Int. Stat. Rev., Vol. 42, No. 1, 1974, pp. 9–15/*Longman Group Ltd/Printed in Great Britain*

The History of Likelihood

A. W. F. Edwards
Gonville and Caius College, Cambridge

Introduction

One of R. A. Fisher's most influential papers was "On the mathematical foundations of theoretical statistics" (1922), in which he propounded the *method of maximum likelihood* as a means of point estimation, and hence established a whole branch of statistical reasoning. Yet this paper does not contain his original statement of the method, which was published in 1912, nor does it contain his original definition of *likelihood*, which appeared in 1921. The great innovation of the 1922 paper was, rather, the clear specification of one approach to the problem of estimation, and the elucidation of the properties of maximum-likelihood estimators.

Methods similar to the method of maximum likelihood have a history prior to the work of Fisher, but the definition of likelihood itself, and the realization that it could be used independently as a measure of relative support for hypotheses, appears to be entirely his own, and the main purpose of the present paper is to investigate the background to his introduction of the new concept.

With the decline in the esteem with which repeated-sampling methods are held, in the face of the Bayesian revival, the concept of likelihood has come to the fore as offering an approach to statistical inference that is neither repeated-sampling nor Bayesian (see Edwards, 1972). Although it is too early to predict the extent to which pure likelihood methods can supersede other approaches, the concept of likelihood is, in addition, so fundamental to both Bayesian inference and Neyman-Pearson methods that an account of its origins needs no further justification. In the former, the likelihood is the vehicle which carries the observational or experimental results in Bayes' Theorem, whilst in the latter, likelihood was one of the key concepts in the original formulation by Neyman and Pearson (1928). The present paper traces the history of both *likelihood* and the *method of maximum likelihood*; it is essential to keep the distinction between the two clearly in mind.

The Method of Maximum Likelihood

The problem of the estimation of a parameter was first put into something like its modern form in the fourth part of James Bernoulli's posthumous work *Ars Conjectandi*, published in 1713. Bernoulli considered the estimation of a binomial parameter, and after propounding his famous theorem, was able to use it to make "confidence interval" statements which, however, depended upon the unknown parameter (see Hacking, 1971). Fifty years later Bayes' *Essay* was published, also posthumously, containing the solution to Bernoulli's estimation problem in terms of inverse probability, and making quite explicit the postulate required. It was not, however, in connection with the binomial that alternative methods were first advanced, for in science a more pressing problem was to find the " best" value for a parameter whose determination was subject to "error". It was appreciated that the assumed error distribution would be effectively continuous, and some of the early suggestions for estimating the parameter (as we should now say) of a continuous distribution were close to the method of maximum likelihood.

It is well known that if a uniform prior distribution is adopted in a Bayesian formulation of the estimation problem, maximizing the posterior probability is equivalent to maximizing the likelihood. This circumstance has tended to obscure the distinction between the Bayesian and the non-Bayesian justification of the method, and in order to emphasize the distinction I shall refer to the method of maximum *likelihood* when no Bayesian connection is offered, and otherwise to the method of maximum *probability*.

In 1760 Lambert proposed a method of estimation, in his *Photometria*, which is analytically identical to the method of maximum likelihood. The work is not directly mentioned in Todhunter's *History*, but Sheynin (1971) gives a full account. Lambert finds the multinomial probability for four observations from a continuous distribution. Then, dropping the irrelevant multinomial coefficient, he proceeds to maximize the "likelihood" by differentiating the log-likelihood and equating it to zero. Lambert's method was referred to by John Bernoulli in his article on the mean ("milieu") in the second of the volumes on "Mathématiques" of the *Encyclopédie Méthodique*, published in Paris in 1785 (but presumably written much earlier). John Bernoulli did not there describe it, on the grounds that it was already published, but instead devoted most of his article to the unpublished methods of Daniel Bernoulli and Lagrange. John refers to the former as "un petit écrit latin de M. Daniel Bernoulli, qu'il me communiqua, en 1769". It was later published (1778) and is given in English translation by Kendall (1961). Todhunter devotes Articles 424–427 to it, but does not appreciate that the method proposed by Bernoulli may be interpreted as being free of a Bayesian justification. Bernoulli simply states (in translation): "I think that of all the innumerable ways of dealing with errors of observation one should choose the one that has the highest degree of probability for the complex of observations as a whole". Unfortunately he then proposes a semi-elliptic error distribution, which results in an intractable maximization problem. Euler (1778) was less than enthusiastic about Bernoulli's admittedly "purely metaphysical speculation". His paper follows Bernoulli's, both in the original and in translation.

Lagrange's method was also subsequently published (1770–73). As Plackett (1958) observes, Lagrange "purports to show that the mode of the distribution of sample means is the same as the population mean". But he is in difficulty over the fact that in the discrete cases with which he worked, the most probable sample, being integral, might have a mean which could not match that of the population. However, in Problem VI Lagrange tackles the practical difficulty that one does not know the population distribution, by *estimating* the proportions with which the various errors occur; and this he does by maximizing the likelihood, finding the maximum-likelihood estimators of multinomial proportions. Todhunter (Art. 562) says that Lagrange "[takes] it as evident that the most probable values" are these, but there seems to be no justification for implying that Lagrange was using an inverse probability argument, though it is true, as Todhunter also remarked, that in the subsequent investigations into the probability that the estimates do not differ from the true values by more than assigned quantities, the inverse argument is used; for without its use this probability must generally involve the unknown true values, which, in Lagrange's treatment, it does not.

In passing we may note that the copy of Todhunter in the Library of Gonville and Caius College, Cambridge, contains a correction in Fisher's handwriting to Art. 561: where Todhunter writes "It appears from common Algebra that the greatest value of μ is when . . .", in connection with Lagrange's determination of the mode, Fisher has (quite correctly) replaced "greatest" by "most frequent".

There are hardly any references to these contributions until modern times, though Daniel Bernoulli's paper is mentioned by Keynes (1921) and discussed briefly by Levy and Roth (1936), who thought he was using an inverse probability argument. Lambert's contribution seems to have vanished for 200 years, partly, no doubt, because Ostwald's *Klassiker* translation omitted the part of interest (Sheynin, 1971). They seem to have been without influence on the subsequent development of estimation theory, probably because the method of least squares, which was shortly to be introduced, provided a practical procedure of such utility that logical qualms were reserved until later.

The Method of Maximum Probability

In *Theoria Motus Corporum Coelestium* (1809), Gauss introduced the method of least squares as a necessary consequence of (1) accepting the arithmetic mean as the optimum estimator, and (2) applying the method of maximum probability, using a uniform prior distribution. He shows that the error distribution, for which the mean is the maximum-probability estimator, is the normal, and that in the normal case maximizing the posterior probability is equivalent to minimizing the sum of squares. The subsequent history of least squares is not our present concern; but Gauss's original treatment is explicitly Bayesian, and therefore does not constitute an application of the method of maximum likelihood.

Similarly Laplace, in *Théorie Analytique des Probabilités* (1820), frequently used maximum-probability values for parameters, having adopted a uniform prior probability distribution. It is true that he writes of "values which render the observed event the most probable", but only in connection with a prior distribution; furthermore, when considering transformation to a new parameter, the differential element is taken into account, indicating the use of a probability density rather than a likelihood. Nor does Laplace always use the maximum probability argument: for example, when considering the classical problem of errors he thinks that the median of the posterior probability distribution should be used: "Some celebrated mathematicians", he writes, "have suggested that one should choose, for the mean, that which renders the observed result the most probable, and hence the abscissa which corresponds to the greatest ordinate of the [posterior probability] curve; but our mean is evidently the one indicated by probability theory".

Of particular interest is Laplace's approximate evaluation of the posterior probability that the true parameter value should lie in a certain interval about the most probable value, for he uses an expansion whose first and most important term is none other than minus the second differential of the log-likelihood evaluated at the maximum, or the *observed information* in my terminology (Edwards, 1972). At the end of the century Karl Pearson and L. N. G. Filon (1898) offered a very similar calculation, finding "the frequency distribution for errors in the values of the frequency constants". I conclude from this phrase that the authors were thinking in terms of inverse probability; E. S. Pearson (1967) concurs. In 1896 Karl Pearson had written: "Thus, it appears that the observed result is the most probable, when r (the correlation coefficient as a parameter) is given the value $S(xy)/(n\sigma_1\sigma_2)$. This value presents no practical difficulty in calculation, and therefore we shall adopt it". Pearson adds the footnote: "It seems desirable to draw special attention to this best value of the correlation coefficient . . .".

That Pearson was following Gauss's method of maximum probability is clear from a letter he wrote to Fisher twenty years later (26 June 1916; quoted by E. S. Pearson, 1968). Fisher had drafted a note pointing out that a contributor to the May 1916 issue of *Biometrika* had used a method of estimation (minimum χ^2) which was arbitrary to the extent that it depended on the way the data were grouped: Pearson replied, declining to publish the note; though "If you will write me a defence of the Gaussian method, I will certainly consider its publication. I frankly confess I approved the Gaussian method in 1897, but I think it logically at fault now". The reason that Pearson found fault with it at that time seems to have been that he could not accept the arbitrary nature of the prior probability distribution needed for completion of the method. Later, after Fisher had sought justification for the method of maximum likelihood in the repeated sampling properties of the estimators, Pearson, by then clear that inverse probability was not involved, attacked it as lacking justification in small samples.

Before recording Pearson's criticism of "the indiscriminate use of Bayes' Theorem" we must sketch the situation that led him to make it. Fisher advanced the method of maximum likelihood in 1912, but unfortunately he used the phrase "inverse probability", that same phrase by which Todhunter had described the Bayesian or Laplacian method. It turned out to be one of the most influential errors of terminology in statistics, for it led directly to his first quarrel with Pearson, who did not look beyond the phrase to Fisher's account and subsequent use of the method, which was non-Bayesian. We shall examine Fisher's 1912 account in detail below; his first *use* of the method was in 1915, in his paper on the correlation coefficient: "I have given elsewhere a criterion, independent of scaling, suitable for obtaining the relation between an observed correlation of a sample and the most probable value of the correlation of the whole population" (p. 520). On the following page he wrote not of the most *probable* value, but of the most *likely* value. Was this the germ of the later definition of *likelihood*? That Fisher's criterion was *independent of scaling* should perhaps have given a clue to Pearson, but it did not. Fisher's paper, in which the distribution of r in samples from a normal bivariate population was first obtained, sparked off an extensive investigation by Pearson and his collaborators. The results appeared in *Biometrika* the following year under the title *On the distribution of the correlation coefficient in small samples. Appendix II to the papers of "Student" and R. A. Fisher. A Cooperative Study*. The authors were Soper, Young, Cave, Lee and Pearson. There is a section "On the determination of the 'most likely' value of the correlation in the sampled population" in which the authors ask "What is the most reasonable value?" It seems likely that Fisher knew nothing of this section until the paper was in proof (E. S. Pearson, 1968), if then, in spite of the description of the paper as an appendix to one of his own. It contained a criticism of what the authors supposed was Fisher's use of a prior distribution for ρ, and concluded "It will thus be evident that in problems like the present the indiscriminate use of Bayes' Theorem is to be deprecated. It has unfortunately been made into a fetish by certain purely mathematical writers on the theory of probability, who have not adequately appreciated the limits of Edgeworth's justification of the theorem by appealing to *general* experience".

Pearson, then, did not use maximum likelihood; nor did he originally appreciate its difference from maximum probability; and his use of maximum probability was short-lived. Thus Haldane's (1957) suggestion that Pearson's was the method "which was developed by Edgeworth as 'the method of maximum credibility', and by Fisher as 'the method of maximum likelihood'" is a little disingenuous. Elsewhere Haldane also found it difficult to dissociate Fisher's method from inverse probability (see Haldane, 1932, and Fisher's reply, 1932). Wright (1968) repeats Haldane's suggestion, though attributing the method to Gauss. Edgeworth was clearly appealing to inverse probability (Pearson, 1967), though latterly he wished to justify prior distributions by general experience rather than to assume them as representations of ignorance: "I submit that very generally we are justified in assuming an equal distribution of *a priori* probabilities over that tract of the measurable with which we are concerned" (1908).

Likelihood-ratio Arguments

It has long been the practice to argue that we may distinguish between hypotheses by comparing the probability with which each would lead to the observed result. Thus de Moivre (1756), writing in 1717: "Further, the same Arguments which explode the Notion of Luck, may, on the other side, be useful in some Cases to establish a due comparison between Chance and Design: We may imagine Chance and Design to be, as it were, in Competition with each other, for the production of some sorts of Events, and may calculate what Probability there is, that those Events should be rather owing to one than to the other".

Laplace (1820), in the sixth principle enunciated in the *Essai philosophique*, writes: "Each of the causes to which an observed event may be attributed is indicated with just as much likelihood as there is probability that the event will take place, supposing the event to be constant". The translation is that of Truscott and Emery (1951), but in the original we find that the French word is "vraisemblance", and that the above sentence is only the first clause of a longer sentence, being concluded by a semi-colon and not a full stop. In translation Laplace continues: "The probability of the existence of any one of these causes is then a fraction whose numerator is the probability of the event resulting from this cause and whose denominator is the sum of the similar probabilities relative to all the causes; if these various causes, considered *a priori*, are unequally probable, it is necessary, in place of the probability of the event resulting from each cause, to employ the product of this probability by the possibility of the cause itself". The word "possibility" is also a loose translation, for Laplace wrote "celle", referring to *probability*. Laplace was evidently being purely " Bayesian".

Perhaps the first example of a likelihood-ratio argument (as we now call such comparisons of probabilities) to come from a writer who explicitly rejected inverse probability is that of Venn (1876): "To decide this question, what we have to do is to compare the relative frequency with which the two kinds of cause would produce such a result".

Fisher and the Definition of Likelihood

On 1 October 1909, R. A. Fisher was admitted to Gonville and Caius College, Cambridge, as an entrance scholar. John Venn was President of the College, and F. J. M. Stratton had been a Fellow for three years. In his third year as an undergraduate (1912) Fisher published a paper "On an absolute criterion for fitting frequency curves". He observes that the similarity of the results obtained by various methods "is harmful from the theoretical standpoint as tending to obscure the practical discrepancies, and the theoretical indefiniteness which actually exist". He mentions the method of least squares, but dismisses it as "obviously inapplicable to frequency curves" because "an arbitrariness arises in the scaling of the abscissa line, for if ξ, any function of x, were substituted for x, the criterion would be modified". Of the method of moments, it "is possibly of more value, though its arbitrary nature is more apparent", for "a choice has been made without theoretical justification in selecting [this] set of r equations of the general form $\Sigma x^p f = \sum_1^n x^p$". Thus does Fisher clear the decks.

"But we may solve the real problem directly." "The most probable set of values for the θ's will make P a maximum", where P is the probability of the observations, given the parameter set θ, less the differential elements dx. Here Fisher seems to propose the method of maximum probability, but without specific reference to a prior distribution. He then considers the case of estimating the mean and *modulus of precision* of a normal distribution, and observes that "the inverse probability system may be represented by the surface traced out by a point at a height P above the point on a plane, of which m [the mean] and h [the modulus] are co-ordinates". Again, Fisher seems to be relying on inverse probability, but this impression is soon to be dispelled: "We shall see (in § 6) that the integration with respect to m is illegitimate and has no definite meaning with respect to inverse probability". After some discussion of bias, we come to § 6, which I quote in full: "We have now obtained an absolute criterion for finding the relative probabilities of different sets of values for the elements of a probability system of known form. It would now seem natural to obtain an expression for the probability that the true values of the elements should lie within any given range. Unfortunately we cannot do so. The quantity P must be considered as the relative probability of the set of values θ_1, θ_2, ..., θ_r; but it would be illegitimate to multiply this quantity by the variations $d\theta_1$, $d\theta_2$, ..., $d\theta_r$, and integrate through a region, and to compare the integral over this region with the integral over all possible values of the θ's. P is a relative probability only, suitable to compare point with point, but incapable of being interpreted as a probability distribution over a region, or of giving any estimate of absolute probability.

"This may be easily seen, since the same frequency curve might equally be specified by any r independent functions of the θ's, say $\phi_1, \phi_2, \ldots, \phi_r$, and the relative values of P would be unchanged by such a transformation; but the probability that the true values lie within a region must be the same whether it is expressed in terms of θ or ϕ, so that we could have for all values

$$\frac{\partial(\theta_1, \theta_2, \ldots, \theta_r)}{\partial(\phi_1, \phi_2, \ldots, \phi_r)} = 1,$$

a condition which is manifestly not satisfied by the general transformation.

"In conclusion I should like to acknowledge the great kindness of Mr J. F. M. Stratton [*sic*], to whose criticism and encouragement the present form of this note is due."

In view of § 6 we must conclude that Fisher was using the phrase "inverse probability" incorrectly, as he later admitted (1922). He was manifestly describing the method of maximum likelihood, and pointing out its invariance to parameter transformation. It would have been a remarkable achievement for an undergraduate if Fisher had done no more than criticize Gauss's method of least squares for its dependence on the choice of a scale for the variate, Pearson's method of moments for its arbitrariness, and, by implication, Laplace's method of maximum probability for its dependence on the choice of parametric form; but Fisher went further and gave us a concept "suitable to compare point with point, but incapable of being interpreted as a probability distribution over a region". He introduced the concept of a likelihood surface for two parameters, and, by being prepared to use his "relative probability" to compare "point with point", gave likelihood a precise meaning, though not yet a name.

The confusion that was caused by Fisher introducing a new concept without a new name has been described above, and we will not pursue the matter much further here. It suffices to add that Fisher, seeing the importance of distinguishing his new concept from probability, introduced the word *likelihood* in a paper written in October 1920 as his reply to the "Cooperative Study" and published the following year. The last section of this is entitled "Note on the confusion between BAYES' Rule and my method of the evaluation of the optimum". *Likelihood* is defined, and, "so defined, probability and likelihood are quantities of an entirely different nature". Fisher discusses the invariance of likelihood to parameter transformation, in contrast to probability, and clearly indicates that likelihood may be used not only as a quantity to be maximized, but also as a relative measure applied to different hypotheses or parameter values.

From § 6 of Fisher (1912) one might feel that Stratton suggested the emphasis in this section to Fisher. I can find no support for this idea in the earlier statistical writings of Stratton (Stratton and Compton, 1910; Wood and Stratton, 1910), and we may conclude that Stratton did not regard Fisher's method as being of particular importance because Brunt (1917) does not mention it, although he acknowledges his "indebtedness to Mr F. J. M. Stratton, of Gonville and Caius College, Cambridge, to whose University lectures I owe most of my knowledge of the subjects discussed in this book, and upon whose notes I have drawn freely". E. S. Pearson, who attended lectures by Stratton, informs me that neither did they provide any evidence of Stratton thinking Fisher's contribution important.

References

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.* **53**, 370–418. Reprinted in *Biometrika*, **45**, 293–315 (1958), and by Pearson and Kendall (1970).

Bernoulli, D. (1778). The most probable choice between several discrepant observations and the formation therefrom of the most likely induction (with a comment by Euler). *Acta Acad. Petrop.* 3–33. English translation in *Biometrika*, **48**, 3–13 (1961), reprinted by Pearson and Kendall (1970).

Bernoulli, J. (1713). *Ars Conjectandi*. Basel.

Brunt, D. (1917). *The Combination of Observations*. Cambridge University Press.
de Moivre, A. (1756). *Doctrine of Chances*, 3rd ed. London.
Euler, L. (1778). See Bernoulli, D. (1778).
Edgeworth, F. Y. (1908). On the probable errors of frequency constants. *J. Roy. statist. Soc.* **71**, 381–397, 499–512, 651–678.
Edwards, A. W. F. (1972). *Likelihood*. Cambridge University Press.
Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Mess. math.* **41**, 155–160. Reprinted in Fisher (1971).
Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, **10**, 507–521. Reprinted in Fisher (1971).
Fisher, R. A. (1921). On the "Probable Error" of a coefficient of correlation deduced from a small sample. *Metron*, **1**, part **4**, 3–32. Reprinted in Fisher (1971).
Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc.* A, 222, 309–368. Reprinted in Fisher (1950) and Fisher (1971).
Fisher, R. A. (1932). Inverse probability and the use of likelihood. *Proc. Camb. Phil. Soc.* **28**, 257–261.
Fisher, R. A. (1950). *Contributions to Mathematical Statistics*. New York: Wiley.
Fisher, R. A. (1971). *Collected Papers*, Vol. 1 (1912–24). University of Adelaide.
Gauss, C. F. (1809). *Theoria Motus Corporum Coelestium*. Hamburg. Werke, **7**, Göttingen (1906). English translation by C. H. Davis. Boston: Little, Brown (1857).
Hacking, I. (1971). Jacques Bernoulli's *Art of Conjecturing*. *Brit. J. Phil. Sci.* **22**, 209–229.
Haldane, J. B. S. (1932). A note on inverse probability. *Proc. Camb. Phil. Soc.* **28**, 55–61.
Haldane, J. B. S. (1957). Karl Pearson, 1857–1957. *Biometrika*, **44**, 303–313.
Kendall, M. G. (1961). Translation of Bernoulli (1778), q.v.
Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.
Lagrange, J. L. (1770–73). Mémoire sur l'utilité de la méthode de prendre le milieu entre les résultats de plusieurs observations. *Miscellanea Taurinensia*, **5**, 167–232. *Œuvres* **2**, 171–234. Paris: Gauthier-Villars (1868).
Lambert, J. H. (1760). *Photometria*. Augustae Vindelicorum.
Laplace, Marquis de (1820). *Théorie analytique des probabilités*, 3rd ed. Paris: Courcier. The introduction is: *Essai philosophique sur les probabilités*. Reprinted in English as *A Philosophical Essay on Probabilities*; New York: Dover Publications (1951).
Levy, H., Roth, L. (1936). *Elements of Probability*. Oxford: Clarendon Press.
Neyman, J., Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, **20**a, part I, 175–240; part II, 263–294. Reprinted in *Joint Statistical Papers*, Cambridge University Press (1967).
Pearson, E. S. (1967). Some reflexions on continuity in the development of mathematical statistics, 1885–1920. *Biometrika*, **54**, 341–355. Reprinted by Pearson and Kendall (1970).
Pearson, E. S. (1968). Some early correspondence between W. S. Gosset, R. A. Fisher and Karl Pearson, with notes and comments. *Biometrika*, **55**, 445–457. Reprinted by Pearson and Kendall (1970).
Pearson, E. S., Kendall, M. G. (1970). *Studies in the History of Statistics and Probability*. London: Griffin.
Pearson, K. (1896). Mathematical contributions to the theory of evolution – III. Regression, Heredity and Panmixia. *Phil. Trans. Roy. Soc.* A, **187**, 253–318. Reprinted in *Karl Pearson; Early Statistical Papers*, Cambridge University Press (1948).
Pearson, K., Filon, L. N. G. (1898). Mathematical contributions to the theory of evolution – IV. On the probable errors of frequency constants and on the influence of random selection on variation and correlation. *Phil. Trans. Roy. Soc.* A. **191**, 229–311. Reprinted in *Karl Pearson; Early Statistical Papers*, Cambridge University Press (1948).
Plackett, R. L. (1958). The principle of the arithmetic mean. *Biometrika*, **45**, 130–135. Reprinted by Pearson and Kendall (1970).
Sheynin, O. B. (1971). J. H. Lambert's work on probability. *Arch. Hist. Exact. Sci.* **7**, 244–256.
Soper, H. E., Young, A. W., Cave, B. M., Lee, A., Pearson, K. (1916). On the distribution of the correlation coefficient in small samples. Appendix II to the papers of "Student" and R. A. Fisher. A cooperative study. *Biometrika*, **11**, 328–413.
Stratton, F. J. M., Compton, R. H. (1910). On accident in heredity with special reference to right- and left-handedness. *Proc. Camb. Phil. Soc.* **15**, 507–512.
Todhunter, I. (1865). *A History of the Mathematical Theory of Probability*. Cambridge and London: Macmillan.
Truscott, F. W., Emery, F. L. (1951). Translation of Laplace's *Essai Philosophique*: see Laplace (1820).
Venn, J. (1876). *The Logic of Chance*, 2nd ed. London: Macmillan.
Wood, J. B., Stratton, F. J. M. (1910). The interpretation of experimental results. *J. Agric. Sci.* **3**, 417–440.
Wright, S. (1968). *Evolution and the Genetics of Populations*. Vol. 1. *Genetic and Biometric Foundations*. Chicago University Press.

Résumé

Le présent article trace l'histoire de l'idée générale de la vraisemblance, de ses origines dans les œuvres de de Moivre (1717), de Lambert (1760) et de D. Bernoulli (1778), jusquà sa formulation explicite par Fisher (1912). On considère les parties des œuvres de Gauss et de Laplace qui s'y rapportent, et en s'occupe surtout de l'évolution des idées de Fisher entre 1912 et 1922.