0.9689\alpha + 0.2476\alpha \sin\theta$	I
12	$\beta - 13°11' = +0.8878\alpha + 0.4602\alpha \sin\theta$	I
13	$\beta - 13°34' = +0.7549\alpha + 0.6558\alpha \sin\theta$	III
14	$\beta - 13°53' = +0.5755\alpha + 0.8178\alpha \sin\theta$	III
15	$\beta - 13°58' = +0.3608\alpha + 0.9326\alpha \sin\theta$	III
16	$\beta - 14°14' = +0.1302\alpha + 0.9915\alpha \sin\theta$	III
17	$\beta - 14°56' = -0.1068\alpha + 0.9943\alpha \sin\theta$	III
18	$\beta - 14°47' = -0.3363\alpha + 0.9418\alpha \sin\theta$	II
19	$\beta - 15°56' = -0.8560\alpha + 0.5170\alpha \sin\theta$	II
20	$\beta - 13°29' = +0.8002\alpha + 0.5997\alpha \sin\theta$	III
21	$\beta - 15°55' = -0.9952\alpha - 0.0982\alpha \sin\theta$	II
22	$\beta - 15°39' = -0.8409\alpha + 0.5412\alpha \sin\theta$	II
23	$\beta - 16°9' = -0.9429\alpha + 0.3330\alpha \sin\theta$	II
24	$\beta - 16°22' = -0.9768\alpha + 0.2141\alpha \sin\theta$	II
25	$\beta - 15°38' = -0.6262\alpha - 0.7797\alpha \sin\theta$	II
26	$\beta - 14°54' = -0.4091\alpha - 0.9125\alpha \sin\theta$	II
27	$\beta - 13°7' = +0.9284\alpha - 0.3716\alpha \sin\theta$	I

Source: Mayer (1750, p. 153).
Note: One misprinted sign in equation 7 has been corrected.

Even though Mayer's description of the third group is not fully accurate (compare equation 8, in group II, with equation 13 in group III), his specification of the three groups shows insight into the geometry of the situation and reveals that his choice of the crater Manilius was perhaps motivated by at least dimly perceived notions of experimental design. The coefficients of α and $\alpha \sin\theta$ are $\sin(g-k)$, and $-\cos(g-k)$, which are related by $[\sin(g-k)]^2 + [-\cos(g-k)]^2 = 1$. The first group consists of those nine equations whose coefficients of α are nearest 1.0, the second group of those nearest -1.0, leaving the third group as those equations with $\sin(g-k)$ "near" zero, that is, with large (although not without exception the largest) $-\cos(g-k)$'s (which all happen, because of the choice

of crater, to be positive). This way of choosing equations for aggregation tends, subject to the restriction to equal groups, to maximize the contrast among the coefficients of α and produce good estimates of α, and, with the present selection of crater, good estimates of $\alpha \sin\theta$ as well. Mayer seems to have understood this because he wrote, "These equations [Table 1.2] can take the place of the foregoing totality of equations [Table 1.1] because each of these three equations has been formed in the most advantageous manner *(die vortheilhaftigste Art).* The advantage consists in the fact that through the above division into three classes, the differences between the three sums are made as large as is possible. The greater these differences are, the more accurately *(richtiger)* one may determine the unknown values of α, β, and θ" (Mayer, 1750, p. 154).

Mayer solved the three equations he found, getting $\alpha = 89'.90 \cong 1°30'$, $\theta = -3°45'$, and $\beta = 14°33'$, and he went on to consider the accuracy of these values. He noted that even under favorable conditions an individual observation of an arc could only be counted as accurate to within 10 or 15 minutes, and he claimed that the effect of an error of this magnitude in g and h on the final determinations could be traced through the formulas he had given. He did not attempt this kind of nonstatistical error analysis, however. Instead he presented an empirical assessment of accuracy.

Earlier in his paper Mayer had illustrated how α, β, and θ were calculated on the basis of only three observational equations (equations 9, 16, and 19 of Table 1.1). The value he had found for α based on those three equations was $\alpha = 1°40'$. Now, he noted, "Because these last values [based on all twenty-seven equations] were derived from nine times as many observations, one can therefore conclude that they are nine times more correct *(neunmal richtiger)*; therefore the error *(Fehler)* in each of the constants is in inverse relationship to the number of their observations" (Mayer, 1750, p. 155). Mayer turned this statement into an interval description of the most important of the unknowns, α, as follows:

> Let the true value *(wahre Wehrt)* be $\alpha = 1°30' \pm x$; then x is the difference or the error *(der Unterschied oder Irrthum)*: how far the quantity α, determined

Table 1.2. Mayer's three equations, as derived from Table 1.1 by adding equations 1, 2, 3, 6, 9, 10, 11, 12, and 27 in group I, equations 8, 18, 19, 21, 22, 23, 24, 25, and 26 in group II, and the rest in group III.

Group	Equation
I	$9\beta - 118°8' = +8.4987\alpha - 0.7932\alpha \sin\theta$
II	$9\beta - 140°17' = -6.1404\alpha + 1.7443\alpha \sin\theta$
III	$9\beta - 127°32' = +2.7977\alpha + 7.9649\alpha \sin\theta$

Source: Mayer (1750, p. 154).

from the 27 observations, can deviate from the true value. Since from three observations we found $\alpha = 1°40'$, the error *(der Fehler)* of that determination is found to be $= 10 \pm x$; consequently we are led to conclude that

$$\pm x : \frac{1}{27} = 10 \pm x : \frac{1}{3},$$

from which we find $x = \pm 1' \frac{1}{4}$. The true value of α can therefore be about 1′ or 2′ smaller or larger than 1°30′. (Mayer, 1750, p. 155)

Thus Mayer introduced the symbol $\pm x$ for the error made in taking $\alpha = 1°30'$; and $10 \pm x$ was the error made by taking $\alpha = 1°40'$. Because the determination 1°30′ was based on nine times as many observations as 1°40′ (and was thus nine times more accurate), he supposed that $\pm x \cdot 27 = (10 \pm x) \cdot 3$. His solution to this equation makes it clear that he assumed both sides must have the same sign, for, setting $e = \pm x$, he actually solves $e \cdot 27 = (10 + e) \cdot 3$ to get $e = 30/24 = 1.25$. Thus either he misses the possibility of solving $|e| \cdot 27 = |10 + e| \cdot 3$ to get $e = -1$ (the error in taking $\alpha = 1°40'$ is nine times that of $\alpha = 1°30'$, albeit in a different direction), or he has deliberately taken the larger of the two values, to give a conservative bound to the error.

We now know that Mayer's judgment of the inverse relationship between the number of equations used and the accuracy of the determination was too optimistic; statistical accuracy at most increases only as the square root of the number of equations, subject to various assumptions on the conditions under which the observations are made. It was to be several years before that relationship emerged in the works of Laplace and, later, Gauss. Thus, rather than be surprised at Mayer's overly optimistic view of his procedure's accuracy, we should be surprised at how qualitatively correct this view was. In fact, even to attempt a numerical estimate of the accuracy of an empirical determination was remarkable for the time.

We can express Mayer's error assessment in modern notation, in the special case of determining a mean as follows: Let e be the limit of accuracy (analogous to Mayer's $\pm x$) for a mean $\overline{X}$ (analogous to his observed 1°30′), let X_1 be a single determination (analogous to his 1°40′), and let n be the ratio of the sample sizes entering into $\overline{X}$ and X_1 (analogous to his $9 = 27/3$). Then $|X_1 - \overline{X}|$ is analogous to his 10, and he would take

$$e \div (1/n) = (|X_1 - \overline{X}| + e) \div 1, \qquad \text{or} \qquad e = |X_1 - \overline{X}|/(n-1).$$

Of course, this represents a considerable formal extrapolation of Mayer's intention, but it shows that his approach was at least qualitatively sound, even if far from the best we can do today. The point is not that he found a particularly clever method of combining his twenty-seven equations but that he found it useful to combine the equations at all, instead of, say, being content with selecting three "good" ones and solving for the unknowns

from them, as he did by way of illustration. This aspect of Mayer's approach is best appreciated by comparing Mayer's work with Euler's memoir of a year earlier.

Saturn, Jupiter, and Euler

Leonhard Euler (1707–1783) was and is best known for his work in pure analysis, but he worked in nearly every area of pure and applied mathematics known at the time (or invented in the succeeding century). Euler was the most prolific mathematician of all time; his collected works now run to nearly eighty quarto volumes and are still in the process of publication. If the maxim "Publish or perish" held literally, Euler would be alive today. Yet for all this abundance, the quality of his work did not suffer; and on several occasions he was honored by foreign academies for his solutions to outstanding problems. One such instance was the prize announced by the Academy of Sciences in Paris for the year 1748, when Euler was in Berlin.

The Academy problem of 1748 concerned the second of the major scientific problems we shall consider in this chapter; entrants were invited to prepare memoirs giving "A Theory of Saturn and of Jupiter, by which one can explain the inequalities that the two planets appear to cause in each other's motion, principally near the time of their conjunction" (Euler, 1749, p. 45).

In 1676 Halley had verified an earlier suspicion of Horrocks that the motions of Jupiter and Saturn were subject to slight, apparently secular, inequalities. When the actual positions of Jupiter and Saturn were compared with the tabulated observations of many centuries, it appeared that the mean motion of Jupiter was accelerating, whereas that of Saturn was retarding. Halley was able to improve the accuracy of the tables by an empirical adjustment, and he speculated that the irregularity was somehow due to the mutual attraction of the planets. But he was unable to provide a mathematical theory that would account for this inequality.

This problem, like that of the motions of the moon, was an instance of the three-body problem. Its appearance at this time was also due to improved accuracy of astronomical observations revealing inadequacies of simple two-body theories of attraction. Unlike the problem of the moon, however, that of Jupiter and Saturn was not commercially motivated;[4]

4. The motions of Jupiter and Saturn about the sun are too slow to be useful for determining longitude at sea. In 1737 W. Whiston suggested that if reflecting telescopes were made part of a ship's navigational equipment then longitude could be determined by observing the eclipses of the moons of Jupiter. Although this method would work on land, the suggestion must have been a source of some amusement to naval astronomers who knew the instabilities of a ship's deck as an observational platform (W. R. Martin, "Navigation," in *Encyclopaedia Britannica*, 11th ed., p. 289).

rather it drew its major impetus from the philosophical implications of an unstable solar system. If the observed trends were to continue indefinitely, Jupiter would crash into the sun as Saturn receded into space! The problem posed by the Academy could be (and was) interpreted as requiring the development of an extension of existing theories of attraction to incorporate the mutual attraction of three bodies, in order to see whether such a theory could account for at least the major observed inequalities as, it was hoped, periodic in nature. Thus stability would be restored to the solar system, and Newtonian gravitational theory would have overcome another obstacle.

Euler's memoir on this difficult subject was judged the winner of the prize even though it fell far short of providing a complete resolution of the problem. This 123-page memoir, "Researches on the question of the inequalities in the movement of Saturn and of Jupiter," was published separately in Paris in 1749, and it may still be read today as a model of clear and orderly mathematical exposition. Euler focused his attention primarily on Saturn (which as the smaller of the two is subject to more pronounced perturbations) and developed an equation for the longitudinal position of Saturn that took into account the mutual attractions of the three bodies. He began by assuming that Jupiter and Saturn followed circular orbits about the sun and that the orbits lay in the same plane. Finding that the simple theory that resulted from this assumption would not admit inequalities of the size actually observed, he considered more complicated hypotheses. He first permitted Saturn's orbit to be an ellipse, he then permitted Jupiter to follow an elliptical orbit also, and he finally incorporated the fact that the planes of the two orbits are not coincident, but at a slight inclination, into the calculations.

After he had completed his mathematical analysis, it remained for Euler to check his results empirically. He wrote, "After having determined the derangements that the action of Jupiter should cause in the movement of Saturn, I now pass to an examination of the degree of precision with which they agree with the observations" (Euler, 1749, p. 111). To make this comparison, he developed a formula for the heliocentric longitude φ of Saturn in the following form (p. 121):

$$\begin{aligned}\varphi = \eta &- 23525'' \sin q + 168'' \sin 2q - 32'' \sin 2\omega \\ &- 257'' \sin(\omega - q) - 243'' \sin(2\omega - p) + m'' \\ &- x'' \sin q + y'' \sin 2q - z'' \sin(\omega - p) \\ &- u(\alpha + 360\ v + p)\cos(\omega - p) + Nn'' \\ &- 0.11405k'' \cos q + (1/600)k'' \cos 2q.\end{aligned}$$

Of this impressive array of symbols, some ($\varphi, \eta, q, \omega, p, N, v$) were given by observation and varied from observation to observation, and some ($x, y, m, z, \alpha, k, n, u$) were fixed unknown corrections whose values were not specified by the theory. A full explanation of this equation is not required here. In essence it represents the observed heliocentric longitude of Saturn (φ) as equal to Saturn's mean heliocentric longitude η (what Saturn's longitude would be if we ignored perturbations and assumed an elliptical orbit) and correction terms. These terms depend on the difference between the longitudes of the two planets (ω), the number of years since 1582 (N), and various orbital characteristics (the planet's eccentric anomalies, p and q; and the number of complete orbits by Jupiter since 1582, v).

The problem Euler faced was this: He had available seventy-five complete sets of observations of $\varphi, \eta, q, \omega, p, N$, and v made in the years from 1582 through 1745. From these he first derived values of n and u in which he had confidence. It remained to determine the six unknown corrections x, y, m, z, α, and k and to check whether, when their values were substituted in the equation for φ, the values derived for the right-hand side agreed well enough with the observed values of φ to enable him to say that the theory explained the observed motions of Saturn.

The problem was an extraordinarily difficult one for the time, and Euler's attempts to grope for a solution are most revealing. Euler's work was, in comparison with Mayer's a year later, a statistical failure. After he had found values for n and u, Euler had the data needed to produce seventy-five equations, all linear in x, y, m, z, α, and k. He might have added them together in six groups and solved for the unknowns, but he attempted no such overall combination of the equations.

To see how Euler did work with his data, it is instructive to look at the way he found the first two unknowns, n and u. He noted that the coefficients of all terms except those involving n and u were, to a close approximation, periodic, with a period of fifty-nine years. He subtracted the equation for 1703 from that for 1585 (2×59 years apart) and that for 1732 from that for 1673 (fifty-nine years apart), thereby getting two linear equations in n and u alone. He solved these equations and checked his results by comparing them with another set derived from four other equations, similarly spaced in time.

Euler attempted to evaluate other correction factors by the same method, that is, by looking at small sets of equations taken under astronomically similar conditions, and thus creating a situation in which many of the coefficients would be approximately equal and the difference of two equations would annihilate most terms. But he did not succeed in finding other situations (like that for n and u) where different sets of equations gave the same results. Once he derived six inconsistent linear equations in only two unknowns but stated that, "Now, from these equations we can conclude nothing; and the reason, perhaps, is that I have tried to satisfy several observations exactly, whereas I should have only satisfied them approximately; and this error has then multiplied itself" (Euler, 1749, p.

136). Immediately after this, he presented twenty-one of the equations, involving the six unknowns other than n and u, only to throw up his hands with no real attempt at a solution. The furthest he went was to set five of the unknowns equal to zero (all except the term whose coefficient was unity in all equations, m). He adjusted this remaining term to be halfway between the largest and smallest constants in the twenty-one equations, thus making the maximum discrepancy as small as possible.

The comparison between the approaches of Euler and Mayer is dramatic. In 1750 Mayer, faced with a set of twenty-seven inconsistent equations in three unknowns, devised a sensible method of combining them into three equations and solving for the unknowns. In 1749 Euler, faced with up to seventy-five equations in up to eight unknowns, was reduced to groping for solutions. Euler worked with small sets of equations (usually as many as there were unknowns), and he only accepted numerical answers when different small sets of equations yielded essentially the same results. Euler's problem was similar to Mayer's, yet of the two only Mayer succeeded in finding a statistical solution to his "problem": a "combination of observations" that Euler could not devise (and, we shall argue, would not have accepted).

The two men brought absolutely first-rate intellects to bear on their respective problems, and both problems were in astronomy. Yet there was an essential conceptual difference in their approaches that made it impossible for Euler to adopt a statistical attitude and a subtle difference between their problems that made it extremely unlikely that Euler would overcome this conceptual barrier. The differences were these: Mayer approached his problem as a practical astronomer, dealing with observations that he himself had made under what he considered essentially similar observational conditions, despite the differing astronomical conditions. Euler, on the other hand, approached his problem as a mathematician, dealing with observations made by others over several centuries under unknown observational conditions. Mayer could regard errors or variation in his observations as random (even though no explicit probability considerations were introduced), and he could take the step of aggregating equations without fear that bad observations would contaminate good ones. In fact, he approached his problem with the conviction that a combination of observations increased the accuracy of the result in proportion to the number of equations combined. Euler could not bring himself to accept such a view with respect to his data. He distrusted the combination of equations, taking the mathematician's view that errors actually increase with aggregation rather than taking the statistician's view that random errors tend to cancel one another.

It has long been the practice of mathematicians to think in terms of the *maximum* error that could occur in a complex calculation rather than in terms of the likely error, to think in terms of absolute error bounds (which would typically increase with aggregation) rather than in terms of likely error sizes (which would not). For example, if a quantity is derived from adding together four numbers, any one of which could be in error by two units, then the sum could err by $4 \cdot 2 = 8$ units, and any attempt at an exact calculation would have to allow for ± 8 units of *possible* error. The longer the chain of calculation, the greater the maximum possible error — the more the potential error would tend to multiply. On the other hand, later statistical theory would show that under some conditions the likely error in such a sum could be much less (perhaps ± 2 units, if the likely error in one number was half the maximum possible) and the likely error in averages would actually decrease even though the mathematician's error bounds would not. Although that theory was yet to come (see Chapters 2 and 3), practicing astronomers like Mayer already had, based upon experience, at least a qualitative sense of its results in simple situations. For example, when Nevil Maskelyne wrote in 1762 that "by examining the error of the adjustment in this manner, by at least three trials, and taking a medium of the results, one can scarce err above half a minute in determining the exact error of the quadrant; whereas one may be mistaken a minute, or more, by a single trial" (Maskelyne, 1762; 1763, p. 4), it was based on his experience with his instrument, not his reading of the small amount of theory available by that time. Euler lacked that kind of direct experience. One bit of evidence supporting the idea that Euler took the more conservative mathematician's view is his previously quoted remark that the error made by supposing the equations held exactly has "multiplied itself" *(cette faute s'est ensuite augmentée)* (Euler, 1749, p. 136). An earlier statement was slightly more to the point: "By the combination of two or more equations, the errors of the observations and of the calculations can multiply themselves" (Euler, 1749, p. 135).

At one point Euler did combine two equations by averaging, but only when all corresponding coefficients were approximately equal. By subtracting three pairs of equations (and annihilating all terms except those involving x, y, z, and αu), Euler found three equations for x. He averaged the first two:[5]

$$\begin{aligned} x &= 683'' - 0.153y'' + 0.179z'' + 0.984\alpha u^\circ \\ \underline{x} &\underline{= 673'' - 0.153y'' + 0.187z'' + 0.983\alpha u^\circ} \\ x &= 678'' - 0.153y'' + 0.183z'' + 0.983\alpha u^\circ. \end{aligned}$$

But the third gave

$$x = 154'' + 0.067y'' + 0.192z'' + 0.980\alpha u^\circ,$$

5. This single instance of averaging seems to be the source of the occasional mistaken attribution to Euler of Mayer's method, which was later called the "method of averages."

and he would conclude only that "the value of *y* is quite large" (Euler, 1749, pp. 130–131). At no other point did he average or add equations except to take advantage of an existing periodicity to cancel terms.

In calling attention to Euler's statistical failure, I mean to imply no criticism of Euler as a mathematical scientist. Rather by contrasting his work with Mayer's I want to highlight the extremely subtle conceptual advance that was evident in Mayer's work. Euler's memoir made a significant contribution to the mathematical theory of attraction. Even his crude empirical solution, setting most of the correction terms equal to zero, provided him with an improvement over existing tables of Saturn's motion. His inability to resolve the major inequalities in the planets' motions was in the end due to the inadequacy of his theory as well as to his lack of statistical technique. Euler himself was satisfied, correctly, that no values of his correction factors could adequately account for the planet's motions, but he suggested as a possible cause for this the failure of Newton's inverse square law of attraction to hold exactly over large distances! The problem in fact withstood successive assaults by Lagrange, Lambert, and Laplace before finally yielding to Laplace in 1787. The lesson Euler's work has for the history of statistics is that even though before 1750 mathematical astronomers were willing to average simple measurements (combining observational evidence from several days or observers into a single number), it was only after 1750 that the conceptual advance of combining observational equations (with varying coefficients for several unknowns testifying to the differing circumstances under which the observations had been made) began to appear.

To what extent was Mayer's approach a method that could be generalized and transferred to problems other than its original application? Did his contemporaries or immediate followers attempt such generalizations? Mayer himself used the approach three times in his 1750 memoir: on the twenty-seven equations for Manilius, on nine equations for the crater Dionysius, and on twelve equations for the crater Censorinus.[6] He made no attempt, however, to describe his calculation as a method that would be useful in other problems. He did not do what Legendre did, namely, abstract the method from the application where it first appeared. His work proved to be influential nonetheless.

In a widely read treatise, *Astronomie* (1771), Joseph Jérôme Lalande presented an extensive discussion of Mayer's work for the specific purpose of explaining how large numbers of observational equations could be combined to determine unknown quantities. Indeed, Lalande presented virtually the whole of Mayer's analysis of Manilius, in what amounts to an

6. We can speculate that the numbers of equations were chosen to permit an equal division into three groups, perhaps by discarding one or two equations. Mayer does not comment on this point, however.

only slightly abridged translation (Lalande, 1771, vol. 3, pp. 418–428), saying, "I report the following numbers only to serve as an example of the method that I wish to explain" (p. 419). It seems plausible that it was Lalande's exposition that called the method to Laplace's attention and that it was Laplace who first developed it into the form in which it became widely known in the nineteenth century as "Mayer's method."

Laplace's Rescue of the Solar System

Pierre Simon Laplace was born in Normandy on 23 March 1749, and his life spanned the Napoleonic era. In what most would agree was the golden age of French science, Laplace was France's most illustrious scientist. Upon his death on 5 March 1827, Poisson eulogized him as "the Newton of France," and the phrase seems apt: Laplace was Newtonian in outlook, breadth, and, at least in probability and statistics, in accomplishment. By the age of twenty his mathematical talent had won him the patronage of d'Alembert; by the end of 1773 he was a member of the Academy of Sciences. At one time or another he was professor at the Ecole Militaire (he is said to have examined, and passed, Napoleon in 1785), member of the Bureau des Longitudes, Professor at the Ecole Normale, Minister of the Interior (for six weeks, in 1799, before being displaced by Napoleon's brother), and Chancellor of the Senate. Laplace's scientific work was no less varied than his public career. His scientific memoirs constitute seven of the total of fourteen volumes of his (not quite complete) *Oeuvres complètes*. About half of them were concerned with celestial mechanics, nearly one-quarter with mathematics exclusive of probability; and the remainder were divided between probability and physics (Stigler, 1978b). He is best known in the history of science for two major treatises that were distilled from this work: the *Traité de mécanique céleste* (four volumes, 1799–1805, with a supplementary volume published in 1825) and the *Théorie analytique des probabilités* (1812). John Playfair called the first of these "the highest point to which man has yet ascended in the scale of intellectual attainment" (Playfair, 1808, pp. 277–278), Augustus De Morgan described the second as "the Mont Blanc of mathematical analysis" (De Morgan, 1837).

In 1787, in the course of a memoir on the inequalities in the motions of Saturn and Jupiter, Laplace proposed what amounts to an extension of Mayer's method of reconciling inconsistent linear equations. In this epochal work, Laplace finally laid to rest what was by then a century-old problem by showing that the inequalities were in fact periodic (with a very long period). In the course of his demonstration Laplace was confronted with equations of the type that had stalled Euler's drive toward a solution.

Laplace's success in this celebrated problem was in considerable part a statistical triumph, a model of the ways in which analyses of data suggested by theory may in turn suggest hypotheses requiring further theoretical development and then observational confirmation. In 1773 Lambert had deduced, on the basis of an empirical investigation of contemporary observations, that the retardation in Saturn's motion noticed by Halley had apparently reversed — Saturn was accelerating and apparently had been doing so at least since 1640. This observation suggested (but did not prove) that the inequality was periodic rather than secular as Halley (and even Euler and Lagrange in early work) had thought. Laplace sought to account for the motions within the constraints of Newtonian gravitational theory, to within the limits of observational accuracy. The problem was an exceedingly difficult one. Even if the inequality was periodic, which of the known planets or moons would he need to include in the theory to obtain satisfactory agreement? Even if only two planets were needed, which of the many ways of developing the equations of motion would both be tractable and permit the desired empirical check?

Building upon his own and Lagrange's earlier work on planetary motions, Laplace succeeded in first proving that a remarkably simple conservation property held for the eccentricity of planetary orbits. This property implied in particular that, given the known ratio of the masses of Jupiter and Saturn, the ratio of the maximum retardation of the mean motion of Saturn to the maximum acceleration of the mean motion of Jupiter should be nearly in the ratio of 7 to 3 — if, in fact, only the mutual attractions of these two planets and the sun needed to be taken into account. Laplace found that two quantities in the correct ratio, namely, 9°16′ and 3°58′, differed from the largest values given in Halley's tables by only 9′. Encouraged by this close agreement, he embarked upon the arduous mathematical development of a theoretical formula for Saturn's motion.

It had long been known that the average annual mean motions of Jupiter (now known to be $n = 30^\circ.349043$) and of Saturn (now known to be $m = 12^\circ.221133$) were approximately in the ratio of 5 to 2; in fact $5m - 2n = 0.40758$. In *his* treatment of this problem, Euler had curtailed his expansion of the longitude of Saturn, omitting terms that contributed less than 30″ each on the principle that some of the observations he would be using could only be counted as accurate to 1′ (Euler, 1749, p. 118). Euler, as we have seen, seems not to have realized that unknown quantities could be determined to greater precision than that of the individual observations, and he was here oblivious to the possible cumulative effect of such terms. Laplace, on the other hand, followed a different tack. Although $5m - 2n$ is a very small quantity (only about 1/74 of Jupiter's mean annual motion), Laplace focused his attention on terms involving this difference because work of Lagrange had shown that such near commensurability of planetary motions could produce significant inequalities. Laplace noticed, first, that the periodic inequality due to the planets' motions corresponding to

terms involving $5m - 2n$ would have a period of about 900 years (about the right order of magnitude to explain the observed inequalities) and, second, that whereas the coefficients of these terms are very small in the differential equations of the planets' motions, they become potentially significant after the successive integrations needed to derive formulas for the planets' longitudes.

Encouraged further by a sense of empirical agreement between these theoretical observations and the known characteristics of the yet unexplained inequality, Laplace undertook to develop a formula for Saturn's longitude, in effect out of Euler's discarded scraps — out of a selection of the terms Euler had omitted as neither measurable nor likely to be important. The results of this intricate analysis were assembled into a 127-page memoir, "Théorie de Jupiter et de Saturne," printed separately by the Academy of Sciences in 1787 and reprinted the following year as part of the Academy's *Mémoires* for the year 1785.

The capstone of Laplace's investigation was his comparison of his theory with observations. He made use of the best available data on the "elements" of the planets' motions, but his theory required four quantities that were not readily available with sufficient accuracy for his purposes; he had to determine them from the observations themselves. These were, in Laplace's notation $\delta\epsilon^I$, δn^I, δe^I, and $\delta\tilde{\omega}^I$; they denoted necessary "corrections" (rates of change) for, respectively, the mean longitude of Saturn in 1750, its mean annual motion, its eccentricity, and the position of its aphelion (its most distant position from the sun). Laplace selected twenty-four observations of Saturn, made at times of opposition (when the sun, earth, and Saturn were aligned) over a 200-year period as being particularly likely to be accurately made. In each case he expressed the difference between the observed longitude of Saturn and that given by his theory as an "equation of condition." For example, the equation for the year 1672 was given as

$$\begin{aligned} 0 = -3'32.8'' + \delta\epsilon^I - 77.28\delta n^I - 2\delta e^I\, 0.98890 \\ - 2e^I(\delta\tilde{\omega}^I - \delta\epsilon^I)0.14858. \end{aligned}$$

(The eccentricity e^I was considered as known and was given elsewhere in the memoir). The entire data set is given in Table 1.3, in which $-a_i$ stands for the first term of the ith equation ($a_i = -3'32.8''$ in the equation for the year 1672), and b_i, c_i, d_i are the coefficients of the unknowns δn^I, $2\delta e^I$, and $2e^I(\delta\tilde{\omega}^I - \delta\epsilon^I)$. It should be noted that the coefficients c_i and d_i are strongly related, as were the coefficients in Mayer's investigation, by $c_i^2 + d_i^2 = 1$. In fact, $c_i = -\sin(\varphi_i - \omega_i)$ and $d_i = \cos(\varphi_i - \omega_i)$, where $\varphi_i - \omega_i$ is the difference between Saturn's observed longitude and its aphelion in the ith year; b_i is the number of years from the beginning of 1750 to the time the observation was made.

Table 1.3. Laplace's Saturn data.

Eq. no.	Year (i)	$-a_i$	b_i	c_i	d_i	Laplace residual	Halley residual	L.S. residual
1	1591	1′11.9″	−158.0	0.22041	−0.97541	+1′33″	−0′54″	+1′36″
2	1598	3′32.7″	−151.78	0.99974	−0.02278	−0.07	+0.37	+0.05
3	1660	5′12.0″	−89.67	0.79735	0.60352	−1.36	+2.58	−1.21
4	1664	3′56.7″	−85.54	0.04241	0.99910	−0.35	+3.20	−0.29
5	1667	3′31.7″	−82.45	−0.57924	0.81516	−0.21	+3.50	−0.33
6	1672	3′32.8″	−77.28	−0.98890	−0.14858	−0.58	+3.25	−1.06
7	1679	3′9.9″	−70.01	0.12591	−0.99204	−0.14	−1.57	−0.08
8	1687	4′49.2″	−62.79	0.99476	0.10222	−1.09	−4.54	−0.52
9	1690	3′26.8″	−59.66	0.72246	0.69141	+0.25	−7.59	+0.29
10	1694	2′4.9″	−55.52	−0.07303	0.99733	+1.29	−9.00	+1.23
11	1697	2′37.4″	−52.43	−0.66945	0.74285	+0.25	−9.35	+0.22
12	1701	2′41.2″	−48.29	−0.99902	−0.04435	+0.01	−8.00	−0.07
13	1731	3′31.4″	−18.27	−0.98712	−0.15998	−0.47	−4.50	−0.53
14	1738	4′9.5″	−11.01	0.13759	−0.99049	−1.02	−7.49	−0.56
15	1746	4′58.3″	−3.75	0.99348	0.11401	−1.07	−4.21	−0.50
16	1749	4′3.8″	−0.65	0.71410	0.70004	−0.12	−8.38	+0.03
17	1753	1′58.2″	3.48	−0.08518	0.99637	+1.54	−13.39	+1.41
18	1756	1′35.2″	6.58	−0.67859	0.73452	+1.37	−17.27	+1.35
19	1760	3′14.0″	10.72	−0.99838	−0.05691	−0.23	−22.17	−0.29
20	1767	1′40.2″	17.98	0.03403	−0.99942	+1.29	−13.12	+1.34
21	1775	3′46.0″	25.23	0.99994	0.01065	+0.19	+2.12	+0.26
22	1778	4′32.9″	28.33	0.78255	0.62559	−0.34	+1.21	−0.19
23	1782	4′4.4″	32.46	0.01794	0.99984	−0.23	−5.18	−0.15
24	1785	4′17.6″	35.56	−0.59930	0.80053	−0.56	−12.07	−0.57

Source: Laplace (1788).
Note: Residuals are fitted values minus observed values.

Laplace was faced with twenty-four inconsistent equations of condition, each linear in the four unknowns. It was a situation similar to one that Euler had failed to resolve and to one that Mayer had resolved by breaking the equations into disjoint groups and adding each group together. Laplace dealt with the problem by using a method that bears a superficial resemblance to Mayer's approach; yet it differed from that of Mayer in one subtle respect that marks it as an important advance toward least squares.

Laplace did not provide an algebraic description of his solution, but he did give a detailed description of the steps he followed. What he did, he explained, was to reduce his twenty-four linear equations to four equations: (i) the sum of equations 1 − 24; (ii) the difference between the sum of equations 1 − 12 and the sum of equations 13 − 24; (iii) the linear combination of equations: $-1 + 3 + 4 - 7 + 10 + 11 - 14 + 17 + 18 - 20 + 23 + 24$; and (iv) the linear combination of equations: $+2 - 5 - 6 + 8 + 9 - 12 - 13 + 15 + 16 - 19 + 21 + 22$. He then solved equations (i)–(iv) and checked the degree to which the resulting equations fit the observations by computing each residual, defined as the "excess" of the fitted value over the observed value (the negative of the modern definition of residual).

Laplace did not explain his selection of four linear combinations, but he seems to have based it upon their effect upon the coefficients of the unknowns in equations (i)–(iv). Thus (i) and (ii) are natural linear combinations to consider: (i) maximizes the coefficient of the constant term, whereas (ii) eliminates it. Nearly the reverse is true for the coefficient of δn^I. Evidently, the choice of which equations were included in (iii) and which in (iv) was made according to whether $|c_i| < |d_i|$ or $|c_i| > |d_i|$. The one exception to this rule is the reversal of it with respect to equations 3 and 5, a minor exception that may have been made to reduce the coefficient of the constant term in (iv) by 2. Once the twenty-four equations were divided between (iii) and (iv), the signs + and − were chosen according to the signs of d_i [for equation (iii)] and c_i [for equation (iv)], thus nearly maximizing the contrast between the coefficients of the last two unknowns.

The subtle advance Laplace had made was this: Where Mayer had only added his equations of condition together within *disjoint* groups, Laplace combined the *same* equations together in several different ways. The relationship between the methods of Mayer, Laplace, and Legendre and the importance of Laplace's advance can be best understood if we interpret them all in terms of a uniform notation. If we write an equation of condition involving four unknowns including a constant term as

$$0 = a_i + w + b_i x + c_i y + d_i z, \qquad i = 1, \ldots, n,$$

where a_i, b_i, c_i, and d_i are observable and w, x, y, and z are unknown, then we might write the jth aggregated equation, $j = 1, 2, 3, 4$, as

$$0 = \sum_i k_{ij}\, a_i + \sum_i k_{ij}\, w + \sum_i k_{ij}\, b_i x + \sum_i k_{ij}\, c_i y + \sum_i k_{ij}\, d_i z,$$

where $\{k_{ij}\}$ form a system of multipliers.

All three of the schemes we have discussed fit this description, although it should be borne in mind that viewing them in terms of this notation is an anachronism and that this unifying view was not to appear until later, in Laplace's 1812 work on this subject.

Mayer treated the case where $n = 27$ and $d_i = 0$ for all i (that is, only three unknowns appeared) and determined the multipliers $\{k_{ij}\}$ for ($j = 1, 2, 3$) from the coefficient of the first unknown with a nonconstant coefficient, that is, from the b_i's. If we let $\underline{b} < \bar{b}$ be the ninth largest and ninth smallest b_i, respectively, then Mayer effectively took

$$k_{i1} = \begin{cases} 1 & \text{if } b_i \leq \underline{b} \\ 0 & \text{otherwise} \end{cases}$$

$$k_{i2} = \begin{cases} 1 & \text{if } b_i \geq \bar{b} \\ 0 & \text{otherwise} \end{cases}$$

$$k_{i3} = \begin{cases} 1 & \text{if } \underline{b} < b_i < \bar{b} \\ 0 & \text{otherwise.} \end{cases}$$

Thus each equation of condition influences only one aggregated equation, and the question of which one is influenced was determined by the coefficient of a single unknown. Legendre's "method of least squares" went to another extreme, taking

$$k_{i1} = 1, \qquad k_{i2} = b_i, \qquad k_{i3} = c_i,$$

and, if the fourth unknown is present,

$$k_{i4} = d_i.$$

Here, unless a coefficient of an unknown is zero, all equations of condition influence all aggregated equations and the coefficients of all unknowns are important in the aggregation. Laplace's approach took a middle ground between these approaches and amounted to a "rounded off" version of the least squares aggregation, where the multipliers $\{k_{ij}\}$ are only allowed to take the values $-1, 0, +1$. If we let $B = \text{median}(b_i)$ and note that since in Laplace's case $c_i^2 + d_i^2 = 1$ (so $|c_i| > |d_i|$ if and only if $|c_i| > 2^{-1/2} \cong 0.707$), then (with the single minor exception noted earlier) Laplace's aggregation was

$$k_{i1} = \quad 1 \qquad \text{all } i$$

$$k_{i2} = \begin{cases} -1 & \text{if } b_i < B \\ 1 & \text{if } b_i > B \end{cases}$$

$$k_{i3} = \begin{cases} -1 & \text{if } -1 \leq c_i < -0.707 \\ 0 & \text{if } -0.707 < c_i < 0.707 \\ 1 & \text{if } 0.707 < c_i \leq 1 \end{cases}$$

$$k_{i4} = \begin{cases} -1 & \text{if } -1 \leq d_i < -0.707 \\ 0 & \text{if } -0.707 < d_i < 0.707 \\ 1 & \text{if } 0.707 < d_i \leq 1. \end{cases}$$

Now, the point of this comparison is not to argue that Laplace almost arrived at the method of least squares. In the first place, even describing his method in this algebraic form is a considerable formal extrapolation of what he actually did, which was to present a single worked example. In the second place, Legendre's method, unlike Laplace's, was formally derived from an explicitly stated criterion of best fit. Rather the point is that Laplace had moved forward from Mayer in treating the data set as a whole and in allowing the values of all the coefficients of the unknowns to influence the aggregation. Mayer had broken his data set into disjoint sets which, judged by the values of the coefficient of one unknown, were made under similar circumstances. In this he went beyond Euler, who would have insisted that the equations be alike in all coefficients before he would combine them. Mayer nevertheless clung to the older tradition in that he treated observations made under very different conditions separately, at least until the final stage of his analysis. Laplace went further by combining all the observational equations in the very first stage of his analysis, and more important, by letting all coefficients influence the manner of combination.

Mayer had focused on the coefficient of only one unknown. Because it was indeed the most important unknown in his application,[7] he arrived at what was, for his situation, a good solution. But if we consider his approach as a method to be applied in other situations, it has the serious shortcoming of requiring that there be a single important unknown, that the investigator be able to tell which unknown is important, and that the coefficients of the other unknowns be distributed so as not to confound the solution by, for example, producing a nearly singular set of aggregated equations. Mayer was successful because he knew what he was doing in his particular application, but another investigator mimicking his procedure in another application might not be so lucky. Laplace's generalization of this approach, on the other hand, did not suffer from this drawback. All unknowns could be determined by his method with reasonable accuracy—at least when there was sufficient information in the original equations of condition to determine them accurately by any method. Another investigator imitating Laplace's method would not require the same degree of "good luck" required by a follower of Mayer and usually would be rewarded by greater accuracy as well.

The criterion minimized by the method of least squares—the sum of squared residuals—is not an unexceptionable measure of the success of a fit, but it provides one useful way of assessing these early efforts. When we compare the square root of the sum of the squares of Mayer's residuals to the square root of the residual sum of squares given by least squares, we

7. In fact, because of the small size of θ (and thus of $\alpha \sin \theta$) and the relatively large measurement error in determining $g - k$ (caused by the small angle between the moon's orbit and the ecliptic and the consequent difficulty of estimating the node F in Figure 1.2), the term involving $\alpha \sin \theta$ does not contribute significantly to the regression. Interestingly Mayer was aware of this measurement problem and did not consider his determination of θ to be reliable.

find that for the Manilius data Mayer's method gives a value only 6 percent larger, for the Dionysius data, 5.7 percent larger, and for the Censorinus data, 66 percent larger. For Laplace's Saturn data, the same comparison has Laplace's method with a value only 5.5 percent larger than that for least squares; whereas when Halley's 1719 empirical adjustment is extrapolated to 1786, the square root of the sum of the twenty-four corresponding squared residuals is 90 percent larger than Laplace's. Note (Table 1.3) that the pattern of either Laplace's or the least squares residuals hints correctly that not all of the periodic inequalities in Saturn's motion had been accounted for, even though the most controversial one had been.

Laplace had come quite a bit closer to providing a general method than Mayer had. Indeed, it was Laplace's generalization that enjoyed popularity throughout the first half of the nineteenth century, as a method that provided some of the accuracy expected from least squares, with much less labor. Because the multipliers were all -1, 0, or 1, no multiplication, only addition, was required.

Among the writers to present Laplace's version of the method as an alternative to least squares was Mary Somerville, who described it in her 1831 book *Mechanism of the Heavens* and attributed it to Mayer. In telling how a set of equations of condition of the form

$$\text{Error} = \epsilon + 0''3133P + 0''2969e,$$

but with varying coefficients, could be combined to solve for three unknowns P, e, and ϵ, Somerville wrote:

> For example, in finding the value of P before the other two, the numerous equations must be so combined, as to render the coefficient of P as great as possible; and the coefficients of e and ϵ as small as may be; this may always be accomplished by changing the signs of all the equations, so as to have the terms containing P positive, and then adding them; for some of the other terms will be positive, and some negative, as they may chance to be; therefore the sum of their coefficients will be less than that of P.
>
> Having determined this equation, in which P has the greatest coefficient possible, two others must be formed on the same principle, in which the coefficients of the other two errors must be respectively as great as possible, and from these three equations values of the three errors will be easily obtained, and their accuracy will be in proportion to the number of observations employed. (Somerville, 1831, p. 409)

In this description, and in our abstraction of Laplace's extension of Mayer, there is an implicit assumption that the signs of the coefficients change, as would be the case if they were centered at their means. In fact, in these early applications this was essentially the case because the unknowns represented corrections to a mean. We note in addition that Mary Somerville follows Mayer in claiming that accuracy is proportional to the number of observations rather than to the square root of the number of observations.

Other writers to mention Laplace's version of the method, crediting it to Mayer, included Francoeur (1830; 1840, p. 432), Bowditch (1832, p. 485), Puissant (1842, vol. 2, p. 344; Puissant does not mention the method in the first edition, 1805), Wolf (1869–1872, vol. 1, p. 279; vol. 2, p. 199), and Whittaker and Robinson (1924, pp. 258–259). After about 1850 the references to the method were of a historical rather than a practical nature, and they describe the method in terms of Mayer's original formulation, with disjoint groups being added.

Roger Boscovich and the Figure of the Earth

Thus far we have seen how problems involving Jupiter, Saturn, and the moon led to the introduction and development of a method of combining inconsistent linear equations in the half-century before the appearance of the method of least squares. "Mayer's method," as it was known, was easy to use and generally led to sensible results, two properties that conspired to keep it as an actively employed tool in astronomers' and geodesists' workshops for a half-century after least squares was introduced. But Mayer's method lacked one quality that the method of least squares was found to have in abundance; and it was this quality that contributed to the eventual eclipse of Mayer's method by the method of least squares. Mayer's method, unlike least squares, was not "best" in the sense that it appeared as the solution to a mathematically posed problem of finding the "best" combination of inconsistent equations. It was an ad hoc method, and its acceptance depended upon its reputation for past successful use, its ease of application, and the investigator's intuitive feeling that by combining the equations in such a way that the coefficients of the unknowns are successively maximized, a mechanically stable (and hence reliable) solution would result. The first of these—a reputation for successful use—was widely believed. Mayer's tables of the moon's motion and his map of the face of the moon were commonly and justly seen as among the most accurate such achievements of eighteenth-century observational astronomy. Even those who knew that the development of Mayer's method into a more general tool was first found in Laplace's work would not be dissuaded from its use—who could not feel confident using the method that had reconciled the motions of Jupiter and Saturn with Newtonian gravitational theory? As the years went by, the vivid impression of these past triumphs faded, however, to be replaced by stories of the triumphs of least squares. Then, as more experience and ways of simplifying computation made least squares easier to use and as succeeding generations of mathematicians made no successful attempt to give formal statement to the vague intuitive notions of the reasonableness of Mayer's method, it faded from view—or rather moved from the workshop of the practitioner to the display case of the statistical museum.

Least squares was the most successful of the early methods of combining inconsistent equations, and the fact that it was based on and derived from an easily understood objective criterion was a major reason for its success. Nevertheless, least squares missed being the first method to be so based by nearly half a century. To trace the genesis and fate of its most famous predecessor, "Boscovich's method," we shall first consider a third major eighteenth-century scientific problem—the problem of the figure of the earth.

The first post-Columbian hint that the earth was not a perfect sphere seems to have been the discovery by Richer in 1672 that a pendulum near the equator was less affected by gravitational attraction than was the same pendulum at Paris. Newton, in the *Principia* (1687), showed how the rotation of the earth could be expected to produce a flattening of the earth at the poles and a bulging at the equator, a shape known as an oblate spheroid. Newton estimated the oblateness or ellipticity (the fraction by which a radius at the equator exceeds the radius at the pole) to be 1/230. This conclusion about the shape or figure of the earth did not go unchallenged, however. Domenico Cassini, director of the Royal Observatory in Paris, thought in fact that the earth was a prolate spheroid, flattened at the equator, not the poles. Several attempts were made over the succeeding century both to determine whether the earth was oblate or prolate and to measure the departure of its shape from spherical.

The two principal methods of determining the figure of the earth were pendulum experiments and arc measurements. We shall be concerned primarily with arc measurements in this chapter, although the statistical problems that arise in the two cases have striking similarities. The determination of the earth's figure from arc measurements required the cooperation of a team of scientists and months of labor under adverse circumstances; it was the perfect sort of challenge for the growing and increasingly adventuresome French scientific community. The idea was to measure the linear length of a degree of latitude at two (or more) widely separated latitudes. If a degree near the equator is found to be shorter than one nearer the pole, then the shape of the earth is oblate; and the difference between the two measurements can be used to calculate the oblateness.

At first thought it may seem paradoxical that a shorter degree near the equator would indicate a *bulging* at the equator, but only because there is a common misapprehension of the definition of a degree of latitude. The latitude of a point on the earth's surface is not, as might be supposed, the angle formed between two rays from the center of the earth, one to the given point and the other at the intersection of the equatorial plane and the point's meridian plane. Indeed, the operational difficulties in taking such a measurement would tax the resources of even the largest geodetical

organization. Latitude in fact is measured as the angle between a ray to the zenith of the given point and the equatorial plane or, alternatively, as the complement of the angle between rays from the given point to the zenith and to the Pole Star. Figure 1.3 shows arcs of 10° latitude for an exaggeratedly oblate earth.

The relationship between arc length and latitude can be derived from the geometry of conic sections, the exact relation being given by an elliptic

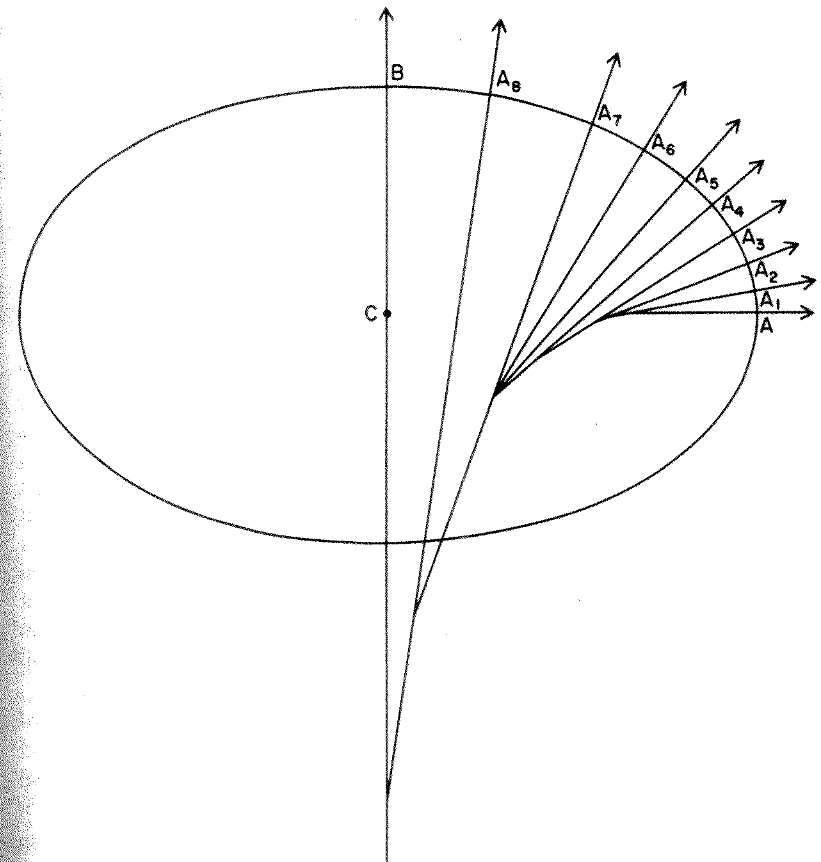

Figure 1.3. A side view of an exaggeratedly oblate earth, illustrating the lengthening of degrees of arc toward the pole. The meridian quadrant AB is broken into nine segments, each of 10° latitude. (Based upon Berry, 1898, p. 277.)

integral. For short arcs, however (the only ones it was practical to measure), a simple approximation will do: If a is the length of 1° of latitude centered at latitude θ, measured along a meridian, then to a good approximation,

$$a = z + y \sin^2 \theta,$$

where z is the length of a degree at the equator, and y is the excess (or deficiency) of a degree at the North Pole over one at the equator. Some early works expressed this as

$$a = z + y \cdot \frac{1}{2} \cdot \text{versed sine } 2\theta,$$

where "versed sine" $2\theta = 1 - \cos 2\theta$. Then the identity $1 - \cos 2\theta = 2 \sin^2 \theta$ would save the bother of squaring $\sin \theta$ in any calculation.

Measurements of a French arc made by Domenico Cassini and his son and successor Jacques before 1720 supported the hypothesis that the earth was prolate, but the narrow range of latitude (9°) and the possibility of a low accuracy in the measurements prevented this anti-Newtonian conclusion from gaining wide acceptance. In 1735 the French Academy launched expeditions by Bouguer to Peru and by Maupertuis to Lapland to measure arcs near the equator and at about 66° latitude for comparison with measurements near Paris. The results effectively refuted the Cassini hypothesis and settled the matter in Newton's favor: The earth was oblate. The only question remaining was the size of the oblateness or ellipticity, because different pairs of arcs gave different values.

In 1755 the results of measuring a length of a meridian arc near Rome were published by the English Jesuit Christopher Maire and the Dalmatian Jesuit Roger Joseph Boscovich (or Rudjer J. Bŏsković) under the title *De Litteraria Expeditione per Pontificiam ditionem ad dimetiendas duas Meridiani gradus*. In successive analyses by Boscovich of these data we find the first successful resolution of the inconsistency of the different arc measurements and the introduction of the statistical procedure that is our immediate concern.

When Boscovich first addressed this problem in 1755—in a chapter of his joint work with Maire for which he took sole responsibility—he met with only limited success. Boscovich was aware, as others before him had been, that to obtain an accurate determination of the figure of the earth it would be necessary to compare measurements widely separated in latitude, as even small errors made in proximate arc measurements would be greatly exaggerated in any pairwise combination of them. Boscovich thus focused his attention on only five determinations that were made at well-separated locations and were likely to be accurate (Table 1.4). Boscovich gave no analytic description of his handling of these data; here as else-

Table 1.4. Boscovich's data on meridian arcs.

Location	Latitude (θ)	Arc length (toises)	Boscovich's $\sin^2 \theta \times 10^4$
(1) Quito	0°0′	56,751	0
(2) Cape of Good Hope	33°18′	57,037	2,987
(3) Rome	42°59′	56,979	4,648
(4) Paris	49°23′	57,074	5,762
(5) Lapland	66°19′	57,422	8,386

Source: Boscovich and Maire (1755, p. 500). Reprinted in Boscovich and Maire (1770, p. 482).

Note: Arc lengths are given as toises per degree measured, where 1 toise ≅ 6.39 feet. The value for $\sin^2 \theta \times 10^4$ for the Cape of Good Hope is erroneous and is evidently based on 33°8′. The correct figure would be 3,014.

where he followed in a Newtonian tradition of giving geometric descriptions rather than analytic ones.[8] It will be easier, however, to relate Boscovich's different efforts to later work if we adopt an analytic formulation from the beginning. In analytic terms, Boscovich was faced with the equivalent of five observational equations,

$$a_i = z + y \sin^2 \theta_i,$$

where a_i and θ_i are the length of an arc (in toise per degree, 1 toise ≅ 6.39 feet) and the latitude of the midpoint of the arc, both at location i. The unknowns y and z are, respectively, the excess of a 1° arc at the pole over one at the equator and the length of a degree at the equator.

In principle any two of the five locations could be used to solve for the polar excess y and the equatorial degree z or, equivalently, for the polar

8. Isaac Todhunter is among those who have experienced frustration because Boscovich retained the cumbersome geometric apparatus of Newton instead of using the more elegantly concise analytic formulations of Clairaut and Euler. Commenting on one of Boscovich's proofs, Todhunter wrote, "Boscovich professes to use Geometry alone: but the Geometry consists chiefly in denoting the length of every straight line by two capital letters instead of a single small letter: this strange notion of Geometry has survived to our own times in the University of Cambridge" (Todhunter 1873, vol. 1, p. 309). Todhunter later took another opportunity to link sarcastically Boscovich and his own University of Cambridge, "In forming an estimate of the treatise we must remember that the author had prescribed to himself the condition of supplying *geometrical* investigations; so the Differential Calculus was not to be introduced. We must consider the treatise rather as the work of a professor for the purposes of instruction than of an investigator for the advancement of science; and then we may award the praise that the task proposed is fairly accomplished. It would have been more desirable to study Clairaut's work than to be confined to Boscovich's geometrical methods: but the experience of our own university shews us that it is possible to find the methods used for teaching occasionally some years in arrear of those used for investigation" (Todhunter, 1873, vol. 1, p. 319).

excess y and the ellipticity (Boscovich computed[9] it here as $1/\text{ellipticity} = 3z/y$). And in fact this is exactly what Boscovich did: He calculated y and the ellipticity based upon each of the $\binom{5}{2} = 10$ pairs and presented the results shown in Table 1.5. This gave him not one, but ten solutions to his problem, and in 1755 he showed himself not quite able to deal with this embarrassment of riches. He did make a weak attempt to combine these findings: He averaged the ten values of the excess[10] and found, using the equatorial degree at Quito ($z = 56{,}751$), the value 1/155 for the ellipticity.

This value must have seemed too large, for he then recomputed the ellipticity after rejecting the pairs (2, 4) and (2, 3) as "so different from the others," possibly because of the close proximity of their degrees of latitude. The mean excess based on the remaining pairs gave, with the Quito degree, an ellipticity of 1/198, but this still seemed unsatisfactory to Boscovich. Instead of accepting either of these figures as a compromise, as an average determination of the ellipticity, Boscovich focused on the discrepancy between this average value and the ten (or eight) components that had made up the average, taking what appeared to him to be large discrepancies as evidence against an ellipsoidal shape for the earth.

> Thus it is evident that the determinations of these degrees cannot be reconciled with the ellipse of Newton, nor with any other ellipse, either more or less oblate. Five degrees, taken arbitrarily, must always give the same ellipse, and we have seen what little agreement there is between those we have chosen. The differences between them are not proportional to the versed sine of double the latitude [that is, versed sine $2\theta = 2\sin^2\theta$]. If they were, each combination of degrees, as we have said, should give the same ellipticity. (Boscovich and Maire, 1755, p. 501; 1770, p. 484)

A modern geodesist would not quarrel with Boscovich's rejection of an ellipsoidal hypothesis, but in the context of his own time he was wrong. If the likely size of measurement errors, even as perceived by Boscovich and his contemporaries, is taken into account and the observational evidence combined in a reasonable way, then Boscovich's data is not wholly incon-

9. A slightly better local approximation would be one Laplace used later, namely $1/\text{ellipticity} = 3z/y + 5/3$, but the difference between the two formulas is negligible in the present application. Other workers used $1/\text{ellipticity} = 3z/y + 3/2$, or $= 3z/y + 2$. All workers of the period wrote the ellipticity in reciprocal form (for example, 1/230), even when it was first calculated in decimal form. We would describe this practice as reparametrization.

10. Actually the text (Boscovich and Maire, 1755, p. 501; 1770, p. 484) gives the average excess as 222 (just one-third of the correct value) even though the ellipticities were correctly calculated. Evidently Boscovich inadvertently inserted the average of the ten values of $y/3$ into the text, a number he would perhaps have found as an intermediate step toward calculating $1/\text{ellipticity} = z/(y/3)$.

Table 1.5. Boscovich's pairwise solutions, based on the data of Table 1.4, for the polar excess y (the amount by which a degree at the pole exceeds a degree at the equator) and the ellipticity (found from the formula $1/\text{ellipticity} = 3z/y$, where z is the length of a degree at the equator as found from the pair of equations).

Pair	Polar excess (y, in toises)	Ellipticity	Pair	Polar excess (y, in toises)	Ellipticity
1, 5	800	1/213	2, 4	133	1/128
2, 5	713	1/239	3, 4	853	1/200
3, 5	1,185	1/144	1, 3	491	1/347
4, 5	1,327	1/128	2, 3	−350	−1/486
1, 4	542	1/314	1, 2	957	1/78

Source: Boscovich and Maire (1755, p. 501). Reprinted in Boscovich and Maire (1770, p. 483).

Note: The ellipticities for pairs (2, 4) and (1, 2) were evidently misprinted in the original; they should be 1/1282 and 1/178. The figures for the pair (1, 4) are erroneous; they should be 560 and 1/304.

sistent with an ellipsoidal hypothesis. The arc at Paris was widely seen as the most accurate of those measured before 1755, and even it was susceptible to large changes whenever it was carefully rescrutinized: From 1738 to 1740, estimates of the Paris degree changed from Picard's original (1671) figure of 57,060 toises to 56,926 toises (Maupertuis in 1738) to 57,183 toises (Maupertuis in 1740) to 57,074 toises (Cassini de Thury in 1740), see Todhunter (1873, vol. 1, p. 127). These successive changes of over 100 toises in the most-studied arc would not justify confidence in much greater accuracy than 100 toises, although individual investigators evidently had higher opinions of their own arcs. Even forty years later, Laplace (1799, vol. 2, p. 448) felt that an error of 97.20 toises "is exactly on the least limit of those which might be considered as possible." If Boscovich's data (Table 1.4) is fit by least squares, the residuals are 13, 83, −95, −80, 78. A slightly better fit is achieved if Boscovich's error in the $\sin^2\theta$ for the Cape is corrected, that is, 15, 82, −94, −80, 78. The accuracy of the Lapland arc is discussed by Todhunter (1879).

Boscovich himself must have felt some uneasiness at his own conclusion, for he did not let the matter rest with his 1755 analysis. Two years later he published a synopsis of the 1755 volume that included a brief statement of a radically new principle for the combination of inconsistent arc measurements. And in 1760, in a prose supplement to a versified treatise on natural philosophy by Benedict Stay, Boscovich gave a full description of his principle, an explanation of how it could be used in practice and a worked example based on the five degrees he had considered in 1755. This 1760 version was later translated into French and appended to Boscovich and Maire (1770) as a part of a note[11] (pp. 501–510).

The principal novelty in Boscovich's approach was its novelty of principle. Where Mayer had proceeded ad hoc, his underlying motivation remaining unformulated, Boscovich began with a generalizable principle, a list of properties that the mean based on a combination of arc measurements should have; and he went on to derive an ingenious geometric algorithm that would find such a mean. He introduced his principle as follows (the italics are Boscovich's):

> The mean we will take will not be a simple arithmetic mean, rather it will be one tied by a certain law to the rules of fortuitous combination and the calculus of probabilities. We are faced here with a problem I have discussed toward the end of a Dissertation inserted in the proceedings of the Institute of *Bologna, volume 4* [that is, Boscovich's 1757 summary], where I contented myself with giving the result of its solution. Here is the problem: *Being given a certain number of degrees, find the correction that must be made to each of them, supposing these three conditions are complied with: the first, that their differences shall be proportional to the differences between the versed sines of twice their latitudes; the second, that the sum of the positive corrections shall be equal to the sum of the negative ones; the third, that the sum of all the corrections, positive as well as negative, shall be the least possible, for the case where the first two conditions will be fulfilled.* The first condition is called for by the law of equilibrium, which requires an elliptical shape; the second, from the fact that deviations of a pendulum, or errors by observers, that augment or diminish degrees have the same degree of probability; the third is necessary in order to approximate the observations as closely as possible, for, as we have observed above, it is clearly very probable that the deviations are quite small, because the scrupulous exactitude of the observers would not permit any suspicion of large errors in their observations. (Boscovich, 1760, as translated from the French of Boscovich and Maire, 1770, p. 501)

If we introduce the symbols $a_1, a_2, \ldots$ for the measured lengths, in toises per degree, of the measured arcs at latitudes $\theta_1, \theta_2, \ldots$, and let δa_i stand for the "correction" that Boscovich would make to the degree a_i, then his first condition is that the corrected degrees, $a_i + \delta a_i$, satisfy

$$a_i + \delta a_i - (a_j + \delta a_j) \propto \text{versed sine } 2\theta_i - \text{versed sine } 2\theta_j.$$

Since versed sine $2\theta = 2\sin^2\theta$, this is equivalent to supposing

$$a_i + \delta a_i = z + y\sin^2\theta_i,$$

11. The 1770 translation also included an additional worked example based on nine arc measurements. Boscovich's original manuscript copy of this additional material is preserved in the Boscovich Archives at the Bancroft Library of the University of California, Berkeley (folder 28).

or

$$\delta a_i = z + y\sin^2\theta_i - a_i,$$

for some choice of z and y. Boscovich's second and third conditions were intended to lead to a best choice of z and y. The second was justified by an appeal to the intuitively plausible notion that positive and negative errors are equally likely, and it stated that the sum of positive corrections should equal the sum of negative corrections. In our notation this is compactly expressed by

$$\sum_{\text{all } i} \delta a_i = 0.$$

Boscovich's third condition was that the sum of the corrections, taken without regard to sign, was to be a minimum; that is,

$$\sum_{\text{all } i} |\delta a_i| \text{ is minimized.}$$

Now, it should be emphasized that this analytic formulation is not found in Boscovich's work. His statement of the conditions was only given in the verbal version we have quoted, and at no place in his work did he give other than a geometric or mechanical description of his solution to the problem. But even though this restriction to a geometric approach was costly in that it limited the generalizability of the approach, there was an important compensating benefit: Boscovich's geometric approach suggested a solution to the problem that would have been far less apparent in an analytic formulation.

Boscovich began by noting that his first condition *could* be expressed analytically—as meaning that the corrected arcs are expressible as an equation of the first degree involving the versed sines—and that if this were done, the second condition would be expressible as a sum involving the unknown coefficients. His statements again were verbal, not symbolic, but he clearly had the equation

$$\sum (z + y\sin^2\theta_i - a_i) = 0$$

in mind. He mistakenly felt, however, that an analytic solution could only be had by differentiating this expression with respect to the unspecified coefficients; and he saw that this would be absurd, writing: "Thus supposing $dz = 0$ we will have nothing: the formula will vanish completely, together with the hopes of the mathematician" (Boscovich, 1760; Boscovich and Maire, 1770, p. 502; Boscovich used the letter x where I use z). Therefore, he moved on to present a simple geometric and mechanical solution.

Boscovich's discussion was accompanied by a diagram in the 1770 translation, reproduced here as Figure 1.4. We may consider AF as being one

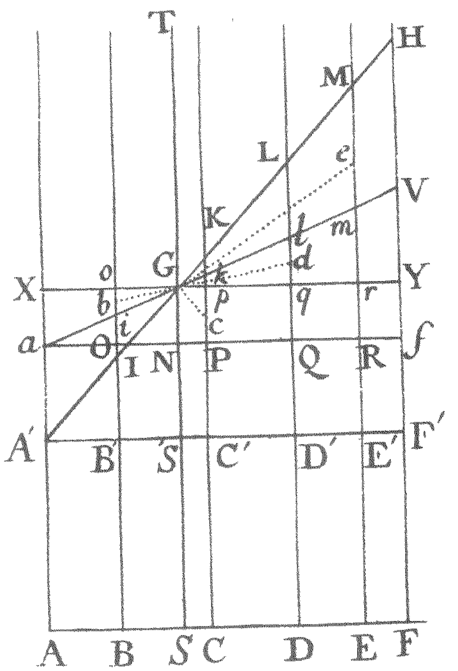

Figure 1.4. Boscovich's algorithm. The horizontal axis AF gives $\sin^2\theta$*, where* θ *is the latitude of the midpoint of the arc; the vertical axis AX gives the arc length in toises per degree. The five arcs are indicated by* a, b, c, d, *and* e*; G is the center of gravity. (From Boscovich and Maire, 1770, plate I.)*

unit in length; A as the origin; and A, B, C, D, and E as representing the five values of the $\sin^2\theta_i$ as marked off on the unit interval. The lengths Aa, Bb, Cc, Dd, and Ee may be considered as representing the lengths (in toises per degree) of the corresponding measured arcs. The problem was to find a straight line A′H such that the corrections aA′, bO, cK, dL, and eM satisfied the second and third conditions. Boscovich began by observing that the second condition, interpreted as describing a condition of mechanical equilibrium, implied that the line must pass through the center of gravity (G) of the points. This reduced the problem to one of finding that line through G such that the third condition was satisfied.

To find the solution, Boscovich imagined a straight line SGT that was rotated clockwise, kept anchored at G. As the line rotated, the sum of the corrections (taken without regard to sign) would decrease until a minimum was achieved, then increase again. Since the corrections for the individual

arcs would change as the line rotated in proportion to the distances AS, BS, SC, SD, and SE, it was only necessary to continue until the line had passed (or just reached) a sufficient number of the five points so that at least half the sum of these five distances, AS + BS + SC + SD + SE, was accounted for by the distances corresponding to the points passed (or just reached). Before such a point was reached, the sum of the corrections would decrease as the line rotated; after that point the sum would increase. He further simplified the solution by noting that the rotating line would encounter the five points in the inverse order to the slopes of the five lines from G to the five points: aX/AS, bo/BS, and so forth.

Boscovich clarified his geometric description of this algorithm by working through the details for the data of Table 1.4. He first located G by calculating the mean of the values for $10^4 \cdot \sin^2\theta_i$ (giving 10^4 AS = 4,356.6) and the mean of the arcs a_i (giving SG = 57,052.6). He then calculated the differences $10^4(\sin^2\theta_i - \text{AS})$ (that is, AS, BS, . . . ; given in column 2 of Table 1.6), the differences $a_i - \text{SG}$ (that is, aX, bo, . . . ; given in column 3 of Table 1.6), and the ratios $10^4(\sin^2\theta_i - \text{AS})/(a_i - \text{SG})$ (that is, AS/aX, BS/bo, . . . ; given in column 4 of Table 1.6). These ratios are the inverses of the slopes described earlier. Thus the numbers in column 4 of Table 1.6 told Boscovich that the rotating line would encounter the five points in the order e, a, d, b, and, finally, c. How far through this list should Boscovich travel? The total of the distances AS, BS, SC, SD, SE ($\times 10^4$, taken without regard to sign) was found from the numbers in column 2 to be twice 5,726.2; thus he should continue until he reached a point such that the sum of the values in column 2 (without regard for sign) corresponding to the passed points reached 5,726.2. Since SE = 4,029.4 < 5,726.2 and SE + SA = 8,386.0 > 5,726.2, the line should be allowed to rotate until it reached a. While the line rotated from e to a, the correction to the point e increased by an amount proportional to SE = 4,029.4 and the total decrease in the other corrections was proportional to SA + SD + SB + SC = 7,423.0. Once a was passed, the total increase was

Table 1.6. Intermediate calculations for Boscovich's 1760 solution.

Arc	$10^4(\sin^2\theta - \text{AS})$	Arc length − SG (toises)	Ratio
(1) = a	−4,356.6	−301.6	14
(2) = b	−1,369.6	−15.6	88
(3) = c	291.4	−73.6	−4
(4) = d	1,405.4	21.4	66
(5) = e	4,029.4	369.4	11

Source: Boscovich and Maire (1770, pp. 505–506).

proportional to SE + SA = 8,386.0, and thus more than counterbalanced the decreasing corrections, proportional to SD + SB + SC = 3,066.4. The solution then was the line passing through G and *a*! This he found corresponded to $z = 56{,}751$ and $y = 692$, from which he calculated the ellipticity as 1/248. Here, in 1760, he used the formula $1/\text{ellipticity} = 3z/y + 2$. Perhaps encouraged by the close proximity of this value to Newton's 1/230 and by the fact that this new value emerged from his own new method of combining observations, Boscovich did not repeat his earlier doubts about the elliptical shape of the earth.

The verbal description quoted earlier and the worked example are really all Boscovich ever wrote about his method. He added an updated discussion of arc measurements to the 1770 translation of his work with Maire. This discussion used his method on nine measured arcs and, after successive reanalyses omitting the three most discordant of the measurements, came to the conclusion that the evidence supported a hypothesis that the earth was somewhat irregular but was formed around an ellipsoidal core. Boscovich gave no further development of the method, no study of its properties, no analytic formulation, and no application of the method to problems other than the figure of the earth. He did briefly indicate the possibility of other applications in 1760, in these words: "Now we see that the method is generally for the correction of any terms which must be in a given ratio, because in substituting this ratio for that of the versed sines, all remains the same" (Boscovich, 1760; Boscovich and Maire, 1770, p. 505). But he appears never to have followed up on this statement by applying the method elsewhere, and he made no further statements regarding its generalizability. In fact, it is tempting to suppose that the method might have faded into obscurity had not a brief reference to its existence, in a 1772 review of the 1770 translation, caught the eye of Laplace.[12]

Laplace and the Method of Situation

In 1789 Laplace took up the question of the figure of the earth for the second time. He had previously approached the question of comparing arc measurements with an ellipsoidal hypothesis in 1783, but on that earlier occasion he had ignored Boscovich's work. He had considered only four of the five arcs studied by Boscovich (omitting the one at Rome measured by Boscovich and Maire), and he had been content to determine the best-fitting ellipsoidal figure as that which minimized the maximum correction needed. For that purpose he introduced an algorithm of his own invention

12. The review was in Jean Bernoulli III's *Recueil pour les astronomes* (Tome II, 1772, pp. 245–249). The relevant portions are quoted, together with evidence that Laplace read the review soon after it appeared, in Stigler (1978b).

(Laplace, 1786a). In his 1789 return to this subject, however, he had Boscovich very much in mind.

After admitting that his earlier approach of minimizing the maximum correction would become too laborious *(très pénible)* to use when many degrees were to be considered, Laplace turned to other approaches to the problem. He first derived an improved version of the earlier algorithm, and he illustrated its use with a set of nine arc measurements (which now included that at Rome); he then turned to Boscovich's method. Laplace felt that the elliptical figure that minimized the maximum correction was

> not the one that the measurements indicated with the greatest likelihood. This latter ellipse must, it appears to me, fulfil the following two conditions: 1° that the sum of the errors be zero; 2° that the sum of the errors taken all with the sign + be a *minimum.* Boscovich has given an ingenious method of achieving this; it is explained at the end of the French edition of his *Voyage astronomique et géographique* [that is, Boscovich and Maire, 1770]. But since he has unnecessarily complicated the method by the consideration of diagrams, I shall present it here in its simplest analytical form. (Laplace, 1793, p. 32)

Laplace proceeded to do just that, giving a precise algebraic statement to Boscovich's algorithm and accompanying his description with a rigorous analytic demonstration that the algorithm solved the problem as stated. He also gave two numerical examples, one involving nine measured arcs, the other involving thirteen observed lengths of seconds pendulums at various latitudes.

In this 1789 treatment Laplace added only an analytic formulation to Boscovich's earlier presentation. Even though this step alone was to prove crucial to Laplace's (and others') later studies of the method's statistical properties (he was to give it the name "Method of Situation"), it was nonetheless only a small conceptual advance—a translation from the language of Newton to the language of Euler. Ten years later, however, Laplace took an additional step, a subtle development of Boscovich's idea that was a symptom perhaps of Laplace's increasingly sharp statistical intuition.

The occasion for Laplace's renewed interest in the figure of the earth was the preparation of the second volume of the *Mécanique céleste.* After developing a mathematical theory of the figures of the heavenly bodies, Laplace returned to the problem he had faced a decade before—that of comparing measured arcs of the meridian with the hypothesis that the earth's figure was ellipsoidal. His analysis now considered only seven arcs; he omitted an ancient measurement in Holland and substituted a French arc measured in 1795 by Delambre and Méchain for two earlier French arcs. Now, all earlier analyses of arc measurements had treated all measurements as equally reliable; that is, all measurements considered good enough to be used at all were allowed an equal opportunity to influence the result of the calculation. Of course, the mechanics of the problem dictated that measurements made at extremes in latitude exerted a greater weight on the combined result than did those made at middle latitudes, but as far as the intrinsic accuracy of the measurement was concerned it was an all-or-nothing proposition. In 1799 Laplace evidently thought this was inappropriate.

In the preceding discussion, I introduced Laplace's notation (but with subscripts for superscripts) for the equations of condition describing the relationship to be tested:

$$a_k - z - p_k y = x_k,$$

where a_k is the arc length of the kth arc (in toises per degree), z the length of a degree at the equator, y the polar excess, p_k the square of the sine of the kth latitude, and x_k the error, due to measurement or failure of the ellipsoidal hypothesis. Now a_k was in fact determined by measuring the length of the arc both on the ground[13] and by astronomical observation of the Pole Star from the arc's two extreme points; the ratio of these two measurements would then give the length a_k in toise per degree. The actual lengths of the measured arcs varied considerably, from just under 1° in Lapland to nearly 10° in France. Surely these lengths would affect accuracy, and they should be incorporated into the analysis. In the *Mécanique céleste* (vol. 2, bk. 3, §40) Laplace therefore modified Boscovich's earlier conditions to the following form:

> First, that the sum of the errors committed in the measures of the whole arcs, ought to be zero. Second, that the sum of all these errors, taken positively, ought to be a *minimum*. By considering, in this manner, the whole arcs, instead of the degrees which have been deduced from them, we shall give to each of these degrees so much more influence, in the computation of the ellipticity of the earth, as the corresponding arc is of greater extent, which ought to be the case. (Laplace, 1799–1805, vol. 2, p. 134; Bowditch, 1832, pp. 434–437)

Following Laplace, let i_k represent the length, in degrees, of the kth arc (so $i_k a_k$ is the arc's length in toises). Then the whole arcs satisfy $i_k a_k - i_k z - i_k p_k y = i_k x_k$, and Laplace's two new conditions become

$$\text{(i)}\quad \sum_k i_k x_k = 0$$

$$\text{(ii)}\quad \sum_k |i_k x_k| = \text{minimum.}$$

13. A baseline (perhaps one-fifth to one-tenth the total distance surveyed) was measured directly by chains. Then the entire length was surveyed by triangulation.

Laplace derived the solution to the problem by a simple modification of the earlier algorithm. The condition (i) was equivalent to

$$A - z - y\,P = 0$$

where

$$A = \frac{\sum i_k a_k}{\sum i_k}, \qquad P = \frac{\sum i_k p_k}{\sum i_k}.$$

Subtracting this equation from each of the equations of condition gave

$$b_k - y q_k = x_k,$$

where

$$b_k = a_k - A, \qquad q_k = P - p_k.$$

Thus Laplace began, following Boscovich, by tying the line to the one point $G^* = (P, A)$, effectively using condition (i) to reduce the problem to one involving a single unknown, the slope y of the line through the weighted center of gravity, G^*.

The next step in the algorithm was to suppose that the equations are labeled to correspond to a decreasing sequence b_k/q_k; that is, so that

$$\frac{b_1}{q_1} \geq \frac{b_2}{q_2} \geq \ldots \geq \frac{b_n}{q_n}.$$

Again, this is an algebraic statement of Boscovich's ordering of the points according to their encounter with the moving line. Finally, let $h_k = |i_k q_k|$, and let $F = h_1 + \ldots + h_n$. Then if r is that integer such that

$$h_1 + \ldots + h_{r-1} < \frac{1}{2}F \quad \text{and} \quad h_1 + \ldots + h_r > \frac{1}{2}F,$$

the solution to the problem was to take $y = b_r/q_r$ and $z = A - Py$. The algorithm is of course just an analytic version of Boscovich's geometric procedure, where the points (p_k, a_k) are replaced by $(i_k p_k, i_k a_k)$. Laplace provided an analytic proof that this was indeed the solution to the stated problem, and he applied the method to seven arc measurements. On the basis of this calculation he concluded that an ellipticity of 1/312 was indicated; but he thought the large error (172.52 toises) this implied in the Lapland arc was evidence that the earth was not ellipsoidal (Laplace, 1799–1805, vol. 2, pp. 138–141; Bowditch, 1832, pp. 443–450). [Ironically, a remeasurement of this arc in 1803 by Svanberg *did* show an error of this magnitude (in fact, larger than 200 toises; Svanberg, 1805, p. 192). Later analyses, however, called Svanberg's work into question (Todhunter, 1879).]

Laplace's weighted analysis may now be seen as a small but significant advance in statistical technique. Previous workers had weighted different measurements differently depending on their realized values. James Short, for example, had in 1763 averaged determinations of the parallax of the sun in such a way as to discount those whose distance from the arithmetic mean was large (Stigler, 1973b). Other workers had discarded discordant measurements. In addition, analyses such as those of Mayer and Boscovich had had the effect of giving greater weight, or greater leverage, to measurements taken under extreme conditions; Cotes's rule, quoted earlier, may be most plausibly read as a statement of this principle. Laplace differed from all of these in weighting the measurements according to an intrinsic measure of the measurements' perceived accuracies, the length of the measured arcs.

Was Laplace's weighting correct from a modern perspective? Two sources of error enter into each a_k: the error in measurement on the ground and the error in the astronomical observations. Only two sets of astronomical determinations are required for each arc, regardless of the length of the arc, so we shall ignore this source of error in evaluating the weighting scheme. (Actually, this is not quite correct, because the astronomically measured arc affects a_k as a divisor and errors in small arcs will have a greater effect than the same error for a large arc. However, the errors from this source were at this time likely to be small relative to those from other sources, and we are not likely to be greatly misled by ignoring them in the present analysis.) The error in the measurement on the ground would have had a variance roughly proportional to the arc length i_k (assuming a constant baseline to arc ratio), so we might then expect the variance of a_k to be *inversely* proportional to i_k. Now Boscovich's original method weighted the measurements as if the a_k had equal variances, and Laplace weighted them as if the a_k had variances inversely proportional to i_k^2. Thus it would seem that Laplace's scheme gave too much weight to long arcs. And in particular, it gave much too much weight to the French arc, a circumstance that would probably not have been deplored by Laplace's colleagues at the Academy. If the errors in the astronomical determinations were not negligible, however, Laplace's weights may have been nearly appropriate. In any event, the fact that he attempted any weighting at all at this early date is most interesting.

The method of Boscovich, as formalized by Laplace, has continued to enjoy occasional use since the publication of the *Mécanique céleste.* Prony described it in detail and applied it to problems of water flow in 1804 (Prony, 1804, pp. xxi–xxxii). Three years later Puissant (1807, p. 63) presented it, again in full analytic detail, and recommended its use in surveying. And in 1809 and 1815 Bowditch published applications of a generalization of the method to cometary data. Bowditch's generalizations

(1809, 1815) were particularly interesting in that they combined Mayer's and Boscovich's methods. For example, faced with fifty-six equations of condition involving five unknowns, Bowditch applied the condition that the equations sum to zero separately to four different (but partially overlapping) subsets of equations; he thus eliminated four of the unknowns and solved for the fifth using Laplace's algorithm. As late as 1832 Bowditch was recommending Boscovich's method over least squares because it gave less weight to defective observations than did least squares (1832, p. 434).

Legendre and the Invention of Least Squares

We have now almost arrived at the method of least squares, both in chronological and conceptual terms. We have seen how by 1800 the principle of combining observational equations had evolved, through work of Mayer and Laplace, to produce a convenient ad hoc procedure for quite general situations. We have also seen how the idea of starting with a mathematical criterion had led, in work of Boscovich and Laplace, to an elegant solution suitable for simple linear relationships involving only two unknowns. The first of these approaches developed through problems in astronomy; the second was (at least in these early years) exclusively employed in connection with attempts to determine the figure of the earth. These two lines came together in the work of a man who, like Laplace, was an excellent mathematician working on problems in both arenas—Adrien Marie Legendre.

Legendre came to deal with empirical problems in astronomy and geodesy at a time when the methods we have discussed had been developed separately in the two fields. It was also a time when a half-century's successful use of these methods had seen a change in the view scientists took of them—from Euler's early belief that combination of observations made under different conditions would be detrimental to the later view of Laplace that such combination was essential to the comparison of theory and experience. Legendre brought a fresh view to these problems; and it was Legendre, and not Laplace, who took the next important step.

Legendre did not hit upon the idea of least squares in his first exposure to observational data. From 1792 on he was associated with the French commission charged with measuring the length of a meridian quadrant (the distance from the equator to the North Pole) through Paris. One of the major projects initiated by the National Convention in the early years after the French Revolution had been the decision in 1792 to change the ancient system of measurement by introducing the metric system as a new order, toppling existing standards of measurement in an action symbolic of the French Revolution itself. The basis of the new system was to be the meter, defined to be 1/10,000,000 of a meridian quadrant. It remained for French science to come up with a new determination of the length of this arc. In keeping with the nationalism that inspired the enterprise, the determination was to be based only on new measurements made on French lands. To this end an arc of nearly 10°, extending from Montjouy (near Barcelona) in the south to Dunkirk in the north, was measured in 1795. By 1799 the complex task of reducing the multitude of angular measurements to arc lengths had been completed by J. B. J. Delambre and P. F. A. Méchain. Although the official reports did not appear until after 1805, a summary of the data was widely circulated by 1799.[14]

In early 1799, before the appearance of the first two volumes of the *Mécanique céleste*, Delambre published an extensive discussion of the theoretical results underlying the reduction of the raw data on this arc. This volume (Delambre, 1799) is prefaced by a short memoir by Legendre that is dated 9 Nivôse, an VII (30 December 1798) and indicates that Legendre did not have the method of least squares at that time. He wrote, in a theoretical discussion of the reduction of arc lengths:

> In this way we obtain four equations of the form
>
> $$0 = fx - gy + hz$$
> $$0 = f'x - g'y + h'z$$
> $$e'' = f''x - g''y + h''z$$
> $$e''' = f'''x - g'''y + h'''z,$$
>
> from which we need to find the values of x, y, z. In this type of analysis, of which astronomical questions offer many examples, it is not necessary to seek to satisfy three of the equations exactly; that would force all the error onto the fourth equation. Rather, we need to try to balance the errors in such a way that they are borne nearly equally by all four equations; this will not be difficult when numerical values have been substituted in the equations. (Legendre, 1798, pp. 9–10)

This little-known comment of Legendre's is revealing: It shows that as early as 1798 he had accepted the notion, evolved from Mayer's early writings, that a balance should be struck between measurements, that all should contribute to the final result. But the comment shows no sense of a need for a general method of striking this balance. Rather it suggests that it will not be difficult to proceed ad hoc: After the numerical values have been substituted in the equations, a balance could be found acceptable for the specific case. Evidently Legendre was to change his mind on this in the next five years.

The occasion for Legendre's reconsideration of observational equa-

14. For example, in France it was published in Laplace's *Mécanique céleste*, vol. 2., bk. 3, §41 (1799), and in Germany in *Allgemeine Geographische Ephemeridenz* vol. 4, p. xxxv (1799).

tions, and for the appearance of the method of least squares, was the preparation in 1805 of a memoir on the determination of cometary orbits (Figure 1.5). The memoir is a scant seventy-one pages (excluding the appendix); and, aside from a few brief remarks at the end of the preface, the method of least squares makes no appearance before page 64. Even this first mention of least squares seems to be an afterthought because, after presenting an arbitrary solution to five linear equations in four unknowns (one that assumed that two equations held exactly and two of the unknowns were zero), Legendre wrote that the resulting errors were of a size "quite tolerable in the theory of comets. But it is possible to reduce them further by seeking the *minimum* of the sum of the squares of the quantities E′, E″, E‴ " (1805, p. 64). He then reworked the solution in line with this principle. It seems plausible that Legendre hit on the method of least squares while his memoir was in the later stages of preparation, a guess that is consistent with the fact that the method is not employed earlier in the memoir, despite several opportunities.

It is clear, however, that Legendre immediately realized the method's potential and that it was not merely applications to the orbits of comets he had in mind. On pages 68 and 69 he explained the method in more detail (with the word *minimum* making five italicized appearances, an emphasis reflecting his apparent excitement), and the memoir is followed on pages 72–80 by the elegant appendix from which the quotation near the beginning of this chapter was taken. The example that concludes the appendix reveals Legendre's depth of understanding of his method (notwithstanding the lack of a formal probabilistic framework). It also suggests that it was because Legendre saw these problems of the orbits of comets as similar to those he had encountered in geodesy that he was inspired to introduce his principle and was able to abstract it from the particular problem he faced. Indeed, the example he chose to discuss was not just given as an illustration, it was a serious return to what must have been the most expensive set of data in France—the 1795 measurements of the French meridian arc from Montjouy to Dunkirk.

To determine the figure of the earth from these data (Table 1.7), Legendre developed the relationship between arc length in degrees and in toises in a form slightly different from that we encountered earlier. Letting L and L' be the astronomically determined latitudes of the end points of an arc (from column 2 of Table 1.7) and S the measured length of the arc (given in column 3 of Table 1.7 in modules, where a module is just 2 toises), Legendre wrote

$$L' - L = \frac{S}{D} + \frac{3}{2}\cdot\alpha\cdot\frac{180}{\pi}\sin(L' - L)\cos(L' + L)$$
$$= \frac{S}{28{,}500} + \mathscr{C}\cdot\frac{S}{28{,}500} + \alpha\frac{270}{\pi}\sin(L' - L)\cos(L' + L).$$

APPENDICE.

Sur la Méthode des moindres quarrés.

Dans la plupart des questions où il s'agit de tirer des mesures données par l'observation, les résultats les plus exacts qu'elles peuvent offrir, on est presque toujours conduit à un système d'équations de la forme

$$E = a + bx + cy + fz + \&c.$$

dans lesquelles a, b, c, f, &c. sont des coëfficiens connus, qui varient d'une équation à l'autre, et x, y, z, &c. sont des inconnues qu'il faut déterminer par la condition que la valeur de E se réduise, pour chaque équation, à une quantité ou nulle ou très-petite.

Si l'on a autant d'équations que d'inconnues x, y, z, &c., il n'y a aucune difficulté pour la détermination de ces inconnues, et on peut rendre les erreurs E absolument nulles. Mais le plus souvent, le nombre des équations est supérieur à celui des inconnues, et il est impossible d'anéantir toutes les erreurs.

Dans cette circonstance, qui est celle de la plupart des problêmes physiques et astronomiques, où l'on cherche à déterminer quelques élémens importans, il entre nécessairement de l'arbitraire dans la distribution des erreurs, et on ne doit pas s'attendre que toutes les hypothèses conduiront exactement aux mêmes résultats; mais il faut sur-tout faire en sorte que les erreurs extrêmes, sans avoir égard à leurs signes, soient renfermées dans les limites les plus étroites qu'il est possible.

De tous les principes qu'on peut proposer pour cet objet, je pense qu'il n'en est pas de plus général, de plus exact, ni d'une application plus facile que celui dont nous avons fait usage dans les recherches précédentes, et qui consiste à rendre

(73)

minimum la somme des quarrés des erreurs. Par ce moyen, il s'établit entre les erreurs une sorte d'équilibre qui empêchant les extrêmes de prévaloir, est très-propre à faire connoître l'état du système le plus proche de la vérité.

La somme des quarrés des erreurs $E^2 + E'^2 + E''^2 +$ &c. étant

$$\begin{aligned} &(a + bx + cy + fz + \&c.)^2 \\ +&(a' + b'x + c'y + f'z + \&c.)^2 \\ +&(a'' + b''x + c''y + f''z + \&c.)^2 \\ +&\ \&c.; \end{aligned}$$

si l'on cherche son *minimum*, en faisant varier x seule, on aura l'équation

$$o = \int ab + x\int b^2 + y\int bc + z\int bf + \&c.,$$

dans laquelle par $\int ab$ on entend la somme des produits semblables $ab + a'b' + a''b'' +$ &c.; par $\int b^2$ la somme des quarrés des coëfficiens de x, savoir $b^2 + b'^2 + b''^2 +$ &c., ainsi de suite.

Le *minimum*, par rapport à y, donnera semblablement

$$o = \int ac + x\int bc + y\int c^2 + z\int fc + \&c.,$$

et le *minimum* par rapport à z,

$$o = \int af + x\int bf + y\int cf + z\int f^2 + \&c.,$$

où l'on voit que les mêmes coëfficiens $\int bc$, $\int bf$, &c. sont communs à deux équations, ce qui contribue à faciliter le calcul.

En général, *pour former l'équation du* minimum *par rapport à l'une des inconnues, il faut multiplier tous les termes de chaque équation proposée par le coëfficient de l'inconnue dans cette équation, pris avec son signe, et faire une somme de tous ces produits.*

On obtiendra de cette manière autant d'équations du *minimum*, qu'il y a d'inconnues, et il faudra résoudre ces équations par les méthodes ordinaires. Mais on aura soin d'abréger tous les calculs, tant des multiplications que de la résolution, en n'admettant dans chaque opération que le nombre de chiffres

10

(74)

entiers ou décimaux que peut exiger le degré d'approximation dont la question est susceptible.

Si par un hasard singulier, il étoit possible de satisfaire à toutes les équations en rendant toutes les erreurs nulles, on obtiendroit également ce résultat par les équations du *minimum*; car si après avoir trouvé les valeurs de x, y, z, &c. qui rendent nulles E, E', &c., on fait varier x, y, z, &c. de $\delta x, \delta y, \delta z$, &c., il est évident que E^2 qui étoit zéro deviendra par cette variation $(a\delta x + b\delta y + c\delta z, \&c.)^2$. Il en sera de même de E'^2, E''^2, &c. D'où l'on voit que la somme des quarrés des erreurs aura pour variation une quantité du second ordre par rapport à $\delta x, \delta y$, &c.; ce qui s'accorde avec la nature du *minimum*.

Si après avoir déterminé toutes les inconnues x, y, z, &c., on substitue leurs valeurs dans les équations proposées, on connoîtra les diverses erreurs E, E', E'', &c. auxquelles ce système donne lieu, et qui ne peuvent être réduites sans augmenter la somme de leurs quarrés. Si parmi ces erreurs il s'en trouve que l'on juge trop grandes pour être admises, alors on rejettera les équations qui ont produit ces erreurs, comme venant d'expériences trop défectueuses, et on déterminera les inconnues par le moyen des équations restantes, qui alors donneront des erreurs beaucoup moindres. Et il est à observer qu'on ne sera pas obligé alors de recommencer tous les calculs; car comme les équations du *minimum* se forment par l'addition des produits faits dans chacune des équations proposées, il suffira d'écarter de l'addition les produits donnés par les équations qui auront conduit à des erreurs trop considérables.

La règle par laquelle on prend le milieu entre les résultats de différentes observations, n'est qu'une conséquence très-simple de notre méthode générale, que nous appellerons *Méthode des moindres quarrés.*

En effet, si l'expérience a donné diverses valeurs a', a'', a''', &c.

(75)

pour une certaine quantité x, la somme des quarrés des erreurs sera $(a' - x)^2 + (a'' - x)^2 + (a''' - x)^2 +$ &c., et en égalant cette somme à un *minimum*, on a

$$o = (a' - x) + (a'' - x) + (a''' - x) + \&c.;$$

d'où résulte $x = \dfrac{a' + a'' + a''' + \&c.}{n}$, n étant le nombre des observations.

Pareillement, si pour déterminer la position d'un point dans l'espace, on a trouvé, par une première expérience, les coordonnées a', b', c'; par une seconde, les coordonnées a'', b'', c'', &c. ainsi de suite; soient x, y, z, les véritables coordonnées de ce point: alors l'erreur de la première expérience sera la distance du point (a', b', c') au point (x, y, z); le quarré de cette distance est

$$(a' - x)^2 + (b' - y)^2 + (c' - z)^2;$$

et la somme des quarrés semblables étant égalée à un *minimum*, on en tire trois équations qui donnent $x = \dfrac{\int a}{n}$, $y = \dfrac{\int b}{n}$, $z = \dfrac{\int c}{n}$, n étant le nombre des points donnés par l'expérience. Ces formules sont les mêmes par lesquelles on trouveroit le centre de gravité commun de plusieurs masses égales, situées dans les points donnés; d'où l'on voit que le centre de gravité d'un corps quelconque jouit de cette propriété générale.

Si on divise la masse d'un corps en molécules égales et assez petites pour être considérées comme des points, la somme des quarrés des distances des molécules au centre de gravité sera un minimum.

On voit donc que la méthode des moindres quarrés fait connoître, en quelque sorte, le centre autour duquel viennent se ranger tous les résultats fournis par l'expérience, de manière à s'en écarter le moins qu'il est possible. L'application que nous allons faire de cette méthode à la mesure de la méridienne, achèvera de mettre dans tout son jour sa simplicité et sa fécondité.

Figure 1.5. Legendre's 1805 appendix, introducing the method of least squares. (From Legendre, 1805, pp. 72–75.)

Table 1.7. Measurements of the French meridian arc, made in 1795 between Montjouy (near Barcelona) and Dunkirk.

Place of observation	Latitude, L	Arc length S	$L' - L$	$L' + L$
Dunkirk	51°2′10″.50	62,472.59	2°11′20″.75	99°53′0″
Pantheon (Paris)	48°50′49″.75	76,145.74	2°40′7″.25	95°1′32″
Evaux	46°10′42″.50	84,424.55	2°57′48″.10	89°23′37″
Carcassonne	43°12′54″.40	52,749.48	1°51′9″.60	84°34′39″
Montjouy	41°21′44″.80			

Source: Legendre (1805, p. 76). Reprinted in Harvey (1822); given in a slightly different form in Laplace (1799–1805, vol. 2, bk. 3, §41) and Bowditch (1832, p. 453).
Note: Arc lengths (S) are in modules; 1 module = 2 toises ≅ 12.78 feet.

Here D is the length in modules of 1° centered at 45° latitude, α is the ellipticity of the earth, and $\mathscr{C}$ is defined by the relationship $D^{-1} = (1 + \mathscr{C})/28{,}500$. At first glance this appears to be quite a change from the earlier relationship $a = z + y \sin^2\theta$; but it is not. Just note that, since $\sin^2 45° = 0.5$, $D = z + y/2$; and, since $\theta = (L' + L)/2$, $2\sin^2\theta = 1 - \cos 2\theta = 1 - \cos(L' + L)$. Then since $3\alpha = y/D$ and $a = S/(L' - L)$, we can easily see that the formulas are equivalent, save that Legendre employs $180\sin(L' - L)/\pi$ instead of its local approximation, $L' - L$. Actually, Legendre's formulation is [except for the local approximation of $\sin(L' - L)$] exactly the same as that used by Laplace for his weighted analysis: In our previous notation, Legendre's equation is equivalent to $i_k a_k = i_k z + i_k y p_k$. The introduction of $\mathscr{C}$ instead of D is just a reparametrization, based on the fact that D is known to be near 28,500; it is both easier and more accurate to work with smaller numbers.

Legendre then let E^i represent the error made in the determination of the ith latitude; and, on the basis of data of Table 1.7, he obtained four equations:

$$\begin{aligned}
E^{I} - E^{II} &= 0.002923 + \mathscr{C}(2.192) - \alpha(0.563) \\
E^{II} - E^{III} &= 0.003100 + \mathscr{C}(2.672) - \alpha(0.351) \\
E^{III} - E^{IV} &= -0.001096 + \mathscr{C}(2.962) + \alpha(0.047) \\
E^{IV} - E^{V} &= -0.001808 + \mathscr{C}(1.851) + \alpha(0.263).
\end{aligned}$$

In forming these equations, Legendre assumed that the effect of these errors on $\sin(L' - L)\cos(L' + L)$ was negligible. Now, he might have applied the method of least squares to these equations directly, but instead he noted that "it is necessary to consider the errors separately." I take this to mean that, despite his lack of any formulation of any probability model, he correctly feared one of the consequences of the correlation of the equation's left-hand sides, namely, that treating these differences as four errors would restrict his choice of solutions. Therefore he introduced a fifth equation, $E^{III} = E^{III}$, which permitted him to reexpress the equations as

$$E^{I} = E^{III} + 0.006023 + \mathscr{C}(4.864) - \alpha(0.914)$$
$$E^{II} = E^{III} + 0.003100 + \mathscr{C}(2.672) - \alpha(0.351)$$
$$E^{III} = E^{III}$$
$$E^{IV} = E^{III} + 0.001096 - \mathscr{C}(2.962) - \alpha(0.047)$$
$$E^{V} = E^{III} + 0.002904 - \mathscr{C}(4.813) - \alpha(0.310).$$

He then solved these equations by the method of least squares, treating E^{III} on the right-hand side as an unknown, together with $\mathscr{C}$ and α. He found $\alpha = 0.00675 = 1/148$ and $\mathscr{C} = 0.0000778$. Thus $D = 28{,}500/(1 + \mathscr{C}) = 28{,}497.78$, and the corresponding length of the meridian quadrant would be $90 \cdot D = 2{,}564{,}800.20$ modules, a value leading to a meter of 0.256480 modules = 0.512960 toises ≅ 3.280 feet.

The actual meter was based upon the value found for D by Laplace in the *Mécanique céleste* (vol. 2, bk. 3, §41; see Bowditch, 1832, p. 465), namely (expressed in terms of a standard degree), $D = 28{,}504.11$, which gave the meridian quadrant as 2,565,370 modules and the meter as 0.256537 modules = 0.513074 toises ≅ 3.281 feet. Laplace's determination incorporated the measurement of the arc at Peru into the calculation of the ellipticity; then he found the value of D from the French arc based on this predetermined ellipticity using his own algorithm to minimize the maximum error. We note that the use of only the French arc would not, because it extended only about 10°, permit a very accurate determination of the ellipticity. The same was not true, however, with respect to D and thus of the meridian quadrant, $90D$. Thus the restriction to French data was made at less cost in efficiency than might be feared, at least as far as the determination of the meter was concerned.

Two observations on Legendre's procedure are in order: Viewed from a later perspective it was not a correct way of dealing with the type of correlation he encountered, and it was not original with Legendre. In fact, Laplace had employed the same sort of approach attempting to untangle errors that appear as differences by treating one as an unknown to be estimated—in his handling of these same data in the *Mécanique céleste* (vol. 2, bk. 3, §41; see Bowditch, 1832, p. 459). It is likely that it was Laplace's analysis that suggested the approach, even though Laplace minimized the maximum error, not the sum of squared errors. This approach is incorrect from a postcorrelation point of view. It does not introduce a bias into the results; but, because it ignores the variability of the artificially designated "unknown," it produces a weighting of the observations that is not of maximum efficiency for any reasonable specification of the errors' stochastic structure. This observation should not be construed as criticism of Legendre or Laplace, however, for it was to be more than a century before efficient methods were developed for dealing with the type of correlation they faced, and in the present case the difference in result is negligible. Rather, it is remarkable that, lacking any explicit probabilistic formulation, they made any attempt at all to deal with the problem. The attempt they did make was a limited one and seems to have been based on a rough intuitive notion of dependence and tied to the explicit notational appearance of the same errors in different equations. It is nonetheless surprising to find even this crude recognition of the dependence at this early time. We shall see that even a half-century after probability was introduced formally into the analysis of such problems, little more understanding of dependence was evident than is found in this, the first published example of the method of least squares.

With Legendre's introduction of least squares, we reach the end of the first stage of the lines of development begun separately by Mayer and Boscovich about a half-century before. The idea of combining different observational equations evolved slowly from Mayer's astronomical work, the idea of an objective criterion of fit was born in Boscovich's geodetic work, and their inspired synthesis was signaled by Legendre's geodetic example, appended to an astronomical memoir. But a key element was missing: There was, in all of this work, no formal appeal to probability and, more to the point, no move to quantify the uncertainty in the derived estimates (save only Mayer's weak attempt in 1750). All of this is the more puzzling because Laplace, who had been writing extensively on probability since 1774, had played a key role in all of this development. In Chapter 4 we shall see how, in the two decades following Legendre's analyses, the next stage was completed by Laplace, with Carl Friedrich Gauss providing a key catalytic agent. To explain this properly, however, we must first examine the major currents of eighteenth-century probability, and it is to this topic I now turn, starting with the work of Jacob Bernoulli.