



## L. The law of error and correlated averages

Professor F. Y. Edgeworth M.A. D.C.L.

To cite this article: Professor F. Y. Edgeworth M.A. D.C.L. (1892) L. The law of error and correlated averages , Philosophical Magazine Series 5, 34:210, 429-438, DOI: [10.1080/14786449208620355](https://doi.org/10.1080/14786449208620355)

To link to this article: <http://dx.doi.org/10.1080/14786449208620355>



Published online: 08 May 2009.



Submit your article to this journal [↗](#)



Article views: 4



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

the little changeful group of interwoven thoughts that is himself—is a *very* small part of the great Autic Universe. We must shift our centre, and exchange the metaphysician's narrow Ptolemaic for a broad Copernican view of existence.

---

L. *The Law of Error and Correlated Averages.*  
By PROFESSOR F. Y. EDGEWORTH, M.A., D.C.L.\*

THIS is a contribution to the investigation of the most general conditions under which the exponential law of error is fulfilled; together with some applications to the theory of "correlated averages" †.

I. The simplest case in which a group of measurable objects range in conformity with the law of error is where every member of the group is a sum of a certain number of items, or elements, each of which has or has not a certain quality, *e. g.* the colour white, or, as it may be expressed, assumes the value 1 or 0, with a certain average frequency (*e. g.* 1 just as often—or half as often—as 0, in the long run) and at random, or in

---

are a part of the autic universe, are shadowed by certain objective changes in my brain; and the term synergos means that other portion of the autic universe which is shadowed *by all the other events* that happen objectively in my brain. It appears from physical considerations that the particular stream effects or other changes in the brain that were the shadows of the perceptions I had at breakfast-time, cannot have occurred alone, but were accompanied by more subtle motions or changes in the brain, which were the shadow of, and thus betokened, certain closely associated events then going on in my synergos. These again were succeeded by motions, changes, or states of strain in the brain during the intervening hours, all of which were a part of the varying shadow of the synergos as it underwent whatever changes took place in it during that interval. Moreover, these intervening events in the brain were of such a kind, as the result has proved, that they have been now followed up by motions or changes in the brain which resemble those that were the shadow of my thoughts at breakfast-time, and which are a part of the group of events now going on in my brain that are the shadow of those thoughts that constitute my mind as it exists at present.

Softening of the brain is the shadow cast within the objective world when very unfortunate events have happened in the autic universe—events which have included a weakening of the power which the synergos and the mind previously had of mutually acting on one another, or else which have prevented the full formation within the synergos of some of the intermediate links of causation spoken of above. Either of these would involve a partial loss of memory.

\* Communicated by the Author.

† See Galton (Proc. Roy. Soc. 1888), "Co-relations and their Measurements;" and Weldon (Proc. Roy. Soc. 1892), "Certain Correlated Variations in *Crangon vulgaris*;" also "Correlated Averages," by the present writer in the Philosophical Magazine for August 1892.

such wise that the value assumed by one item does not affect that of another. This is a case fully discussed by the classical writers on Probabilities; under the guise of problems relating to games of chance. Thus if we take several batches of balls, every batch containing  $n$  balls, each of which is either black or white, the probability of a white being  $p^*$ ; then the frequency of white balls is approximately represented by a curve of error, of which the centre corresponds to  $np$ , the most probable number of white balls in a batch, and the modulus is  $\sqrt{2np(1-p)}$ . I venture to refer to my reproduction, in a former number of this Journal†, of Poisson's reasoning on this problem, in order to recall a proposition which will be required here: namely, that the limits, on either side of  $np$ , up to which the approximation holds good are of the order  $\sqrt{n}$ ‡.

II. An easy transition brings us to a more general case in which each item has any finite limits  $a_r$  and  $\alpha_r$  ( $\alpha$  greater than  $\alpha_r$ ). These limits need not be identical for each item; provided that the range of any one item, say  $a_r - \alpha_r$ , is small in comparison with the sum of the ranges  $S(a_r - \alpha_r)$ . Nor are we confined to the supposition that each item should assume one or other of two values  $a_r$  and  $\alpha_r$ ; it may assume any one of an indefinite number of values indicated by the curve of distribution  $y_r = f_r(x)$ , representing the frequency with which the  $r$ th element assumes each value  $x$ ; where  $f_r$  may have any form whatever, continuous or not, provided that it does not extend beyond  $\alpha_r$  and  $a_r$ , and that the integral between those limits is unity. Moreover, it is allowable to affect each item with a different factor or "weight"; provided that no weight is preponderant—large in comparison with the sum of all the other weights. When we have thus substituted, for a "sum," a linear function of independently varying items, we have reached the extent of generalization to which Laplace thought it necessary to carry the investigation for the purpose of the theory of errors§.

III. We enter on a less trodden path when, following the lead of Mr. Glaisher||, we pass from a "linear" function of items to any function whatever. To make the transition less violent, let us break it up into two steps; and first consider

\*  $n$  being a large number, and  $p$  not a very small fraction.

† Phil. Mag. 1887, xxiv. p. 330.

‡ There is some approximation outside these limits; but not of the degree usually assigned.

§ There is a good account of Laplace's analysis in Todhunter's 'History of Probabilities,' art. 1001 *et seqq.*

|| Memoirs of the Astronomical Society, vol. xl. p. 105.

the case in which each member of a group is—not a linear function—but *some function of* a linear function of numerous elements oscillating in the manner defined above : say  $F(\Xi)$ ; where

$$\Xi = p_1x_1 + p_2x_2 + \&c. + p_nx_n;$$

$p_1, p_2, \&c.$  are of the same order of magnitude;  $x_1, x_2, \&c.$  oscillate respectively between limits which we may write without loss of generality  $0-a_1, 0-a_2, \&c.$ ; while  $Sa$  may be taken as the unit. Now by the usual theory  $\Xi$  fluctuates about its average value—say  $X$ —in conformity with a Probability-curve of which the modulus is of the order  $\frac{1}{\sqrt{n}}$ \*

( $Sa$  being = 1). Put  $\Xi = F(X + \xi)$  and, expanding, write

$$F(\Xi) = F(X) + \xi F'(X) + \frac{\xi^2}{2} F''(X + \theta\xi)$$

(where  $\theta$  is a proper fraction). This expression is approximately equal to its first two terms  $F(X) + \xi F'(X)$ ; provided that, for the values of  $F(\Xi)$  with which we are concerned—at most from 0 to 1 (if  $Sa=1$ )—the function is free from singularity and continuous in senses defined by the condition that  $\frac{\xi^2}{2} F''(X + \theta\xi)$  is small in relation to  $F'(X)$ , for values of  $\xi$  of the order  $\frac{1}{\sqrt{n}}$ †.

That condition holding, we may reason thus:—Of the group formed by the varying values of  $\Xi$  the greater part is, by the usual theory, arranged according to a probability-curve with centre  $X$  (the average value of  $\Xi$ ), up to a distance from that centre,  $\pm \xi$ , where  $\xi$  is a small fraction. But every value of  $F(X + \xi) = F(X) + \xi F'(X)$  nearly. Therefore the greater part of the group  $F(\Xi)$  ranges in conformity with a probability-curve whose centre is  $F(X)$ , and whose modulus is that of  $\xi$  multiplied by  $F'(X)$ .

For example, let  $F(\Xi) = \Xi^2$ ; where  $\Xi$  is the sum of  $m$

\* Above, p. 430, and the article there referred to.

† This is a kind of assumption continually made, I think, by mathematicians. To take an instance cognate to the present subject, Laplace, when introducing the Method of Least Squares (*Theorie analytique*, Book II. ch. iv. art. 20) supposes the datum of observation to be a function of the “element” which it is sought to determine. An approximate value for this value having been obtained; and this value, *plus* a correction  $z$  having been substituted for the element; “expanding in ascending powers of  $z$  [en reduisant en série par rapport à  $z$ ] and neglecting the square of  $z$ , this function will take the form  $h + pz$ .” It will take that form only upon the condition above stated.

independent elements, each element in random fashion assuming either the value 0 or  $+i$  with equal frequency in the long run. Then, according to the magnitude of  $m$ , and the degree of accuracy required, the group formed by the varying values of  $\Xi$  may be regarded as conforming to a probability-curve whose modulus is  $\sqrt{\frac{m}{2}}i$ , up to a distance from the

central value  $\left(\frac{m}{2}i\right)$  amounting to the quartile, octile, decile, &c., as the case may be. Thus of a group of values assumed by  $\Xi$  about a quarter occurs between the limits  $\frac{m}{2}i$  and  $\frac{m}{2}i - q\sqrt{\frac{m}{2}}i$ ; where  $q$  is the "quartile" for modulus unity,  $=.476\dots$ . Another quarter occurs between  $\frac{m}{2}i + q\sqrt{\frac{m}{2}}i$ .

Similarly the octiles, deciles, &c. are

$$\frac{m}{2}i \pm r\sqrt{\frac{m}{2}}i, \quad \frac{m}{2}i \pm s\sqrt{\frac{m}{2}}i, \quad \&c.;$$

where  $r, s$ , &c. are coefficients obtained from the tables; either below, or not much above unity. And so on, up to the largest percentile up to which the approximation is accurate.

Now let this group be deformed by squaring each of the observations. The new median, quartile, octile, &c. will of course be the squares of the respective old ones. The new median will be  $\frac{m^2}{4}i^2$ ; the new quartiles  $\left(\frac{m}{2} \pm q\sqrt{\frac{m}{2}}\right)^2 i^2$ ;

distant respectively from the new median,  $\frac{m^2}{4}i^2$ , by

$$\frac{m^{\frac{3}{2}}}{\sqrt{2}}q\left(1 + \frac{q}{\sqrt{2m}}\right)i^2 \quad \text{and} \quad \frac{m^{\frac{3}{2}}}{\sqrt{2}}q\left(1 - \frac{q}{\sqrt{2m}}\right)i^2.$$

Now the last-written two expressions differ from the expression outside the brackets and from each other by small quantities; so that but a small proportion of the group occurs between the limits

$$\left(\frac{m^2}{4} + \frac{m^{\frac{3}{2}}}{\sqrt{2}}q\right)i^2 \quad \text{and} \quad \left(\frac{m^2}{4} - \frac{m^{\frac{3}{2}}}{\sqrt{2}}q\right)i^2 + q^2\frac{m}{2}i^2;$$

and in the corresponding interval in the neighbourhood of the quartile *below* the Mean. Thus the new quartiles are approximately  $\left(\frac{m^2}{4} \pm \frac{m^{\frac{3}{2}}}{\sqrt{2}}q\right)i^2$ . In like manner it may be shown

that the new octiles, deciles, &c. are approximately

$$\left(\frac{m^2}{4} \pm \frac{m^{\frac{3}{2}}}{\sqrt{2}} r\right) i^2, \quad \left(\frac{m^2}{4} \pm \frac{m^{\frac{3}{2}}}{\sqrt{2}} s\right) i^2, \text{ \&c.}$$

( $r$ ,  $s$ , &c. having the meaning assigned in the last paragraph). Thus the translated curve of distribution is such that its quar-

tiles are approximately  $\left(\frac{m^2}{4} \pm q \frac{m^{\frac{3}{2}}}{\sqrt{2}}\right) i^2$ , its octiles  $\left(\frac{m^2}{4} \pm r \frac{m^{\frac{3}{2}}}{\sqrt{2}}\right) i^2$ ,

its deciles  $\left(\frac{m^2}{4} \pm s \frac{m^{\frac{3}{2}}}{\sqrt{2}}\right) i^2$ ; that is, the new distribution is approximately a probability-curve whose centre is  $\frac{m^2}{4} i^2$ , and modulus  $\frac{m^{\frac{3}{2}}}{\sqrt{2}} i^2$ .

In the preceding example the conformity of  $\Xi$  to a probability-curve is known *à priori* by the usual theory, referred to in our second section\*. Let us take another example in which this conformity is known *à posteriori* by actually observing the measurements of a group: for instance, heights of men. The annexed numerals give the number of men per thousand of each particular height in inches; as ascertained by Mr. Elliott from the measurement of some 25,000 American recruits (International Statistical Congress, 1863). Thus there are 121 men (per thousand) of the height 70 inches, that is, as I understand, between the heights 69.5 and 70.5 inches.

1	0	0	0	1	2	20	48	75	117	134
57	58	59	60	61	62	63	64	65	66	67
157	140	121	80	57	26	13	5	2	0	1
68	69	70	71	72	73	74	75	76	77	78

Now let the curve or locus of distribution thus constituted be translated by squaring each of the measurements; while each compartment thus dislocated carries with it, so to speak, the men appertaining thereto. *E. g.* the 121 men who were originally found between 69.5 and 70.5 are now distributed between  $69.5^2$  and  $70.5^2$ . It will be found that the translated observations fulfil the law of error in the same sense as the original ones; that is, in the sense in which a planet is proved *à posteriori* to move in an ellipse. The parameters being calculated from some of the observations, other observations are found to tally with the curve thus determined.

Let us adopt a uniform method of calculating the para-

\* The case might be brought under our first section by a mere change of unit.

meters (the centre and the modulus) for the primary and for the dislocated curves ; namely, from the Arithmetic mean, and from the distance between certain percentiles which are given by observation. In the primary group, above the point 69·5 occur 305 observations, ·305 of the total thousand ; and below the point 66·5 occur 264 observations, ·264 of the total. Whence between the Mean and the point 69·5 there should occur  $\cdot 5 - \cdot 305 = \cdot 195$  of the total ; and between the Mean and the point 66·5, ·236 of the total. Now it is found from the tables that ·195 of the total number corresponds to ·361 of the Modulus, ·236 of the total to ·447 of the Modulus. Accordingly we have ·808 ( $= \cdot 361 + \cdot 447$ ) modulus, equated to the interval between 66·5 and 69·5 inches = 3 inches. Whence the modulus  $= 3 \div \cdot 808 = 3\cdot 71$  inches.

Let us now apply this result to *predict* the number of observations at particular heights (not too near the extremity, where the fulfilment of the Probability-curve is not to be looked for). To predict the number above 70·5 (which gives the number between 69·5 and 70·5, the number above 69·5 being already taken for granted) we are to employ the Arithmetic mean, which is 68·20.  $70\cdot 5 - 68\cdot 2 = 2\cdot 3$  inches = ·62 modulus. Now, according to the Tables, the proportion outside the point which is at a distance of ·62 modulus from the Mean  $= \frac{1}{2}(1 - \cdot 6194)$  of the total  $= \frac{1}{2} \times \cdot 380 \times 1000 = 190$ . The real number is 184.

To find the number above 71·5 we have  $71\cdot 5 - 68\cdot 2 = 3\cdot 3$  inches = ·89 modulus ; corresponding to  $\frac{1}{2}(1 - \cdot 7918)$  of the total  $= \frac{1}{2} \cdot 2082 \times 1000 = 104$ . The real number is also 104.

Proceeding similarly with the lower limb, to predict the number below 65·5, we have  $68\cdot 2 - 65\cdot 5 = 2\cdot 7 =$  nearly ·728 modulus ; corresponds to  $\frac{1}{2}(1 - \cdot 6971)$  total  $= \frac{1}{2} \cdot 302 \times 1000 = 151$ . Whereas the real number is 147.

By parity we find below 64·5, 79 by calculation, against 72 observed.

The question is now whether we shall fare equally well when we apply the same method to the group which is formed by squaring each observation in the manner explained. By parity of calculation the modulus of the new curve  $\times \cdot 808$

$$= (69\cdot 5^2 - 66\cdot 5^2) \text{ inches} = 408.$$

Whence the new modulus = 505 inches. Also the Arithmetic mean of the new group is 4657·38. Accordingly, to predict the number of observations above  $(70\cdot 5)^2$  we have

$$4970\cdot 25 - 4657\cdot 38 = 312\cdot 87 = \cdot 62 \text{ modulus.}$$

Whence the calculated number is 190, exactly the same result

as before. By parity the number above  $(71.5)^2$  is calculated to be 102.

With regard to the lower limb, in order to predict the number below  $65.5^2$ , we have  $4657.38 - 4290.25 = 367.13 = .727$  modulus, which corresponds to

$$\frac{1}{2}(1 - (.6914 + .7 \times .0067)) \times 1000 = \frac{1}{2}(1 - .6961) \times 1000 = 152 ;$$

a result which is wider than the former result from the true figure 147 by only an unit. For the number below  $(64.5)^2$  there is found by parity a result not much worse than that which was obtained from the primary observations.

I have applied a similar test to the group which is formed by *cubing* the original observations. The results of both verifications are embodied in the annexed Table \* :—

		Below 64.5 in.	Below 65.5 in.	Above 70.5 in.	Above 71.5 in.
Observed.....		72	147	184	104
Calculated.	From the original observations .....	79	151	190	104
	From the squares thereof.....	82	152	190	102
	From the cubes .....	86	151	191	98

This verification might, I think, have been predicted from the circumstance that the Arithmetic means of the squares and cubes differ by very little from the respective medians,  $68.2^2$  and  $68.2^3$  (the square root of the one Arithmetic mean being 68.2, the cube root of the other 68.3, each correct to the first decimal). Now the distortion to be apprehended is the unsymmetrical extension of the upper, and shrinking of the lower limb; but this cannot be considerable, while the Median and Arithmetic mean are nearly coincident.

Accordingly we may add to the verifications above recorded other instances in which the consilience between the Median and Arithmetic mean is preserved. Thus in the case of observations on the height of adult males, recorded by the Anthropometric Committee of the British Association (Report of the Brit. Assoc. 1883), the Arithmetic mean of the primary observations (expressed as *per-milles*, and upon the under-

\* I have to thank Mrs. Bryant, D.Sc., of the North London Collegiate School, for having worked the greater part of the arithmetical examples in this and the preceding paper (Phil. Mag. Aug. 1892).



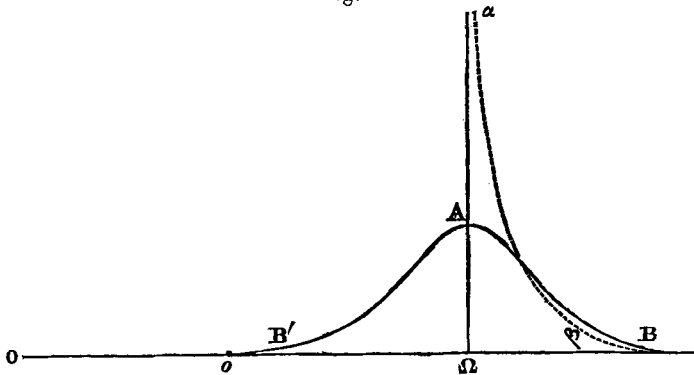
standing \* that each number of men, *e. g.* 155, entered against a certain height, *e. g.* 67 inches, means that there were 155 men between 66·5 and 67·5 inches) is 67·023. The Arithmetic mean of the squares is 4498·671, very nearly the square of the primary Arithmetic mean, viz. 4492·03. Also the Arithmetic mean of the cubes is not far from the cube of the Arithmetic mean.

Similar verifications would no doubt be obtained, if we employed, for  $F$ , any other simple function, *e. g.*  $c \log \Xi$  or  $c e^{-\Xi}$ . The limits within which the rule may be expected to hold will appear, if we consider an exception. Let

$$\Xi = (x - X), \text{ and } F(X) = (x - X)^2 = \xi^2$$

(employing the same notation as on p. 431). If the varying values of  $\xi$  range under a probability-curve, then the squares of these measurements will *not* range under such a curve. The operation of squaring will cause the negative limb of the

Fig. 1.



original curve to be screwed round to the positive side ; and, in addition to this displacement, there will be the distortion caused by substituting, for the error-function  $\frac{1}{\sqrt{\pi c}} e^{-\frac{x^2}{c^2}}$ , the

function  $\frac{1}{2\sqrt{x}\sqrt{\pi c}} e^{-\frac{x}{c^2}}$  †.

\* A misunderstanding: for as I have learnt, since the above was printed, the entry 155 men against 66 inches means that there were 155 men *above* 66 inches ; but our argument is not affected by pushing up the whole set of measurements *en bloc* half an inch higher.

† In general substituting for  $x$ , in the error-function,  $\phi_{-1}(x)$ , and multiplying by  $\frac{d}{dx} \phi_{-1}(x)$  ; where  $\phi(x)$  is the new value of an observation which originally measured  $x$ —or rather something between  $x$  and  $x+dx$ .

This dislocation is exhibited in the annexed diagram (fig. 1), where B'AB is the original curve,  $\alpha\beta$  what it becomes when each of the observations (measured from the centre  $\Omega$ ) is squared.

If for  $F(\Xi)$  be put the *cube* or any odd power of  $(x-X)$ , then the original probability-curve under which the values of  $x$  (and  $x-X$ ) ranged will be transformed to a curve symmetrical on both sides of  $X$ , but not a probability-curve.

These exceptions arise when  $F$  is a function of the *deviation* of an observation from its average value. The rule is fulfilled when  $F$  is a simple continuous function of the observations themselves measured, as is usual with concrete quantities, from an origin below the least possible value—as we measure human heights, or death-rates, or other statistics, from zero; and not from  $67\frac{1}{2}$  inches, or  $\cdot 2$  per cent., or whatever the average may be. If any such function of an observation is substituted for the observation itself in a group obeying the law of error, we may expect that the transformed group will also obey that law.

We have here the explanation of incidents which must have puzzled many students of Probabilities: why Mr. Galton should have found the Arithmetic and Geometric means of observations to give sensibly identical results (Proc. Roy. Soc. 1886); how Quetelet could be justified in affirming that *both* weights and heights of men obey the law of error (*Anthropométrie*); supposing, as is plausible, that the weight of a man is apt to be proportioned to the square of his height.

In fine we have here an answer to the objection which has been made to Quetelet's doctrine of the Mean Man by Cournot and other high authorities on Statistics. The objection is thus stated by the eminent Prof. Westergaard (*Theorie der Statistik*, p. 189):—

“Suppose we had measured for a number of men three lines of the body which make a right-angled triangle, and we wished to determine the corresponding triangle of the Average-man. Then it may be shown [*es zeigt sich dann so fort*] that the three averages do not make a right-angled triangle. Call the sides  $a_1, b_1, c_1, a_2, b_2, c_2$  &c.,  $c_1, c_2$  &c. being the hypotenuses; then the sides of the average-triangle are

$$\frac{1}{n} \Sigma a, \quad \frac{1}{n} \Sigma b, \quad \text{and} \quad \frac{1}{n} \Sigma c = \frac{1}{n} \Sigma \sqrt{a^2 + b^2}.$$

One should have accordingly

$$(\Sigma a)^2 + (\Sigma b)^2 = (\Sigma \sqrt{a^2 + b^2})^2.$$

But this can only occur—except by accident—[*in der Regel nur*  
*Phil. Mag.* S. 5. Vol. 34. No. 210. Nov. 1892. 2 H

dann zutreffen würde] when the quantities  $a, b, c$ , are proportional to  $a_2, b_2, c_2$  &c.; which in general is not the case"\*.

Suppose that the two sets of measurements which we have above adduced as examples had been made not on American recruits and English adult males, but upon two organs  $a$  and  $b$ , so related to a third  $c$  that  $c = \sqrt{a^2 + b^2}$ . The values of  $a^2$ , as we have shown, vary according to a law of error; and so do those of  $b^2$ . Accordingly, by universal admission the sum  $a^2 + b^2$  will vary according to the typical law; and we have shown that in general, if a quantity varies according to this law, so also will its square root. Thus the Arithmetic mean of the observed  $c$ 's will fit the Arithmetic means of the observed  $a$ 's and  $b$ 's.

The gist of the reasoning, it will be remarked, is that the greater part of a group conforming to the law of error is apt to be packed within limits which are narrow relatively to the largest possible member of the group, and even the average member; in the symbols above used  $\xi$  (the deviation) is small relatively to  $X$  (the average); for the greater part of the group, at least, up to some percentile near the extremities. This is true by the Laplace-Poisson theory above adverted to, even in the case most unfavourable for the argument where  $X$  is measured from a point just below the least possible; as in the example given at p. 432. Even then it would be safe to treat  $(X + \xi)^2$ , or  $\sqrt{X + \xi}$ , as equal to the linear function  $X + 2\xi$ , or  $\sqrt{X} + \frac{1}{2} \frac{1}{\sqrt{X}} \xi$ , respectively. But

for natural groups the origin should perhaps be placed at some distance below the smallest possible observation. The smallest possible dwarf must be well above zero. Upon this view the smallness of the modulus in comparison with the distance of the centre from the origin becomes more decided.

As a matter of fact, the ratio of the modulus to the mean value (the order of our  $\xi \div X$ ) is found upon an average of several instances, taken from Mr. Galton's men\* and Mr. Weldon's shrimps†, to be from  $\frac{1}{10}$  to  $\frac{1}{20}$ . Mr. Galton in some authoritative observations on this topic‡ assigns for the ratio in question (in the case of human stature)  $\frac{1}{15}$ .

[To be continued.]

\* Cf. Cournot, '*Théorie des Chances*,' Ch. X, Art. 123. Morselli follows Cournot in attempting to demonstrate *a priori* the impossibility of constructing a type involving numerous means of different organs (*Metodo in Antropologia*, 1880, p. 26).

† Proc. Roy. Soc. 1888 and 1892.

‡ Phil. Mag. vol. xlix, p. 44. He gives for the ratio of the mean to the probable error, "about 30."