
Bayesian Statistical Inference: An Overview

Ludovico Piccinato

Abstract

The Bayesian approach for statistical inference is examined, pointing out also the differences among the many Bayesian philosophies. Moreover comments are given about topics where the Bayesian approach seems (at least to Bayesians) more suitable than the alternatives. At last the decision-theoretic approach is shortly discussed.

1 An Historical Outline

In the second half of the nineteenth century the dominant approach to statistical inference was the framework originated by P.S. Laplace, where an honour place was given to Bayes theorem. Both Bayes and Laplace are sometimes mentioned as supporters of a very strict approach to probability and inference: for the classical problem of the *probability of causes*, they would assume an equal probability for the causes, so that, in a modern language, the final probabilities would be proportional to the likelihoods. As shown by authoritative historians, this picture is not correct. Indeed Bayes in his famous 1763 posthumous paper assumed equal prior *predictive* probabilities (that is probabilities of observables) so that the equiprobability of causes was derived as a consequence (Stigler 1982, 1986). The *Principle of Indifference* formulated by P.S. Laplace in 1774 states that the ratio of the final probabilities of two causes A_i and A_j conditional on an event E equals the likelihood ratio $P(E|A_i)/P(E|A_j)$. This amounts to say that the initial probabilities of the causes are equal. But in many places Laplace himself explains that when the cases at hand are not equally possible, one has to subdivide

L. Piccinato (✉)
Sapienza Università di Roma, Rome, Italy
e-mail: ludovico.piccinato@uniroma1.it

or join them to reach a set of equipossible cases. Therefore, even for Laplace, equiprobability is the result of an elaboration, not an aprioristic assumption. It has also been observed (Stigler 1986, p. 135) that in some occasions Laplace explicitly adopted non-uniform priors. Also Karl Pearson in most of his works is clearly sympathetic with the approach based on the so-called *inverse probabilities* (see, e.g., Dale 1991), and his influence was relevant at least until the first decades of the twentieth century. Moreover many authors suggest the use of an uniform prior because it is approximately justified in the case of large sample size, a position that was deepened in modern times.

The break in the tradition is due to the work of Fisher. His main theoretical contribution in this period is Fisher (1922), where Sect. 1 starts with a severe criticism of a famous paper by Pearson (1920). Fisher's offensive, followed after a few years by the well-known contributions of J. Neyman and E.S. Pearson, provoked an eclipse of the Bayesian approach for about three decades (Zabell 1989). This does not mean that in the same period remarkable developments in the Bayesian framework did not occur. On the contrary, in the same period, important works by H. Jeffreys, I.J. Good, F.P. Ramsey and B.de Finetti were published; the point is that the statistical community paid very few attention to such arguments, whose relevance was recognized only many decades later. The change occurred since the 1950s (of the twentieth century) with the work of some scholars, including L.J. Savage, H. Raiffa and R. Schlaifer, D.V. Lindley (de Finetti 1959; Lindley 1965; Ramsey 1926; Savage 1954, 1962). English translations of some key works of de Finetti were provided. In the 1970s several books were published where a definitive setting was given to the Bayesian theory. These include DeGroot (1970), de Finetti (1970), Box and Tiao (1973), Lindley (1972), Berger (1985) and, for the Italian literature, Daboni and Wedlin (1982) and Cifarelli and Muliere (1989). For more bibliographical details see Fienberg (2005); extensive historical information is given in Fienberg (1992) and Fienberg (2006).

Let us assume a standard statistical model, say $\{p(x|\theta), x \in \mathcal{X}, \theta \in \Theta\}$, where x is the possible result, θ is the unknown parameter, p is a density or a mass function, and let x_{obs} the observed result. If the goal is to make inferences on the unknown parameter θ , a Bayesian statistician of any century should first of all complete the model adding a probability law (the *prior distribution*) for the parameter, say $\pi(\theta)$. Then he/she can use the celebrated Bayes' formula $\pi(\theta|x_{\text{obs}}) \propto \pi(\theta)p(x_{\text{obs}}|\theta)$ which provides the *posterior* probability distribution for the parameter, i.e. the probability distribution updated with the acquired information. The use of priors is the most evident difference between Bayesian and non-Bayesian methods and a more detailed analysis will be given in the next section.

Automatically, the use of Bayes' formula implies that the values $p(x|\theta)$ with $x \neq x_{\text{obs}}$ have no effect on the analysis. On the contrary the frequentist approach to statistical inference produces conclusions which depend on the whole statistical model, not only on the likelihood function $L(\theta|x_{\text{obs}}) = p(x_{\text{obs}}|\theta)$. In this writer's opinion this aspect, that is the violation of the so-called *Likelihood Principle*, is what mostly moves the frequentist approach away from the Bayesian approach. More comments will be given in Sect. 3.

2 Prior Probabilities

In the first decades of the twentieth century the concept of probability was studied in great depth. One well-known approach sees the probability as a limit of observable frequencies, an interpretation which is not suitable for the Bayesian methodology because of its lack of generality. On the contrary, the probability as a measure of belief in the occurrence of an uncertain event (*subjective* or *personal* probability) has clearly a general applicability. A way to evaluate a probability is a comparison with a standard (see [Bertrand 1907](#), and for a more recent version [Lindley 2006](#)). Another approach, developed independently by F.P. [Ramsey \(1926\)](#) and by B. [de Finetti \(1931\)](#) specifies the probability as the fair price of an unitary stake in a bet on an uncertain event. Then de Finetti introduced the principle of *coherence*, i.e. that a subject must avoid bets where he would lose whatever the result and showed that such principle is equivalent to the standard Kolmogorov axioms of probability (not considering the complete additivity, which turns out to be a possible but non necessary choice). Moreover de Finetti formulated the problem of inference as prediction of future results given a partial initial trajectory of the stochastic process of the observations ([de Finetti 1937](#)). This formulation does not introduce unknown parameters and replaces the standard notion of random sampling with the concept of *exchangeability*. At a first sight this approach is radically different from the standard one, popularized by Fisher and Neyman and based on the usual statistical models. However the celebrated de Finetti representation theorem shows that exchangeability corresponds to conditional independence so that the procedure based on prior plus likelihood is essentially equivalent to the completely predictive approach, since the assumptions on the process also determine prior and likelihood.

The collaboration between de Finetti and Savage in the 1950s contributed very much to the revival of the Bayesian approach in general, in particular to the acceptance of subjective probabilities. Note that in the Chap. XII of [de Finetti \(1970\)](#), after a short premise about its connection with the predictive approach, the problem of inference is directly treated in the current model-based framework. A common way to respect the original predictive approach by de Finetti is to “justify” the model-based approach through the representation theorem; see, e.g., [Dawid \(1982\)](#). The most systematic treatment in this framework was given by [Bernardo and Smith \(1994\)](#). A claim in favour of the predictivistic approach was presented by [Cifarelli and Regazzini \(1982\)](#) and remarkable methodological researches were conducted under this perspective. For instance, in [Regazzini \(1999\)](#) such approach is explored in a nonparametric context. Less common are treatments oriented to applied problems; the exceptions include [Muliere and Petrone \(1993\)](#) and [Spizzichino \(2001\)](#). A general analysis of de Finetti’s work in mathematical statistics cannot be given here, and we refer to [Piccinato \(1986\)](#), [Cifarelli and Regazzini \(1996\)](#), [Bernardo \(1998\)](#) and the references therein. The implementation of the subjectivistic paradigm requires a new interest for the problem of elicitation, i.e. how to put in a probabilistic form the knowledge owned by the experts. Many papers were dedicated to this topic, starting with [de Finetti and Savage \(1962\)](#); a

basic remark is that it is often easier to give a probability to the observables than to unknown parameters. For a recent systematic review see [O'Hagan et al. \(2006\)](#).

A concept by de Finetti which found only a limited acceptance among statisticians was finite additivity (for a deepening see [Cifarelli and Regazzini 1996](#)). It is known, however, that complete additivity allows to use properties which hold in the finite problems (for instance, conglomerability) so that its adoption is natural when infinity appears essentially as an approximation of large or unprecised numbers. For special problems, when infinity has its own specific role, resorting to finite additivity can be clarifying also in practical settings (see, e.g., [Scozzafava 1984](#)).

In the 1960s the classical argument about the almost irrelevance of the prior in the presence of a significant experimental information (the *Principle of Precise Measurement*) was reconsidered and clarified ([Edwards et al. 1963](#); [Savage 1962](#)). This topic has a connection with the recurrent idea of using *noninformative* priors. In the classic period the uniform distribution was often and naively used in this sense, though many authors remarked that the uniformity is not maintained under one-to-one transformations, while on the contrary noninformativity should remain. H. Jeffreys introduced in the 1930s his *invariant rule*, which satisfies this property. The concept of noninformativeness, to take it seriously, has surely very weak bases: a probability distribution always represent an information. This explains why a multiplicity of different proposals were advanced in the years (see [Kass and Wasserman 1993](#)). One of these proposals, cautiously named *reference prior*, is based on the idea of minimizing the missing information; it partially extends Jeffrey's rule and is now almost a standard. The proposal originated by a paper by J.M. [Bernardo \(1979\)](#) and was later developed in particular by J. Berger. For more recent treatments see [Berger and Bernardo \(1992\)](#) and [Berger et al. \(2009\)](#). A criticism to the method is that it entails a violation of the Likelihood Principle, since the posterior distribution depends not only on the likelihood function but also on the model (see, e.g., the discussion by Lindley of [Bernardo 1979](#)). It could be remarked that this kind of prior (as Jeffrey's) is necessarily connected with the model since it is obviously impossible to speak of minimal information in an absolute sense; compatibility with the Likelihood Principle is, however, hold by [Bernardo \(2005, Sect. 3.6\)](#).

The availability of an agreed default rule, where no effort of elicitation is required, suggested an approach which is now called *Objective Bayesian Analysis*. For a comparison of the contrasting arguments see [Berger \(2006\)](#), [Goldstein \(2006\)](#) and the related discussion. Authoritative proponents of the objective approach ([Berger et al. 2009](#)) remark that the term "objective" means that the procedure only depends on the model assumed and the data obtained, so that the kind of objectivity is simply the same of the frequentist statistics. There are significant practical and logical differences with a pure subjectivistic approach, but, in the present writer's opinion, these are only variants of a more general Bayesian framework. As mentioned before, I think that the qualification "Bayesian" is due when we assume that any uncertain event has a probability. It is not necessary that there exists a subject who has actually such information. In any case the Bayes theorem explains how to update an information, be it effective or conventional.

In order to simplify the elicitation process many suitable partial formalizations are in use. Among the most known tools there are the conjugate classes of priors. The concept had a systematic treatment by [Raiffa and Schlaifer \(1961\)](#) but the same idea (often limited to the binomial model) appeared many times much before. Until the availability of the MCMC techniques it was often difficult to get the posterior distributions unless the prior was a member of a conjugate class. Using the de Finetti theorem [Lindley \(1965\)](#) represented exchangeable parameters through a hierarchical model. This allowed a very convenient Bayesian treatment of the general linear model (see [Lindley and Smith 1972](#) for a generalization).

In the last decades procedures pointing at conventional priors were suggested and proved useful in applications. For instance, the book ([Spiegelhalter et al. 2004](#)) made popular the use of sceptical priors, mainly in a clinical context. Another technique of modelling the prior is the use of power priors, initially proposed in [Ibrahim and Chen \(2000\)](#), that is suitable, for instance, when there are historical data similar to those at hand but not such to justify the assumption of exchangeability (see also [De Santis 2007](#)).

Another departure from an ideal subjectivistic practice is the distinction, now very much used, between *design* prior and *analysis* prior. This idea appears from the first time in [Tsutakawa \(1972\)](#) and was developed with various motivations, as the necessity of having a proper prior in the stage of design (while in the stage of analysis an improper prior is often preferred) or of privileging the region of the parameter space which could make the results more interesting (see [Etzioni and Kadane 1993](#); [Wang and Gelfand 2002](#)).

3 Statistical Models and Likelihood Principle

In the framework of a standard statistical model the Likelihood Principle has its own strength even without reference to the Bayesian paradigm. Savage, in the discussion of [Birnbaum \(1962\)](#), writes that he came to Bayesian statistics seriously only through recognition of the Likelihood Principle. The issue is, however, controversial; for instance, [Cox \(2006, p. 47\)](#) comments that the principle is convincing in its weak version (two results under the same model are equivalent when the likelihood functions are proportional) but qualifies “less compelling” the strong version (when it is not required that the model is fixed). It is well known that among the merits of Fisher there is the introduction of the likelihood function ([Fisher 1922](#)). His attitude about the Likelihood Principle has been largely discussed; for a thorough analysis see [Savage \(1976\)](#). The formal definition is due to [Birnbaum \(1962\)](#), but the argument was already informally in use. Many Bayesian authors stress the relevance of the principle in the context of a Bayesian analysis, see, e.g., [Edwards et al. \(1963\)](#) and [Lindley \(1972\)](#); moreover the likelihood literature is a source of interest for the Bayesian school (we could mention at least [Basu 1975](#), [Royall 1997](#)). A definitive treatment is [Berger and Wolpert \(1988\)](#).

A Bayesian attitude is also indirectly favoured by the apparent necessity of using sometimes at least a partial conditioning. One of the most famous examples, the

case of the two laboratories, was published by D.R. Cox (1958). The example shows that a rigid applications of the frequentist rule, which implies an exclusive attention to the long run performances of the statistics, can be untenable, while if one conditions on a suitable ancillary statistic the paradox disappears, as it automatically happens in a Bayesian analysis. This kind of examples took many years to become popular (with the exception of the Bayesian literature), at least in the textbooks. I can just mention that E.L. Lehmann, in the second edition of his classic *Testing Statistical Hypotheses*, added a last chapter where the topic is thoroughly examined and a serious comment on the suitability of the unconditional approach is provided (Lehmann 1986, p. 541): “if repetitions [...] are potential rather than actual interest will focus on the particular event at hand, and conditioning seems more appropriate”. Therefore the comparison among the main theories of inference involves more the comparison between choosing a conditioning statistic and choosing a prior distribution, than adopting an objective or a subjective approach. Let us finally mention that recent researches by Bayesian authors about the relationships between the different inferential approaches give a special role just to the conditional frequentist approach (Bayarri and Berger 2004; Berger 2003).

The main advantage of the model-based approach is the possibility of separating the different sources of information, i.e. the pre-experimental information, inbedded in the prior, and the experimental information, inbedded in the likelihood function (in the framework of the model). However, this approach is not completely general for inference problems. Difficulties in finding an agreed specification of the likelihood function were, for instance, considered in Bayarri et al. (1988). In any case, however, a Bayesian can resort to the completely predictive approach in the sense of de Finetti, though this could force to reformulate inferential problems.

4 The Development of Bayesian Methodology

Many hints to the development of Bayesian methodology were provided by the existing frequentist methodology: problems having a solution in a non-Bayesian approach had to be revised and reformulated. I shall comment some examples.

One of these themes is *robustness*, that was initially considered in the Bayesian literature mainly in relation to the choice of the prior. Instead of considering a single prior, classes of priors were taken into account in order to check the resulting differences. Beyond parametric classes, attention was drawn also to nonparametric or to partially nonparametric classes, as the class of monotone distributions, of symmetric distributions, or contaminated distributions, quantile classes and so on (for reviews see Berger et al. 1996, Ríos Insua and Ruggeri 2000). This rich literature allowed to move from mathematical convenience to much more realistic formulations of prior uncertainty. The proposal of interactive procedures (as in Liseo et al. 1996) was a further step in this direction.

Another topic inherited by the frequentist statistic is the issue of *model testing and selection*. In a controversial paper Box (1980a,b) claimed that the Bayesian analysis is fully adequate within a given model but is not useful for model criticism.

His proposal for model criticism is based on the prior predictive distribution and has a clear frequentist flavour, together with an analogy with the classical p -value. This proposal suggested many developments in different directions. From one hand, letting aside the traditional criticism to the theory of significance from a Bayesian viewpoint (a seminal paper is [Berger 1986](#)), new concepts of Bayesian p -values were introduced, with application to the case of composite hypotheses and to the model criticism ([Bayarri and Berger 2000](#)). When the goal is to choose one model many authors suggest an explicit decision setting; see, for instance, [San Martini and Spezzaferri \(1984\)](#), [Key et al. \(2001\)](#), [Walker et al. \(2001\)](#), [Barbieri and Berger \(2004\)](#). The most natural Bayesian approach to compare many models, when one is considered “true” (the so called M-closed setup), is however to attach probabilities to every model, in order to account for model uncertainty, and proceed with the standard probability rules. A general exposition of the Bayesian model averaging is [Hoeting et al. \(1999\)](#). An alternative path is the use of Bayes factors for comparing models without assigning prior probabilities to the model themselves. This was, for instance, proposed by O’Hagan in the discussion of [Box \(1980b\)](#). Problems associated with the Bayes factors should, however, be considered; see [Lavine and Schervish \(1999\)](#) about their use as measures of evidence and [Carota and Parmigiani \(1996\)](#) about their use with nonparametric models. For general treatments refer to [Kass and Raftery \(1995\)](#) and [Berger \(1999\)](#). In the comparison of models it may be desirable to assign improper priors to the parameters of each model. Apart from special situations (e.g., [Consonni and Veronese 1991](#)), a general solution is resorting to the so-called partial Bayes factors; for different proposals and discussions see [Berger and Pericchi \(1996\)](#), [O’Hagan \(1995\)](#) and [De Santis and Spezzaferri \(1997\)](#). For the general topic of model selection, including also the assumption of a Model-open setting, that is when it is not assumed that the set of models contains the “true” model, see [Racugno \(1997\)](#), [Lahiri \(2001\)](#), [Kadane and Lazar \(2004\)](#) and [Clyde and George \(2004\)](#).

As another example, let us mention *nonparametric inference*. The mathematical modeling of the problem requires the use of probability measures on function spaces so that the practical understanding of the prior assumptions is quite demanding and a Bayesian treatment was delayed for a long time. [Lindley \(1972, p. 66\)](#) wrote “this is a subject about which the Bayesian method is embarrassingly silent”. This was true, at those times, although in a very short paper, many years before, [de Finetti \(1935\)](#) outlined the issue in a Bayesian framework (comments on this in [Cifarelli and Regazzini 1996](#)). A turning point was the approach by [Ferguson \(1973\)](#) through the so-called Dirichlet process, which gave rise to many of the contemporary researches. Many different extensions and alternatives were since then proposed (see, e.g., [Walker and Muliere 1997](#), [Lijoi and Prünster 2000](#)).

At last we mention some problems that the Bayesian approach can handle in a particularly easy way, differently from the other approaches. These include the elimination of *nuisance parameters*, the possibility of a direct treatment of *prediction problems*, the possibility of a complete treatment of the *design of experiments*. Let us suppose that the parameter θ is a vector, say $\theta = (\lambda, \gamma)$, and that the inference concerns only the component λ . The likelihood function depends

of course on both the components, but, given the posterior distribution, we can get a posterior distribution for the parameter of interest alone by a simple marginalization, that is $\pi(\lambda|x_{\text{obs}}) = \int \pi(\lambda, \gamma|x_{\text{obs}})d\gamma$. For a recent review see [Liseo \(2006\)](#).

A problem of prediction is characterized by a statistical model $(q(y|\theta), y \in \mathcal{Y}, \theta \in \Theta)$, for the future result Y , where the “true” parameter θ is the same of the statistical model of the observation X . Under the only assumption of the independence of X and Y (for a given θ) it is impossible to represent how the knowledge of X provides information on Y . On the contrary, the introduction of a prior distribution $\pi(\theta)$ for the parameter allows us to calculate the conditional distribution of Y given X , that is $m(y|x_{\text{obs}}) = \int q(y|\theta)\pi(\theta|x_{\text{obs}})d\theta$ which is the most natural base for a prediction of the value y . A general reference across the different approaches, is [Geisser \(1993\)](#).

In a problem of design of an experiment we have a class \mathcal{E} of possible experiments, which can differ, for instance, for size of the sample, sequential stopping rule, choice of the controlled variables and so on. Any choice $e \in \mathcal{E}$ will get an evaluation depending in general on the result x and the parameter θ , both not known in advance. Under these conditions, general methods of eliminating θ without an integration with respect to a prior distribution are unreasonable or unavailable, unless there are particular patterns as it may occur with linear models (Kiefer’s theory). The case of linear models were reformulated also in a Bayesian setting, see, e.g., [Smith and Verdinelli \(1980\)](#) and [Giovagnoli and Verdinelli \(1983\)](#). A particular problem, of a great practical importance, is the determination of an optimal sample size. The diffusion of the Bayesian approach produced a lot of new methods; a starting point for the more recent researches in this field is the issue 2, 1997 of the journal *The Statistician*, entirely devoted to the subject; see also [De Santis \(2006\)](#) for a robust approach. Note that the choice of a design s primarily a decision problem, though the final goal could be an inferential statement. An excellent framework also for this particular problem is therefore given in the classic text by [Raiffa and Schlaifer \(1961\)](#). For further reviews and treatments always in a Bayesian setting see [Chaloner and Verdinelli \(1995\)](#) and [Piccinato \(2009\)](#).

5 Relations with the Decision-Theoretic Approach

The decision-theoretic approach has many merits in clarifying the differences between the approaches. One may ask whether the reformulation in decision-theoretic terms does not modify or restrict the aims of inference. This is surely not true for the Neyman–Pearson–Wald school, because in that case the idea of optimization is intrinsic to the theory. It is well known that many times Neyman explained how inductive reasoning is often impossible since it involves events lacking a probability; inductive behaviour, i.e. optimizing the long run performance of procedures, would be instead the operational solution (see, e.g., [Neyman 1957](#)). For the Bayesian approach the situation is different, since both paths are possible; either completing the model involving also a specification of the available terminal

acts with the corresponding utilities/losses or performing a purely probabilistic analysis, without a formal implication of any decision.

If we explicitly adopt a complete decision setting, it is natural that the Bayesian approach aims at minimizing the expected loss of the terminal actions conditional on a specific result, while the frequentist approach aims at minimizing the risk of a procedure unconditionally on the result but conditionally on the unknown parameter. Wald (1950) proved that there is a strong connection between the two optimalities (the complete class theorem). Loosely speaking, the theorem shows that any reasonable decision is formally Bayesian and vice versa. In the Wald's approach prior probabilities are only weighting devices, but this result can be commented as a partial conciliation between the two approaches (see, e.g., de Finetti 1951, Raiffa and Schlaifer 1961, p. 16). The calculation of the risk requires, however, an integration on the sample space, and this is a violation of the likelihood principle; this can imply contradictions with the basic goal of minimizing losses for terminal acts (see, e.g., Piccinato 1980). A good long run performance is in itself a sensible characteristic for a statistical procedure but it should not be achieved by omitting to take into account the actual result, when available.

6 Final Remarks

The aim of the present paper is to outline the many theories and proposals framed in the Bayesian setting (for further information see Berger 2000). We hold that all this is a richness of the approach and does not preclude its fundamental unitarity, based on the formal representation of the process of learning from experience. Limitations of space and knowledge prevent here any hope of completeness. It is worth noting that the development of simulation methods allows now to deal with complex models, with thousands of parameters, as it occurs in the modern applications in genomics and in environmental analysis (see, e.g., Chen et al. 2010), while in the past mathematical tractability played a serious limiting role. In the next future, overviews of the Bayesian approach will surely focus much more on these aspects.

References

- Barbieri, M.M., Berger, J.O.: Optimal predictive model selection. *Ann. Stat.* **32**, 870–897 (2004)
- Basu, D.: Statistical information and likelihood. *Sankhyā A* **37**, 1–71 (1975)
- Bayarri, M.J., Berger, J.O.: P-values for composite null models (with discussion). *J. Am. Stat. Assoc.* **95**, 1127–1170 (2000)
- Bayarri, M.J., Berger, J.O.: The interplay of Bayesian and frequentist analysis. *Stat. Sci.* **19**, 58–80 (2004)
- Bayarri, M.J., DeGroot, M.H., Kadane, J.B.: What is a likelihood function? In: Gupta, S.S., Berger, J.O. (eds.) *Statistical Decision Theory and Related Topics*, vol. 1, pp. 3–16. Springer, New York (1988)
- Berger, J.O.: *Statistical Decision Theory and Bayesian Analysis*. Springer, New York (1985)

- Berger, J.O.: Are P-values reasonable measures of accuracy? In: Francis, I.S. et al. (eds.) *Pacific Statistical Congress*. Elsevier, Amsterdam (1986)
- Berger, J.O.: Bayes factors. In: Kotz, S., Read, C.B., Banks, D.L. (eds.) *Encyclopedia of Statistical Sciences*. Update, vol. 3, pp. 20–29. Wiley, New York (1999)
- Berger, J.O.: Bayesian analysis: a look at today and thoughts of tomorrow. In: Raftery, A.E., et al. (eds.) *Statistics in the 21th Century*, pp. 275–290. Chapman and Hall/CRC, Boca Raton (2000)
- Berger, J.O.: Could Fisher, Jeffreys and Neyman have agreed on testing? *Stat. Sci.* **18**, 1–32 (2003)
- Berger, J.O.: The case for objective Bayesian analysis. *Bayesian Anal.* **3**, 385–402 (2006)
- Berger, J.O., Bernardo, J.M.: On the development of reference priors. In: Bernardo, J.M., et al. (eds.) *Bayesian Statistics 4*. Clarendon Press, Oxford (1992)
- Berger, J.O., Pericchi, L.R.: The intrinsic Bayes factor for model selection and prediction. *J. Am. Stat. Assoc.* **91**, 109–122 (1996)
- Berger, J.O., Wolpert, R.: *The Likelihood Principle*, 2nd edn. Institute of Mathematical Statistics, Hayward (1988)
- Berger, J.O., Betrò, B., Moreno, E., Pericchi, L.R., Ruggeri, F., Salinetti, G., Wasserman, L. (eds.): *Bayesian Robustness*. Institute of Mathematical Statistics, Hayward (1996)
- Berger, J.O., Bernardo, J.M., Sun, D.: The formal definition of reference priors. *Ann. Stat.* **37**, 905–938 (2009)
- Bernardo, J.M.: Reference posterior distributions for Bayesian inference (with discussion). *J. Roy. Stat. Soc. Ser. B* **41**, 113–147 (1979)
- Bernardo, J.M.: Bruno de Finetti en la Estadística Contemporánea. In: Rios, S. (ed.) *Historia de la Matematica en el siglo XX*, pp. 63–80. Real Academia de Ciencias, Madrid (x 1998)
- Bernardo, J.M.: Reference analysis. In: Dey, D.K., Rao, C.R. (eds.) *Handbook of Statistics*, pp. 17–90. Elsevier, Amsterdam (2005)
- Bernardo, J.M., Smith, A.F.M.: *Bayesian Theory*. Wiley, Chichester (1994)
- Bertrand, J.: *Calcul des Probabilités*. Chelsea, New York (1907)
- Birnbaum, A.: On the foundations of statistical inference. *J. Am. Stat. Assoc.* **57**, 269–306 (1962)
- Box, G.E.P.: Sampling and Bayes' inference in scientific modelling and robustness (with discussion). *J. Roy. Stat. Soc. Ser. A* **143**, 383–430 (1980a)
- Box, G.E.P.: Sampling and Bayes' inference and robustness in the advancement of learning (extended abstract with discussion). In: Bernardo, J.M., et al. (eds.) *Bayesian Statistics*. University Press, Valencia (1980b)
- Box, G.E.P., Tiao, G.C.: *Bayesian inference in statistical analysis*. Addison-Wesley, Reading (1973)
- Carota, C., Parmigiani, G.: On Bayes factors for nonparametric alternatives. In: Bernardo, J.M., et al. (eds.) *Bayesian Statistics 5*, pp. 507–511. Clarendon Press, Oxford (1996)
- Chaloner, K., Verdinelli, I.: Bayesian experimental design: a review. *Stat. Sci.* **10**, 273–304 (1995)
- Chen, M-H, Dey, D.K., Miller, P., Sun, D., Ye, K. (eds.): *Frontiers of Statistical Decision Making and Bayesian Analysis*. In Honor of James O. Berger. Springer, New York (2010)
- Cifarelli, M., Muliere, P.: *Statistica Bayesiana*. G. Iuculano, Pavia (1989)
- Cifarelli, D.M., Regazzini, E.: Some considerations about mathematical statistics teaching methodology suggested by the concept of exchangeability. In: Koch, G., Spizzichino, F., (eds.) *Exchangeability in Probability and Statistics*. North Holland, Amsterdam (1982)
- Cifarelli, D.M., Regazzini, E.: de Finetti's contribution to probability and statistics. *Stat. Sci.* **11**, 253–282 (1996)
- Clyde, M., George, E.J.: Model uncertainty. *Stat. Sci.* **19**, 81–94 (2004)
- Consonni, G., Veronese, P.: Bayes factors for linear models with improper priors. In: Bernardo, J.M., et al. (eds.) *Bayesian Statistics 4*, pp. 587–594. Clarendon Press, Oxford (1991)
- Cox, D.R.: Some problems connected with statistical inference. *Ann. Math. Stat.* **29**, 357–372 (1958)
- Cox, D.R.: *Principles of Statistical Inference*. Cambridge University Press, Cambridge (2006)
- Daboni, L., Wedlin, A.: *Statistica. Un'introduzione alla statistica neo-bayesiana*. UTET, Torino (1982)

- Dale, A.I.: *A History of Inverse Probability. From Thomas Bayes to Karl Pearson*. Springer, New York (1991)
- Dawid, A.P.: Intersubjective statistical models. In: Koch, G., Spizzichino, F. (eds.) *Exchangeability in Probability and Statistics*. North-Holland, Amsterdam (1982)
- de Finetti, B.: Sul significato soggettivo della probabilità. *Fundamenta Mathematicae* **17** (1931). Reprinted in: *Opere Scelte*, vol. I, pp. 191–222. Cremonese, Roma (2006)
- de Finetti, B.: Il problema della perequazione. *Atti della Soc Ital per il Progresso delle Scienze, XXIII riunione* (1935). Reprinted in: *Opere Scelte*, vol. I, pp. 287–288, Cremonese, Roma (2006)
- de Finetti, B.: La prévision: ses lois logiques, ses sources subjectives. *Ann. Inst. H. Poincaré* **7**, 1–68 (1937) [English translation in *Studies in Subjective Probability* (Kyburg, H.E. and Smokler, H.E. eds.), Wiley, New York, 1964]
- de Finetti, B.: L'opera di Abraham Wald e l'assetamento concettuale della statistica matematica moderna. *Statistica* **11**, 185–192 (1951)
- de Finetti, B.: *La probabilità e la statistica nei rapporti con l'induzione secondo i diversi punti di vista*. Cremonese, Roma (1959) (English translation in de Finetti, B.: *Probability, Induction and Statistics*. Wiley, London, 1972)
- de Finetti, B.: *Teoria della probabilità*. Einaudi, Torino (1970) (English translation in *Theory of Probability*, Wiley, New York, 1974)
- de Finetti, B., Savage, L.J.: Sul modo di scegliere le probabilità iniziali. In: *Sui Fondamenti della Statistica*. Biblioteca di Metron, pp. 81–154. Università di Roma, Roma (1962)
- DeGroot, M.H.: *Optimal Statistical Decisions*. McGraw-Hill, New York (1970)
- De Santis, F.: Sample size determination for robust Bayesian analysis. *J. Am. Stat. Assoc.* **101**, 278–291 (2006)
- De Santis, F.: Using historical data for Bayesian sample size determination. *J. Roy. Stat. Soc. Ser. A* **70**, 95–113 (2007)
- De Santis, F., Spezzaferri, F.: Alternative Bayes factors for model selection. *Can. J. Stat.* **25**, 503–515 (1997)
- Edwards, W., Lindman, H., Savage, L.J.: Bayesian statistical inference for psychological research. *Psychol. Rev.* **70**, 193–242 (1963)
- Etzioni, R., Kadane, J.B.: Optimal experimental design for another's analysis. *J. Am. Stat. Assoc.* **88**, 1404–1411 (1993)
- Ferguson, T.: A Bayesian analysis of some nonparametric problems. *Ann. Stat.* **1**, 209–230 (1973)
- Fienberg, S.E.: A brief history of statistics in three and one-half chapters. *Stat. Sci.* **7**, 208–225 (1992)
- Fienberg, S.E.: A “Bayesian classics” reading list. *ISBA Bull.* **12**, 9–14 (2005)
- Fienberg, S.E.: When did Bayesian inference become “Bayesian”? *Bayesian Anal.* **1**, 1–40 (2006)
- Fisher, R.A.: On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Ser. A* 309–368 (1922). Reprinted in: Fisher, R.A.: *Contributions to Mathematical Statistics*. Wiley, New York (1950)
- Geisser, S.: *Predictive Inference: An Introduction*. Chapman and Hall, London (1993)
- Giovagnoli, A., Verdinelli, I.: Bayes D-optimal and E-optimal block designs. *Biometrika* **79**, 695–706 (1983)
- Goldstein, M.: Subjective Bayesian analysis: principles and practice. *Bayesian Anal.* **1**, 403–420 (2006)
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.: Bayesian model averaging: a tutorial (with discussion). *Stat. Sci.* **14**, 382–417 (1999)
- Ibrahim, J.G., Chen, M.H.: Power prior distributions in regression models. *Stat. Sci.* **15**, 46–60 (2000)
- Kadane, J.B., Lazar, N.A.: Methods and criteria for model selection. *J. Am. Stat. Assoc.* **99**, 279–290 (2004)
- Kass, R.E., Raftery, A.E.: Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795 (1995)
- Kass, R.E., Wasserman, L.: The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.* **91**, 1343–1370 (1993)

- Key, J.T., Pericchi, L.R., Smith, A.M.F.: Bayesian model choice: What and why? In: Bernardo, J.M., et al. (eds.) *Bayesian Statistics 6*. Clarendon Press, Oxford (2001)
- Lahiri, P. (ed.): *Model Selection*. Institute of Mathematical Statistics, Beachwood (2001)
- Lavine, M., Schervish, M.J.: Bayes factors: what they are and what they are not. *Am. Statistician* **53**, 119–122 (1999)
- Lehmann, E.L.: *Testing Statistical Hypotheses*, 2nd edn. Wiley, New York (1986)
- Lijoi, A., Prünster, I.: Models beyond the Dirichlet process. In: Hjort, N.L., et al. (eds.) *Bayesian Nonparametrics*, pp. 80–136. Cambridge University Press, Cambridge (2000)
- Lindley, D.V.: *Probability and Statistics*. Cambridge University Press, Cambridge (1965)
- Lindley, D.V.: The estimation of many parameters. In: Godambe, V.P., Sprott, D.A., (eds.) *Foundations of Statistical Inference* pp. 435–447, Holt Rinehart and Winston, Toronto (1971)
- Lindley, D.V.: *Bayesian Statistics, A Review*. SIAM, Philadelphia (1972)
- Lindley, D.V.: *Understanding Uncertainty*. Wiley, Hoboken (2006)
- Lindley, D.V., Smith, A.M.F.: Bayes estimates for the linear model. *J. Roy. Stat. Soc. Ser. B* **34**, 1–18 (1972)
- Liseo, B.: The elimination of nuisance parameters. In: Dey, D., Rao, C.R. (eds.) *Handbook of Statistics*, vol. 25, pp. 193–219. Elsevier, Amsterdam (2006)
- Liseo, B., Petrella, L., Salinetti, G.: Robust Bayesian analysis; an interactive approach. In: Bernardo, J.M., et al. (eds.) *Bayesian Statistics 5*, pp. 661–666. Clarendon Press, Oxford (1996)
- Muliere, P., Petrone, S.: A Bayesian predictive approach to sequential search for an optimal dose, parametric and nonparametric models. *J. Ital. Stat. Soc.* **3**, 349–364 (1993)
- Neyman, J.: “Inductive behavior” as a basic concept of philosophy of science. *Rev. Int. Stat.* **25**, 7–22 (1957)
- O’Hagan, A.: Fractional Bayes factors for model comparisons. *J. Roy. Stat. Soc. Ser. B* **57**, 99–138 (1995)
- O’Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakov, T.: *Uncertain Judgements. Eliciting Experts’ Probabilities*. Wiley, Chichester (2006)
- Pearson, K.: The fundamental problem of practical statistics. *Biometrika* **13**, 1–16 (1920)
- Piccinato, L.: On the orderings of decision functions. *Symposia Mathematica (Istituto Nazionale di Alta Matematica)* **XXV**, 61–70 (1980)
- Piccinato, L.: de Finetti’s logic of uncertainty and its impact on statistical thinking and practice. In: Goel, P.K., Zellner, A. (eds.) *Bayesian Inference and Decision Techniques. Essays in Honor of Bruno de Finetti* pp. 13–30, North Holland, Amsterdam (1986)
- Piccinato, L.: *Metodi per le Decisioni Statistiche*, 2nd edn. Springer-Italia, Milano (2009)
- Racugno, W. (ed.): *Proceedings of the Workshop on Model Selection*. Pitagora, Bologna (1997)
- Raiffa, H., Schlaifer, R.: *Applied Statistical Decision Theory*. MIT Press, Cambridge (1961)
- Ramsey, F.P.: Truth and probability. In: *The Foundations of Mathematics and Other Logical Essays*. Kegan, London (1926). Reprinted in: *Studies in Subjective Probability* (Kyburg, H.E., Smokler, H.E. eds.) Wiley, New York (1964)
- Regazzini, E.: Old and recent results on the relationship between predictive inference and statistical modelling either in nonparametric or parametric form (with discussion). In: Bernardo, J.M., et al. (eds.) *Bayesian Statistics 6*, pp. 571–588. Clarendon Press, Oxford (1999)
- Ríos Insua, D., Ruggeri, F. (eds.): *Robust Bayesian Analysis*. Springer, New York (2000)
- Royall, R.: *Statistical Evidence. A Likelihood Paradigm*. Chapman and Hall, London (1997)
- San Martini A., Spezzaferri, F.: A predictive model selection criterion. *J. Roy. Stat. Soc. Ser. B* **46**, 296–303 (1984)
- Savage, L.J.: *The Foundations of Statistics*. Wiley, New York (1954) (2nd edn., Dover, New York, 1972)
- Savage, L.J.: Subjective probability and statistical practice. In: Barnard, G.A., Cox, D.R. (eds.) *The Foundations of Statistical Inference*. Methuen, London (1962)
- Savage, L.J.: On rereading R.A. Fisher. *Ann. Stat.* **4**, 441–500 (1976)
- Scozzafava, R.: A survey of some common misunderstandings concerning the role and meaning of finitely additive probabilities in statistical inference. *Statistica* **44**, 21–45 (1984)

- Smith, A.M.F., Verdinelli, I.: A note on Bayes designs for inference using a hierarchical linear model. *Biometrika* **7**, 613–619 (1980)
- Spiegelhalter, D.G., Abrams, K.R., Myles, J.P.: Bayesian approaches to clinical trials and health-care evaluations. Wiley, New York (2004)
- Spizzichino, F.: Subjective Probability Models for Lifetimes. Chapman and Hall/CRC, Boca Raton (2001)
- Stigler, S.M.: Thomas Bayes' Bayesian inference. *J. Roy. Stat. Soc. Ser. A* **145**, 250–258 (1982)
- Stigler, S.M.: The History of Statistics: The Measurement of Uncertainty Before 1900. Harvard University Press, Cambridge (1986)
- Tsutakawa, R.K.: Design of Experiment for Bioassay. *J. Am. Stat. Assoc.* **67**, 584–590 (1972)
- Wald, A.: Statistical Decision Functions. Wiley, New York (1950)
- Walker, S.G., Muliere, P.: Beta-Stacy processes and a generalisation of the Pólya-urn scheme. *Ann. Stat.* **25**, 1762–1780 (1997)
- Walker, S.G., Gutierrez-Peña, E., Muliere, P.: A decision theoretic approach to model averaging. *Statistician* **50**, 31–39 (2001)
- Wang, F., Gelfand, A.E.: A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Stat. Sci.* **17**, 193–208 (2002)
- Zabell, S.: R.A.Fisher on the history of inverse probability (with discussion). *Stat. Sci.* **4**, 247–263 (1989)