



## Darwin, Galton and the Statistical Enlightenment

Stephen M. Stigler

*University of Chicago, USA*

[Received November 2009]

**Summary.** On September 10th, 1885, Francis Galton ushered in a new era of Statistical Enlightenment with an address to the British Association for the Advancement of Science in Aberdeen. In the process of solving a puzzle that had lain dormant in Darwin's *Origin of Species*, Galton introduced multivariate analysis and paved the way towards modern Bayesian statistics. The background to this work is recounted, including the recognition of a failed attempt by Galton in 1877 as providing the first use of a rejection sampling algorithm for the simulation of a posterior distribution, and the first appearance of a proper Bayesian analysis for the normal distribution.

**Keywords:** C. Darwin; F. Galton; Hiram Stanley; History of statistics; Quincunx

### 1. Introduction

On the morning of Thursday, September 10th, at the 1885 meeting of the British Association for the Advancement of Science in Aberdeen, Francis Galton addressed the assembled scientists as President of Section H, Anthropology. That date may be taken as the beginning of a half-century period that I shall call the Statistical Enlightenment, a period marked by Galton's address at one end, and the publication of Fisher's book *The Design of Experiments* on the other. It was a remarkable period that encompassed the major works of Galton, Francis Edgeworth, Karl Pearson, Ronald A. Fisher, Jerzy Neyman and Egon S. Pearson.

The characterization of that period as one of Statistical Enlightenment may seem unusual, but I believe that it is apt. The term is not intended to suggest a pair of 'dark ages' on either side, since the histories of statistics (including my own) show no lack of published brilliance and widely adopted methods in the earlier period, and no one could deny the exciting explosion of statistical theory and methodology in the period since 1935. Nonetheless, the years 1885–1935 formed a distinctive epoch in the annals of statistical thought. To adapt phrases that were used to characterize the earlier European Enlightenment, the era of Statistical Enlightenment produced new understandings that fundamentally changed the way in which people thought, that brought a synthesis based on reason to a broad array of methods, gaining wide assent and leading to revolutionary changes in several sciences. It brought a unity of conception even while permitting a diversity of interpretations. I shall return to the character of these new understandings after explaining how all of this evolved (if I may use the term) from a question that Darwin (1859) overlooked in his pursuit of other goals in his magisterial *Origin of Species*.

*Address for correspondence:* Stephen M. Stigler, Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, IL 60637, USA.  
E-mail: stigler@uchicago.edu

## 2. The *Origin of Species*

Most readers of Darwin's *Origin of Species* come to the book with an eye expectantly looking for his understanding of the idea of evolution by natural selection, and they find what they are looking for. To a statistician it can present a different picture. To a statistical eye, the structure of the argument is seen to be keyed to some interesting statistical ideas, and these ideas themselves were both basic to Darwin's own purpose, and pose questions that were peripheral to his goals. This may come as a surprise. After all, Darwin loved data but was famously not enamoured of the higher reaches of our science, and Karl Pearson chose to adorn the front of his journal the *Annals of Eugenics* (which was founded in 1925) with Darwin's epigram 'I have no Faith in anything short of actual measurement and the rule of three'. Still, from one point of view the first chapters of Darwin (1859) were fundamentally statistical; perhaps only statistical.

Darwin's goal was clearly announced in his title: to demonstrate convincingly the origin of species by means of natural selection. If natural selection was to be the means, then the first order of business was to demonstrate that there was sufficient heritable variation in any biological population to provide material to select *from*. Accordingly, chapters 1, 2 and 5 were exclusively concerned with variation, starting with variation in domestic plants and animals. Darwin presented a wealth of information on dogs, pigeons, fruit and flowers. By starting with domestic populations he could exploit his readers' knowledge of widespread experience in selective breeding and horticulture to improve the breed or the crop. The substantial variation in material was convincingly argued and, what's more, the variations that he presented were demonstrably heritable.

Many statistical ideas were present in this discussion, implicitly if not in recognizable modern form. These included comparison of within- and between-species variation, and correlation, using that name to describe what we might now call 'linkage', as in his statement 'Some instances of correlation are quite whimsical; thus cats with blue eyes are invariably deaf' (Darwin (1859), first edition; by the fourth edition the claim had become 'cats which are entirely white and have blue eyes are generally deaf'). In all this, the information was not in quantitative form, but if ever the plural of anecdote was data it was so with Darwin.

There was one point, however, that Darwin and most of his early readers overlooked; it was a consequence of Darwin's argument that I shall call Galton's puzzle. Consider this question: if fresh variability is introduced with each generation, would not there have to be a limit to its accumulation, lest the variation in the population increase indefinitely? In that form the issue was raised in an early anonymously published critique of Darwin (1859) by a British engineer, Fleeming Jenkin (Jenkin (1867), which was reprinted in Colvin and Ewing (1887), with a memoir by Robert Louis Stevenson). In the 1870s Galton was to put a different spin to it. He asked, how could intergenerational variation ever be reconciled with the approximate stability that we find in many populations' dispersion over short time periods?

Galton's interest in Darwinian questions dated from the 1860s, when he published his book *Hereditary Genius*, a study of the inheritance of intellectual ability (Galton, 1869). That book generated little enthusiasm at the time, but Galton persevered and in the 1870s pursued a study of the inheritance of physical traits, such as stature in humans and size in plants. Those studies reached a critical point in 1877.

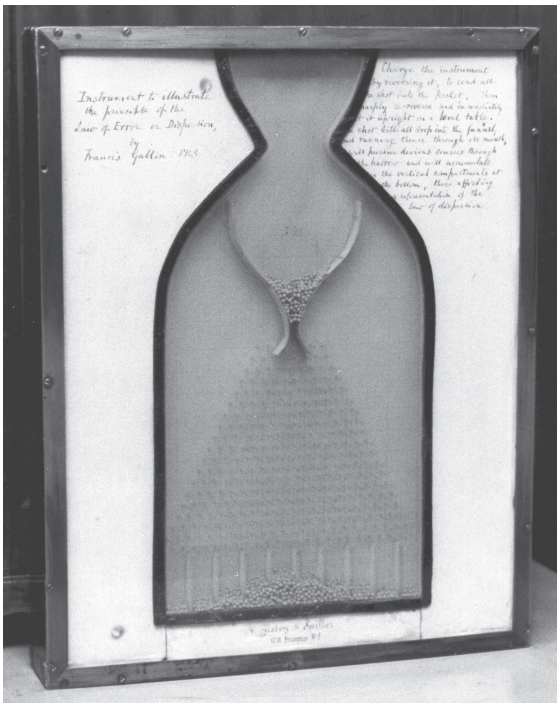
## 3. Francis Galton, February 9th, 1877

On Friday, February 9th, 1877, Francis Galton delivered an ambitious lecture at the Royal Institution in London. The title he gave was 'Typical laws of heredity', and his clear goal was

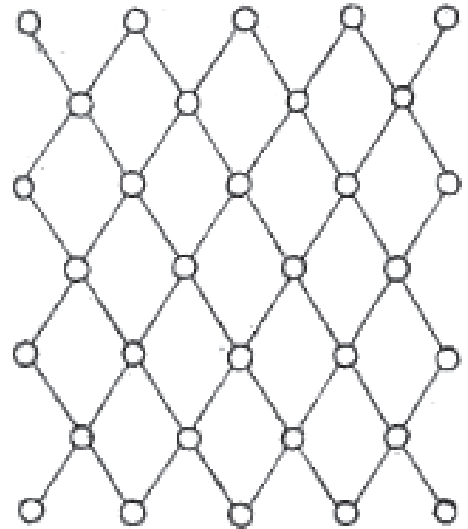
to give a mathematical formulation for at least a part of Darwin's theory—that part relating to heredity under ordinary or equilibrium conditions (Galton (1877a), also published in Galton (1877b)). He was particularly keen to reconcile two aspects of Darwin's theory that seemed in conflict. On the one hand, there was intergenerational variation—this was absolutely crucial for Darwin; if offspring were always identical to parents, then evolution by natural selection was not possible. But, on the other hand, there was also intergenerational stability—all experience under fairly constant environmental conditions showed that the range of variability on short timescales, as between two generations, was essentially constant. Indeed, the very possibility of *defining* species depended on this stability; if all is constantly in flux in a major way, classification is impossible. That these two requirements were in potential conflict was behind the criticism of Darwin by Fleeming Jenkin, but the crisp articulation in this way was due to Galton. I call this Galton's *puzzle*, since Galton did not view it as a criticism of Darwin (although others might have); rather it was a challenge to be overcome.

Galton's argument took two points of departure. The first was the quincunx, the second was the notion of 'reversion'. The quincunx was the machine that he had conceived and built in 1873, and had used to illustrate a previous lecture at the Royal Institution, in 1874. The original version from that lecture is in the Galton collection at University College London (Fig. 1).

The quincunx was used to model intergenerational variation: lead shot are dropped from the top and cascade randomly through rows of alternating offset pins, landing in compartments at the bottom as a binomial or approximately normal distribution. The name that he gave it



(a)



*Quid Quincunce Speciosius, qui, in  
quam cunq; partem Spectaveris  
rectus est. Quintilian: 7*

(b)

**Fig. 1.** (a) Galton's 1873 version of the quincunx, with 17 rows of alternatingly offset pins and Galton's pellets of lead shot (described in more detail in Stigler (1986a), page 277), and (b) the version of the same pattern (termed 'Quincunce' there) in Browne (1658)

had been used in 1658 by Thomas Browne to describe a pattern like that of the pins, which is a pattern found in the planting of fruit trees in an ancient Persian garden (Browne, 1658); the term was also in use in the 1800s to describe the mesh pattern of some fishermen's nets (Bathurst, 1837). For Galton, the quincunx clearly demonstrated the potential problem of increasing dispersion in even a single generation; that to maintain population stability would require some counteracting force. This is where reversion would enter the picture.

The idea of reversion was not new with Galton; it had been a familiar observation of breeders and horticulturalists, who found that there was a recurring tendency for offspring of selected plants or animals to revert towards past conditions, and in the *Origin of Species* Darwin had discussed reversion (using that term) as a problem to be overcome in the creation of new species. Darwin's worry was that naturally selected individuals could revert to their original form, weakening the usefulness of his data on domestic plants and animals, or even preventing the establishment of a new species. Darwin recognized that reversion was a real phenomenon, and he even dedicated a full chapter to it in his 1868 book on variation (Darwin (1868), chapter 14). But he thought that the threat of this tendency was 'exaggerated' by some critics; that it was insufficient as a force to interfere with the effects of natural selection.

What was new with Galton came from an experiment that he had performed before his lecture, where he found that, in sweet peas, the reversion of size of pea between generations had two marked characteristics: it was *linear* in the distance from the population mean, and the variation of offspring of selected groups of parents was *constant*—the dispersion in size of offspring was the same whether the parental group was near or far from the population mean.

In his 1877 lecture Galton proposed that reversion was the solution to the puzzle: it was the counteracting force that prevented increasing dispersion on an intergenerational timescale. He offered a new version of the quincunx as illustration (Fig. 2).

The population distribution in the previous generation is at the top layer, the pellets then pass down 'inclined shoots' (chutes) (these represented reversion), and then their passage through the pins displays 'family variability', producing at the bottom a population distribution composed of offspring of different types (the small 'hillocks'), in the aggregate identical to the one at the top—*voilà*, stability. Galton's puzzle was evidently solved—or was it? At first glance this looks suspect; what on Earth are those inclined chutes—what is the mechanism behind them? And why should we expect an *exact* cancellation of effects? It had the appearance of a 'just-so' story, an *ad hoc* unverifiable hypothesis that guarantees the conclusion. Actually Galton gave arguments to meet both points: why there were chutes and why there was exact cancellation. They were clever arguments, even if they were ultimately unsatisfactory.

To produce the inclined chutes, Galton offered nothing less than Darwin's own mechanism of natural selection and survival of the fittest. The more extreme parents, Galton suggested, would produce fewer offspring (lower productivity) and the more extreme offspring would have a lower survival rate (natural selection). These actions together could produce precisely the effect shown! He offered a mathematical argument and illustrated it with a third quincunx to show specifically how reversion occurred (Fig. 3).

In this example, the top level pellets fall through a 'natural selection' screen shaped like a normal curve; those that pass between the vertical curved screen and the viewer form below a more compact distribution that is also precisely a normal distribution, whereas those that fall behind the screen are rejected. I propose to call this *Galton's theorem* and I shall return to it later, giving Galton's proof in Appendix A. Galton's final model actually had this device operating twice: once as a 'productivity' screen and once as a 'natural selection' screen. But why are the screens precisely normal, and why are their standard deviations such as would exactly balance the

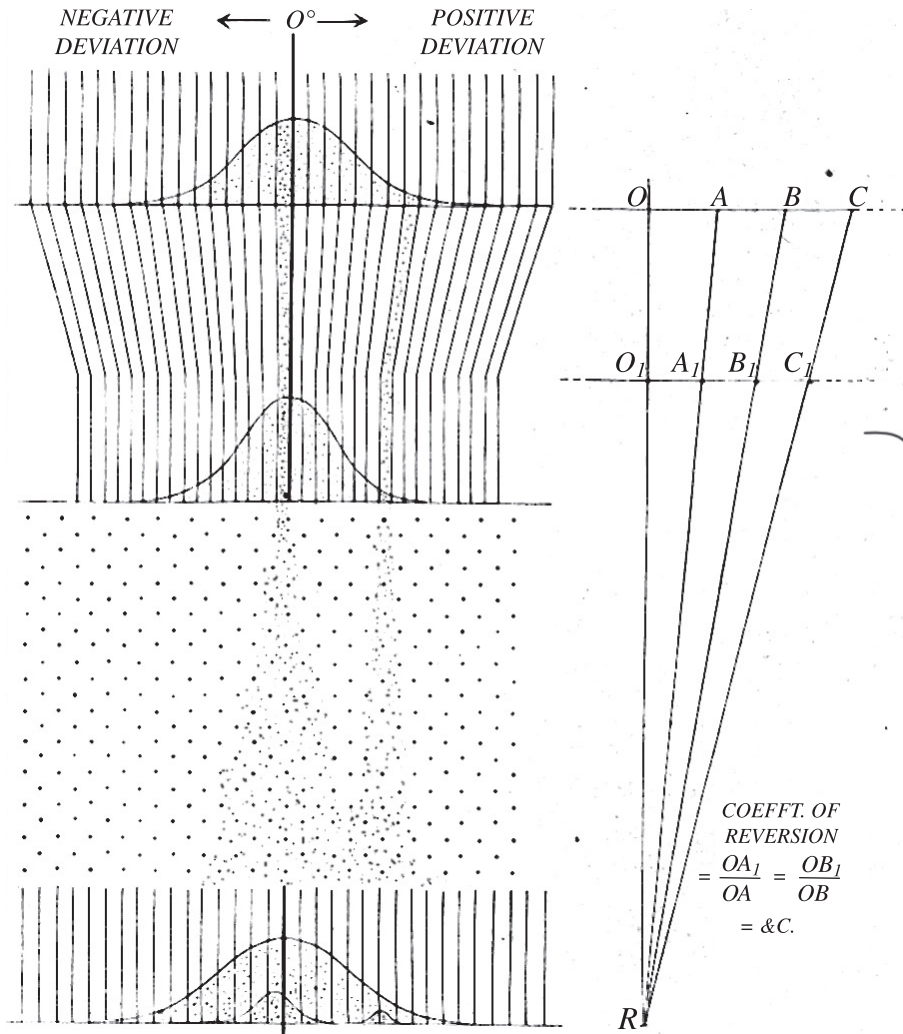
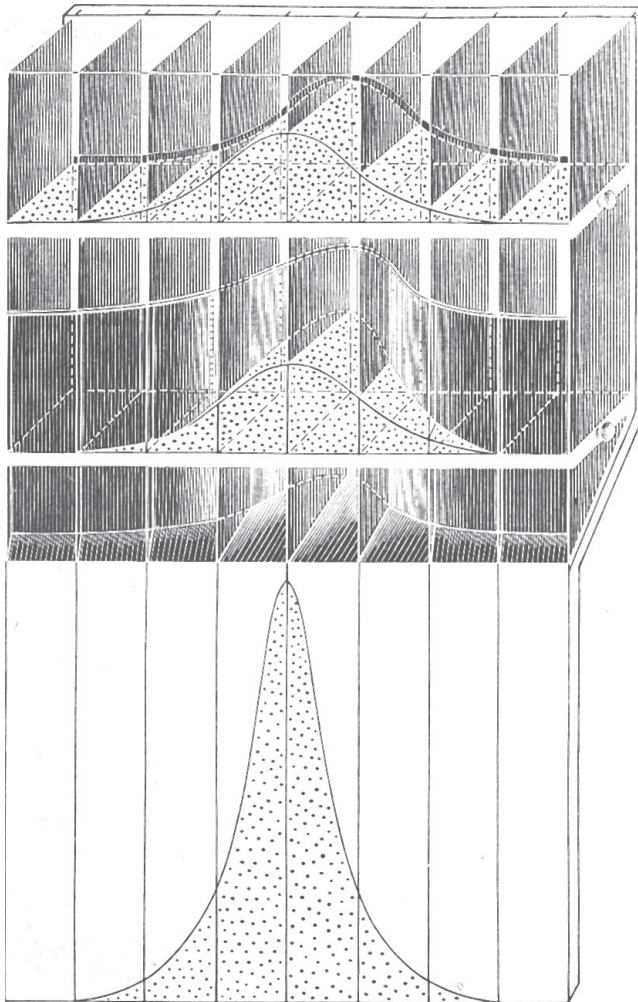


Fig. 2. Galton's 1877 version of the quincunx (Galton, 1877a,b)

dispersion between the generations? To the first of these questions, the normal distribution, he gave a speculative explanation: essentially, the net effect was an aggregation of small variations, and thus nature mimicked an error process in its screening; a process such as would produce normally distributed observations in astronomy. To the second question, the exact balance between reversion and family variability, Galton slightly more convincingly invoked Darwinian evolution: the model was describing a process in equilibrium after a long period of evolution, so of course it would be in balance. In an appendix he summarized all this mathematically, describing exactly what the relationship between the reversion coefficient  $r$  (which is proportional to the slope of the chutes to the perpendicular) and the variances (population  $c^2$ , productivity, natural selection and in-family variability  $v^2$ ) would have to be to produce stability. In the case of 'simple descent', with productivity and natural selection operating uniformly (and hence ignorable), this gave the relationship  $c^2 = v^2/(1 - r^2)$  and thus gave the population variance as  $c^2 = r^2 c^2 + v^2$ , or



**Fig. 3.** Galton's upper stage for his third quincunx, demonstrating the working of natural selection, from his 1877 address: note that the curved screen at the second level is vertical; below that screen the accepted pellets then flow towards the front, to a thin compartment of uniform thickness; the text indicates that he had a working model at the lecture, but that has not survived (Galton, 1877a, b)

population variation = variation of reverted parent means + within-family variation.

Galton's lecture was remarkable; it had encompassed a part of most of Darwin's ideas in a single mathematical structure, and it seemed to have solved a puzzle that could have been interpreted as troubling for Darwin's theory. But it would not have taken much reflection for Galton to realize that it just did not work. His full model (with productivity and natural selection) required a much stronger degree of intergenerational natural selection than was plausible. Also, the case for the particular (normal) form of the screen was very weak, and Galton's own empirical finding of a constant within-family variance seemed to contradict the hypothesis of more severe screening in the extremes. And indeed Galton soon went back to the drawing board to rethink the case. It was what he found in this reconsideration that would launch the Statistical Enlightenment in 1885.



#### 4. Francis Galton, September 10th, 1885

In the late 1800s, the annual meetings of the British Association for the Advancement of Science were *the* major national scientific events of the year. By the 1880s the British Association for the Advancement of Science had over 4000 members, with Sections devoted to mathematical and physical science, chemical science, geology, biology, geography, economic science and statistics, mechanical science and anthropology. The meetings were grand affairs—the 1892 meeting in Edinburgh was preceded with the circulation of an *Excursion Handbook* of 168 pages describing all manner of social events and side trips for members and accompanying parties. Each meeting would be followed by publication of a *Report* in the form of a four-inch-thick volume with committee reports, abstracts and full papers.

The 1885 meeting in Aberdeen lasted from September 9th to 16th. The attendance totalled 2203, including 697 ‘Members’, 1053 ‘Associates’, 447 ‘Ladies’ and six ‘Foreigners’. The programme featured 49 *Reports on the State of Science*, seven addresses and 380 transactions, including several by leading British scientists (Lord Rayleigh, William Thomson (the future Lord Kelvin), Oliver Lodge and D’Arcy Wentworth Thompson). There were papers ranging from the latest theories of electricity, to the intelligence of the dog (describing an experiment that taught a dog to read), to the history of the game of hopscotch (which, notwithstanding its modern name, was invented in the pre-Christian era, and not in Scotland) (British Association for the Advancement of Science, 1886).

Galton was present in several official roles, including as President of Section H, Anthropology, which was then in its second year. He gave his Presidential address on September 10th; it was published 2 weeks later both in *Nature* and in *Science* (Galton, 1885a, b), and in British Association for the Advancement of Science (1886), pages 1206–1214, and in expanded form with figures in Galton (1886b). In November 1885 and January 1886 he published further developments: one with more technical detail titled ‘Regression towards mediocrity in hereditary stature’ (Galton, 1885c), and one with slight refinements, ‘Hereditary stature’ (Galton, 1886a). The fullest technical exposition, with an appendix by J. Hamilton Dickson (who helped a little with the mathematics at a late stage) was reserved for another paper in the *Proceedings of the Royal Society*, ‘Family likeness in stature’ (Galton, 1886c). All of this work together constituted a remarkable extension of his 1877 address; one that both revised the approach in a really fundamental way and also silently retracted a significant part of the 1877 analytical framework.

What was new and revolutionary was Galton’s single most important and best known contribution to statistics: he introduced there the idea of thinking of the two generations as a bivariate normal pair, with two different and conceptually distinct lines of conditional expectation, the ‘regression’ lines, to use the term that he introduced here in place of reversion. In a real sense Galton had invented multivariate analysis. What was silently missing was Darwin and all that was Darwinian—the inclined chutes were gone and with them any substantive mention of survival of the fittest or natural selection. Galton stated that he could now ‘get rid of all these complications’. In supreme irony, in what had started out as an attempt to mathematize the framework of the *Origin of Species* ended with the essence of that great work being discarded as unnecessary!

Galton’s moment of epiphany had evidently come from considering a new data set that he had generated, data on the stature of two generations of a large number of British families, and from a new way of thinking about his analytical tool, the quincunx. To begin with, he had constructed a two-way table of the counts for pairs of adult children’s heights and the average heights of their parents. This he could fit in with his 1877 framework, with the right-hand marginal total column being the frequency distribution at the top (the previous generation) and the bottom

TABLE I.

NUMBER OF ADULT CHILDREN OF VARIOUS STATURES BORN OF 205 MID-PARENTS OF VARIOUS STATURES.  
(All Female heights have been multiplied by 1.08).

Heights of the Mid-parents in inches.	Heights of the Adult Children.														Total Number of		Medians.
	Below	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	Above	Adult Children.	Mid-parents.	
Above ..	..	..	..	..	..	..	..	..	..	..	..	1	3	..	4	5	..
72.5	..	..	..	..	..	..	..	1	2	1	2	7	2	4	19	6	72.2
71.5	..	..	..	..	1	3	4	3	5	10	4	9	2	2	43	11	69.9
70.5	1	..	1	..	1	1	3	12	18	14	7	4	3	3	68	22	69.5
69.5	..	..	1	16	4	17	27	20	33	25	20	11	4	5	183	41	68.9
68.5	1	..	7	11	16	25	31	34	48	21	18	4	3	..	219	49	68.2
67.5	..	3	5	14	15	36	38	28	38	19	11	4	..	..	211	33	67.6
66.5	..	3	3	5	2	17	17	14	13	4	..	..	..	..	78	20	67.2
65.5	1	..	9	5	7	11	11	7	7	5	2	1	..	..	66	12	66.7
64.5	1	1	4	4	1	5	5	..	2	..	..	..	..	..	23	5	65.8
Below ..	1	..	2	4	1	2	2	1	1	..	..	..	..	..	14	1	..
Totals ..	5	7	32	59	48	117	138	120	167	99	64	41	17	14	928	205	..
Medians ..	..	..	66.3	67.8	67.9	67.7	67.9	68.3	68.5	69.0	69.0	70.0	..	..	..	..	..

NOTE.—In calculating the Medians, the entries have been taken as referring to the middle of the squares in which they stand. The reason why the headings run 62.2, 63.2, &c., instead of 62.5, 63.5, &c., is that the observations are unequally distributed between 62 and 63, 63 and 64, &c., there being a strong bias in favour of integral inches. After careful consideration, I concluded that the headings, as adopted, best satisfied the conditions. This inequality was not apparent in the case of the Mid-parents.

(a)

Table IV (R.F.F. Data).

Relative number of Brothers of various Heights to Men of various Heights, Families of Six Brothers and upwards being excluded.

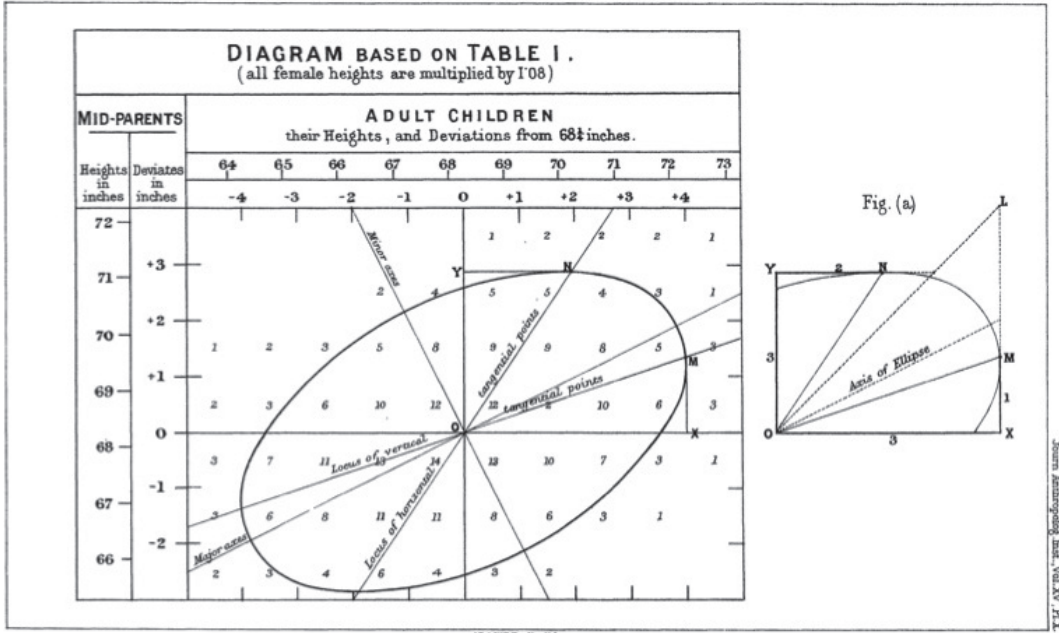
Heights of the men in inches.	Heights of their brothers in inches.														Total cases.	Medians.
	Below 61.7	62.2	63.2	64.2	65.2	66.2	67.2	68.2	69.2	70.2	71.2	72.2	73.2	Above 73.7		
Above 73.7 ....	..	..	..	..	1	..	1	..	1	4	3	3	3	2	18	
73.2 ....	1	..	..	..	..	..	1	1	2	1	3	4	..	3	16	
72.2 ....	1	..	1	2	1	1	..	8	6	8	11	5	4	3	51	70.3
71.2 ....	..	..	..	4	4	4	9	11	15	12	8	11	3	3	84	69.3
70.2 ....	1	..	2	4	3	7	6	12	25	18	11	8	1	3	101	69.3
69.2 ....	..	..	4	6	13	12	18	29	29	24	15	6	2	1	159	68.6
68.2 ....	1	..	..	3	6	7	15	16	29	12	11	8	1	..	109	69.9
67.2 ....	1	..	4	3	8	14	21	15	19	6	9	..	1	1	102	67.7
66.2 ....	..	..	1	7	10	12	14	7	12	7	4	1	..	..	75	67.2
65.2 ....	..	1	1	4	13	9	8	6	13	3	4	1	..	1	64	67.2
64.2 ....	..	1	..	6	4	7	3	3	6	4	4	2	..	..	40	67.3
63.2 ....	..	..	..	..	1	1	4	..	4	2	..	1	..	..	13	
62.2 ....	..	..	..	..	1	..	..	..	..	..	..	..	..	..	1	
Below 61.7 ....	..	..	..	..	..	..	1	1	..	1	..	1	1	..	5	
..	5	2	13	39	65	74	101	109	161	102	83	51	16	17	838	

(b)

Fig. 4. (a), (b) Galton's tables and (c) the bivariate normal contour derived from the first of them (Galton, 1885c)



Plate X.



J.P. & W. Emble. Lith.

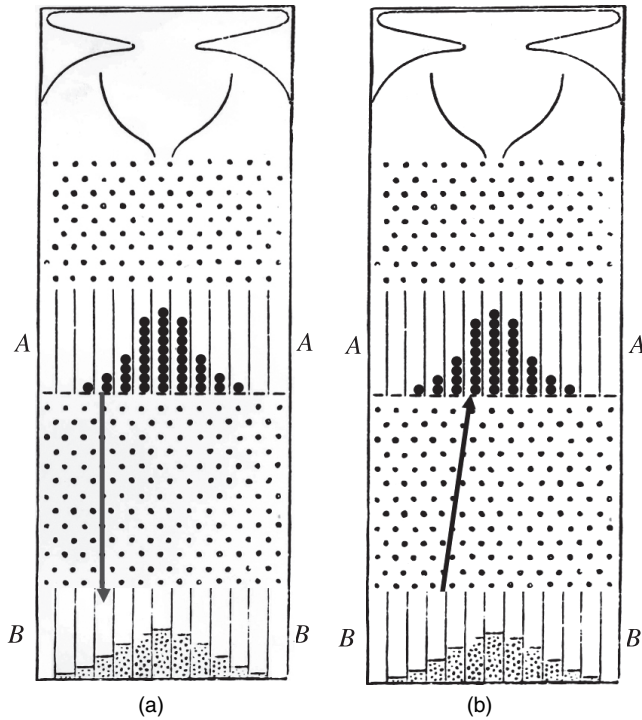
(c)

Fig. 4. (continued)

marginal total row being the frequency distribution at the bottom of the quincunx (the following generation), and the separate row counts representing the progeny of the selected family groups (these were the little hillocks of 1877, which in aggregate would comprise the population, two of which are shown at the bottom level of Fig. 2). But, when he constructed a similar table of the heights of pairs of brothers, almost exactly the same in form, it was clear that this relationship was not a question of descent. These tables together led directly to his famous construction of the bivariate normal distribution with its lines of conditional expectation (Fig. 4).

Galton could tell an implausible story about natural selection at work between parent and child; he could not do the same between brother and brother! The symmetry in the fraternal relationship prevented such a Darwinian excuse and had led him to think about running his quincunx backwards. Fig. 5 illustrates the way that he presented it in his book *Natural Inheritance* (Galton, 1889). When the pellets in an upper compartment are released, their average final position is directly below. But what if we ask of a compartment at the lower level, from where did these pellets come? The answer was not 'on average, directly above'. Rather, it was 'on average, more towards the middle', for the simple reason that there were more pellets above it towards the middle that could wander left than there were in the left extreme that could wander to the right, inwards (Fig. 5).

Galton's great insight from this new approach was that stability implied reversion or, as he now called it, regression. He no longer needed the *ad hoc* arguments of 1877, when he claimed that a particular amount of reversion would fortuitously imply stability. With this reversal of stance the argument based on equilibrium could now carry the entire weight. He could, and did, introduce interclass correlation as a model for variability, but now the entire puzzle was resolved by the one fundamental insight. (Stigler (1986a), chapter 8, discusses many other statistical insights that Galton derived from the quincunx.)



**Fig. 5.** Depiction of the regression phenomenon, based on a figure from Galton (1889), page 63, showing (a) the expected landing position of pellets from a specific upper level compartment and (b) the expected point of origin of pellets landing in a specific lower level compartment

There was a hint in the lecture that Galton was moving towards a mechanism for inheritance in his thinking. He seemed to approach Mendelian genetics when he speculated (Galton, 1885a, b, 1886a, b)

‘There can be no doubt that heredity proceeds to a considerable extent, perhaps principally, in a piecemeal or piebald fashion, causing the person of the child to be to that extent a mosaic of independent ancestral heritages, one part coming with more or less variation from this progenitor, and another from that. To express this aspect of inheritance, where particle proceeds from particle, we may conveniently describe it as “particulate”.’

But this and other such statements in his work were never developed; we can only wonder what he would have made of Mendel’s work if he had encountered it.

### 5. Reception to the address

Galton’s lecture of September 10th was well received. The September 19th, 1885, issue of *The Lancet* described it as ‘one of the most interesting of the year’ and captured the gist of his idea, writing

‘Galton holds that the number of individuals in a population who differ little from mediocrity is so preponderant, that it is more frequently the case that an exceptional man is the somewhat exceptional son of rather mediocre parents than the average son of very exceptional parents’

(The Lancet (1885), page 538). Alfred Russel Wallace wrote to Galton, ‘I was delighted with your address at the Brit. Ass. On Hereditary Stature’ (March 7th, 1886).

But not all who encountered this work were so favourable in their judgements. In 1889 when Galton published his definitive account of this work in the book *Natural Inheritance*, an astute

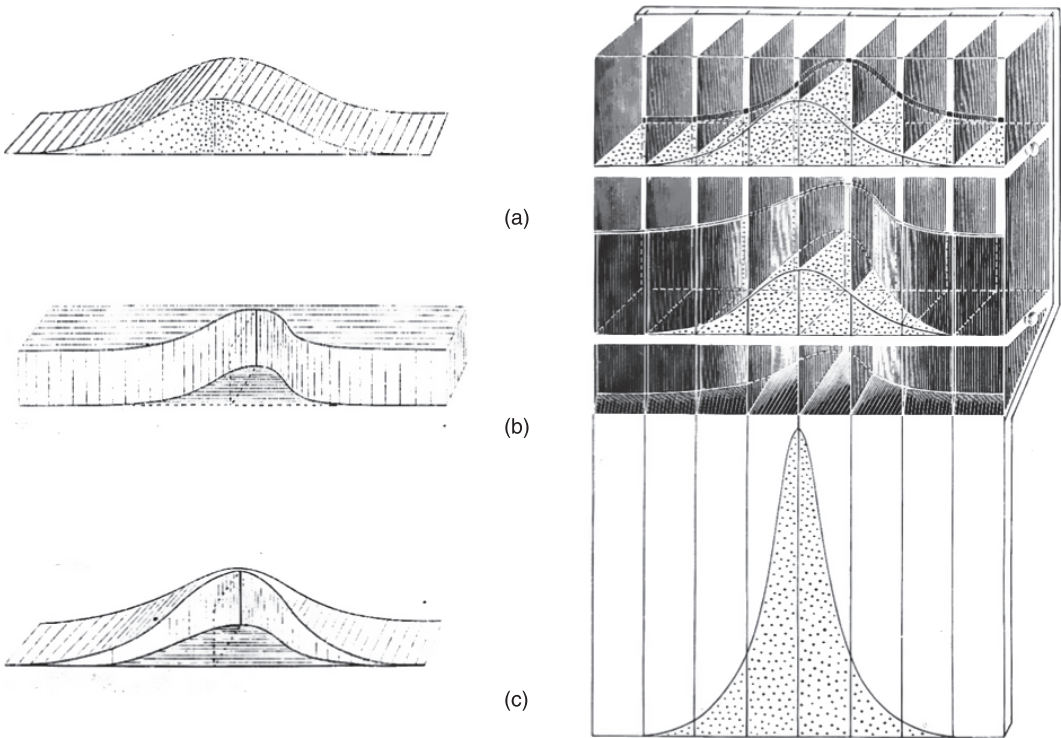
reader in Chicago wrote to *Nature* with a telling criticism. That reader was Hiram M. Stanley, a philosopher and psychologist at Lake Forest University. Stanley pointed out that in fact the tie between Galton's theoretical structure and heritability was tenuous, and even unsupported. Galton's structure, Stanley noted, was quite general and would apply equally well if the entire source of transmission was environmental. In effect, Stanley argued that heredity and environment were inextricably confounded in Galton's data (Stanley, 1889). Galton himself recognized no such doubts, and I think never replied to Stanley or his criticism. Ironically, Stanley was correct at the time, but it was exactly the generality of Galton's structure that R. A. Fisher was to exploit in 1918 in what would become the fundamental mathematical basis for modern Mendelian genetics. This was not Stanley's only early attempt to reign in scientific speculation from Chicago; 3 years later he attempted to deflate the hoopla about there being canals on Mars in a letter to *Science* (Stanley, 1892).

Among those at the Aberdeen meeting was Francis Ysidro Edgeworth. 2 days after Galton spoke, Edgeworth presented a paper on the estimation of variance components for a two-way additive effects model, work that attracted little attention even when published in the *Journal of the Statistical Society of London* later that year (Edgeworth (1885) (with an abstract in British Association for the Advancement of Science (1886)) and Stigler (1999), chapter 5). We do not know whether Edgeworth attended Galton's talk, but it surely caught his attention. In 1892–1893 Edgeworth published a series of papers based on a deep investigation of Galton's structure; they included Edgeworth's discovery of the relationship between the coefficients in the quadratic form in the multivariate normal density and the inverse of the covariance matrix. This work caught Karl Pearson's eye; in 1895 Pearson expanded on it, including deriving the product moment estimate of the correlation coefficient. Galton's advance of 1885 was in broad circulation within a decade after his address (Stigler (1986a), chapters 8–10).

## 6. The Statistical Enlightenment

The half-century after 1885 saw a remarkable series of developments in statistical theory and methodology—truly the building of the foundation of our modern science. Much of this was associated with Karl Pearson, Ronald A. Fisher and Jerzy Neyman and Egon Pearson. Many areas were so little represented in earlier work that they could be described as entirely new: the analysis of correlation and association, the identification and separation of causes in multiply classified data, the validation of this through testing and the design of experiments as a foundation for inference. Old subjects such as Bayesian inference that had survived without important development were reborn in the new light. It would be foolish to attribute all this to Galton's influence; it would be equally foolish to deny the importance of that influence.

The idea of regression was what most interested Galton and his audience, but the enlightenment that I have spoken of that followed was based on this general framework that Galton had invented—the multivariate analysis he gave in terms of different conditional distributions for the bivariate normal distribution. It is true that if you dig deeply in earlier literature you can find instances of multivariate normal densities: Robert Adrain in 1808, Laplace about 1812, French work on artillery fire in the 1820s, the crystallographer Auguste Bravais in 1846 and Erastus De Forest in the 1870s, among these. But all of these are as simple generalizations of univariate frequency functions and none examined or exploited the conditional distributions—the multivariate structure—as Galton did. Indeed, there seems to be no one before Galton who even asked what the conditional distributions were in this setting, and much less found them and commented on the deep statistical messages that they implied.



**Fig. 6.** Galton's rejection sampling algorithm for simulating a posterior distribution: (a) prior; (b) likelihood; (c) posterior

Galton had conceived of pairs of generations or other pairs of quantities as a true multivariate statistical object that could be sliced and diced, and examined both marginally and conditionally from any point of view, and this had a signal effect on the theory and practice of inference. This was a radically new statistical perspective, and it gave us a new type of question to ask, and a new way to think about statistical association: and, more fundamentally, a new way to think about inference.

Earlier statistical scientists—Laplace and Cournot being important examples—had been acutely aware that ‘before data’ and ‘after data’ represented different conceptual views, and they developed these views in important ways. Laplace distinguished sampling distributions from posterior distributions and gave elegant asymptotic approximations for both. But in their work, and with that of Gauss and others, the prior distributions were, as with Bayes, uniform, and generally improper. Andrew Dale’s extensive scholarly history of ‘inverse probability’ (Dale, 1999), most of it covering work before Galton, may be read as showing how barren the Bayesian cupboard was in that era as far as statistical inference was concerned: before Galton, Bayesian statistics consisted almost entirely of the use of flat priors to deduce what in another age would be called maximum likelihood estimates. Conditional or relative probabilities have a long and vigorous history, but awareness of conditional distributions and their role in inference does not. This led in some instances to confusion and error, e.g. by Laplace in 1774 (Stigler, 1986b). Indeed, before the 20th century there was no generally accepted notation for conditional probability, and much less for conditional distributions; our most common notation (the vertical bar) evidently dates only from 1931 (Jeffreys (1931), page 15, and Shafer and Vovk (2006)).

Take another look at Galton's discarded 1877 model for natural selection (Fig. 6). It is nothing less than a workable simulation algorithm for taking a normal prior (the top level) and a normal likelihood (the natural selection vertical screen) and finding a normal posterior (the lower level, including the rescaling as a probability density with the thin front compartment of uniform thickness). As far as I know, this is the first appearance of this calculation in statistics, and Galton did give a proof in an appendix in 1877. In 1885 he took this much further and for the first time gave the full calculus of the bivariate normal distribution, with all marginal and conditional distributions. In one sense Bayesian inference preceded Galton by two centuries; in another sense it hardly existed at all until Galton's framework had been digested and developed by others.

Even if he would only have contributed this work, this framework for modern statistical analyses, Galton should be remembered as a great scientist. His cousin Charles Darwin was certainly more famous in Galton's lifetime and ever since. But statisticians could be excused if they raised the question of which cousin was ultimately more influential. Whatever the answer, September 10th should rank as a day of celebration on every statistician's calendar. We might be excused if we paraphrase Galton's cousin Charles's final sentence in the *Origin of Species* and state

'There is a grandeur in this view of statistical relations; from so simple a beginning endless forms most beautiful and wonderful have been and are being evolved'.

## Appendix A: Galton's theorem

Galton's natural selection upper stage for the quincunx of 1877 can be succinctly described as the following theorem, where the distribution of  $Y$  represents the initial population distribution,  $f(x)$  represents the 'natural selection' vertical screen and  $U$  the depth (the distance from the front plate).

*Theorem.* Suppose that  $Y$  has an  $N(0, A^2)$  distribution. Let  $f(x)$  be the  $N(0, B^2)$  density. Let  $U$  be uniformly distributed over  $(0, f(0))$ , independent of  $Y$ . Then the distribution of  $Y$  given  $U \leq f(Y)$  is  $N(0, C^2)$ , where  $1/C^2 = 1/A^2 + 1/B^2$ .

Galton's proof was simple and direct. The frequency of an initial value  $Y = y$  is proportional to  $\exp(-y^2/2A^2) dy$ . Given  $y$ , the chance of making it through to the next stage is proportional to  $f(y) = \exp(-y^2/2B^2)$ , so the frequency of  $Y = y$  at the next stage is proportional to

$$\exp(-y^2/2A^2) \exp(-y^2/2B^2) dy = \exp(-y^2/2C^2) dy.$$

When rescaled to give a relative frequency, this gives  $N(0, C^2)$ .

Galton worked in terms of population distributions, thinking literally of the pellets as a population and the vertical screen as censoring differentially depending on how extreme the values were. We can put the theorem in a more familiar modern form as follows: if the prior distribution of location parameter  $Y$  is  $N(0, A^2)$  and the distribution of  $X = y$  given  $Y = y$  (the likelihood) is  $N(0, B^2)$ , then the posterior distribution of  $Y$  given  $X = 0$  is  $N(0, C^2)$ . (To relate this to Galton's motivating example we might think of  $Y$  as stature and suppose that, given  $Y = y$ , each individual is assigned a fitness  $X$ , where  $X$  is  $N(y, B^2)$ . An equilibrium value for  $X$  would be  $X = 0$ ; given  $X = 0$ , the stature distribution would then be  $N(0, C^2)$ . But this would be an unwarranted extension of Galton's 1877 intentions.)

Galton was not thinking in explicit Bayesian terms, of course, but mathematically he has posterior  $N(0, C^2) \propto \text{prior } N(0, A^2) \times \text{likelihood } f(x = 0|y)$ . This may be the earliest appearance of this calculation; the now standard derivation of a posterior distribution in a normal setting with a proper normal prior. Galton gave the general version of this result as part of his 1885 development, but the 1877 version can be seen as an algorithm employing rejection sampling that could be used for the generation of values from a posterior distribution. If we replace  $f(x)$  above by the density  $N(a, B^2)$ , his algorithm would generate the posterior distribution of  $Y$  given  $X = a$ , namely  $N(aC^2/B^2, C^2)$ . The assumption of normality is of course needed for the particular formulae here, but as an algorithm the normality is not essential; posterior values for any prior and any location parameter likelihood could in principle be generated by extending this algorithm.

## Acknowledgements

I am grateful to Robert J. Richards, Xiao-Li Meng, Radford Neal and the Joint Editor for helpful comments.

## References

- Bathurst, C. (1837) *Notes on Nets; the Quincunx Practically Considered*. London: Van Voorst.
- British Association for the Advancement of Science (1886) *Report of the Fifty-Fifth Meeting of the British Association for the Advancement of Science; held at Aberdeen in September 1885*. London: Murray.
- Browne, T. (1658) *The Garden of Cyrus, or, the Quincunciall, Lozenge, or Net-work Plantations of the Ancients, Artificially, Naturally, Mystically Considered*. London: Brome.
- Colvin, S. and Ewing, J. A. (eds) (1887) *Papers Literary, Scientific, &c. by the Late Fleeming Jenkin, F.R.S., LL.D.*, vol. I, pp. 215–263. London: Longman, Green.
- Dale, A. I. (1999) *A History of Inverse Probability, from Thomas Bayes to Karl Pearson*, 2nd edn. New York: Springer.
- Darwin, C. R. (1859) *The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. London: Murray.
- Darwin, C. (1868) *The Variation of Animals and Plants under Domestication*, vols 1, 2. London: Murray.
- Edgeworth, F. Y. (1885) On methods of ascertaining variations in the rate of births, deaths, and marriages. *J. Statist. Soc. Lond.*, **48**, 628–649.
- Galton, F. (1869) *Hereditary Genius*. London: Macmillan.
- Galton, F. (1877a) Typical laws of heredity. *Nature*, **15**, 492–495, 512–514, 532–533.
- Galton, F. (1877b) Typical laws of heredity. *Proc. R. Inst. Gt Br.*, **8**, 282–301.
- Galton, F. (1885a) Opening address as President of the Anthropology Section of the British Association for the Advancement of Science, September 10th, 1885, at Aberdeen. *Nature*, **32**, 507–510.
- Galton, F. (1885b) Types and their inheritance. *Science*, **6**, 268–274.
- Galton, F. (1885c) Regression towards mediocrity in hereditary stature. *J. Anthr. Inst. Gt Br. Ireln*, **15**, 246–263.
- Galton, F. (1886a) Hereditary stature. *Nature*, **33**, 295–298; correction, **33**, 317.
- Galton, F. (1886b) President's Address. *J. Anthr. Inst. Gt Br. Ireln*, **15**, 489–499.
- Galton, F. (1886c) Family likeness in stature. *Proc. R. Soc. Lond.*, **40**, 42–73.
- Galton, F. (1889) *Natural Inheritance*. London: Macmillan.
- Jeffreys, H. (1931) *Scientific Inference*. Cambridge: Cambridge University Press.
- Jenkin, F. (1867) Darwin and the Origin of Species. *Nth Br. Rev.*, **44**, 277–318.
- Shafer, G. and Vovk, V. (2006) The sources of Kolmogorov's *Grundbegriffe*. *Statist. Sci.*, **21**, 70–98.
- Stanley, H. M. (1889) Mr. Galton on natural inheritance. *Nature*, **40**, 642–643.
- Stanley, H. M. (1892) On the interpretation of the markings on Mars. *Science*, **20**, 235.
- Stigler, S. M. (1986a) *The History of Statistics: the Measurement of Uncertainty before 1900*. Cambridge: Harvard University Press.
- Stigler, S. M. (1986b) Laplace's 1774 memoir on inverse probability. *Statist. Sci.*, **1**, 359–378.
- Stigler, S. M. (1999) *Statistics on the Table*. Cambridge: Harvard University Press.
- The Lancet (1885) Inheritance of typical characteristics. *Lancet*, Sept. 19th.