Bayesian Inference

Data Evaluation and Decisions

Introductory chapter.

For the full text see the book published by Springer Verlag, Heidelberg 2003

Hanns Ludwig Harney

Heidelberg, April 2003

Contents

Preface

The present book deals with experience although it is theoretical. The question how to draw conclusions from random events, is treated. Combining ideas by Bayes and Laplace with concepts of modern physics, we answer some aspects of the question.

In the present text, features of a textbook and a monography are combined. The arguments are given as explicitely as possible — partly by help of appendices containing the more lengthy derivations. There are numerous examples and illustrations — often from recent physics. Problems are posed and their solutions are given. The present theory is conservative in that the most widely known methods of error estimation are retrieved.

On the other hand, some material is unconventional. The non-informative prior is considered the basis of statistical inference, and a unique definition is given and defended. Not only does the prior allow to find the posterior distribution, it also provides the measure that one needs to construct error intervals and to make decisions. Criteria to judge the quality of a fit are derived for histograms and multinomial distributions. They complement the conventional chi-squared test applicable to Gaussian events. An attempt is made to classify measurable parameters according to their dependence on each other. This has lead to a description of especially well defined parameters — in the sense that they refer to logically exclusive aspects of the data.

The binomial distribution — sketched on the book-cover — stands for three hundred years of research on statistics. It is the first clearly formulated statistical model. It is the first example of statistical inference. We hope to convince the reader of this book that the subject is not yet closed.

Hanns L. Harney

Heidelberg, November 2002

Chapter 1

Knowledge and Logic

Science does not prove anything. Science infers statements about reality. Sometimes the statements are of stunning precision, sometimes they are rather vague. Science never reaches exact results. Mathematics provides proofs but it is void of reality. The present book shall show in mathematical terms how to express uncertain experience in scientific statements.

Every observation leads to randomly fluctuating results. Therefore the conclusions drawn from them, must be accompanied by an estimate of their truth expressed as a probability. Such a conclusion typically has the form: "The quantity ξ inferred from the present experiment has the value $\alpha \pm \sigma$." An experiment never yields the true value of ξ. Rather the result is characterised by an interval in which the true value should lie. It does not even lie with certainty in that interval. A more precise interpretation of the above interval is: "The quantity ξ is with the probability $K = 0.68$ in the interval $[\alpha - \sigma, \alpha + \sigma]$." Trying to be even more precise one should say:

"We assign a Gaussian distribution to the parameter ξ. The distribution is centered at α and has the standard deviation σ. The shortest interval containing ξ with the probability $K = 0.68$ is then $\alpha \pm \sigma$." In a simplifying manner, the standard deviation of the assumed Gaussian distribution is called "the error" of the result although "the" error of the result cannot be specified. One is free to choose the length of the error interval because one is free to choose the probability K.

The present book deals with the generalisation of the well-known rules of Gaussian error assignments to cases, where the Gaussian model does not apply. Of course, the Gaussian model is treated, too. But the book is animated by the question: How to estimate the error interval when the data follow a distribution other than Gaussian, e.g. a Poissonian one? This requires to answer the general questions: What is — in any case — the definition of an error interval? How do we understand probability?

1.1 Knowledge

The parameter that one wants to know is never measured directly and immediately. The true length of a stick is hidden behind the random fluctuations of the value that one reads on a meter. The true position of a spectral line is hidden in the line width that one observes with the spectrograph. The fluctuations have different causes in these two cases but they cannot be avoided. One does not observe the interesting parameter ξ. One rather observes events x that have a distribution depending on ξ. Data analysis means to infer ξ from the event x — usually on the basis of a distribution p

that parametrically depends on ξ. This parameter is also called the hypothesis that conditions the distribution of x. The connection between x and ξ is given by $p(x|\xi)$ — in words: The distribution p of x, given ξ. It must depend on ξ in such a way that different hypotheses entail different distributions of x so that one can learn from x about ξ.

Inferring ξ is incomplete induction. It is induction because it is based on observation — as opposed to logical deduction based on first principles. It is incomplete because it is based on one event. Note that even an experiment that produces a huge amount of data, yields one event in the sense that it does not yield all possible data, and its repetition would produce a different event. For this reason, no experiment yields the true value of ξ and inference of ξ is achieved by assigning a distribution to ξ although one assumes that all the events are in fact conditioned by one and the same true value of ξ. Thus the distribution $P(\xi|x)$ assigned to ξ, is a representation of the limited knowledge of ξ.

There has been a long debate whether this procedure — the Bayesian inference — is justified and is covered by the notion of probability. The key question was: Can one consider probability not only as the relative frequency of events but also as a value of truth assigned to a statement? We take for granted that the answer is "yes", and consider the debate as historical.

For the founders of statistical inference, Bayes[1] [9] and Laplace[2] [89, 92],

[1]Thomas Bayes, 1702–1761, English mathematician and Anglican clergyman. In a posthumously published treatise, he formulated for the first time a solution to the problem of statistical inference.

[2]Pierre Simon Marquis de Laplace, 1749–1827, French mathematician and physicist. He contributed to celestial and general mechanics. His work "Mécanique céleste" has

the notion of probability has carried both concepts: The probability attached to a statement ξ can mean the relative frequency of its occurrence or the state of knowledge on ξ.

This "or" is not exclusive; it is not an "either or". It allows statements ξ that cannot be subjected to a "quality test" revealing how often they come true. The test is possible for the statement "The probability that the coin falls with head upward is $1/2$". However, the statement "It is very probable that it rains tomorrow" is not amenable to the frequency interpreation — not because the qualitative value "very probable" is vague but because "tomorrow" always exists only once. So the latter statement can only be interpreted as evaluating the available knowledge. A fine description of these different interpretations has been given by Cox [35, 36]. See also Chaps. 13 and 14 of Howson and Urbach [69]. A taste of the above-mentioned debate is given by the polemics in [77].

We do speak of probability in connection with statements that do not allow the frequency interpretation. We shall, however, require a mathematical model that quantifies the probability attached to a statement.

The above distinction is only an apparent one because the interpretation of probability as a value of the available knowledge is always possible and is thus the broader one. This is seen from the following examples.

Somebody buys a car. The salesman claims to be 95% sure that the car will run the first 100 000 km without even a minor trouble. This praise states his knowledge on or his belief in the quality of the product at hand.

---

been considered to rival Newtons "Principia". He invented the spherical harmonics. He formulated and applied Bayes' theorem independently of him.

Yet there could be a statistical quality test which turns this personal belief into a known frequency of break-down. But even if the praise by the salesman is objective in this sense, it becomes a personal belief for the interested client into the quality of his car — the one car that he decides to buy.

Let us try to translate this into the language of measurement! The quantity ξ was measured as $\alpha \pm \sigma$. Hence, with 68% probability it is in that interval. Setting aside systematic errors, one can offer a frequency interpretation to this statement: If one were to repeat the measurement — say — 100 times, the result should fall into the interval with the frequency of 68%. This is right but does not well describe the situation. If one actually had 100 more measurements, one would reasonably use them to state one final result of considerably higher precision than the first one. How to do this, is described in Chap. 2. The final result would again be a single one [128, 129].

There does not seem to be a clear distinction between the cases which allow the frequency interpretation of probability and the cases which allow only its interpretation as a value of knowledge. The latter one is the broader one and we accept it. But we keep in mind that it is the frequency interpretation that leads to mathematically formulated distributions. Some of them are presented in Chaps. 4 and 5.

As a consequence, it may be practical but it is not necessary to distinguish the statistical from the systematic error of an experiment. The statistical error is the consequence of the finite amount of data and can in principle be demonstrated by repeating the experiment. The systematic error results from parameters that are not precisely known although they are not fluctuating randomly. These two types of error correspond rather well to the above two interpretations of probability. Accepting thtm both as possible interpretations of a unique concept, one can combine both errors into a single one. This is indeed done for the graphical representation of a result or its use in related experiments; see Sec.4.2.1 of the *Review of Particle Physics* [55]

1.2 Logic

Since probability can be interpreted as the value of the available knowledge, it can also be considered the implementation of non-Aristotelian logic into scientific communication [120]. Jaynes has simply termed it the logic of Science [79, 72]. It even serves everyday communication as certain weather forecasts show. In philosophy, it is called the logic of temporal statements [152]. “Temporal” because the value of truth estimates the future confirmation of the statement. Without this relation to time, a statement must be either true or false.

The probability attached to the statement ξ, can be considered the value of truth assigned to ξ . The continuum of probability values is then a manifold of values of truth situated between “false” and “true”. This introduces the *tertium* which is excluded in Aristotelian logic by the principle *tertium non datur* saying that a third qualification — other than “true” and “false” — is not available.

Logical operations in a situation where other qualifications are available, must be done in such a way that they are consistent with Aristotelian logic in the following sense: Probabilities shall be quantified. The calculus of probability shall be part of mathematics. Mathematics is based on Aristotelian logic. The rules of mathematical logic can be laid down in terms of symbolic logic. Therefore the rules of handling probabilities — i.e. continuous values of truth — must be consistent with symbolic logic.

From this follow certain conditions which must be observed when values of truth are assigned to statements like "from ξ follows x". This value of truth is the conditional probability $p(x|\xi)$, i.e. the probability to find x when ξ is given. Cox [35] has shown in 1946 that consistency between non-Aristotelian and mathematical logic requires the following two rules.

(i) Let α, ξ, x be three statements that possibly imply each other. The values of truth p of the implications "ξ follows from α" and "x follows from $\xi \wedge \alpha$" and "$x \wedge \xi$ follows from α" must be defined such that the product rule

$$p(x \wedge \xi \mid \alpha) = p(x \mid \xi \wedge \alpha)\, p(\xi \mid \alpha) \tag{1.1}$$

holds. Here, the operator $\wedge$ means the logical "and". Cox' result seems obvious to every person having only a little experience with probabilities. It takes an effort to realize that it is not trivial. Note that probabilistic values of truth need not be positive numbers. There are realms of physics where the positive numbers p are replaced by complex probability amplitudes a. Hence, values of truth can be more complicated objects than positive numbers. In any case, they must respect relation (1.1). The symmetry principle introduced in Chap. 6, indeed leads in Chap. 8 to consider probability amplitudes. Although we restrict ourselves to real a, they allow richer logical combinations than the positive numbers p. Amplitudes can be positive or negative. Thus the probability attached to "ξ_1 or ξ_2" can be smaller than the probabilities attached to either one statement. One summarizes this phenomenon by saying that the statements can interfere or that the alternative

"ξ_1 or ξ_2" is coherent. We shall encounter this in Chap. 15.

(ii) Conditional distributions — such as $P(\xi|x)$ — must be proper and normalized so that

$$\int d\xi \, P(\xi|x) = 1 \, . \tag{1.2}$$

Here, the integral without indication of the limits of integration extends over the entire domain of definition of ξ. This rule is necessary in order to assign a probability to a negation. The probability of the assertion "ξ is not in the interval $[\xi_<, \xi_>]$" is the integral over the complement of the interval $[\xi_<, \xi_>]$. The integral over the complement exists since P shall be proper. The assignment of unit probability to the statement that ξ is somewhere in its domain of definition, is a convention. Not only the posterior P, all conditional distributions shall be normalized. Equation (1.2) analogously holds for the model $p(x|\xi)$. Without this requirement, its dependence on the parameter ξ would not be clearly defined. One could multiply it with any non-negative function of ξ without changing the distribution of x. Hence, inferring ξ from the event x is possible only if (1.2) holds. Nevertheless, in the present book, distributions will be admitted that cannot be normalized — provided that they do not depend on a parameter to be inferred. Such distributions are called improper. One cannot assign a value of truth to a negation that involves a quantity with an improper distribution. Even then, however, one can assign a value of truth to a statement that contains the logical "and". We return to this in Chap. 2.

The joint distribution of the multiple event $x_1 \wedge x_2 \wedge \ldots \wedge x_N$ will often be
A.1.1 discussed. The interested reader should derive it from the logical rule (1.1) under the assumption that x_k follows the distribution $p(x_k|\xi)$. The logical

operator $\wedge$ is never written in the sequel. Instead, the multiple event is called $x_1 \dots x_N$ or simply $x = (x_1 \dots x_N)$.

An immediate consequence of the rules (1.1, 1.2) is Bayes' theorem discussed in Chap. 2. It specifies the posterior probability P of x, given ξ. By the same token, the error interval of ξ is given. It is the smallest interval in which ξ lies with probability K. We call it the Bayesian interval $\mathcal{B}(K)$. To find the smallest interval, one needs a measure in the space of ξ. The measure is identified with the prior distribution appearing in Bayes' theorem.

1.3 Ignorance

Into the definition of $P(\xi|x)$ enters a distribution $\mu(\xi)$ which is independent of the event x. It can be interpreted as the description of ignorance about ξ, and is called the a priori distribution. All methods of inference described in the present book, rely on Bayes' theorem and a unique definition of μ.

The definition starts from a symmetry principle. In Chaps. 6, 8, 11, models $p(x|\xi)$ are considered that connect the parameter ξ with the event x by way of a group of transformations. This symmetry is called form invariance. The invariant measure of the symmetry group — we explain it in Chap. 6 — is the prior distribution μ. This procedure is inspired by the ideas of Hartigan [63], Stein [138] and Jaynes [75].

The invariant measure is not necessarily a proper distribution, see Chap. 2.5. It can be obtained — without any analysis of the group — as a functional of the model p. The functional is known as Jeffreys' rule [80]. Here, it is introduced in Chap. 9.

By accepting the interpretation of probability as a value of truth, we include the "subjective" or "personal" interpretations presented in [126, 127, 98, 42, 43]. However, we do not go so far as to leave the prior distribution at the disposal of the person or the community analyzing given data. This is done in chapter 14 of Howson and Urbach [69] and in the work by D'Agostini [38, 37, 39]. In contrast, we adhere to a formal and general definition of the prior distribution in order to avoid arbitrariness.

Form invariant distributions offer more than a plausible definition of the prior distribution. Form invariance helps to clarify the dependence of parameters on each other. This allows to devise a scheme, where one parameter ξ_1 is inferred independently of the other parameters $\xi_2 \dots \xi_N$ in the sense that ξ_1 refers to an aspect of the event x that is separate from the aspects described by the other parameters; see Chap. 12. This scheme is useful because the extraction of ξ_1 is often linked to and dependent on other parameters that must be included in the model although they are not interesting. The intensity of a signal depends on the determination of the background although the interest is focussed on the signal.

Form invariance is usually considered to occur so rarely that one cannot found the definition of the prior distribution onto it. See $ 6.9 of [12] as well as [122]. The present Chap. 11 shows that there are more form invariant distributions than previously believed.

Still Bayesian inference cannot be restricted to form invariant distributions. When this symmetry is lacking, one considers the square root of the probability $p(x|\xi)$ — i.e. the amplitude a_x — as the component of a vector that parametrically depends on ξ. This is the parameter representation of a surface. The measure on the surface is the prior distribution. To understand this, one needs some differential geometry [70, 123, 124] which is explained in Chap. 9. The differential geometrical measure is again given by Jeffreys' rule [80].

Differential geometry by itself cannot establish Jeffreys' rule as the generally valid measure. One must show that the surface $a(\xi)$ is to be considered in the space of the amplitudes — not the probabilities or a function other than the square root of the probabilities. This, however, becomes obvious from the form invariant models.

Beyond the observed event x, information on ξ is often available that should be incorporated into Bayesian inference and that will let shrink the Bayesian interval. The order of magnitude of ξ is usually known. A fly is neither as small as a microbe nor as large as an elephant. One knows this before measuring a fly. Such information can be built into the prior distribution which thereby changes from the ignorance prior μ to an informed prior μ^{ent}. An informed prior may simply be posterior of a foregoing experiment. It may also be generated by entropy maximisation, given previous information. Jaynes [73, 74] has transferred this method from thermodynamics into the analysis of data. This idea has found much interest and has led to a series of conferences [135, 133, 46, 47, 131, 50, 82, 134, 106, 66, 132, 60, 151] and many publications [33]. We take this method for well-known and do not treat it in the present book. Note, however, that entropy maximisation cannot replace the definition of the ignorance prior μ. According to Jaynes [75], it uses μ.

1.4 Decisions

Bayesian inference chooses from the family of distributions $p(x|\xi)$ the ones that best reproduce the observed event x. This does not mean that any one of the distributions is satisfactory. How does one decide whether the model $p(x|\xi)$ is satisfactory in the sense that it contains distributions consistent with the available data?

When x follows a Gaussian distribution this is decided by the chi-squared criterion of Chap. 14. In Chap. 16, generalisations to the histogram and the multinomial model are given. It turns out that for the decision, one needs a measure in the space of ξ. We have identifid the measure with the prior distribution μ.

Hence, the definition of a measure is essential for practically all conclusions from statistical data. One needs a measure — the prior distribution — in order to infer a parameter and to construct an error interval; see Chap. 2. One needs a measure in order to decide whether the given value of a parameter is probable or rather improbable; see Chap. 3. One needs a measure in order to decide whether a given set of events is compatible with a predicted distribution; see Chaps. 14, 16.

Bibliography

[1] C.E. Aalseth, F.T. Avignone III, A. Barabash, F. Boehm, R.L. Brodzinski, J.I. Collar, P.J. Doe, H. Ejiri, S.R. Elliott, E. Fiorini, R.J. Gaitskell, G. Gratta, R. Hazama, K. Kazkaz, G.S. King III, R.T. Kouzes, H.S. Miley, M.K. Moe, A. Morales, J. Morales, A. Piepke, R.G.H. Robertson, W. Tornow, P. Vogel, R.A. Warner, and J.F. Wilkerson. Comment on 'Evidence for neutrinoless double beta decay'. *Mod. Phys. Lett. A*, 17:1475–1478, 2002.

[2] A.Y. Abul-Magd. Wealth distribution in an ancient Egyptian society, 2002.

[3] O.-A. Al-Hujaj and H.L. Harney. Objective Bayesian statistics. Technical report, Max-Planck-Institut für Kernphysik, 1997. See arXiv:physics/9706025.

[4] O.A. Al-Hujaj. Objektive Bayessche Statistik. Theorie und Anwendung. Master's thesis, Fakultät für Physik und Astronomie der Universität Heidelberg, Max-Planck-Institut für Kernphysik, D-69029 Heidelberg, 1997.

[5] V.P. Alfimenkov, S.B. Borzakov, Vo Van Tkhuan, Y.D. Mareev, L.B. Pikel'ner, A.S. Khrykin, and E.I. Sharapov. Parity violation at the 0.75-ev neutron resonance of ^{139}La. *JETP Lett.*, 35(1):51–54, 1982.

[6] V.P. Alfimenkov, S.B. Borzakov, Vo Van Tkhuan, Y.D. Mareev, L.B. Pikel'ner, A.S. Khrykin, and E.I. Sharapov. Parity nonconservation in neutron resonances. *Nucl. Phys. A*, 398:93–106, 1983.

[7] P.W. Anderson. The Reverend Thomas Bayes, needles in haystacks, and the fifth force. *Physics Today*, page 9, January 1992.

[8] Roger Barlow. *Statistics*. Wiley, Chichester, 1989.

[9] Thomas Bayes. An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.*, 53:330–418, 1763. Reprinted in Biometrika 45, 293–315 (1958) and in 'Studies in the History of Statistics and Probability' E.S. Pearson and M.G. Kendall eds., C.Griffin and Co. Ltd., London 1970 and in 'Two papers by Bayes with commentaries' W.E. Deming ed., Hafner Publishing Co., New York, 1963.

[10] Frank Benford. The law of anomalous numbers. *Proc. American Philos. Soc.*, 78:551–572, March 1938.

[11] W.A. Bentley and W.J. Humphreys. *Snow Crystals*. Dover, New York, 1931. Reprinted in 1962 and 1980.

[12] James O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer Series in Statistics. Springer, New York, second edition, 1985.

[13] J.O. Berger. Estimating a product of means: Bayesian analysis with reference priors. *J. Am. Statist. Assoc.*, 84:200–207, 1989.

[14] J.O. Berger. On the development of the reference prior method. In Bernardo et al. [17], pages 35–49.

[15] J.M. Bernardo. Unacceptable implications of the left Haar measure in a standard normal theory inference problem. *Trab. Est. Invest. Operat.*, 29:3–9, 1978.

[16] J.M. Bernardo. Reference posterior distributions for Bayesian inference. *J. R. Statist. Soc. B*, 41:113–147, 1979.

[17] J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors. *Bayesian Statistics 4. Proceedings of the Fourth International Meeting held at Valencia on April 15–20, 1991*, Oxford, 1992. Oxford University Press.

[18] J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors. *Bayesian Statistics 5*, Oxford, 1996. Clarendon Press. Proceedings of the Fifth International Meeting held at Valencia on June 5–9, 1994.

[19] J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, editors. *Bayesian Statistics 6. Proceedings of the Sixth Valencia International Meeting June 6–10, 1998*, Oxford, 1999. Clarendon Press.

[20] J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M Smith, editors. *Bayesian Statistics. Proceedings of the International Meeting held at Valencia on May, 28 to June 2, 1979*, Valencia, 1980. Valencia University Press.

[21] J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M Smith, editors. *Bayesian Statistics 2. Proceedings of the Second International Meeting held at Valencia onSeptember 6–10, 1983*, Amsterdam, 1985. North-Holland.

[22] J.M. Bernardo, M.H. DeGroot, D.V. Lindley, and A.F.M Smith, editors. *Bayesian Statistics 3. Proceedings of the third International Meeting held at Valencia on June 1–5, 1987*, Oxford, 1988. Oxford University Press.

[23] J.M. Bernardo and A.F. M. Smith. *Bayesian Theory.* Wiley, Chichester, 1993.

[24] Jacob Bernoulli. *Wahrscheinlichkeitsrechnung.* Harri Deutsch, Thun, 1999. Reprint of vols. 107 and 108 of the series *Ostwalds Klassiker der exakten Wissenschaften.* The Latin original appeared in 1713 under the title of *Ars conjectandi.*

[25] P.R. Bevington and D.K. Robinson. *Data Reduction and Error Analysis for the Physical Sciences.* McGraw-Hill, New York, 3rd edition, 2002.

[26] E.G. Bilpuch, A.M. Lane, G.E. Mitchell, and J.D. Moses. Fine structure of analogue resonances. *Phys. Rep.*, 28(2):145–244, 1976.

[27] E. Blanke, H. Driller, W. Glöckle, H. Genz, A. Richter, and G. Schrieder. Improved experimental test of detailed balance and time reversibility in the reactions $^{27}\mathrm{al} + p \leftrightarrow ^{24}$ mg. *Phys. Rev. Lett.*, 51:355–358, 1983.

[28] V. Blobel and E. Lohrmann. *Statistische und numerische Methoden der Datenanalyse.* Teubner, Stuttgart, 1998.

[29] K.H. Böckhoff, editor. *Nuclear Data for Science and Technology. Proceedings of the International Conference Antwerp September 1982*, Dordrecht, 1983. Reidel.

[30] O. Bohigas and M.-J. Giannoni. Chaotic motion and random matrix theories. In Dehesa et al. [44], pages 1–99.

[31] O. Bohigas, R.U. Haq, and A. Pandey. Fluctuation properties of nuclear energy levels and widths: Comparison of theory with experiment. In Böckhoff [29], pages 809–813.

[32] G.E.P. Box and G.C. Tiao. *Bayesian Inference in Statistical Analysis.* Wiley, New York, 1992. Originally published in 1973 by Addison-Wesley, Reading.

[33] B. Buck and V.A. Macaulay, editors. *Maximum Entropy in Action. A Collection of Expository Essays*, Oxford, 1991. Clarendon Press.

[34] Ted Chang and Cesaro Villegas. On a theorem of Stein relating Bayesian and classical inferences in group models. *Can. J. Statist.*, 14(4):289–296, 1986.

[35] R.T. Cox. Probability, frequency and reasonable expectation. *Am. J. Phys.*, 14:1–13, 1946.

[36] R.T. Cox. *The Algebra of Probable Inference.* Johns Hopkins Press, Baltimore, Maryland, 1961.

[37] Giulio D'Agostini. Jeffreys priors versus experienced physicist priors. Arguments against objective Bayesian theory. physics/9811045 and http://www-zeus.roma1.infn.it/ agostini/.

[38] Giulio D'Agostini. Probability and measurement uncertainty in physics — a Bayesian primer, 1995. Lecture Notes, preprint DESY 95-242, hep-ph/9512295v2.

[39] Giulio D'Agostini. Bayesian reasoning versus conventional statistics in high energy physics. In von der Linden et al. [151], pages 157–170.

[40] Gauri Sankar Datta and Jayanta Kumar Gosh. On prior providing frequentist validity for bayesian inference. *Biometrika*, 82:37–45, 1995.

[41] A.P. Dawid, N. Stone, and J.V. Zidek. Marginalisation paradoxes in Bayesian and structural inference. *J. Roy. Statist. Soc. B*, 35:189–233, 1973.

[42] Bruno de Finetti. *Probability, Induction and Statistics. The Art of Guessing.* Wiley, New York, 1972.

[43] Bruno de Finetti. *Theory of Probability.* Wiley, 1974.

[44] J.S. Dehesa, J.M.G. Gomez, and A. Polls, editors. *Mathematical and Computational Methods in Nuclear Physics: Proceedings of the 6. Granada workshop held in Granada, Spain, Oct. 3–8, 1983*, volume 209 of *Lecture Notes in Physics*, Heidelberg, 1984. Springer.

[45] Manfred Eigen and Ruthild Winkler. *Das Spiel.* Piper, München, 1975. See p. 125 for the snowflakes.

[46] G.J. Erickson and C.R. Smith, editors. *Maximum Entropy and Bayesian Methods in Science and Engineering, Laramie, 1987*, volume 1. Foundations, Dordrecht, 1988. Kluwer.

[47] G.J. Erickson and C.R. Smith, editors. *Maximum Entropy and Bayesian Methods in Science and Engineering, Seattle, Wyoming 1987*, volume 2. Applications, Dordrecht, 1988. Kluwer.

[48] F. Feruglio, A. Strumia, and F. Vissani. Neutrino oscillations and signals in β and $0\nu 2\beta$ experiments. *Nucl. Phys. B*, 637:345–377, 2002.

[49] R.A. Fisher. Theory of statistical information. *Proc. Cambridge Philos. Soc.*, 22:700–725, 1925.

[50] P.F. Fougère, editor. *Maximum Entropy and Bayesian Methods, Dartmouth, 1989*, Dordrecht, 1990. Kluwer.

[51] R. Gilmore. *Lie Groups, Lie Algebras, and Some of Their Applications.* Wiley, New York, 1974.

[52] V.P. Godambe and D.A. Sprott, editors. *Foundations of Statistical Inference. Waterloo, Ontario 1970*, Toronto, 1971. Holt, Rinehart & Winston.

[53] I.S. Gradshteyn and I.M. Ryzhik. *Table of Integrals, Series, and Products.* Academic Press, New York, 1980.

[54] H.-D. Graef, H.L. Harney, H. Lengeler, C.H. Lewenkopf, C. Rangacharyulu, A. Richter, P. Schardt, and H.A. Weidenmüller. Distribu-

tion of eigenmodes in a superconducting stadium billiard with chaotic dynamics. *Phys. Rev. Lett.*, 69:1296–99, 1992.

[55] Particle Data Group. The review of particle physics. *Eur. Phys. J. C*, 15:1–878, 2000.

[56] TRIUMF Computing Services Group. PHYSICA data analysis and visualisation software. www.triumf.ca/physica/html/homepage.html. PHYSICA is available via anonymous ftp from csftp.triumf.ca.

[57] Th. Guhr, A Müller-Groeling, and H.A. Weidenmüller. Random-matrix theories in quantum physics: Common concepts. *Physics Reports*, 299:189–425, 1998.

[58] A. Haar. Der Maßbegriff in der Theorie der kontinuierlichen Gruppen. *Ann. Math.*, 34:147–169, 1933.

[59] Morton Hamermesh. *Group Theory and its Application to Physical Problems.* Addison-Wesley, Reading, Massachusetts, 1962. Reprinted by Dover Publications, New York, 1989.

[60] K.M. Hanson and R.N. Silver, editors. *Maximum Entropy and Bayesian Methods, Santa Fe, 1995*, Dordrecht, 1996. Kluwer.

[61] H.L. Harney, A. Hüpper, and A. Richter. Ericson fluctuations, detailed balance and time-reversal invariance. *Nucl. Phys. A*, 518:35–57, 1999.

[62] W.L. Harper and C.A. Hooker, editors. *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Dordrecht, 1976. Reidel.

[63] J. Hartigan. Invariant prior distributions. *Ann. Math. Statist.*, 35:836–845, 1964.

[64] J.A. Hartigan. Note on confidence-prior of Welch and Peers. *J. Royal Statist. Soc. B*, 28:32–44, 1966.

[65] J.A. Hartigan. *Bayes Theory.* Springer, New York, 1983.

[66] G.R. Heidbreder, editor. *Maximum Entropy and Bayesian Methods, Santa Barbara 1993*, Dordrecht, 1996. Kluwer.

[67] D.A. Hejhal, J. Friedman, M.C. Gutzwiller, and A.M. Odlyzko, editors. *Emerging Applications of Number Theory. Workshop held July 15–26, 1996 at the University of Minnesota*, volume 109 of *The IMA Volumes in Mathematics and its Applications*, New York, 1999. Springer.

[68] J. Hilgert and K.H. Neeb. *Lie-Gruppen und Lie-Algebren.* Vieweg, Braunschweig, 1991.

[69] C. Howson and P. Urbach. *Scientific Reasoning: The Bayesian Approach.* Open Court, Chicago, second edition, 1993.

[70] Shun ichi Amari. *Differential Geometrical Methods in Statistics*, volume 28 of *Lecture Notes in Statistics.* Springer, Heidelberg, 1985.

[71] Shun ichi Amari, editor. *Differential Geometry in Statistical Inference.* Number 10 in Lecture Notes and Monograph Series. Institute of Mathematical Statistics, Hayward, California, 1987.

[72] E.T. Jaynes. Probability theory: The logic of science. Unfinished book, see http://omega.albany.edu:8008/JaynesBook.html.

[73] E.T. Jaynes. Information theory and statistical mechanics I. *Phys. Rev.*, 106:620–630, 1957.

[74] E.T. Jaynes. Information theory and statistical mechanics II. *Phys. Rev.*, 108:171–190, 1957.

[75] E.T. Jaynes. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics*, SSC-4(3):227–241, September 1968.

[76] E.T. Jaynes. Marginalization and prior probabilities. In Zellner [163], pages 43–87.

[77] E.T. Jaynes. Confidence intervals vs. Bayesian intervals. In R.D. Rosenkrantz, editor, *Papers on Probability, Statistics and Statistical Physics*, volume 158 of *Synthese Library*, pages 149–209, Dordrecht, 1983. Reidel. The original article is in [62].

[78] E.T. Jaynes. Bayesian methods: General background. In Justice [83], pages 1–25.

[79] E.T. Jaynes. Probability theory as logic. In Fougère [50], pages 1–16.

[80] H. Jeffreys. *Theory of Probability.* Oxford University Press, Oxford, 1939. 2nd edition 1948; 3rd edition 1961, here Jeffreys' rule is found in III §3.10.

[81] H. Jeffreys. An invariant form of the prior probability in estimation problems. *Proc. of the Roy. Soc.*, 186 A:453–461, 1946.

[82] W.T. Grandy Jr. and L.H. Schick, editors. *Maximum Entropy and Bayesian Methods, Laramie, 1990*, Dordrecht, 1991. Kluwer.

[83] J.H. Justice, editor. *Maximum Entropy and Bayesian Methods in Applied Statistics, Calgary, 1984*, Cambridge, 1986. Cambridge University Press.

[84] R. E. Kass. The geometry of asymptotic inference. *Statistical Science*, 4:188–219, 1989.

[85] R.E. Kass. *The Riemannian Structure of Model Spaces: A Geometrical Approach to Inference*. PhD thesis, University of Chicago, 1980. See especially pages 94–95.

[86] R.E. Kass and L. Wasserman. The selection of prior distributions by formal rules. *J. Am. Statist. Assoc.*, 91:1343–1370, 1996.

[87] H.V. Klapdor-Kleingrothaus. Reply to a comment of article 'evidence for neutrinoless double beta decay'. arXiv:hep-ph/0205228.

[88] H.V. Klapdor-Kleingrothaus, A. Dietz, H.L. Harney, and I.V. Krivosheina. Evidence for neutrinoless double beta decay. *Mod. Phys. Lett. A*, 16:2409–2420, 2001.

[89] Pierre Simon de Laplace. Mémoire sur la probabilité des causes par les événements. *Mém. de math. et phys. présentés à l'Acad. roy. des sci.*, 6:621–656, 1774. Reprinted in [91], vol. 8, pages 27–65. An English translation can be found in [144].

[90] Pierre Simon de Laplace. *Théorie analytique des probabilités*. Courcier, Paris, 1812. Second edition 1814; third edition 1820. Reprinted as vol. 7 of [91].

[91] Pierre Simon de Laplace. *Œuvres complètes de Laplace.* Gauthier-Villars, Paris, 1886–1912. 14 volumes.

[92] Pierre Simon de Laplace. *A Philosophical Essay on Probabilities.* Dover, New York, 1951. Original title *Essay philosophique sur les probabilités.*

[93] P.M. Lee. *Bayesian Statistics: An Introduction.* Arnold, London, second edition, 1997.

[94] Don S. Lemons. On the numbers of things and the distribution of first digits. *Am. J. Phys.*, 54:816–817, September 1986.

[95] J. Levin, D. Kella, and Z. Vager. Interpretation of data in multidimensional spaces and its application to Coulomb-explosion imaging. *Phys. Rev.*, A 53:1469–1475, 1996.

[96] R.D. Levine. Geometry in classical statistical thermodynamics. *J. Chem. Phys.*, 84:910–916, 1986.

[97] H.W. Lewis. What is an experiment? *Am. J. Phys.*, 50:1164–1165, 1982.

[98] D.V. Lindley. A statistical paradox. *Biometrika*, 44:187–192, 1957.

[99] W. Lucha and F. Schöberl. *Gruppentheorie – Eine elementare Einführung für Physiker.* B.I. Hochschultaschenbuch, Mannheim, 1993.

[100] Benoît Mandelbrot. The variation of certain speculative prices. *J. Business*, 36:394–419, 1963.

[101] R.N. Mantegna and H.E. Stanley. Scaling behaviour in the dynamics of an economic index. *Nature*, 376:46–49, 1996.

[102] W.L. Martinez and A.R. Martinez. *Computational Statistics Handbook with MATLAB*. Chapman & Hall, Boca Raton, 2002.

[103] Microcal™. Origin, user's manual. Microcal Software, Inc., Northampton, MA, 1997.

[104] G.E. Mitchell, J.D. Bowman, S.I. Penttilä, and E.I. Sharapov. Parity violation in compound nuclei: Experimental methods and recent results. *Phys. Rep. — Rev. Sec. Phys. Lett.*, 354:157–241, 2001. The original publication on parity violation in ^{115}In is [140].

[105] G.E. Mitchell, J.D. Bowman, and H.A. Weidenmüller. Parity violation in the compound nucleus. *Rev. Mod. Phys.*, 71:445–457, 1999.

[106] A. Mohammad-Djafari and G. Demoment, editors. *Maximum Entropy and Bayesian Methods, Paris 1992*, Dordrecht, 1993. Kluwer.

[107] Rahul Mukerjee and Dipak K. Dey. Frequentist validity of posteriori quantiles in the presence of nuisance parameter: Higher order asymptotic. *Biometrika*, 80(3):499–505, 1993.

[108] J. Neyman and L.M. Le Cam, editors. *Bernoulli, Bayes, Laplace. Proceedings of an International Research Seminar. Statistical Laboratory.* Springer, New York, 1965.

[109] Anna Nicolaou. Bayesian intervals with good frequentist behaviour in the presence of nuisance parameters. *J. R. Statist. Soc. B*, 55:377–390, 1993.

[110] Antony O'Hagan. *Bayesian Inference*, volume 2B of *Kendall's Advanced Theory of Statistics.* Arnold, London, 1994.

[111] Maria Overbeck-Larisch and Wolfgang Dolejski. *Stochastik mit Mathematica.* Vieweg, Braunschweig, 1998.

[112] H. W. Peers. On confidence points and Bayesian probability points in the case of several parameters. *J. R. Statist. Soc. B*, 27:9–16, 1965.

[113] H. W. Peers. Confidence properties of Bayesian interval estimators. *J. R. Statist. Soc. B*, 30:535–544, 1968.

[114] V. Plerou, P. Gopikrishnan, B. Rosenow, L.A.N. Amaral, and H.E. Stanley. Econophysics: Financial time series from a statistical physics point of view. *Physica A*, 279:443–456, 2000.

[115] N.G. Polson and G.C. Tiao, editors. *Bayesian Inference.* Cambridge University Press, Cambridge, 1995. Two volumes.

[116] W.H Press. *Numerical Recipes in FORTRAN: The Art of Scientific Computing.* Cambridge University Press, Cambridge, second edition, 1992.

[117] W.H. Press, S.A. Teukolski, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C.* Cambridge University Press, Cambridge, 2nd edition, 1992.

[118] C. Radakrishna Rao. Differential metrics in probability spaces based on entropy and divergence measures. Technical Report Report No. AD-A160301, AFOSR-TR-85-0864, Air Force Office of Scientific Research, Bolling AFB, DC, 1985.

[119] C. Radakrishna Rao. Differential metrics in probability spaces. In ichi Amari [71], pages 218–240.

[120] Hans Reichenbach. *Wahrscheinlichkeitslehre.* Vieweg, Braunschweig, 1994.

[121] A. Richter. Playing billiards with microwaves – quantum manifestations of classical chaos. In Hejhal et al. [67], pages 479–523.

[122] C.P. Robert. *The Bayesian Choice.* Springer, Berlin, Heidelberg, 2nd edition, 2001.

[123] C.C. Rodriguez. Objective Bayesianism and geometry. In Fougère [50], pages 31–39.

[124] C.C. Rodriguez. Are we cruising a hypothesis space? In von der Linden et al. [151], pages 131–139.

[125] B.P. Roe. *Probability and Statistics in Experimental Physics.* Springer, Berlin, Heidelberg, 1992.

[126] J.L. Savage. *The Foundations of Statistics.* Wiley, New York, 1954. There is a second edition by Dover, New York 1972.

[127] J.L. Savage. Elicitation of personal probabilities and expectations. *J. Am. Statist. Assoc.*, 66:783–801, 1971.

[128] E. Schrödinger. The foundation of the theory of probability I. *Proc. R. Irish Acad.*, 51A:51–66, 1947. Reprinted in *Gesammelte Abhandlungen*, Vienna 1984, vol. 1, pages 463–478.

[129] E. Schrödinger. The foundation of the theory of probability II. *Proc. R. Irish Acad.*, 51A:141–146, 1947. Reprinted in *Gesammelte Abhandlungen*, Vienna 1984, vol. 1, pages 479–484.

[130] D.S. Sivia. *Data Analysis. A Bayesian Tutorial.* Clarendon, Oxford, 1998.

[131] J. Skilling, editor. *Maximum Entropy and Bayesian Methods, Cambridge, 1988*, Dordrecht, 1989. Kluwer.

[132] J. Skilling and S. Sibisi, editors. *Maximum Entropy and Bayesian Methods, Cambridge, 1994*, Dordrecht, 1996. Kluwer.

[133] C.R. Smith and G.J. Erickson, editors. *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems, Laramie, 1983*, Dordrecht, 1987. Kluwer.

[134] C.R. Smith, G.J. Erickson, and P.O. Neudorfer, editors. *Maximum Entropy and Bayesian Methods, Seattle 1991*, Dordrecht, 1992. Kluwer.

[135] C.R. Smith and W.T. Grandy Jr., editors. *Maximum-Entropy and Bayesian Methods in Inverse Problems [Proceedings of two workshops, Laramie 1981 and 1982]*, Dordrecht, 1985. Reidel.

[136] H.E. Stanley, L.A.N. Amaral, D. Canning, P. Gropokrishnan, Y. Lee, and Y. Liu. Econophysics: Can physicists contribute to the science of economics? *Physica A*, 269:156–169, 1999.

[137] S. Steenstrup. Experiments, prior probabilities, and experimental prior probabilities. *Am. J. Phys.*, 52:1146–1147, December 1984.

[138] C.M. Stein. Approximation of improper prior measures by proper probability measures. In Neyman and Le Cam [108], pages 217–240.

[139] W. Stephen. *The Mathematica Book.* Cambridge University Press, Cambridge, 4th edition, 1999.

[140] S.L. Stephenson, J.D. Bowman, F. Corvi, B.E. Crawford, P.P.J. Delheij, C.M. Frankle, M. Iinuma, J.N. Knudson, L.Y. Lowie, A. Masaike, Y. Masuda, Y. Matsuda, G.E. Mitchell, S.I. Pentilä, H. Postma, N.R. Roberson, S.J. Seestrom, E.I. Sharapov, H.M. Shimizu, Y.-F. Yen, V.W. Yuan, and L. Zanini. Parity violation in neutron resonances in ^{115}in. *Phys. Rev. C*, 61:045501/1–11, 2000.

[141] S.L. Stephenson, J.D. Bowman, B.E. Crawford, P.P.J. Delheij, C.M. Frankle, M. Iinuma, J.N. Knudson, L.Y. Lowie, A. Masaike, Y. Matsuda, G.E. Mitchell, S.I. Pentilä, H. Postma, N.R. Roberson, S.J. Seestrom, E.I. Sharapov, Y.-F. Yen, and V.W. Yuan. Parity nonconservation in neutron resonances in ^{232}th. *Phys. Rev. C*, 58:1236–1246, 1998.

[142] S. Stigler. *American Contributions to Mathematical Statistics in the Nineteenth Century.* Arno, New York, 1980. Two volumes.

[143] S.M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900.* Harvard University Press, Cambridge, MA, 1986.

[144] S.M. Stigler. Laplace's 1774 memoir on inverse probability. *Statist. Sci.*, 1:359–378, 1986.

[145] M. Stone. Right Haar measure for convergence in probability to quasi posteriori distributions. *Ann. Math. Statist.*, 36:440–453, 1965.

[146] C. Villegas. On Haar priors. In Godambe and Sprott [52], pages 409–414.

[147] C. Villegas. Inner statistical inference. *J. Am. Statist. Assoc.*, 72:453–458, 1977.

[148] C. Villegas. On the representation of ignorance. *J. Am. Statist. Assoc.*, 72:651–654, 1977.

[149] C. Villegas. Inner statistical inference II. *Ann. Statist.*, 9:768–776, 1981.

[150] C. Villegas. Bayesian inference in models with Euclidean structures. *J. Am. Statist. Assoc.*, 85(412):1159–1164, 1990.

[151] W. von der Linden, V. Dose, R. Fischer, and R. Preuss, editors. *Maximum Entropy and Bayesian Methods, Garching, 1998*, Dordrecht, 1999. Kluwer.

[152] C.F. von Weizsäcker. *Die Einheit der Natur.* Hanser, Munich, 1971. See page 241.

[153] B.L. Welch. On comparisons between confidence points procedures in the case of a single parameter. *J. Royal Statist. Soc. B*, 27:1–8, 1965.

[154] B.L. Welch and H.W. Peers. On formulae for confidence points based on integrals of weighted likelihood. *J. R. Statist. Soc. B*, 25:318–329, 1963.

[155] Hermann Weyl. *Symmetry.* Princeton University Press, Princeton, NJ, 1952. A German edition was published 1955 by Birkhäuser, Basel, under the title of *Symmetrie.*

[156] Hermann Weyl. *The Classical Groups: Their Invariants and Representations.* Princeton University Press, Princeton, NJ, 1953. Reprinted 1964.

[157] Hermann Weyl. *Gruppentheorie und Quantenmechanik.* Wissenschaftliche Buchgesellschaft, Darmstadt, 1967. Reprint of the 2nd edition, Hirzel, Leipzig, 1931.

[158] E.P. Wigner. Results and theory of resonance absorption. In C.E. Porter, editor, *Statistical Theory of Spectra: Fluctuations*, page 199. Academic Press, New York, 1957. The original work is in Oak Ridge National Laboratory Report No. ORNL-2309, 1957.

[159] E.P. Wigner. *Group Theory and Its Application to the Quantum Mechanics of Atomic Spectra.* Academic Press, New York, 1959. Translated from the German by J.J. Griffin.– Expanded and improved edition.

[160] W.von Witsch, A. Richter, and P.von Brentano. Test of time-reversal invariance through the reactions $^{24}\mathrm{mg} + \alpha \leftrightarrow^{27} \mathrm{al} + p$. *Phys. Rev.*, 169:923–932, 1968.

[161] E. Witten. High-energy physics: The mass question. *Nature*, 414:969–971, 2002.

[162] B.G. Wybourne. *Classical Groups for Physicists.* Wiley, New York, 1974.

[163] A. Zellner, editor. *Bayesian Analysis in Econometrics and Statistics: Essays in Honour of H. Jeffreys*, volume 1 of *Studies in Bayesian Econometrics*, Amsterdam, 1980. North Holland.