THE THEORY OF PROBABILITY: SOME COMMENTS ON LAPLACE'S THÉORIE ANALYTIQUE*

BY E. C. MOLINA

1. *Introduction*. Laplace's work on probability has been surrounded by a great deal of misinformation. This situation is due to the mistakes of one historian having been passed on by others who have not gone to the originals. My purpose is to point out some of these errors before they come to be regarded as unquestionable from constant reiteration. Such reiteration not only does injustice to Laplace but leads to confusion of thought in a subject wherein clear thinking is of paramount importance.

The contributions of the author of the *Mécanique Céleste* to the theory of probability and its applications began with a memoir published in 1774. Extending over a period of nearly half a century they terminated in the fourth supplement to the third edition of the *Théorie Analytique des Probabilités*; the first edition appeared in 1812.

The comments which I will submit today for your consideration have reference particularly to the mature work of Laplace as expounded in the two books which constitute the *Théorie Analytique*. Except where one of the three editions of this great classic is explicitly mentioned, the comments apply equally to the first and third editions and, presumably, to the second edition also.

In his eulogy of Laplace, Fourier said:†

"On ne peut pas affirmer qu'il lui eût été donné de créer une science entièrement nouvelle, · · · : mais Laplace était né pour tout perfectionner, pour tout approfondir, pour reculer toutes les limites, pour résoudre ce que l'on aurait pu croire insoluble. Il aurait achevé la science du ciel, si cette science pouvait être achevée.

"On retrouve ce même caractère dans ses recherches sur l'analyse des probabilités, science toute moderne, immense, dont l'objet souvent méconnu a donné lieu aux interprétations les plus fausses, mais dont les applications embrasseront un jour tout le champ des connaissances humaines."

---

* Presented to the Society, February 22, 1930, by invitation of the Program Committee.

† Fourier, *Éloge Historique de M. Le Marquis de Laplace*, prononcé dans la séance publique de l'Académie Royale des Sciences, le 15 Juin 1829.

Then, after a few words on the part played by Pascal, Fermat, Huygens, Jacques Bernoulli, Stirling, Euler, Lagrange, D'Alembert, and Condorcet in the creation and development of probability theory, Fourier summarized Laplace's contributions and the future of probability theory in these words:

"Laplace en a réuni et fixé les principes. Alors elle est devenue une science nouvelle, soumise à une seule méthode analytique, et d'une étendue prodigieuse. Féconde en applications usuelles, elle éclairera un jour d'une vive lumière toutes les branches de la philosophie naturelle."

This eulogy raises three questions which bear on the theory of probability:

(1) What are the actual contents of the *Théorie Analytique*?

(2) How are the results, methods and views therein expounded related to the present status of probability and statistical theory?

(3) How much did Laplace owe to his predecessors?

An exhaustive treatment of any one of these questions is impossible in the short space of an hour. My comments on the third will be limited to the work of Bayes; work which must be carefully differentiated from that of Laplace when one deals with the probability of causes. I will deal with the first question at least sufficiently to make my comments on the second intelligible. The second question may be reworded as follows: *to what extent will one conversant with the Théorie Analytique be in touch with the present status of probability theory, and how sound a foundation will he have found therein for statistical applications of the theory*? This question will be the keynote of the comments which I now proceed to make.

2. *Generating Functions.* The central analytical concept of the *Théorie Analytique*, introduced in Book I, Chapter I, and appearing explicitly on the last page of the last supplement of the third edition, is that of the *fonction génératrice*:

"Soit y_x une fonction quelconque de x; si l'on forme la suite infinie

$$y_0+y_1t+y_2t^2+y_3t^3+ \cdots +y_xt^x+y_{x+1}t^{x+1}+ \cdots +y_\infty t^\infty,$$

on peut toujours concevoir une fonction de t qui, développée suivant les puissances de t, donne cette suite: cette fonction est ce que je nomme *fonction génératrice de y_x*."

The development of this concept and its applications in pure mathematics, such as, for example, the interpolation and transformation of series, the solution of finite difference equations and the expression of functions in terms of definite integrals occupy the whole of Book I. Throughout Book II the concept is made use of in the solutions of probability problems.

Boole* expressed little sympathy toward generating functions and Todhunter† adopted Boole's attitude. My opinion is that these mathematicians made a mistake; the very high esteem in which their predecessor Lacroix‡ held Laplace's concept is justified by MacMahon's summary§ of the part it plays in combinatory analysis. But to appreciate the full significance of the generating function it is essential to consider two aspects of the matter with which MacMahon is not concerned. These are:

3. *The "Fonction Caractéristique."* Using the substitution $t^x = e^{i\omega x}$, Laplace transforms‖ the generating function into a second form to which he does not assign a new name. If Laplace had not restricted the exponent x to integral values, the second form of his generating function would be identical with the function introduced by Lévy¶ as follows:

"Nous appellerons *fonction caractéristique*, et désignerons par $\phi(t)$, la valeur probable de

$$e^{itx},$$

c'est-à-dire

$$\phi(t) = \sum \alpha_h e^{itx_h}."$$

Here $i = \sqrt{-1}$. Obviously Lévy's footnote:

"La notion de fonction caractéristique semble avoir été introduite par Cauchy dans plusieurs notes présentées à l'Académie des Sciences en 1853; le mot de fonction caractéristique a été introduit par Poincaré, qui appelait d'ailleurs ainsi la valeur probable de e^{tx}, et non celle de e^{itx}. On verra aisément

---

* Boole, *A Treatise on the Calculus of Finite Differences*, 3d edition, p. 15.

† Todhunter, *A History of the Mathematical Theory of Probability*, §951.

‡ Lacroix, *Traité du Calcul Différentiel et du Calcul Intégral*, 2d edition, Paris, 1819, vol. 3, Chapter IV.

§ MacMahon, *Combinatory Analysis*, 1915, vol. I, *Introduction*, p. vi.

‖ Laplace, *Théorie Analytique*, pp. 83 and 84. The page references given below for the *Théorie Analytique* are in accord with the page numbers of the third edition as published by the Académie des Sciences in 1886.

¶ Lévy, *Calcul des Probabilités*, p. 161.

l'avantage qu'il y a à modifier comme nous l'avons fait la définition de Poincaré."

calls for the substitution $\begin{pmatrix}\text{Cauchy}\\ \text{Laplace}\end{pmatrix}\begin{pmatrix}1853\\ d\end{pmatrix}$, where $d \leqq 1812$.

For evidence that this historical statement has been accepted and for a vivid illustration of how the contents of the *Théorie Analytique* have been forgotten let us turn to Darmois' charming and significant little volume *Statistique Mathématique*. On page 45 we find:

"Cette fonction a été introduite dans la théorie des probabilités, par Cauchy en 1853, puis par Poincaré (voir P. Lévy, *Calcul des Probabilités*, p. 161, note)."

On page 75 of Darmois, in his chapter on *Repeated Trials and the Law of Large Numbers*, we find

" $\cdots$ fonction caractéristique:

$$\lambda(t) = (q_1 + p_1e^t)(q_2 + p_2e^t) \cdots (q_s + p_se^t).\text{''} \tag{55}$$

Now, noting that Lévy would use *it* instead of *t* on the right-hand side of this equation, let us compare (55) with the following abstract from Chapter IX, Book II, of the *Théorie Analytique*:

"Cela posé, considérons le produit

$$(1 - q + qc^{\nu\omega\sqrt{-1}})(1 - q^{(1)} + q^{(1)}c^{\overset{(1)}{\nu}\omega\sqrt{-1}}) \cdots (1 - q^{(s-1)} + q^{(s-1)}c^{\overset{(s-1)}{\nu}\omega\sqrt{-1}}).\text{''}$$

Giving each of the quantities ν, $\nu^{(1)}, \cdots, \nu^{(s-1)}$ the particular value 1 we obtain Darmois' expression (55) and, hence, Poisson's generalization of the Bernoulli theorem; in other words, the law of large numbers.

4. *Fourier Reciprocal Integrals*. Laplace's treatment of generating functions embodies explicitly the pair of equations (the notation is mine)*

$$U(\omega) = \sum_{x=0}^{\infty} y(x)e^{i\omega x},$$

$$y(x) = \frac{1}{2\pi}\int_{-\pi}^{+\pi} U(\omega)e^{-i\omega x}d\omega,$$

* Laplace, *Théorie Analytique*, pp. 83 and 84.

where x takes the discrete values

$$0,1,2,3,\cdots,\infty.$$

It is of interest to note how closely his subsequent treatment of these equations in the famous fourth chapter of Book II came to the development of the Fourier reciprocal equations

$$U(\omega) = \int_{-\infty}^{\infty} y(x)e^{i\omega x}dx,$$

$$y(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} U(\omega)e^{-i\omega x}d\omega.$$

Todhunter says,* in this connection,

"We shall begin our account of Laplace's fourth chapter by giving Poisson's solution of a very general problem, as we shall then be able to render our analysis of Laplace's processes more intelligible. But at the same time it must be remembered that the merit is due almost entirely to Laplace; although his processes are obscure and repulsive, yet they contain all that is essential in the theory: Poisson follows closely in the steps of his illustrious guide, but renders the path easier and safer for future travelers."

How many times the "obscure and repulsive" processes of Laplace have intrigued me into reading Chapter IV, Book II, shall be kept a secret; I admit depraved tastes but do not seek notoriety. Therefore, following the example of the great historian, abstracts from a paper published by Poisson† will be submitted for your consideration instead of the equivalent analysis of Laplace. Under the heading

"Sur la Probabilité des résultats moyens des Observations"

we find

"La question que je me propose de traiter dans ce Mémoire a déjà été l'objet des travaux de plusieurs géomètres, et particulièrement de M. Laplace, dont les recherches sur cette matière intéressante sont réunies dans la *Théorie analytique des Probabilités* (liv. II, chap. IV), et dans les trois suppléments à ce grand ouvrage. La généralité de l'analyse de M. Laplace, la variété et l'importance des objets auxquels il en a fait l'application, ne laissent sans doute rien à désirer; mais il m'a semblé que quelques points de cette théorie pouvaient encore être développés; et j'ai pensé que les remarques que j'ai eu l'occasion de

* Todhunter, op. cit., §1001.

† Poisson, *Connaisance des Temps de l'année* 1827.

faire en l'étudiant, seraient propres à en éclaircir les difficultés, et pourraient aussi n'être pas sans utilité dans la pratique. "

$$\text{“}p = \frac{1}{\pi}\int_{-\infty}^{\infty}\left(\int_{-a}^{a} fxe^{x\alpha\sqrt{-1}}dx\right)^{s} e^{-b\alpha\sqrt{-1}} \sin c\alpha \frac{d\alpha}{\alpha}\cdot \qquad (1)\text{”}$$

"En faisant $s=1$ dans l'équation (1), on aura la probabilité que l'erreur d'une seule observation est comprise entre $b-c$ et $b+c$, laquelle sera

$$p = \frac{1}{\pi}\int_{-\infty}^{\infty}\left(\int_{-a}^{a} fxe^{x\alpha\sqrt{-1}}dx\right)e^{-b\alpha\sqrt{-1}} \sin c\alpha \frac{d\alpha}{\alpha}\cdot\text{”}$$

Then, in this case, "d'une seule observation" Poisson shows, by interchanging the order of integrations, that

$$\text{“}p = \int_{b-c}^{b+c} fxdx.\text{”}$$

It is now only a short step to Fourier. Poisson's two expressions for p give

$$\int_{b-c}^{b+c} f(x)\,dx = \frac{1}{\pi}\int_{-\infty}^{\infty}\left(\int_{-a}^{a} f(x)e^{x\alpha\sqrt{-1}}d\alpha\right)e^{-b\alpha\sqrt{-1}} \sin c\alpha\frac{d\alpha}{\alpha};$$

therefore, if we take $c=db/2$, we obtain at once, since we have $\sin (c\alpha)\rightarrow(c\alpha)$ as $c\rightarrow 0$,

$$f(b) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-a}^{a} f(x)e^{x\alpha\sqrt{-1}}\,dx\right)e^{-b\alpha\sqrt{-1}}\,d\alpha$$

and, if the limits of error are $\pm\infty$ instead of $\pm a$,

$$f(b) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(x)e^{x\alpha\sqrt{-1}}\,dx\right)e^{-b\alpha\sqrt{-1}}\,d\alpha;$$

the well known Fourier relationship.

It is of interest to note that the analyses of Laplace and Poisson contain the exact equivalent of the *discontinuity factor*

$$\operatorname{sgn}(\nu) = \frac{2}{\pi}\int_{0}^{\infty}\sin(\nu t)\frac{dt}{t}\cdot$$

They therefore anticipated Dirichlet's* use of this operator by

* Dirichlet, *Sur une nouvelle méthode pour la détermination des intégrales multiples*, Journal de Mathématiques Pures et Appliquées, vol. 4 (1839), pp. 164–168.

at least 27 years and 12 years, respectively. E. L. Dodd* has recently emphasized the utility of the function sgn (ν) in probability theory.

5. *Evaluation of Definite Integrals.* In the theory of probability and its applications we are often confronted with definite integrals whose integrands involve factors raised to high powers. This situation arises particularly when one is solving a problem of causes. Frequently but little information is at hand concerning the a priori existence probabilities involved and, therefore, large numbers of observations must be taken into account if reliable conclusions are to be drawn.

A method of approximation, which is peculiarly efficacious when applied to integrands of the type under consideration, is given by Laplace in the first chapter of the second part of Book I, under the heading "De l'intégration par approximation des différentielles qui renferment des facteurs élevés à des grandes puissances." No one familiar with this method can fail to agree with the opinion of it expressed by Todhunter:†

> "The method of approximation to the values of definite integrals, which is here expounded, must be esteemed a great contribution to mathematics in general and to our special department in particular."

It is pertinent to mention here a point which will be emphasized later in connection with the probability of causes. That is, if one desires to be conversant with the evolution of Laplace's thoughts on probability he must, of course, read the memoir of 1774. But one who wishes to be conversant with the mature and final work of Laplace must read the later memoirs and the *Théorie Analytique*. The point may be illustrated by the following comments quoted from Todhunter. Speaking of the 1774 memoir he says "Laplace by a rude process of approximation $\cdots$," while with reference to the same question in the *Théorie Analytique*, Todhunter says " $\cdots$: the present demonstration is much superior to the former."‡

Book I, Second Part, contains in addition to the chapter entitled "De l'intégration par approximation des différentielles

* Dodd, *The frequency law of a function of variables with given frequency laws*, Annals of Mathematics, Sept. 1925.

† Todhunter, op. cit., §905, p. 484.

‡ Todhunter, op. cit., §§871 and 894, pp. 467 and 478.

qui renferment des facteurs élevés à de grandes puissances" two more chapters, entitled "De l'intégration par approximation des équations linéaires aux différences finies et infiniment petites" and "Application des méthodes précédentes à l'approximation de diverses fonctions de très grands nombres," respectively. Among these "diverses fonctions de très grands nombres" we find an exact expression in terms of incomplete beta functions for the incomplete binomial summation. This result was published in 1924 by Karl Pearson as an original contribution and as late as 1928 he implies that Laplace gave only an approximate result.*

With reference to the method of evaluating definite integrals consider, to fix ideas, a frequency function $y=f(x)$ having a single mode at $x=a$. Suppose we want its integral,

$$\int_{\theta_1}^{\theta_2} ydx$$

between assigned limits. For this purpose Laplace has given us two formulas. We are to use his first formula based on the transformation

$$y(x) = y(\theta)e^{-t}, \quad \theta = \theta_1 \quad \text{or} \quad \theta_2,$$

if the limits θ_1 and θ_2 do not include the mode $x=a$. If the limits do include the mode, or one of them is near it, we are to use his second formula based on the change of variable

$$y(x) = y(a)e^{-t^{\mu+1}};$$

Laplace states that he developed his second formula to cover the case where y satisfies the differential equation

$$-\frac{1}{y}\cdot\frac{dy}{dx} = (x-a)^{\mu}f(x).$$

After developing the general formula for any value of μ, Laplace says

"Le cas de $\mu+1=2$ étant le plus ordinaire, nous allons exposer ici les formules qui y sont relatives."

---

* Pearson, Biometrika, vol. 16 (1924), pp. 202–203; also vol. 20A, Parts I and II, July, 1928, p. 170.

It is manifest that a student of Laplace is quite prepared to take under consideration the Pearsonian system of frequency curves built around the particular differential equation

$$-\frac{1}{y}\cdot\frac{dy}{dx} = (x - a)f(x),$$

where

$$f(x) = 1/(A + Bx + Cx^2).$$

No mere outline of his method for evaluating definite integrals does justice to Laplace. You will be well repaid if you read the entire chapter of the *Théorie Analytique* wherein it is expounded.

Anticipating my remarks on the probability of causes I find it convenient to draw attention here to an important matter which is involved implicitly in the formulas of Book I, Part II. The matter is covered explicitly in Laplace's *Mémoire sur les Probabilités*, of 1780.

Laplace sets $y = \phi u^s u'^{s'} u''^{s''} \cdots$ in his approximation formulas and introduces the equation*

$$v = -y\frac{dx}{dy} = \frac{-1}{\dfrac{sdu}{udx} + \dfrac{s'du'}{u'dx} + \dfrac{s''du''}{u''dx} + \cdots + \dfrac{d\phi}{\phi dx}}.$$

He then says:

"Ainsi, dans le cas où $s, s', s'', \cdots$ sont de très grands nombres, v sera fort petit, et si l'on fait $1/s = \alpha$, α étant une fraction très petite, la fonction v sera de l'ordre α, et les termes successifs de la formule (A) seront respectivement des ordres $\alpha, \alpha^2, \alpha^3, \cdots$."

But what is the significance of the factor ϕ in y? If we Compare Laplace's discussion in the memoir of 1780 with Poincaré's and Edgeworth's† it is evident that in problems of causes ϕ, if used, would symbolize the a priori probability function. Laplace in

* Laplace, *Théorie Analytique*, Book I, Art. 22.

† Laplace, Oeuvres, vol. 9, p. 470.

Poincaré, *Calcul des Probabilités*, 2d edition, p. 255.

Bowley, *F. Y. Edgeworth's Contributions to Mathematical Statistics*, pp. 11–12.

his discussion of the matter points out, as do Poincaré and Edgeworth later, that the a priori factor is of negligible importance when the productive probability involves factors raised to high powers. We have here a good illustration of the fact that if one reads beyond the memoir of 1774 he will find little, if anything, of fundamental importance left untouched by Laplace.

6. *Probability of Causes.* The probability of causes is a subject which calls for careful consideration. One may easily become lost in a maze of contradictions and misleading interpretations which exist in spite of the fact that the subject has been lucidly exhibited. The chapters on the Bayes theorem in some of the older works and in, for example, the recent treatises by Burnside and Coolidge are entirely satisfactory. Moreover, not only engineering, but all students, should read the treatment of the question in *Probability and Its Engineering Uses* by T. C. Fry.

It will be helpful to bear in mind that in this paper the formula

$$P_i = \omega_i p_i / \sum \omega_i p_i$$

will be referred to as *Laplace's generalization of Bayes' formula* (or theorem) for the probability in favor of the ith cause when the a priori existence probabilities, ω_i, are not all equal. By *Bayes' formula* (or theorem) will be meant the restricted formula (or theorem)

$$P_i = p_i / \sum p_i$$

based on the postulate that the a priori existence probabilities are all equal. Although, as written, these formulas apply only to discrete values of ω and p, no such restriction will be assumed in the following discussion; both Bayes* and Laplace apply their formulas freely to continuously varying quantities.

Now let us examine those sections of the *Théorie Analytique* that have reference to the probability of causes. The abstracts from them given in Appendix I to this paper contain:

(1) The Bayes formula (or theorem) for both discrete and continuous variables.

* Bayes, *An essay towards solving a problem in the doctrine of chances*, Philosophical Transactions, vol. 53 (1763).

(2) The Laplacian generalization of the Bayes formula (or theorem) for both discrete and continuous variables. The generalization has reference to the a priori existence probabilities.
(3) A proof of the a posteriori probability formula based on the properties of the probability formula for a compound event. In other words, a proof embodying the fundamental element of Bayes' original proof.
(4) A definite statement to the effect that in his applications Laplace will work with the restricted formula because, as we can always represent by a single function the product of two given functions, the generalized formula may be reduced in form to the restricted formula.
(5) Evidence that Laplace avoids small samples.

One who has digested the contents of the *Théorie Analytique* on the probability of causes may be trusted to steer a straight course through the maze of contradictions and misunderstandings mentioned at the beginning of this section. However, "to be forewarned is to be forearmed."

Compare, for example, the abstracts from the *Théorie Analytique* given in Appendix I with the abstracts from Keynes which I have placed in Appendix II. We have here a good example of Laplace's contributions being judged by his first memoir of 1774 instead of by the contents of the *Théorie Analytique*. The case is paralleled by Keynes' interpretation of De Morgan's views on inverse probabilities. This interpretation is based solely on a passage contained in the introduction to DeMorgan's *Essay on Probabilities*; Keynes ignores the general rule given on page 59 of the Essay in the chapter entitled *On Inverse Probabilities* (the italics are mine):

"The rule to which the preceding reasoning conducts us is as follows: *When the different states under which an event may have happened are not equally likely to have existed*, then having found the probability which each state would give to the observed event, *multiply each by the probability of the state itself before using the rule in page 55.*"

Karl Pearson's writings do not help one to a clear understanding of Laplace's contributions to probability theory. He writes,* as a principle due to Laplace:

* Pearson, *Grammar of Science*, 3d edition, 1911, p. 145.

"If a result might flow from any one of a certain number of different constitutions, all equally probable before experience, then the several probabilities of each constitution after experience being the real constitution, are proportional to the probabilities that the result would flow from each of these constitutions."

But this is not the Laplacian generalization; it is the restricted theorem of 1774.

Consider now Pearson's paper entitled *On the influence of past experience on future expectation.** He says:

"(2) Let the chance of a given event occurring be supposed to lie between x and $x+\delta x$, then if on $n=p+q$ trials an event has been observed to occur p times and fail q times, the probability that the true chance lies between x and $x+\delta x$ is, on the equal distribution of our ignorance

$$P_x = \frac{x^p(1-x)^q dx}{\int_0^1 x^p(1-x)^q dx}. \tag{i}$$

"This is Bayes' Theorem.

"Now suppose that a second trial of $m=r+s$ instances be made, then the probability that the given event will occur r times and fails s, is on the *a priori* chance being between x and $x+\delta x$

$$= P_x \frac{\underline{|m}}{\underline{|r}\,\underline{|s}} x^r(1-x)^s,$$

and accordingly the total chance C_r, whatever x may be, of the event occurring r times in the second series, is

$$C_r = \frac{\underline{|m}}{\underline{|r}\,\underline{|s}} \frac{\int_0^1 x^{p+r}(1-x)^{q+s}dx}{\int_0^1 x^p(1-x)^q dx}. \tag{ii}$$

"This is, with a slight correction, Laplace's extension of Bayes' Theorem."

This last sentence is doubly misleading. In the first place it has given the erroneous impression that equation (ii) should be credited to Pearson.† Now the exact equivalent of (ii) is given not only in §30, Book II, of the *Théorie Analytique* but also in a memoir dated 1780. The formula omitting the factor $m!/(r!s!)$ appears in the memoir of 1774. But on what ground can this formula of 1774 be called incorrect? Todhunter, com-

---

* Pearson, Philosophical Magazine, March, 1907, p. 366.

† G. F. McEwen, Bulletin of the Scripps Institution of Oceanography, vol. 2, No. 1, p. 20.

menting on it, says "so that of course the m white tickets and n black tickets are supposed to be drawn in an assigned order."* In the second place the expression "Laplace's extension of Bayes' theorem" is likely to lead to confusing equation (ii) with Laplace's generalization of Bayes' theorem, a confusion fatal to progress.

The problem Pearson dealt with in 1907 constitutes the subject matter of his later paper entitled *The fundamental problem of practical statistics.*† This paper gives a still poorer picture of Laplace's contributions and grasp of the subject. Consider this statement:

> "None of the early writers on this topic—all approaching the subject from the mathematical theory of games of chance—seem to have had the least inkling of the enormous extension of their ideas, which would result in recent times from the application of the theory of random sampling to every phase of our knowledge and experience—economic, social, medical, and anthropological—and to all branches of observation whether astronomical, physical or psychical."

Evidently Pearson has not read Fourier's eulogy on Laplace or, for example, the following item from the table of contents of the *Théorie Analytique*:

> "Application à la variation diurne du baromètre et à la rotation de la Terre, déduite des expériences sur la chute des corps. La même analyse est applicable aux questions les plus délicates de l'Astronomie, de l'Économie politique, de la Médecine, etc."

As to psychical matters Bayle‡ says of Laplace:

> "Le chapitre dans lequel il trace les plans d'une science psychologique est des plus remarquables."

Unfortunately, the paper on the *Fundamental Problem* begins with a quite fallacious mathematical analysis of the part played therein by the a priori existence probabilities. Had Pearson started with the Laplacian generalization of the Bayes theorem, Burnside§ would not have had occasion to state:

---

* Todhunter, op. cit., p. 467.

† Pearson, Biometrika, Oct., 1920.

‡ Laplace, *Essai Philosophique sur les Probabilités—Édition Nouvelle avec une Introduction par X. Torau Bayle*, Paris, 1920, preface, p. XIII.

§ Burnside, Biometrika, vol. 16 (1924), p. 189.

There is therefore no reason for supposing that any conclusions drawn from the investigation on p. 5 will hold with respect to the statistical problem stated at the beginning of Professor Pearson's paper.

The name of Laplace has been introduced in endless discussions regarding the equal distribution of ignorance in the probability of causes. Now, if I remember correctly, Laplace does not put forth any argument in favor of this distribution. It was Bayes who did so and his argument is worthy of serious consideration.

In transmitting Bayes' paper to the Royal Society, Price pointed out that Bayes first postulates that the existence probabilities are equally likely in order to establish his restricted formula on a sound basis and then presents in a scholium an argument to justify the now usual procedure regarding existence probabilities when no a priori information is available.

Let p be the (unknown) probability of an event happening in a single trial. In a corollary preceding his theorem, Bayes shows that the probability of the event happening x times in n trials is independent of x, provided all values of p between 0 and 1 are a priori equally likely. Likewise, consider a box containing N balls, some of them white and the rest black, the ratio of white to total balls having with equal likelihood any value from $0/N$ to N/N. If a sample of n balls is to be drawn from the box the probability that the sample will contain an assigned number of white balls is independent of the number assigned.

Now, says Bayes in his scholium, what can be more descriptive of total ignorance regarding the value of p than that, a priori, a sample is as likely to give one result as any other? But if total ignorance is so described then Bayes' corollary justifies the assumption that, a priori, all values of p are equally likely.

The significance of Bayes' scholium seems to have been overlooked by R. A. Fisher in his discussion of Bayes' theorem versus the idea of *maximum likelihood*. Fisher argues* that instead of the probability p for an event happening in a single trial we might just as well use some other variable, say θ, where

$$\sin \theta = 2p - 1$$

* R. A. Fisher, Philosophical Transactions, series A, vol. 222, p. 325.

and assume all values of θ to be a priori equally likely. But this assumption regarding all values of θ (between the limits corresponding to $p=0$ and $p=1$) will *not* give for the probability of the event happening x times in n trials a result independent of x. This negative result rather weakens Fisher's argument against Bayes' rule.

Unfortunately, the presentation in Todhunter's *History* of the work of Bayes and Laplace on the probability of causes is most inadequate. Todhunter gives Bayes' corollary without mentioning the particular point which Bayes wished to make with it; that the probability of the event happening x times in n trials is independent of x. Failure to appreciate this point kills the significance of Bayes' scholium; the scholium is also omitted by Todhunter. Again, although the work of Laplace calls for high tribute, Todhunter places the tribute in the wrong place. Speaking of the 1774 memoir he says:

> "The memoir is remarkable in the history of the subject, as being the first which distinctly enunciated the principle for estimating the probabilities of the causes by which an observed event may have been produced. Bayes must have had a notion of the principle · · · ."

But this memoir does not contain the Laplacian generalization of the Bayes' rule. On the other hand when describing the contents of the *Mêmoire sur les Probabilitês** and the *Thêorie Analytique*, Todhunter fails to mention the generalization which is embodied therein.

In view of this situation it is lamentable that some authorities, in their discussions on the probability of causes, seem to have taken Todhunter's *History* as a substitute for Laplace.

This is perhaps a good place to point out that the notorious example about the sun rising a million times in succession appears in Price's publication of Bayes' Essay and that its republication by Laplace is immediately followed by the statement "Mais ce nombre est incomparablement plus fort pour celui qui, connaissant par l'ensemble des phénomènes le principe régulateur des jours et des saisons, voit que rien dans le moment actuel ne peut en arrêter le cours."

Boole in his *Laws of Thought* says: "The following is a summary, chiefly taken from Laplace, of the principles which have

* Laplace, Oeuvres, vol. 9.

been applied to the solution of questions of probability." The summary contains the restricted rule of 1774 but *not* Laplace's generalization; this should be borne in mind when one quotes Boole on the probability of causes.

It is of particular interest to note that Poincaré, like the great master, frequently discards the a priori existence functions with the distinct understanding that the number of observations is large.

7. *Hermite Polynomials.* In Book II, Chapter III, Art. 17, Laplace introduces the partial differential equation

$$\frac{\partial U}{\partial r'} = 2U + 2\mu\frac{\partial U}{\partial \mu} + \frac{\partial^2 U}{\partial \mu^2}$$

as approximately equivalent to the finite difference equation which gives the solution of a problem stated as follows:

"Considérons maintenant deux urnes A et B renfermant chacune le nombre n de boules, et supposons que, dans le nombre total $2n$ des boules, il y en ait autant de blanches que de noires. Concevons que l'on tire en même temps une boule de chaque urne, et qu'ensuite on mette dans une urne la boule extraite de l'autre. Supposons que l'on répète cette opération un nombre quelconque r de fois, en agitant à chaque fois les urnes, pour en bien mêler les boules; et cherchons la probabilité qu'après ce nombre r d'opérations, il y aura x boules blanches dans l'urne A."

Laplace* refers to this problem as "remarquable en ce que sa solution offre le premier exemple de l'emploi du calcul aux différences infiniment petites partielles, dans les questions de probabilités." The problem has recently been used by Dr. A. J. Lotka in building up an illustration of *The Statistical Meaning of Irreversibility.*†

Commenting on the partial differential equation given above Todhunter says:

"Laplace proceeds to integrate his approximate equation by the aid of definite integrals · · · , and then he passes on to other theorems which bear an analogy to those which occur in connection with what are called *Laplace's Functions.*"

The "other theorems" to which Todhunter refers in 1865 (one

---

* Laplace, Oeuvres, vol. 12, p. 361.

† Lotka, *Elements of Physical Biology*, Baltimore, 1925, p. 31.

year after the paper by Hermite entitled *Sur un nouveau développement en série des fonctions* was published) anticipated by half a century polynomials* which today occupy a central position in the literature on probability and statistical theory.

Laplace expresses the solution of his partial differential equation in terms of functions U_i, U_i' which are related to the Hermite polynomials as follows:

$$(-1)^n(2n)!U_n(\mu) = n!H_{2n}(\mu) = n!2^nA_{2n}(\mu(2)^{1/2}),$$

$$2(-1)^n(2n+1)!U_n'(\mu) = n!(-1)H_{2n+1}(\mu) = n!2^{n+1/2}A_{2n+1}(2^{1/2}\mu).$$

In these equations H and A represent the Hermite polynomials as given† by Hermite and Appell, respectively. Hermite actually used U and Appell, to avoid confusion with another use of the symbol U, states that he will use H instead of Hermite's symbol. But Appell does more than this: he modifies Hermite's original expressions by a change of variable as indicated in the second and third terms of the equations given above.

Laplace defines his functions in terms of definite integrals and, also, in series analogous to the second series form in which Hermite gives his polynomials. The orthogonal properties of U_i, U_i' are proved. Then these properties are used by Laplace to determine the coefficients $Q^{(1)}$, $Q^{(2)}$, etc., and $L^{(0)}$, $L^{(1)}$, etc., in the expansion

$$X = \frac{2e^{-\mu^2}}{(n\pi)^{1/2}}\Big[1 + Q^{(1)}(1-2\mu^2) + \cdots + L^{(0)}\mu + L^{(1)}\mu\left(1 - \frac{3}{2}\mu^2\right) + \cdots\Big],$$

"X étant une fonction donnée de μ." This, of course, completely anticipates the so-called Gram-Charlier A series.

8. *Limits in Probability*. Article 17 is by no means the only

* Uspensky, *On the development of arbitrary functions in series of Hermite's and Laguerre's polynomials*, Annals of Mathematics, Sept., 1927, p. 593, footnote.

† Hermite, Oeuvres, vol. 2, p. 295. Appell, *Fonctions Hypergéométriques et Hypersphériques, Polynomes d'Hermite*, 1926, p. 333.

important section of Chapter III, Book II. I would invite your attention particularly to its opening paragraph which says:

"A mesure que les événements se multiplient, leurs probabilités respectives se développent de plus en plus; leurs résultats moyens et les bénéfices ou les pertes qui en dépendent convergent vers des limites dont ils approchent avec des probabilités toujours croissantes."

Note that the word "toujours" has reference to "probabilités" and not to "limites." There must be no misundertanding on this point so let us turn to the Oeuvres, volume 10, page 308, where we find:

"On voit ainsi comment les événements, en se multipliant, nous découvrent leur possibilité respective; mais on doit observer qu'il y a dans cette analyse deux approximations, dont l'une est relative aux limites qui comprennent la valeur de x et qui se resserrent de plus en plus, et dont l'autre est relative à la probabilité que x se trouve entre ces limites, probabilité qui approche sans cesse de l'unité ou de la certitude. C'est en cela que ces approximations diffèrent des approximations ordinaires, dans lesquelles on est toujours assuré que le résultat est compris dans les limites qu'on lui assigne."

It is quite evident that Laplace had in mind the fundamental difference between the idea of a limit as used in pure mathematics and the limit concept upon which frequency definitions of probability have been based. This distinction has been brought out vividly by T. C. Fry* and has also been discussed by Castelnuovo and Du Pasquier.

Frequency definitions of probability remind us of Venn. Therefore, I will make a short digression on his *Logic of Chance.* Venn claims therein† that:

"Probability has been very much abandoned to mathematicians, who as mathematicians have generally been unwilling to treat it thoroughly.

. . . .

Any subject which has been discussed by such men as Laplace and Poisson, and on which they have exhausted all their powers of analysis, could not fail to be profoundly treated, so far as it fell within their province. But from this province the real principles of the science have generally been excluded, or so meagerly discussed that they had better have been omitted altogether."

In my opinion the contributions of Laplace to probability theory imply a philosophic outlook of the highest order.

---

* Fry, *Probability and its Engineering Uses*, §36, pp. 88–91.

† Venn, *Logic of Chance*, 3d edition, 1888, p. viii.

Moreover, if bound in one volume, the number of paragraphs of the *Essai Philosophique*, *Mêmoires*, and *Thêorie Analytique* in which fundamental ideas are discussed would compare favorably with the *Logic of Chance* in size. How the two would compare in other respects is a matter of personal point of view.

9. *Summary.* In answer to the question—*to what extent will one conversant with the "Thêorie Analytique" be in touch with the present status of probability theory*?—I have submitted for your consideration:

(1) The virtual identity between the Laplacian generating function and the Cauchy-Poincaré characteristic function, functions for whose introduction in recent probability literature we are indebted to Paul Lévy;

(2) The close approach of Laplace's analysis to the form of the Fourier reciprocal equations;

(3) The explicit presentation (à une constant près) by Laplace of the Hermite polynomials and the related Gram-Charlier expansion.

With reference to the second half of my keynote question —*how sound a foundation will he have found therein for statistical applications of the theory*?—I have presented:

(1) Laplace's contributions to the theory of inverse probability, and pointed out

(2) The distinction he draws between the meaning of the word *limit* when used outside the domain of probability theory, and its meaning when the word is attached to the observed frequency with which an event happens.

As evidence that the *Thêorie Analytique* is in advance of much recent probability literature, and on account of its great practical value, I have outlined the Laplacian method of dealing with integrands involving factors raised to high powers.

Naturally I have been unable to carry out the purpose of this Address without revealing to some extent my great admiration for Laplace. But I offer no apology for admiring the author of what Herschel has called "*the ne plus ultra of mathematical skill and power.*"

Appendix I

Abstracts from the Théorie Analytique, 3d Edition,

Book II, Chapter I

Pages 183–185:

"La probabilité d'un événement futur, tirée d'un événement observé, est le quotient de la division de la probabilité de l'événement composé de ces deux événements, et déterminée *a priori*, par la probabilité de l'événement observé, déterminée pareillement *a priori*.

"De là découle encore cet autre principe relatif à la probabilité des causes, tirée des événements observés.

"Si un événement observé peut résulter de n causes différentes, leurs probabilités sont respectivement comme les probabilités de l'événement, tirées de leurs existence; et la probabilité de chacune d'elles est une fraction dont le numérateur est la probabilité de l'événement, dans l'hypothèse de l'existence de la cause, et dont le dénominateur est la somme des probabilités semblables, relatives à toutes les causes.

"Considérons, en effet, comme événement composé l'événement observé, résultant d'une de ces causes. La probabilité de cet événement composé, probabilité que nous désignerons par E, sera, par ce qui précède, égale au produit de la probabilité de l'événement observé, déterminée *a priori* et que nous nommerons F, par la probabilité que, cet événement ayant lieu, la cause dont il s'agit existe, probabilité qui est celle de la cause, tirée de l'événement observé, et que nous nommerons P. On aura donc

$$P = \frac{E}{F}\cdot$$

La probabilité de l'événement composé est le produit de la probabilité de la cause par la probabilité que, cette cause ayant lieu, l'événement arrivera, probabilité que nous désignerons par H. Toutes les causes étant supposées *a priori* également possibles, la probabilité de chacune d'elles est $1/n$; on a donc

$$E = \frac{H}{n}\cdot$$

La probabilité de l'événement observé est la somme de tous les E relatifs à chaque cause; en désignant donc par $S(H/n)$ la somme de toutes les valeurs de H/n, on aura

$$F = S\frac{H}{n};$$

l'équation

$$P = \frac{E}{F}$$

deviendra donc

$$P = \frac{H}{SH}$$

ce qui est le principe énoncé ci-dessus, lorsque toutes les causes sont *a priori* également possibles. Si cela n'est pas, en nommant p la probabilité *a priori* de la cause que nous venons de considérer, on aura

$$E = Hp,$$

et, en suivant le raisonnement précédent, on trouvera

$$P = \frac{Hp}{SHp}$$

ce qui donne les probabilités des diverses causes, lorsqu'elles ne sont pas toutes également possibles *a priori.*"

Book II, Chapter VI

Page 370:

"26. La probabilité de la plupart des événements simples est inconnue: en la considérant *a priori*, elle nous paraît susceptible de toutes les valeurs comprises entre zéro et l'unité; mais, si l'on a observé un résultat composé de plusieurs de ces événements, la manière dont ils y entrent rend quelques-unes de ces valeurs plus probables que les autres. Ainsi, à mesure que le résultat observé se compose par le développement des événements simples, leur vraie possibilité se fait de plus en plus connaître, et il devient de plus en plus probable qu'elle tombe dans des limites qui, se resserrant sans cesse, finiraient par coïncider, si le nombre des événements simples devenait infini. Pour déterminer les lois suivant lesquelles cette possibilité se découvre, nous la nommerons x. La théorie exposée dans les Chapitres précédents donnera la probabilité du résultat observé, en fonction de x. Soit y cette fonction; si l'on considère les différentes valeurs de x comme autant de causes de ce résultat, la probabilité de x sera, par le troisième principe du n° 1, égale à une fraction dont le numérateur est y, et dont le dénominateur est la somme de toutes les valeurs de y; en multipliant donc le numérateur et le dénominateur de cette fraction par dx, cette probabilité sera

$$\frac{y\,dx}{\int y\,dx},$$

l'intégrale du dénominateur étant prise depuis $x=0$ jusqu'à $x=1$. La probabilité que la valeur de x est comprise dans les limites $x=\theta$ et $x=\theta'$ est par conséquent égale à

$$\text{(I)} \qquad \frac{\int y\,dx}{\int y\,dx},$$

l'intégrale du numérateur étant prise depuis $x=\theta$ jusqu'à $x=\theta'$, et celle du dénominateur étant prise depuis $x=0$ jusqu'à $x=1$.

"La valeur de x la plus probable est celle qui rend y un maximum. Nous la désignerons par a. Si aux limites de x, y est nul, alors chaque valeur de y a une valeur égale correspondante de l'autre côté du maximum.

"Quand les valeurs de x, considérées indépendamment du résultat observé, ne sont pas également possibles, en nommant z la fonction de x qui exprime leur probabilité, il est facile de voir, par ce qui a été dit dans le Chapitre I^er de ce

Livre, qu'en changeant dans la formule (I), y dans yz, on aura la probabilité que la valeur de x est comprise dans les limites $x=\theta$ et $x=\theta'$. Cela revient à supposer toutes les valeurs de x également possibles *a priori*, et à considérer le résultat observé comme étant formé de deux résultats indépendants, dont les probabilités sont y et z. On peut donc ramener ainsi tous les cas à celui où l'on suppose *a priori*, avant l'événement, une égale possibilité aux différentes valeurs de x, et, par cette raison, nous adopterons cette hypothèse dans ce qui va suivre.

"Nous avons donné dans les n^{os} 22 et suivants du Livre I^{er} les formules nécessaires pour déterminer, par des approximations convergentes, les intégrales du numérateur et du dénominateur de la formule (I), lorsque les événements simples dont se compose l'événement observé sont répétés un très grand nombre de fois; car alors y a pour facteurs des fonctions de x élevées à de grandes puissances. Nous allons, au moyen de ces formules, déterminer la loi de probabilité des valeurs de x, à mesure qu'elles s'éloignent de la valeur a, la plus probable, ou qui rend y un maximum."

Page 376:

"Ainsi les valeurs les plus probables sont proportionnelles aux nombres des arrivées des couleurs, et lorsque le nombre n est un grand nombre, les pròbabilités respectives des couleurs sont à très peu près égales aux nombres de fois qu'elles sont arrivées divisés par le nombre des tirages."

Page 384:

"28. C'est principalement aux naissances que l'analyse précédente est applicable, et l'on peut en déduire, non seulement pour l'espèce humaine, mais pour toutes les espèces d'êtres organisés, des résultats intéressants. Jusqu'ici les observations de ce genre n'ont été faites en grand nombre que sur l'espèce humaine; nous allons soumettre au calcul les principales."

Pages 392 and 393:

"Soit x la probabilité d'un individu de l'âge A, pour vivre à l'âge $A+a$; la probabilité de l'événement observé est alors le terme du binôme $[x+(1-x)]^p$ qui a x^q pour facteur; cette probabilité est donc

$$\frac{1\cdot 2\cdot 3\cdots p}{1\cdot 2\cdot 3\cdots (p-q)1\cdot 2\cdot 3\cdots q}x^q(1-x)^{p-q};$$

ainsi la probabilité de la valeur de x, prise de l'événement observé, est

$$\frac{x^q dx(1-x)^{p-q}}{\int x^q dx(1-x)^{p-q}},$$

l'intégrale du dénominateur étant prise depuis $x=0$ jusqu'à $x=1$.

"La probabilité que, sur les p' individus de l'âge A, $q'+z$ vivront à l'âge $A+a$ est

$$\frac{1\cdot 2\cdot 3\cdots p'}{1\cdot 2\cdot 3\cdots (q'+z)1\cdot 2\cdot 3\cdots (p'-q'-z)}x^{q'+z}(1-x)^{p'-q'-z}.$$

En multipliant cette probabilité par la probabilité précédente de la valeur de x, le produit intégré depuis $x=0$ jusqu'à $x=1$ sera la probabilité de l'existence de

$q'+z$ personnes à l'âge $A+a$. En nommant donc P cette probabilité, on aura

$$P = \frac{1\cdot 2\cdot 3 \cdots p' \int x^{q+q'+z}dx(1-x)^{p+p'-q-q'-z}}{1\cdot 2\cdot 3\cdots (q'+z) 1\cdot 2\cdot 3 \cdots (p'-q'-z)\int x^{q}dx(1-x)^{p-q}},$$

les intégrales du numérateur et du dénominateur étant prises depuis $x=0$ jusqu'à $x=1$."

Appendix II

Abstracts from a Treatise on Probability by J. M. Keynes

Pages 175 and 176:

"And in 1774 the rule was clearly, though not quite accurately, enunciated by Laplace in his 'Mémoire sur la probabilité des causes par les événements' (*Mémoires presentés à l'Académie des Sciences*, vol. vi, 1774). He states the principle as follows (p. 623):

"'Si un événement peut être produit par un nombre n de causes différentes, les probabilités de l'existence de ces causes prises de l'événement sont entre elles comme les probabilités de l'événement prises de ces causes; et la probabilité de l'existence de chacune d'elles est égale à la probabilité de l'événement prise de cette cause, divisée par la somme de toutes les probabilités de l'événement prises de chacune de ces causes.'

"He speaks as if he intended to prove this principle, but he only gives explanations and instances without proof. The principle is not strictly true in the form in which he enunciates it, as will be seen on reference to theorems (38) of Chapter XIV; and the omission of the necessary qualification has led to a number of fallacious arguments, some of which will be considered in Chapter XXX.

"The value and originality of Bayes' Memoir are considerable, and Laplace's method probably owes much more to it than is generally recognized or than was acknowledged by Laplace. The principle, often called by Bayes's name, does not appear in his *Memoir* in the shape given it by Laplace and usually adopted since; but Bayes's enunciation is strictly correct and his method of arriving at it shows its true logical connection with more fundamental principles, whereas Laplace's enunciation gives it the appearance of a *new* principle especially introduced for the solution of causal problems. The following passage* gives in my opinion, a right method of approaching the problem: 'If there be two subsequent events, the probability of the second b/N and the probability of both together P/N, and, it being first discovered that the second event has happened, from hence I guess that the first event has also happened, the probability I am in the right is P/b.' If the occurrence of the first event is denoted by a and of the second by b, this corresponds to $ab/h=a/bh\cdot b/h$ and therefore

$$a/bh = \frac{ab/h}{b/h}; \text{ for } ab/h = P/N,\ b/h = b/N,\ a/bh = P/b.$$

* Quoted by Todhunter, op. cit., p. 296. Todhunter underrates the importance of this passage, which he finds unoriginal, yet obscure.

The direct and indeed fundamental dependence of the inverse principle on the rule for compound probabilities was not appreciated by Laplace."

Page 178:

"It will be noticed that in the formula (38 · 2) the *a priori* probabilities of the hypotheses a_1 and a_2 drop out if $p_1 = p_2$, and the results can then be expressed in a much simpler shape. This is the shape in which the principle is enunciated by Laplace for the *general* case,* and represents the uninstructed view expressed with great clearness by De Morgan:† 'Causes are likely or unlikely, just in the same proportion that it is likely or unlikely that observed events should follow from them. The most probable cause is that from which the observed event could most easily have arisen.' If this were true the principle of Inverse Probability would certainly be a most powerful weapon of proof, even equal, perhaps, to the heavy burdens which have been laid on it. But the proof given in Chapter XIV makes plain the necessity in general of taking into account the *a priori* probabilities of the possible causes. Apart from formal proof this necessity commends itself to careful reflection. If a cause is very improbable in itself, the occurrence of an event, which might very easily follow from it, is not necessarily, so long as there are other possible causes, strong evidence in its favour. Amongst the many writers who, forgetting the theoretic qualification, have been led into actual error, are philosophers as diverse as Laplace, De Morgan, Jevons, and Sigwart, Jevons‡ going so far as to maintain that the fallacious principle he enunciates is 'that which common sense leads us to adopt almost instinctively.' "

THE AMERICAN TELEPHONE AND TELEGRAPH COMPANY

---

* See the passage quoted above, p. 175.

† 'Essay on Probabilities,' in the *Cabinet Encyclopaedia*, p. 27.

‡ *Principles of Science*, vol. 1, p. 280.