TITIAN AND THE GEOMETRY OF COLOUR

TRISTRAM DE PIRO

In this fragment, which continues the previous chapter, we consider how the aesthetics of asymmetry and colour, which we found in some of Titian's paintings, can, again, be used to guide the analysis of functions, this time, in terms of randomly evolving patterns and averaging effects, within the framework of probability theory. We will continue with Laplace's work as a motivating example. Although, as we have seen, Laplace's early published work of 1771 started with solving differential equations and the finite difference method, by 1774 he was already starting to think about the mathematical and philosophical concepts of probability and statistics, and, in 1774, published his first paper on the subject, "Memoire sur la probabilite des causes par les evenements", "Savants etranges 6", (pp621-656), Oeuvres 8, p27-65. In Section 2, p29, Laplace formulates the following principle with an application;

"PRINCIPE. -Si un evenement peut etre produit par un nombre n de causes differentes, les probabilites de l'existence de ces causes prises de l'evenement sont entre elles comme les probabilites de l'evenement prises de ces causes, et la probabilite de l'existence de chacune d'elles est egale a la probabilite de l'evenement prise de cette cause, divisee par la somme de totes les probabilites de l'evenement prises de chacune de ces causes.

La question suivante eclaircira ce principe, en meme temps quelle en fera voir l'usage:

Je suppose que l'on me presente deux urnes A et B, dont la premiere contienne p billets blancs et q billets noirs, et la seconde contienne p' billets blancs et q' billets noirs; je tire de l'une de ces urnes (j'ignore de laquelle) $f + h$ billets, dont f sont blancs et h sont noirs; on demande, cela pose, quelle est la probabiliteque l'urne dont j'ai tire ces billets est A ou qu'elle est B.

En supposant que cette urne soit A, la probabilite d'en tirer f billets blancs et h billets noirs est

...

Soit K cette quantite; si l'on suppose maintenant que l'urne dont j'ai tire les billete est B, la probabilites d'en tirer f billets blancs et h billete noirs se determinera en changeant, dans K, p et q en p' et q'; soit K' ce quo devient alors cette expression. Cela pose, les probabilites que l'urne dont j'ai tire les billete est A ou B sont entre elles, par le principe enonce ci-dessus, comme K est a K'; la probabilite que cette urne est A est egale a $\frac{K}{K+K'}$, et celle qu'elle est B est egale a $\frac{K'}{K+K'}$".

Here, Laplace explains the idea of Bayes' Theorem, ([1]), and how the probability of an event is determined by its conditional probabilities on a set of mutually exclusive causes, ([2]). He illustrates this concept with the problem of drawing white and black tickets from two urns A and B, concluding that if the probability of drawing f white tickets and h black tickets from A is K, the the probability of drawing f white tickets and h black tickets from B is K', then the probability that the the urn was A, given the draw of f white tickets and h black tickets, is $\frac{K}{K+K'}$, assuming an equal probability of drawing from urn A and from B, ([3]). In Section 5, p42, Laplace defines the probability density function, illustrating his definition with figure 23, and recognises that the notion of independence between observations, see figure 22, amounts to taking the product of the probability density functions;

"Probleme III.- Determiner le milieu que l'on doit prendre entre trois observations donnees d'un meme phenomene.

Solution.- Representons le temps par une droite indefinie AB (fig. 1), et supposons que la premiere observation fixe l'instant du phenomee au point a, la seconde au point b et la troisieme (figure 22) au point c;

...

nous representerons l'equation par celle-ci: $y = \phi(x)$. Or voici les proprietes de cette courbe:

1° Elle doit etre partagee en deux parties entierement semblables par la droite VR, car il est tant aussi probable que l'observation s'ecartera de la verite a droite comme a gauche: (figure 23)

2° Elle doit avoir pour asymptote la ligne KP, parce que la probabilite que l'observation s'eloigne de la verite a une distance infinie est evidemment nulle;

3° L'aire entiere de cette courbe doit etre egale a l'unite, puisqu'il est certain que l'observation tombera sur un des points de la droite KP. Supposons maintenant (fig. 1) que le veritable instant du phenomene soit au point V, a la distance x du point a, la probabilite que les trois

[1]Named after the English mathematician, Thomas Bayes, 1701-1761, specifically, that;

$P(A|B)P(B) = P(B|A)P(A)$

for events A and B.

[2]$P(A) = P(A|B_1)P(B_1) + \ldots + P(A|B_n)P(B_n)$, for mutually exclusive and exhaustive events $\{B_1, \ldots, B_n\}$

[3]Specifically, denoting the event of drawing f white tickets and h black tickets by C, and the events of drawing from urn A and urn B by A and B, we have, if $P(C|A) = K$, $P(C|B) = K'$, then;

$P(A|C) = P(C|A)\frac{P(A)}{P(C)} = P(C|A)\frac{P(A)}{P(C|A)P(A)+P(C|B)P(B)} = \frac{K.\frac{1}{2}}{\frac{K}{2}+\frac{K'}{2}} = \frac{K}{K+K'}$

by the two parts of Laplace's proposition and the assumption on equal probability.

observations a, b et c s'ecarteront aux distances Va, Vb et Vc sera;

$\phi(x)\phi(p-x)\phi(p+q-x)$."

The intuitive idea behind independence is that the probability of an event A is unaffected by the occurrence (or not) of an event B. Laplace naturally extends this idea to continuous random variables, which encode the possible (continuous) set of values of an event. Laplace is able to formalise his intuition, due to his command of the analytic method, using infinitesimals. ([4]).

In 1778, Laplace continued his research into probability with the "Memoire sur les probabilites", (Memoires de lAcademie des Sciences 1778,(1781), p227-237, Oeuvres 9,p383-485). Here, in Section 18, p423, he formulates a continuous version of the urns problem, finding the probability x of being a boy, given a sample of p boys and q girls, from a binomial distribution, with probability x, and, assuming the distribution of probabilities is uniform, as $\frac{x^p(1-x)^q dx}{\int_0^1 x^p(1-x)^q dx}$, ([5]);

"Soient x la possibilite de la naissance d'un garcon et $1-x$ celle de la naissance d'une fille; la probabititu que, sur $p+q$ enfants, il y aura p garons et q filles, sera, comme on l'a vu dans l'article precedent, egale a $\lambda x^p(1-x)^q$; or, si l'on regarde x comme une cause particulicre de

[4] The formal definition of independence for events, is that $P(A \cap B) = P(A)P(B)$, or, $P(A|B) = P(A)$. The modern definition of a probability density function f_X for a random variable X can be found in [5], p60, so that the probability $P(a \leq X \leq b) = \int_a^b f(x)dx$. As Laplace notes in 3°, we then have that $P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 1$, and, in 2°, $lim_{n\rightarrow\infty} \int_{|x|\geq n} f(x)dx = 0$, (a formal proof of this point follows from the integrability of the pdf and the dominated convergence theorem). A definition of a joint pdf, $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$, for variables $\{X_1,\ldots,X_n\}$, can be found in [5], p78, with Laplace's generalisation of independence that $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)X_n$ appearing on p81 (Theorem 6B). A nonstandard formulation of independence, with its relation to the standard definition, can be found in [8], chapter 7, Definition 7.1 and Theorem 7.2, see also [1]

[5] Laplace construes the upper integral in the text as varying indefinitely between limits in the interval $[0, 1]$

cet evenement, $\frac{\int x^p(1-x)^q dx}{\int x^p(1-x)^q dx}$ sera, par l'article XV, la probabitite de cette cause, ... ", ([6]).

Laplace use his method of finite differences to calculate $\int x^p(1-x)^q dx$, comparing the two expressions (λ), (p424), and (γ), (p433);

"Cette suite est, dans les differences tinies, ce qu'est la suite (λ) de l'article XVIII dans les differences infiniment petites."

In his "Mmoire sur divers points dAnalyse", of 1809, (Journal de lcol Polytechnique, Tome VIII., 229-265, 1809, Oeuvres Completes, XIV, 178-214), which we considered above, Laplace develops the calculus of generating functions in Section 1, "Sur le calcul des fonctions generatrices". These represent functions as the coefficients of power series, see Chapter 5, and are required for his work in probability. He makes the important observation on p187;

"Soit u une fonction des deux quantites t et t', et concevons qu'en la developpant dans une serie ordonnee par rapport aux puissances de t et de t', $y_{x,x'}$ soit le coefficient de $t^x t^{x'}$ dans cette serie; u sera la fonction generatrice de $y_{x,x'}$; $u[(\frac{1}{t}-1)^2-(\frac{1}{t'}-1)]$ sera la fonction generatrice de $\Delta^2 y_{x,x'} - \Delta' y_{x,x'}$, la caracteristique Δ etant relative a la variable x, et la caracteristique Δ' a la variable x'."

relating generating functions to the method of solving the heat equation, by finite differences.

At about the same time, Laplace finished his "Memoire les Approximations des Formules qui sont Fonctions de Tres Grands Nombres et sur leur Application aux Probabilites", (Memoires de lInstitut de France, 1st Series, T. X, year 1809, (1810), pp. 383389. Oeuvres completes XII, pp. 301353), in which he proves the Central Limit Theorem. Laplace introduces the idea of a characteristic function for a discrete random variable, in Section 6, p322, observing that the corresponding function for n independent variables, can be obtained by

[6]Formally, we can generalise the principle of Laplace's first paper to continuous distributions;

$p(x|y)p(y) = p(y|x)p(x)$ (Bayes's Theorem)

$p(y) = \int p(y|x)p(x)dx$ (Conditioning on Events)

to obtain $p(x|y) = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$. Then, letting $p(y,z|x)$ denote the probability of there being y boys and z girls, in a sample of size $p+q$, given the probability x of a boy, $p(x)$ the distribution of x, and $p(x|y,z)$ the probability of being a boy, given y boys and z girls, in a sample of size $p+q$, we have that;

$p(p,q|x) = \lambda x^p(1-x)^q$, where $(\lambda = C_p^{p+q})$, $p(x) = 1_{[0,1]}$

and

$p(p,q) = \int_0^1 \lambda x^p(1-x)^q dx$. Then;

$p(x|p,q) = \frac{p(p,q|x)p(x)}{\int_0^1 p(p,q|x)p(x)dx} = \frac{\lambda x^p(1-x)^q}{\int_0^1 \lambda x^p(1-x)^q dx} = \frac{x^p(1-x)^q}{\int_0^1 x^p(1-x)^q dx}$

taking the $n'th$ power;

"Cela pose, representons par $\phi(\frac{s}{i+i'})$ la probabilite de l'erreur s pour chaque observation, et considerons la fonction

$$\phi(\frac{-i}{i+i'})e^{-i\omega\sqrt{-1}}+\phi(\frac{-(i-1)}{i+i'})e^{-(i-1)\omega\sqrt{-1}}+\ldots\phi(\frac{0}{i+i'})+\ldots+\phi(\frac{i'-1}{i+i'})e^{(i'-1)\omega\sqrt{-1}}+\phi(\frac{i'}{i+i'})e^{i'\omega\sqrt{-1}}$$

En elevant cette fonction a la puissance n, le coefficient de $e^{r\omega\sqrt{-1}}$ dans le developpement de cette puissance sera la probabilite que la somme des erreurs de n observations sera r, ... ", ([7]). Using the infinitesimal method, he can also use this definition for a continuous random variable, defining, on p323, the nonstandard versions $\frac{q'}{h}$ and k' of the expectation and variance;

"q' est la abcisse correspondante a l'ordonnee du centre de gravite de l'aire de la courbe"

...

$\frac{k'=\int\frac{(x'-q')^2}{h^2}\phi(x}{h)}$", ([8]), and, obtaining an expression for the characteristic function on $p324$;

"$e^{-\frac{k'}{2k}nt^2}(1+Ant^4+\ldots)$"

On $p324$, Laplace uses the inversion theorem, which, in the previous section, I have suggested was known by him at this time;

"Si, conformement a l'analyse de l'article IV, on multiplie la fonction (o) par $2cosl\omega$, le terme independant de ω dans le produit exprimera la probabilite que la somme des erreurs sera ou $nq-l$ ou $nq+l$"

to obtain the probability distribution of the sum $X_1+\ldots+X_n$, and, then, on p325, finds the normal probability distribution of $\frac{X_1+\ldots+X_n}{\sqrt{n}}$, for infinite n;

"Si l'on multiplie cette fonction par dl, en integrant on aura la probabilite que la somme des erreurs sera comprise dans les limites $nq\pm l$ ou $nq\pm(i+i')r\sqrt{n}$; or on a;

$dl=(i+i')dr\sqrt{n}$;

cette probabilite sera donc

[7] Formally, the characteristic function of a random variable X, is given by $c_X(t)=E(e^{itX})$, where E denotes the expectation or mean value. For n independent random variables, with the same probability distribution, $c_{X_1+\ldots+X_n}(t)=E(e^{it(X_1+\ldots X_n)})=[E(e^{itX})]^n$, as $E(Y_1\ldots Y_n)=E(Y_1)\cdot\ldots\cdot E(Y_n)$, for n independent random variables $\{Y_1,\ldots,Y_n\}$. The corresponding property for n independent nonstandard random variables can be found in Lemma 3 of [9].

[8] The nonstandard definitions of expectation and variance can again be found in Lemma 3 of [9]

$\frac{2}{\sqrt{\pi}}\sqrt{\frac{k}{2k'}}\int e^{\frac{-k}{2k'r^2}}dr$", ([9]) Laplace announced his result to the Academie in the same year.

In the same paper, Section 1, p305, Laplace considers the question of of the probability distribution of the average inclination of n independent satellite orbits, defining the distribution of the independent sum, ([10]);

"et l'on demande la probabilite que l'inclinaison moyenne de n orbites sera comprise dans des limites donnees"

...

Cela pose, nommons $t, t_1, t_2, \ldots$ les inclinations des n orbites, et supposons leur somme egale a s, nous aurons

$t + t_1 + \ldots t_{n-1} = s$".

After deriving the formula (a), on p307, for this defined distribution, Laplace recognises some connection with the generatrice functions used in his "Mmoire sur divers points dAnalyse", and obtains the probability distribution, on p317 for the orbit inclinations, presumably by solving the associated nonstandard difference heat equation;

"Si l'on neglige les termes de l'ordre $\frac{1}{n}$, l'integrale $\frac{2}{\sqrt{\pi}}\int dse^{-t^2}$ ou $\frac{3}{\sqrt{\pi}}\sqrt{\frac{3}{2}}\int dre^{\frac{-3}{2}}r^2$ exprime la probabilite que la somme des inclinations des orbites sera comprise dans les limites $\frac{h}{2} - \frac{rh}{2\sqrt{n}}$ et $\frac{h}{2} + \frac{rh}{2\sqrt{n}}$."

a particular case of his later general formula in the central limit theorem. Laplace is, here, and in his proof of the central limit theorem, generalising the result that Brownian motion can be obtain from a random walk, with coin tossings as the individual steps, ([11]). Figure 23 clarifies this idea. The first part of the diagram illustrates a possible path $R(x_0, t)$, obtained as the "average" of discrete independent observations $\{\omega_1, \ldots, \omega_n\}$ over time, the randomness of the motion resulting from the assumption of independence. For discrete times t_i, with a fixed number of observations, the possible values $R(x, t_i)$, corresponding to the final values of possible paths of duration t_i, display an increasingly random and irregular pattern, but, as each segment splits over successive times, the total area is preserved; in this sense, there is one rather than two degrees of freedom, for the partition, at each step. The second part shows the continuous analogue of this process, known as Brownian motion, obtained over continuous time t, after using infinitesimally small time steps, with a possible paths $W(x_0, t)$ and time slices $W(x, t_i)$. An important property is that "almost every" path

[9] Laplace's nonstandard proof of the central limit theorem is very similar to the accepted standard proof, see p130 of [5]. A statement of the nonstandard version is given in Theorem 7.4 of [8], see also [1], though its proof uses the standard formulation of the result.

[10] Here, as in the proof of the central limit theorem, Laplace construes the average of a sample of n elements, by dividing through with $\sqrt{n}$

[11] See Definition 7.7 of [8]. A nonstandard proof of this last result can be found in [8], Theorem 7.8, based on [1].

is continuous but nowhere differentiable. The averaging behaviour continues, the defining charateristic of a more general class of processes called martingales, ([12]). These, in fact, originated as betting strategies in France, during the 18th century, ([13]), and it seems reasonable that Laplace knew of their existence, ([14]). The fact that martingales, arising from gambling, are equivalent to processes with the above averaging behaviour follows from Theorem 0.5 and Theorem 0.13 of [9];

Theorem 0.1. *Any martingale X is representable as a stochastic integral, and, conversely, every stochastic integral is equivalent to a martingale.*

Proof: The idea is to lift a standard martingale X to a nonstandard martingale $\overline{X}$, and show that for infinite numbers $\{\eta, \nu\}$, with $\eta = 2^\nu$;

$\overline{X}_t(x) = \sum_{j=0}^{[\nu t]} c_j(t,x)\omega_j(x)$

where the functions c_j depend only on the previous information. Under certain technical restrictions, one can then show that the right hand side sum corresponds to a stochastic integral, defined in terms of Brownian motion, rather than the steps of a random walk. The converse involves showing that such sums have the required averaging behaviour, using the fact that $E(\omega_j|F_{j-1}) = 0$, for F_{j-1} representing the past information; that is each ω_j is obtained as an innovation from 0 at time t_{j-1}. Observe the similarity with Theorems 1.5 and 1.6 from the previous section, ([15]).

Laplace's alternative derivation of the distribution, using the heat equation, is a precursor to a general method of finding the (time-evolving) probability distributions of martingales, or more complex processes, ([16]), and investigation of his papers could still lead to interesting new insights into the subject. Indeed, in the subsequent "Memoire sur les Integrales Definies

[12]See Footnote 2 and Definition 0.7 of [9] for the standard and nonstandard definitions

[13]Based on a fair game with independent, identically distributed random variables, $\{\omega_1, \ldots \omega_n\}$, a martingale strategy is to bet on ω_n at time n as a function e_n of the previous results $\{\omega_1, \ldots, \omega_{n-1}\}$, the payoff $Y_n = e_1\omega_1 + e_2(\omega_1)\omega_2 + \ldots + e_n(\omega_1, \ldots, \omega_{n-1})\omega_n$ is known as a martingale. The continuous analogues of such payoffs are known as stochastic integrals, denoted by $Y_t = \int e(t,x)dW_t$

[14]Indeed, in his later "Essai Philosophique sur les Probabilites", (1814), within the chapter "Applications du Calcul des Probabilites Des Jeux", (p40), Laplace formulates the idea behind martingale strategies for fair games;

"Que l'on jette dans une urne, cent numeros depuis un jusqu'a cent, dans l'ordre de la numeration, et qu'apres avoir agite l'urne, pour meler ces numeros, on et tire un; il est clair que si le melange a ete bien fait, les probabilites de sorte des numeros, seront les memes. Mais si l'on craint qu'il n'y ait entre elles, de petites differences dependentes de l'ordre suivant lequel les numeros ont ete jetes dans l'urne; on diminuera considerablement ces differences, en jetant dans une seconde urne, ces numeros suivant leur ordre de sortie de la premier urne, et en agitant ensuite, cette seconde, pour meler ces numeros. Une troisieme unrne, une quatrieme, etc., diminueraient de plus ces differences, deja insensibles dans la seconde urne"

[15]A standard proof of this result can be found in [13], for other general results on martingales, see [15]. Viewing an algebraic curve as a probability space, one can also hope to generalise the theorem to this setting.

[16]The Fokker-Planck Theorem gives a partial differential equation $\frac{\partial p}{\partial t} = \frac{\partial^2 Dp}{\partial x^2}$ for the probability distributions $p(x,t)$ of random variables X_t, satisfying $X_t = \int_0^t \sqrt{2D(X_s,s)}dW_s$, usually abbreviated as a "stochastic differential equation", $dX_t = \sqrt{2D(X_t,t)}dW_t$. In the case that $D = 1$, we obtain the heat equation for the probability distribution of rescaled Brownian motion, as $\sqrt{2}\int_0^t dW_s = \sqrt{2}W_t$. In certain cases, when the

et leur Application aux Probabilites", (Memoires de lAcademie des Sciences, 1810 (1811), Ouevres Completes XII,pp357-412), Section 5, Laplace succesfully finds a new differential equation for the following probability distribution;

"Considerons deux urnes A et B, renfermant chacune le memenombre n de boules; et supposons que, dans le nombre total $2n$ de boules, il y en ait autant de blanches que de noires. Concevons que l'on tire en meme temps une boule de chaque urne, et qu'ensuito on mette dans une urne la boule extrait de l'autre. Supposons que l'on repete cette operation un nombre quelconque r de fois, en agitant a chaque fois les urnes pour en bien mcler les boules: et chercons la probabilite qu'apres ce nombre r d'operations il y aura x boules blanches dans l'urne A. Soit $z_{x,r}$ cette probahilite;", ([17]).

After obtaining a difference equation for this probability, he claims;

" $\ldots\ z_{x,r} = U$;

the preceding equation in the partial differences will become, by neglecting the terms of order $\frac{1}{n^2}$,

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

Laplace immediately followed his work on the Central Limit Theorem, with the "Supplement au Memoire sur les Approximations des Formules qui sont Fonctions Dde Tres Grands Nombres", (Memoires de lAcademie des Sciences, 1st Series, T. X, 1809, (1810), Oeuvres Completes T. XII. pp. 349353). In this short work, Laplace begins, on p351, by formulating a principle of maximum likelihood;

"Soient donc l la distance du point qu'il faut choisir a l'origine de la courbe des probabilites, et z l'abscisse correspondante a y et comptee de la meme origine; le produit de chaque erreur par sa probabilite, abstraction faite du signe, sera $(l - z)y$, depuis $z == 0$ jusqu'a $z == l$, et ce produit sera $(z - l)y$ depuis $z == l$ jusqu'a l'extremite de la courbe. On aura donc

integrand of a martingale representation is a function of the original martingale, or, when a stochastic equation of the above form is given, the theorem provides a method of calculating the required distributions. In formulating a nonstandard analogue of Fokker-Planck, one just needs to follow the method of [1] and [8], Definition 7.20, Remarks 7.21, in giving a nonstandard version of the stochastic integral, and replacing the differential equation with a difference equation. The reader might be interested in deciding whether the nonstandard formulation is true. An alternative nonstandard method of solving SDE's directly, in terms of the random variable, again generalising Laplace's finite difference method, can be found in [6].

[17]"We consider two urns A and B, each containing the same number n of balls; and we suppose that, in the total number $2n$ of balls, there are as many of them white as of black. We imagine that we draw at the same time a ball from each urn, and that next we put into one urn the ball extracted from the other. We suppose that we repeat this operation any number r times, by agitating at each time the urns in order to well mix the balls; and we seek the probability that after this number r of operations there will be x white balls in urn A. Let $z_{x,r}$ be this probability;"

$\int(l-z)ydz + \int(z-l)ydz$

pour la somme de tous ces produits, la premiere integrale etant prise depuis z nul jusqu'a $z = l$, et la seconde etant prise depuis $z == l$ jusqu'a la derniere valeur de z. En differentiant la somme precedente par rapport a l, il est facile de s'assurer que l'on aura

$dl \int ydz - dl \int ydz$

pour cette diffcrcntielle, qui doit etre nulle dans le cas du minimum; on adonc alors

$\int ydz = \int ydz$", ([18]). Laplace formulates the likelihood function for n independent observations on p352, using his result from above;

"Dans le cas present, on a, en faisant $x = X + z$,

$y = pp'p'' \ldots e^{-p^2\pi_1 X + z_1^2 - p'^2\pi q - X - z_1^2 - p''^2\pi q' - X - z_1^2 - \ldots}$

p etant egal a $\frac{a\sqrt{n}}{\sqrt{pi}}$, et par consequent exprimant la plus grande probabilite du resultat donne par les observations n; p' exprime pareillement la plus grande ordonnee relative aux observations n', et ainsi du reste: r pouvant, sans erreur sensible, s'etendre depuis -∞ jusqu'a $+\infty$, comme on l'a vu dans l'article VII du memoire cite, on peut prendre dans les memes limites, et alors si l'on choisit X de maniere que la premiere puissance de z disparaisse de l'exposant de c, l'ordonnee y correspondante a z nul divisera l'aire de la courbe en deux parties egales, et sera en meme temps la plus grande ordonnee. En effet, on a, dans ce cas,

$X = \frac{p'^2 q + p''^2 q' + \ldots}{p^2 + p'^2 + p''^2 + \ldots}$".

Laplace then finds the maximum of this function, to obtain a parameter estimate X, as a function of the independent observations $\{q, q', \ldots, q_i, \ldots, q_n : 1 \leq i \leq n\}$, ([19]). On p353, Laplace makes the observation;

---

[18] Laplace minimises the error, relative to the unknown true mean l_0, of a set of observations $\{z_i : 1 \leq i \leq n\}$, (corresponding to independent random variables $\{Y_i : 1 \leq i \leq n\}$), with a known probability distribution $y(z,l)$. Laplace maximizes the function $f(l) = \int_{z \leq l}(l-z)ydz + \int_{z \geq l}(z-l)ydz$, obtaining the equation $\int_{z \leq l} ydz = \int_{z \geq l} ydz$, to estimate l_0. It is important to realise that the estimators $(\hat{l})_n$ should be functions of the random variables $\{Y_i : 1 \leq i \leq n\}$; some work is required to show this, here, interpreting the integral as a sum over the observations. An estimator is said to be unbiased if $E(\hat{l}) = l_0$, and consistent if $lim_{n\rightarrow\infty}(\hat{l})_n = l_0$; Laplace may be trying to estimate the median rather the mode here, and, for a certain class of distributions, these properties may hold.

[19] Laplace expresses the the likelihood $y_n(q_1, \ldots, q_n, l) = \prod_{i=1}^n y(q_i, l)$, and solves $\frac{dy_n}{dl} = 0$, to obtain the estimator X. A more efficient, but equivalent, method, is to solve $\frac{dlog(y_n)}{dl} = 0$, or, equivalently, $\sum_{i=1}^n logy(q_i, l) = 0$, to obtain the estimators $(\hat{l})_n$. Under certain assumptions on the probability distribution, the crucial one being that it has a unique local maximum, the estimators are consistent, see footnote 18 and [14], but, are not necessarily unbiased. The proof of consistency in [14] relies on the Strong Law of Large Numbers, a particular case of the Ergodic Theorem, see [11], and [8] for a nonstandard proof.

"La valeur precedente de X est celle qui rend un minimum la fonction

$$(pX)^2 + [p'(q - X)]^2] + [p''(q' - X)]^2 + \ldots$$

c'est-a-dire la somme des carres des erreurs de chaque resultat, multipliee respectivement par la plus grande ordonnee de la courbe de facilite de ses erreurs."

that the above procedure is equivalent to estimating the slope of a "line of best fit" by minimising the sum of squares of the residuals. Figure 24 illustrates the idea of least squares, [20], with a collection of data points $\{(p_i, q_i) : 1 \leq i \leq n\}$, obeying the relation $q_i = (\lambda_0 p_i + c_0) + u_i$, where the slope λ_0 and intercept c_0 are fixed, and the residuals, or error terms, u_i, are drawn from a fixed distribution, with unknown variance, ([21]). If the errors terms are drawn from a normal distribution, which is assumed here, then, as Laplace points out correctly, the method of maximum likelihood yields the same estimators $\{(\hat{\lambda})_n, (\hat{c})_n\}$ for the slope and intercept, as least squares, in particularly such estimators are consistent. These estimators are also unbiased, ([22]), and, in the final paragraph, Laplace applies his central limit theorem of the previous paper, observing;

"Ainsi cette propriete, qui n'est qu'hypothetique lorsqu'on ne considere que des resultats donnes par une seule observation on par un petit nombre d'observations, devient necessaire lorsque les resultats outre lesquels on doit prendre un milieu sont donnes chacun par un tres grandes nombre d'observations, quelles que soient d'ailleurs les lois de facilite des erreurs de ces observations.

...

la probabilite que l'erreur de resultat choisi $A+X$ sera comprise dans les limites $\pm\frac{T}{\sqrt{n}}$ sera;

$\frac{2\int dte^{-t^2}}{\sqrt{\pi}}$"

that the estimators are asymptotically normal, ([23])

[20]Originally introduced by Legendre in 1805, but not in the context of probability

[21]Minimising $\sum_{i=1}^{n} u_i^2$, one obtains the estimators $(\hat{\lambda})_n = \frac{s_{p,q,n}}{s_{p,n}^2}$, where $s_{p,q,n} = \frac{\sum_{i=1}^{n}(p_i - \overline{p}_n)(q_i - \overline{q}_n)}{n}$, $s_{p,n}^2 = \frac{\sum_{i=1}^{n}(p_i - \overline{p}_n)}{n}$, $\overline{p}_n = \frac{\sum_{i=1}^{n} p_i}{n}$, and $\overline{q}_n = \frac{\sum_{i=1}^{n} q_i}{n}$ and $(\hat{c})_n = \overline{q}_n - \frac{(\hat{\lambda})_n s_{q,n} \overline{p}_n}{s_{p,n}}$, where $s_{q,n}^2 = \frac{\sum_{i=1}^{n}(q_i - \overline{q}_n)}{n}$ which depend only on the data $\{(p_i, q_i) : 1 \leq i \leq n\}$. Laplace obtains a similar result, but his notation is still unclear.

[22]This was shown in Gauss's contemporary astronomical paper "Theoria motus corporum coelestium in sectionibus conicis solem ambientum" of 1809, in which he also derives the normal distribution. However, interestingly, the estimator obtained using maximimum likelihood for the variance of the residuals is biased, differing from the unbiased estimator by a factor of $\frac{n-1}{n}$

[23]In the sense that $lim_{n\rightarrow\infty}(\sqrt{n}(\hat{\lambda})_n)$ follows a normal distribution. This would also be true for the least squares estimators if the error terms were independently drawn from another distribution, it is not clear whether Laplace realises this.

In 1812, Laplace continued his work in probability, by publishing his "Theorie analytique des probabilites". The first part of this text essentially gives a more mature presentation of his previous work on the theory of generatrices, with some possibly new remarks on the passage from finite to infinitely small quantities, in Chapter 2, Section 19, "Considerations sur le passage du fini a l'infinement petit". In the second part, entitled a "Theorie General des Probabilites", Laplace recalls 4 central principles of probability, used in his earlier work,([24]), which establish his Bayesian philosophy, in the theory of probability. In his "Essai Philosophique sur les Probabilites" of 1814, Laplace reiterates this principles among a more comprehensive list of 10, in the chapter "Principes generaux du Calcul des Probabilites", ([25]).

As we have seen Laplace's major work in physics was his consideration of planetary and satellite motion in the "Mechanique Celeste". Undoubtedly influenced by Fourier's work on the heat equation, he also published the important "Mmoire sur les mouvements de la lumire dans les milieux diaphanes" in 1810, (Mmoires de lAcadmie des Sciences, Ist Srie, Tome X,

---

[24]Namely;

"I^{er} PRINCIPE. La probabilite d'un evenement compose de deux evenements simples, est la produit de la probabilite d'un de ces evenements, par la probabilite que cet evenement etant arrive, l'autre evenement aura lieu"

"II^{e} PRINCIPE. Le probabilte d'un evenement futur, tiree d'un evenement observe, est la quotient de la division de la probabilite de evenement compose de ces deux evenemens, et determinee a priori, par la probailite de l'evenement observe, determinee pareillement a priori"

"III^{e} PRINCIPE. Si un evenement observe peut resulter de n causes differentes; leurs probabilites sont respectivement, comme les probabilites de l'evenement, tirees de leur existence; et la probabilite de chacune d'elles est une fraction dont le numerateur est la probabilite de l'evenement, dans l'hypothese de l'existence de la cause, et dont le denominateur est la somme des probabilites semblables, relatives a toutes les causes"

"IV^{e} La probabilite d'un evenement futur est la somme des produits de la probabilite de chaque cause, tiree de l'evenement observe, par la probabilite de chaque cause, tiree de l'evenement observe, par la probabilite que cette cause existant, l'evenement future aura lieu"

The first two principles with two events $A$and B being that (i). $P(A) = P(A|B)P(B)$ and (ii). $P(A|B) = \frac{P(A\cap B)}{P(B)}$

The second two principles, for an event B and exclusive events $\{A_i : 1 \leq i \leq n\}$, that (iii). $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$, for $(1 \leq i \leq n)$ and (iv). $P(B) = \sum_{j=1}^{n} P(B|A_j)P(A_j)$.

[25]The position of Bayesianism is usually contrasted with that of frequentism, in which events are usually construed as having a well defined, predetermined probability, rather than being continually updated by new evidence, based on Bayes's theorem. For the most part, in the modern theory, these approaches seem to have been reconciled, for example the simultaneous treatment of the true value of a parameter in a distribution versus its estimates as random variables obtained by maximum likelihood. There still appear to be some remaining issues in the interpretation of hypotheses, modern Bayesianists preferring to construe these as random variables, rather than either true or false, the latter; I believe, still being the usual approach. It is an interesting point that, although this is a question mainly of interpreting evidence, rather than any paradox in mathematics, the two interpretations can lead to very different decisions on whether to accept or reject a particular hypotheses, as in, for example, "Lindley's Paradox".

Oeuvres Completes de Laplace, Vol 12, 267-298). Clearly aware of Huygen's wave theory of light in "Traite de la Lumiere", Laplace discusses the idea of refraction in crystals, and, on p287, advocating Newton's theory of light as a particle, explains this effect in terms of the reflections from the molecular surface;

"Les memes resultats ont lieu relativement aux rayons extraordinaires; car, sans connaitre la cause de la refraction extraordinaire, on peut cependant assurer qu'elle est due a des forces attractives et repulsives qui agissent de molecule a molecule, suivant des fonctions quelconques de la distance, et qui, dans les cristaux, sont modifiees par la figure de leurs molecules integrantes, par celle des molecules de la lumiere et par la maniere dont ces molecules se presentent les unes aux autres."

He goes on to derive the heat equation for the propagation of temperature by the transmission of light molecules, quoting Newton's principle;

"On est parti, dans la theorie de l'equilibre et du mouvement de la chaleur, de ce principe, donne par Newton, savoir que la chaleur communiquee par un corps a un autre qui lui est contigu est proportionnelle a la difference de leurs temperatures"

It is interesting that Laplace was unable to incorporate his ideas from probability theory in this work, although this featured heavily in his discussion of the inclinations of orbits, ([26]).

References

[1] A Non-Standard Representation for Brownian Motion and Ito Integration, Robert M. Anderson, Israel Journal of Mathematics, Vol. 25, (1976).

[2] Statistical Inference, G. Casella, R. Berger, Thomson Learning, (2002).

[3] Macroscopic and Staistical Thermodynamics, Yi-Chen Cheng, World Scientific Publishing, (2006).

[4] Thermodynamics, E. Fermi, Dover, (2000).

[5] Probability, An Introduction, Geoffrey Grimmett and Dominic Walsh, OUP, (1986).

[6] An Infinitesimal Approach to Stochastic Analysis, Jerome Keisler, AMS, (1984).

[7] Quantum Mechanics, Volume 2, Albert Messiah, Elsevier, (1961).

[8] Applications of Nonstandard Analysis to Probability Theory, Tristram de Piro, available at http://www.magneticstrix.net, (2013)

[26]The application of ideas of probability theory to random particle motion could still be an interesting research topic. The process of reversing martingales to model such effects, and as an efficient method of propagating heat, or concentration, can be found in [10]. The use of stochastic differential equations can be found in [3], and [7] has a discussion of statistical results in physics. Older approaches ,including the laws of thermodynamics, can be found in [4] and [12].

[9] A Simple Proof of the Martingale Representation Theorem using Nonstandard Analysis, Tristram de Piro, available at http://www.magneticstrix.net, (2013).

[10] Electron Bunching, Tristram de Piro, available at http://www.magneticstrix.net, (2014).

[11] Ergodic Theory, K. Peterson, CUP, (2012).

[12] Elements of Classical Thermodynamics, A.B. Pippard, CUP, (1957).

[13] Stochastic Calculus and Financial Applications, J.M. Steele, Springer, (2010).

[14] Note on the Consistency of the Maximum Likelihood Estimate, Abraham Wald, The Annals of Mathematical Statistics 20, (1949) .

[15] Probability with Martingales, D. Williams, CUP, (2000).