7 Laplace's probability

0 Texts

Stigler, Stephen(1986) *The History of Statistics : The Measurement of Uncertainty Before 1900*
Laplace, Pierre Simon(1774a) Mémoire sur les suites récurro-récurrentes et sur les usages dans la théorie des hasards. *Mémoires de l'Académie royale des sciences presentés par divers savans* 6:353-371
Laplace, Pierre Simon(1774) Mémoire sur la probabilité des causes par les évènemens. *Mémoires de l'Académie royale des sciences presentés par divers savans* 6:621-56. Translated in Stigler(1986).

Stigler, Stephen (1986) Laplace's 1774 memoir on inverse probability. *Statistical Science* 1
Laplace, Pierre Simon(1776) Recherches sur 1'intégration des équations differentielles aux différences fillies et sur leur usage dans la théorie des hazards . *Mémoires de l'Académie royale des sciences presentés par divers savans* 7:113-é163
Laplace, Pierre Simon(1778) Mémoire sur les probabilités. *Mémoires de l'Académie royale des Sciences de Paris*, 227-332.
Laplace, Pierre Simon (1812) *Théorie analytique des probabilités.* Paris: Courcier
Laplace, Pierre Simon (1814) *Essai philosophique sur les probabilités*. Paris:Courcier. Reprinted 1951 New York:Dover

Todhunter, I (1865) *A history of the mathematical theory of probability* New York, Chelsea (1949).
Dale, Andrew I. (1999) *A history of inverse probability* Springer, New York
Daston, Lorraine (1988) *Classical probability in the Enlightenment* Princeton University Press, Princeton.
www.Britannica.com

1 Probability definitions

The inverse probability statement of Bayes, which gives the posterior distribution of a binomial parameter, requires two senses of probability; The binomial parameter p is an *aleatory* probability in Hacking's terminology, a fixed physical quantity determined by the gambling equipment; the probability that this parameter lies in a given range is an *epistemic* probability, dependent on our knowledge about the gambling equipment obtained in a sequence of trials. Bayes uses the word *chance* to describe this second form of probability, though he asserts the two words mean the same thing. Today, the word chance would be used to describe the physical parameter rather than the epistemic probability.

Let us see how Laplace's views evolved on these two uses of probability
I found Dale's book on the history of inverse probability invaluable for this purpose.

The first definition of probability given by Laplace is in a paper appearing in 1774 prior to his 'causes' paper:

La probabilité d'un événement est égale a la somme des produits de chaque cas favorable par sa probabilité divisé par la somme des produits de chaque cas possible par sa probabilité, et si chaque cas est également probable, la probabilité de l'évènement est égale au nombre des cas favorables divisé par le nombre de tous les cas possibles.

Laplace states this as a principle rather than a definition. The first clause translates as the probability of an event is the sum of probabilities of favourable events divided by the sum of probabilities of possible events. The second clause states that the probability of an event is the number of favourable cases divided by the number of possible cases.

The first clause can hardly be taken as a definition of probability, since it requires the probabilities of the cases.

In Laplace(1774) ' causes', he gives a justification for assignment of equal probabilities:

On suppose dans la théorie [i.e. des probabilités1 que les différents cas qui améun événement sont également probables, ou, s'ils ne le sont pas, que leur probabilité est dans un rapport donné. Quand on veut ensuite faire usage de cette théorie, on regarde *deux événements comme également probables, lorsqu'on ne voit aucune raison qui rende l'un plus probable que l'autre, parce que, quand bien meme il y aurait une inégale possibilité entre eux, comme nous ignorons de quel côtéla plus grande, cette incertitude nous fait regarder l'un comme aussi probable que l'autre. Lorsqu'il n 'est question que de probabilités simples, il paraît que cette inégalité de probabilités ne nuit en rien à la justesse de l'application du calcul aux objets physiques. ..mais, lorsqu'il s'agit de probabilité composé, il me semble que l'application que l'on fait de la théorie aux évènements physiques demande à être modifièe.*

This passage contains the first expression of the principle of insufficient reason, stating that two events are equally probable if we see no reason that makes one more probable than another. Such an idea of probability is epistemic, referring to what we know about the events.
In the latter part of the paragraph, Laplace says that for simple probabilities, this inequality of probabilities does not seem to invalidat applying the calculus to physical objects, but for composite probabilities, the theory may need to be modified. I have no idea what he means.

In Laplace(1776):

Nous regardons une chose comme l'effet du hasard, lorsqu'elle n'offre à nos yeux rien de régulier, ou qui allnonce ull dessein, et que nous ignorans d 'ailleurs les causes qui 1 'ont produite. Le hasard n'a donc aucune réalité en lui-même; ce n'est qu'un terme propre à désigner notre ignorance sur la maniére dont les différentes parties d 'ull phénoméne se coordonnent entre elles et avec le reste de la Nature.
La notion de probabilité tient à cette ignorance. Si nous sommes assurés que, sur deux évènements qui ne peuvent exister ensemble, I 'ull ou 1 'autre doit nécessairement arriver, et que nous ne voyons aucune raison pour laquelle l'ull arriverait plutôt que \'autre, l'existence et la non-existence de chacun d'eux est également probable.

Translated, *hasard*, chance means nothing in itself, it only expresses our ignorance about the relationships between the different parts of the phenomenon and the rest of nature. DeMoivre and Hume say something like this also, that chance does not really exist, saying something is caused by chance means only that we do not know the causes.

And then, two events are equally probable if we see no reason why one should happen rather than the other.
And these ideas are then expressed in an explicit definition:

la probabilité de l'existence d'un événement n'est ainsi que le rapport du nombre des cases favorables à celui de tous les cas possibles, lorsque nous ne voyons d'ailleurs aucune raison pour laquelle l'un de ces cases arriverait plut\^ot que l'autre.

2 Possibilities

Laplace(1778): *je me propose de traiter dans ce Mémoire deux points importants de l'analyse des hasards qui ne paraissent point avoir encore été suffisamment approfondis: le premier a pour objet la maniére de calculer la probabilité des événements composés d'événements simples dont on ignore des possibilités respectives; l'objet du second est l'influence des événements passés sur la probabilité des événements futurs, et la loi suivant laquelle, en be développant, ils nous font connaître les causes qui les onts produits.*

This passage is notable for the first appearance of 'possibility', Laplace's other word for probability. He will show us how to determine the probabilities of events when the possibilities are unknown.
Laplace proposes three ways to determine the probability of events:

1. a priori, lorsque, par la nature même des événements, on voit qu'ils sont possibles dans un rapport donné

2. a posteriori, en répétant un grand nombre de fois l'expérience qui peut amener l'événement dont il s'agit, et en examinant combien de fois il est arrivé

3.enfin, par la considération des motifs qui peuvent nor déterminer à prononcer sur l'existence de cet événement.

Laplace applies these methods to the outcomes of a series of games between two gamblers A and B, in which one of the gamblers , say A, has probability $(1+\alpha)/2$ of winning. If α has probability $\varphi(\alpha)$, then the probability of A winning the first n matches is $\int_0^1 (\frac{1+\alpha}{2})^n \varphi(\alpha) d\alpha$, the first appearance of non-uniform prior probabilities.
If " *en sorte que toutes les valeurs de α soient également possibles* ", then the Bayesian probability of the event A winning the first n matches , $\frac{1}{n+1}$ comes from integrating with $\varphi(\alpha) = 1$. It seems as if equally probable and equally possible are being used interchangeably.

Yet Laplace notes that the " law of possibility" , $(1+\alpha)/2$, can be determined only by a long sequence of observations, from which it seems as if possibility is the *aleatory* sense of probability.
(In this same paper, Laplace introduces his method for approximating $\int x^p (1-x)^q dx$, which essentially approximates the whole integral by an integral in a small interval about the maximum.)

3 Circularities

In the Théorie, he defines probability circularly:

la probabilité d 'un événement est le rapport du nombre des cas qui lui sont favorables au nombre de tous les cas possibles,lorsque rien ne porte à croire que 1 'un de ces cas doit arriver plutôt que les autres, ce qui les rend, pour nous, également possibles ....

si tous les cas ne sont pas egalement possibles, on determinera leurs possibilites respectives, et alors la probabilite de l'événements, sera la somme des probabilités de chaque cas favorable.

and he goes on to give an inverse theorem circularly,

VI. Chacune des causes, auxquelles un événement observé peutêtre attribue, est indiquee avec d 'autant plus de vraisemblance, qu 'il est plus probable que, cette cause étant supposée exister, 1'événement aura lieu; la probabilité de 1'existence d 'une quelconque de ces causes est donc une fraction dont le numerateur est la probabilité de 1'événement , resultante de cette cause, et dont le denominateur est la somme des probabilités semblables relatives a toutes les causes: si ces diverses causes considérés a priori sont inegalement probables, il faut au lieu de la probabilité de l'événement, result ante de chaque cause, employer ,le produit de cette probabilité, par la possibilité de la cause j'l elle-meme. C'est le principe fondamental de cette branche de 1'Analyse des hasards, qui consiste a remonter des événements aux causes.

So, it seems to me, Laplace is using a circular definition, switching between probabilities and possibilities, sometimes using one in an aleatory sense as an unknown probability, sometimes using it in an epistemic sense as a probability relative to knowledge.

4 Combination of observations

Problem 3 *Dèterminer le milieu que l'on doit prendre entre trois observations donnès d'un même phènoméne.*

Laplace's milieu is actually the median, which Laplace showed minimized the expected absolute error.

From Stigler, p106

Laplace's discussion was conducted in reference to three figures (see Figure 3.1). Laplace considered the problem to be as follows: Given three observed times of a phenomenon (a, b, and c) along the time axis AB, find the point V that we should take as the true time of the phenomenon. In determining this point he supposed that if V is the true instant of the phenomenon, then the probability of an observation differing from the truth by an amount x was given by a curve y = φ(x), shown as ORM in the middle diagram. Two points remained to be settled. What curve should be taken as the error curve φ(x) And given φ(x), how should the mean be determined?

Laplace listed three conditions that φ(x) should satisfy: First, the curve should be symmetrical about V "because it is just as probable that the observation deviates from the truth to the right as to the left. ' Second, the curve must decrease toward the axis KP "because the probability that the observation differs from the truth by an infinite distance is evidently zero." Third, the area under the curve must be one "because it is certain that the observation will fall on one of the points of the line KP." These conditions did not determine φ(x), but they did permit some general reflections on the problem… As Laplace wrote, "But of an infinite number of possible functions, which choice is to be preferred?"

Laplace argued that there was "no reason to suppose a different law for the ordinates than for their differences", and this leads to the Laplace or double exponential error distribution $\varphi(x) = \exp(-m\,|\,x\,|)/2m$.

To solve the problem of finding the posterior median of V|a,b,c, Laplace first computes $\varphi(V\,|\,a,b,c,m) = \dfrac{\varphi(a-V)\varphi(b-V)\varphi(c-V)}{\int \varphi(a-V)\varphi(b-V)\varphi(c-V)dV}$ ***which depends on the unknown scale parameter m. Laplace finds the posterior distribution of m given a,b,c,*** $\varphi(m\,|\,a,b,c)$. ***Then he computes the marginal posterior integrating out m,***

$$\varphi(V\,|\,a,b,c) = \int \varphi(V\,|\,a,b,c,m)\varphi(m\,|\,a,b,c)dm$$

and looks for the median of this posterior distribution as "the mean between the three observations".

This is what he intends to do, although Stigler points out that his actual computation was slightly different. We see that Laplace is using probability in a consistent way to handle all the quantities he is uncertain about, a,b,c,V,m, that is carrying out the characteristic Bayesian analysis of uncertainty.

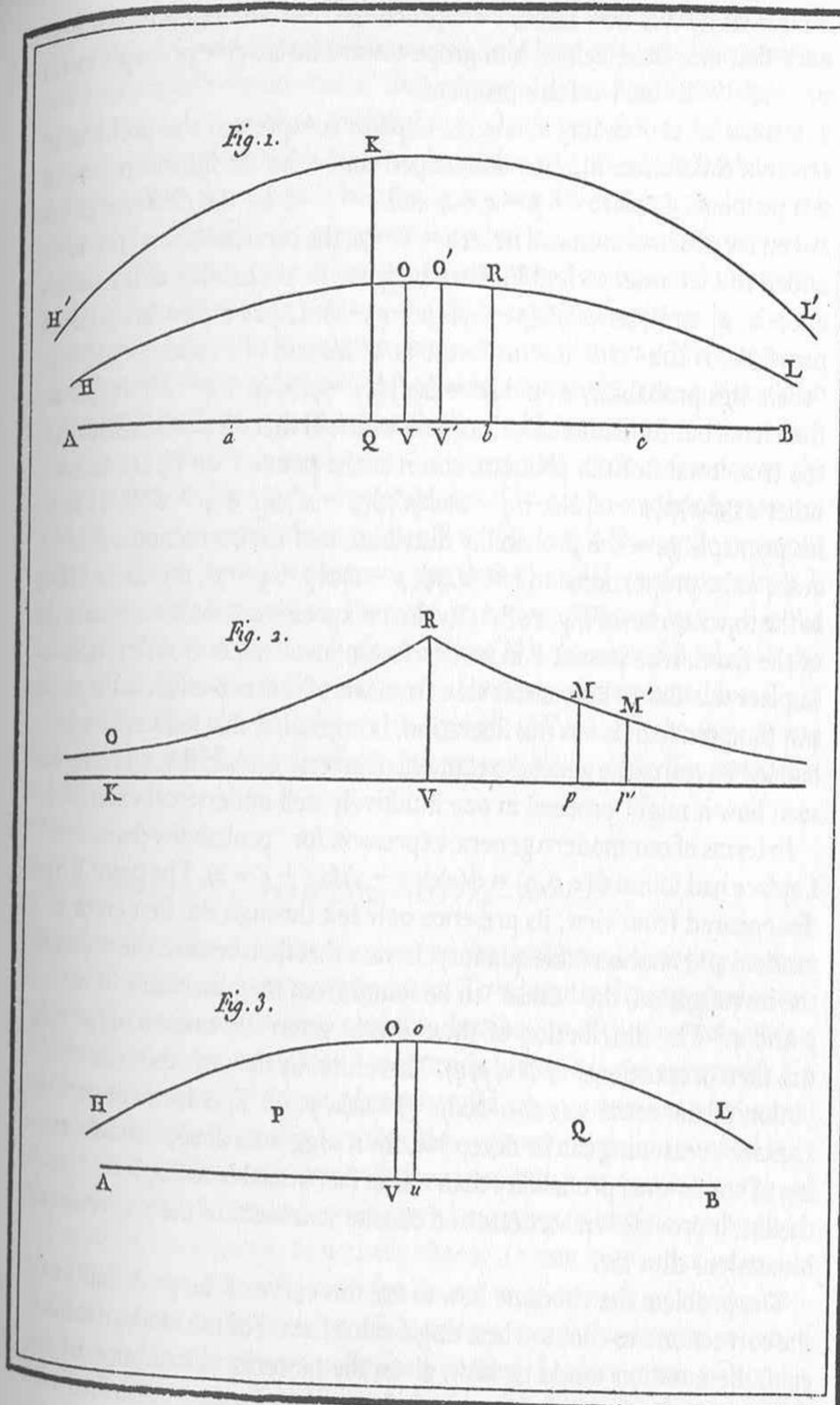

Figure 3.1. Diagrams accompanying Laplace's 1774 memoir on inverse probability. His Figure 2 shows the double exponential density, and his Figures 1 and 3 show posterior distributions, the latter illustrating his argument that the posterior median minimizes the posterior expected error. (From Laplace, 1774, facing p. 656.)