Simpson's first writings on this topic), then Boscovich could be regarded as a precursor of Simpson (and of Lagrange) on this matter[6]. Sheynin [1973b, p. 318] framed Boscovich's problem as follows.

> In terms of probability (not chances) BOSCOVICH's problem could be stated thus: equiprobable values of each of n random quantities $\xi_1, \xi_2, \ldots, \xi_n$ are 0, 1, and -1. To find, the law of distribution of their sum (ξ). The generally known formula, the equivalent of which BOSCOVICH actually knew, is
>
> $$P\{\xi = a\} = \sum[P\{\xi_1 = x\}\, P\{\xi_2 = y\} \ldots P\{\xi_n = w\}],$$
>
> the summation extending over all the values of $x, y, \ldots, w$ complying with the conditions
>
> $$x + y + \cdots + w = a, \quad x, y, \ldots, w = 1, \text{or } 0, \text{or } -1.$$

In 1984 Stigler discussed a manuscript fragment by Simpson headed 'A problem proposed to me by M. Boscowitze', that problem being one in least absolute deviations regression. This fragment Stigler dated as 1760, the year in which Boscovich visited London, and in 1990 Farebrother presented evidence to show that Boscovich had personally received Simpson's solution in June 1760.

Although it would be inappropriate to spend too much time on Simpson's paper and tract themselves, a few general comments might not come amiss[7]. First of all, one should note that Simpson essentially used de Moivre's method of generating functions[8] in what was tantamount to the derivation of the distribution of a sum of independent errors, each having a discrete probability distribution, this latter being uniform in Proposition I and triangular in Proposition II[9]. In the latter portion of his tract Simpson introduced a continuous triangular distribution as an approximation to the earlier discrete one, and it is worth noting that he is sometimes described as the first to consider not only the concept of error distributions, but also that of continuous distributions[10]; Shoesmith [1985], however, notes that, before Simpson, J. Bernoulli, N. Bernoulli, and de Moivre had all derived continuous approximations to the binomial distribution[11]. It is important to note that estimation theory arose from the consideration of problems in which most of, if not all, the statistical variability was occasioned by errors of measurement, and not from problems in which the data themselves exhibit a large measure of internal variability (Huber [1972, p. 1042]). Simpson's work was continued later by Lagrange and still later by Laplace[12].

There are those who have perhaps read somewhat more into Simpson's work than was actually achieved. Sheynin [1968] notes that Simpson could have derived the Normal distribution (later than de Moivre) and could have been the first to sketch this distribution (the sketch of his limiting distribution is not in fact that of a Normal density). Lancaster [1994] in fact goes further, stating that Simpson was the first to consider the Normal distribution as a possible error density function.

In his discussion of Boscovich's determinism, Sheynin [1973b, p. 321] suggests that in his *Philosophiæ naturalis theoria* of 1758 Boscovich might have been considering a discrete uniform distribution of velocities

$$v, v \pm \Delta v, v \pm 2\Delta v, \ldots, v \pm n\Delta v.$$

If this indeed be so, then this consideration may be viewed as a generalization of the distribution considered in the undated manuscript by Boscovich mentioned before, and hence the latter might well pre-date the 1758 book.

In discussing Boscovich's 1757 criteria for the determination of the best-fitting straight line to a set of data, Eisenhart [1961, p. 200] instances the following conditions imposed by Boscovich.

> (I) The sums of the positive and negative corrections ... shall be equal.
> (II) The sum of (the absolute values of) all of the corrections, positive and negative, shall be as small as possible.

Although different to Simpson's (as is not surprising, in view of the difference in the problems addressed), these criteria are of interest in showing the direction being taken by those investigating error theory at that time. One might also note Eisenhart's statement that Boscovich believed his first criterion to be required by

> the *traditional*[13] assumption that positive and negative errors are equally probable. [1961, p. 209]

The fact that the arithmetic mean of a set of observations is not always the best summarizing statistic was discussed by Daniel Bernoulli in 1777. Here Bernoulli took it as axiomatic that errors in excess or in defect of the true value were equally possible, whereas values nearer the true value were more probable than those removed from it. Errors greater than some maximum value were regarded as impossible. And although Simpson assumed a triangular distribution, Bernoulli took the probability curve to be a semi-circle of given radius.[14]

An interesting woodcut from a 1535 book by Jacob Köbel on surveying is reproduced in Chapter 20 of Stigler [1999], an illustration showing the feet of sixteen church-goers being used to determine the length of a 'right and lawful rood'. Stigler notes this woodcut demonstrates (a) the act of taking a sample where chance enters into the selection, (b) the recognition of the benefit to be derived from compensatory errors, and (c) a rudimentary acquaintance with the well-known concepts of statistical sufficiency and exchangeability. For our purposes it is (b) that is of greatest relevance.