“Additivity” versus “Maxitivity” at the heart of the paradoxical and efficient nature of Statistics

M. Rémon

Department of Mathematics, Namur University, B-5000 Namur, Belgium

marcel.remon@fundp.ac.be

Summary.

Unlike the Probability Theory based on additivity, Statistical Inference seems to hesitate between ”Additivity” and a so-called ”Maxitivity” approach. After a brief overview of three types of principles for any (parametric) statistical theory and the proof that these principles are mutually exclusive, the paper shows that two kinds of support measures are conceivable, an *additive* one and a *maxitive* one (based on maximization operators). Unfortunately, none of them is able to cope with the ignorance part of the statistical experiment and, in the meantime, with the partial information given through the structure of the data. To conclude, the author promotes the combined use of both approaches, as an efficient middle-of-the-road position for the statistician.

Résumé.

Contrairement à la théorie de probabilité qui est fondée sur l’additivité, l’inférence statistique semble hésiter entre ”l’Additivité” et ce que d’aucun appelle la ”Maxitivité”. Après un bref survol des trois catégories de principes applicables à toute théorie (paramétrique) statistique et la démonstration que ceux-ci sont mutuellement exclusifs, le papier montre que deux types de mesure de support sont envisageables, à savoir une mesure *additive* ou une mesure *maxitive* (basée sur des opérateurs de maximisation). Malheureusement, aucune n’est capable d'appréhender correctement la part d'ignorance contenue dans l'expérience statistique et, dans le même temps, l'information partielle délivrée par la structure des données. En conclusion, l'auteur propose l'utilisation combinée des deux approches comme une position efficace et médiane pour le statisticien.

Keywords: Principles of statistics, Support measures, Maxitivity property, Statistical paradoxes

1. Introduction

In this paper, we restrict ourselves to parametric statistical models. Doing that, we know that we leave aside a large part of statistics. But we think that equivalent reflections can be done for non parametric statistics. Each statistical theory tries to answer the same basic question : *"What can we say about the underlying hypotheses, from the observed data information we get ?"* The different schools of inference have succeeded in giving an answer to the question, as long as one accepts some principles related to these schools. Classical approach is best if one looks for long-run properties. Bayesian inference should be chosen if one has meaningful proper prior information over the parameter space. Structural inference is to be used for transformation models, etc. But there is not always evident *prior* to choose and the Bayesian approach is therefore difficult to apply; or the data come from a unique and non replicable experiment and the Classical approach is no longer appropriate.

It would be naive to believe that a single inference theory could be suitable for all inference problems. In that sense, we totally agree with Kalbfleisch and Sprott (1970) : *"In fact, the main criticism to be directed at the study of statistical inference today is the slavish adherence to rigid dogmas and principles (e.g. Bayes theorem, likelihood principle, admissibility, etc.) which is characteristic of the various schools of inference $\cdots$ To claim that all problems of inference have been, or even can be, solved by one overriding principle seems to us naive."*

This appears to close definitively the search for a general statistical inference school. What we try to do in this paper, is to go deeper in the formal understanding of such a failure. We do that by focusing on the opposition between the way to handle ignorance on one side and structural data information on the other side. For that, it is important to look at the principles underlying the various schools of inference.

2. Three sets of principles : a brief overview of Statistical principles

2.1. Notations

We represent a parametric statistical experiment by means of the following model :

$-\mathcal{M}(X, \Theta, p_\theta(x), \mu(x))$ where X is the sample space,

Θ the parameter space,

$p_\theta(x)$ the density family with respect to $\mu(x)$, and

$\mu(x)$ a σ-finite measure over X

(usually the Lebesgue or the countable measure)

$-$Plus the knowledge of the observed data "$x \in E$".

Many principles will not be mentioned here because they are mere consequences of general principles such as the Likelihood or Invariance ones. We think, for instance, of the Mathematical Equivalence principle (Birnbaum (1964)), which states that our inference should be independent from any one-to-one transformation of the sample space. This principle is a corollary of the Likelihood principle.

2.2. First set : invariance concerning the parameter space

2.2.1. The (Strong) Invariance principle $\mathcal{I}$

$\mathcal{I}$: Let $\mathcal{M}_1(X, \Theta, p_\theta(x), \mu(x))$ and $\mathcal{M}_2(X^{'}, \Theta^{'}, p^{'}_{\theta^{'}}(x^{'}), \mu^{'}(x^{'}))$ be two different models for the same experiment, connected by two functions $f : X \to X^{'}$ and $g : \Theta \to \Theta^{'}$ such that $p_\theta(\{x : f(x) = x^{'}_o\}) = p^{'}_{g(\theta)}(x^{'}_o)$ and $\mu(\{x : f(x) = x^{'}_o\}) = \mu^{'}(x^{'}_o)$ $\forall \theta \in \Theta$ and $\forall x^{'}_o \in X^{'}$. The Invariance principle states that equivalent inference about $g(\theta)$ should be made from the first model given the knowledge "$f(x) = x^{'}_o$" as from the second model given the same observation "$x^{'} = x^{'}_o$".

To better understand the Invariance principle, let us consider $N(\mu, \sigma^2)$, the Normal model. Suppose we only observe the value of the standard deviation $s^2 = s^2_o$. The Invariance principle states that inference about σ given the observed s^2_o should be the same whether one uses the Normal model $N(\mu, \sigma^2)$ or the Chi-square distribution for $\frac{s^2}{\sigma^2}$. The Invariance principle requires that statistical inference does not depend on the choice of parameterization for the model. A consequence of this principle is the invariance of inference under one-to-one transformation of the parameter. This principle is advocated by many authors, see for instance Hartigan (1967). It is at the heart of the paradoxes studied by Dawid et al. (1973) against Bayesian and Structural inference. See also the old Bertrand-Von Mises paradox about the choice of the ratio "wine-water" versus "water-wine" as our parameterization within an uniform model (Von Mises (1939)).

2.3. Second set : invariance concerning the sample space

2.3.1. The Censoring principle $\mathcal{CE}$

$\mathcal{CE}$: For any specified outcome x_o of an experiment $\mathcal{M}(X, \Theta, p_\theta(x), \mu(x))$, our statistical evidence is fully characterized by the function $p_\theta(x_o)$, $\theta \in \Theta$, without further reference to $\mathcal{M}$ or x_o, i.e. all our statistical information is contained in the likelihood function (Birnbaum (1964)).

$\mathcal{CE}$ was first proposed by Pratt (1962) by means of an example : if an accurate voltmeter gave a reading of 87, does it matter, for the interpretation of this reading (assumed error-free), whether the meter's range was bounded by 1,000 or by 100 ?

2.3.2. The Stopping Rule principle $\mathcal{ST}$

$\mathcal{ST}$: The Stopping Rule principle states that the sampling design is irrelevant to statistical inference at the stage of data analysis.

This principle is formally equivalent to the Censoring principle, if one considers the following stopping rule : stop the experiment as soon as “x_o” is observed. $\mathcal{ST}$ can be accepted if one is working with an experiment which will be performed once only. It is certainly not a satisfying principle for long-run sampling experiment, which is the basis of classical inference.

2.3.3. The (Strong) Likelihood principle $\mathcal{L}$

$\mathcal{L}$: Suppose a statistical experiment is characterized by two different models with common parameter space : $\mathcal{M}_1(X, \Theta, p_\theta(x), \mu(x))$ and $\mathcal{M}_2(X^{'}, \Theta, p^{'}_\theta(x^{'}), \mu^{'}(x^{'}))$ such that $p_\theta(x_o) = c \cdot p^{'}_\theta(x^{'}_o)$ for each θ in Θ, for some x_o in X, $x^{'}_o$ in $X^{'}$ and for constant $c \neq 0$. Then $\mathcal{L}$ states that the same inference should be made about θ whatever x_o or $x^{'}_o$ is observed.

The Likelihood principle says that all the relevant information for inference about θ is contained in the sole knowledge of the relative likelihood function. This principle is advocated and criticized by many statisticians. See for instance Fisher (1950), Birnbaum (1962, 1964), Barnard (1967, 1973), Basu (1973), Berger (1985) or Berger and Wolpert (1988). Birnbaum (1964) proved that $\mathcal{L}$ is equivalent to the Sufficiency principle $\mathcal{S}$ (see next section) & $\mathcal{CE}$ or to $\mathcal{S}$ & $\mathcal{ST}$.

2.4. *Third set : Reduction-type principles*

2.4.1. The Reduction principle $\mathcal{R}$

$\mathcal{R}$: In logic, if $A \Rightarrow C$ and $B \Rightarrow C$, then $(A \cup B) \Rightarrow C$. In statistics, one has a similar principle. Let $I(A)$ be the inferential information contained in the observation A [or some statistical inference made from A]. If the data A and B lead to the same inference $I(A) = I(B)$, one should perform equivalent inference from the observation of $A \cup B$: i.e. if $A \Rightarrow I(A)$ and $B \Rightarrow I(B) = I(A)$, then $(A \cup B) \Rightarrow I(A \cup B) = I(A) = I(B)$.

This Reduction principle was proposed by Dawid (1977). It gives a general framework for all the (partial) Sufficiency or Conditionality principles.

2.4.2. The Sufficiency principle $\mathcal{S}$

$\mathcal{S}$: In an experiment $\mathcal{M}(X, \Theta, p_\theta(x), \mu(x))$, we get the same information about θ, if we observe the realization x_o or only its realization through a sufficient statistic $T(x_o) = t_o$.

This principle is certainly the most widely accepted principle in statistics. Together with $\mathcal{CE}$ or $\mathcal{ST}$, it implies that the likelihood function is only relevant for inference up to a proportional constant.

2.4.3. *The Conditionality principle $\mathcal{CO}$*

$\mathcal{CO}$: Suppose we have an experiment $\mathcal{M}(X, \Theta, p_\theta(x), \mu(x))$ and a maximal ancillary statistic $T(x)$. T is ancillary if $p_\theta(T(x))$ is independent of θ. Then our inference about θ should be done through the conditional probability $p_\theta(x|T(x))$.

$\mathcal{CO}$ was studied, among others, by Cox (1958) and Barndorff-Nielsen (1971, 1973). Birnbaum (1962) proved that the Sufficiency principle $\mathcal{S}$ together with the Conditionality principle $\mathcal{CO}$ implies the Likelihood principle $\mathcal{L}$.

2.4.4. *The Partial Nonformation principles $\mathcal{PN}$*

$\mathcal{PS}$ [Partial Sufficiency principles] : Let $T(x)$ be a partial sufficient statistic, in some specified sense, like *B-, S-, M-, K-, I-* or *L*-sufficiency. See Barndorff-Nielsen (1971), Rémon (1984), Cano Sanchez et al. (1989) or Jorgensen (1993) for definitions of partial sufficiency. All the Partial Sufficiency principles state that one gets the same inferential information about some parameter of interest from the knowledge of "x_o" or "$T(x_o)$", and that one has to do inference through the marginal distribution $P_\theta(T(x))$.

Equivalent Partial Conditionality principles $\mathcal{PC}$ require that our inference should be done through the conditional distribution given the observation of some *B-, S-, ...* ancillary statistic. Barndorff-Nielsen (1978) introduced the concept of nonformation which generalizes both notions of partial sufficiency and partial ancillarity, and leads to Partial Nonformation principles $\mathcal{PN}$.

2.5. *Summary*

All these statistical principles can be summarized in three principles :

- The *Invariance principle* $\mathcal{I}$ about the choice of parameterization;

- The *Likelihood principle* $\mathcal{L}$ about the choice for the reference sample space, which is equivalent to the Censoring $\mathcal{CE}$ or Stopping Rule principle $\mathcal{ST}$ together with the Sufficiency principle $\mathcal{S}$;

- The *Reduction principle* $\mathcal{R}$ which generalizes the Partial Nonformation principles $\mathcal{PN}$ about the kind of information one has to consider in the data ("x_o" or "$T(x_o)$" ?).

The next section will discuss the logic of *ignorance* versus *structural information* with respect to the best choice for a support measure over the hypotheses space Θ.

3. Is there a good choice for a support measure with respect to the hypotheses ?

3.1. Introduction

In this paper, we choose the general terms of *support measure* to express the support the observation data give to some unknown hypothesis. When looking for *support measures* in the theories of ignorance or uncertainty, one finds a lot of propositions. Let us just mention here the ancient Laplace's (1812) inverse probability theory (see Dale (1999) or Fienberg (2006)), Dempster-Shafer's belief function (Shafer, 1976), the classical Bayesian *a posteriori* probability, the structural inference (Fraser, 1968), the theory of possibility (Zadeh, 1978; Dubois and Prade, 2007), the plausibility measures (Friedman and Halpern, 1995) or the recent general uncertainty theory (Zadeh, 2005).

3.2. The case of the non informative Bayesian priors

It is well known that additive priors, as proposed by the Bayesian theory, are not suitable for expressing absence of knowledge about hypotheses. See Shafer (1976) : if we have no information about three hypotheses H_1, H_2 and H_3, we cannot say that we have a better knowledge about $H_1 \cup H_2$ with respect to H_3 because we can add these small pieces of (non)information. Another example is the one proposed by Bernardo (1979) : we toss a coin and we wish to do inference about its bias through the parameter of interest $\phi = |\theta - \frac{1}{2}|$, where θ is the probability of observing *"Head"*. We know that the coin is either fair ($H_1 : \theta = \frac{1}{2}$), double-headed ($H_2 : \theta = 1$) or double-tailed ($H_3 : \theta = 0$). We observe $x_0 =$*"Head"* $\cup$ *"Tail"*, i.e. we have no information coming from the data. The likelihood function is then $l(\theta|x_0) = 1 \quad \forall\theta \in \Theta$. If we express our ignorance about θ through an additive uniform measure : $p(H_i) = \frac{1}{3}, i = 1, 2, 3$, we therefore state that the hypothesis $H_1 : \phi{=}0$ is twice less likely than the hypothesis $H_A : \phi{\neq}0$. This contradicts the situation of ignorance.

Dawid et al. (1973) and Stone (1976) have proposed many paradoxes against the additive nature of the Bayesian prior, especially in the context of lack of information. Jeffreys (1939) worked a lot to find non informative Bayesian priors. In his paper about the history of Bayesian Inference, Fienberg (2006) writes that trying to *'derive "objective" priors that expressed ignorance or lack of knowledge'* is like trying *'to grasp the holy grail that had eluded statisticians since the days of Laplace'*. In fact, we can broaden the scope of the incompatibility between ignorance and additivity, to situations where partial information is available, i.e. to any kind of support measure.

3.3. *The incompatibility between "additivity" and the logic of ignorance*

Our knowledge (*a priori* or *a posteriori*) about the "true" unknown hypothesis $\theta_0 \in \Theta$ can be in some way informative. This does not mean that our support measure about this hypothesis behaves like a probability measure. Let us define the *support measure* describing the likelihood the observation E gives to the hypothesis $\theta \in \Theta_1$ by $S[\theta \in \Theta_1|E]$. A support measure, like any plausibility measure (Friedman and Halpern, 1995), has to satisfy three "axioms" :

- $S[\theta \in \Theta_1|E] = 0$ if $E \Rightarrow (\theta \notin \Theta_1)$
- $S[\theta \in \Theta_1|E] = 1$ if $E \Rightarrow (\theta \in \Theta_1)$
- $\Theta_2 \subseteq \Theta_1 \;\Rightarrow\; S[\theta \in \Theta_2|E] \leq S[\theta \in \Theta_1|E]$

  [monotonicity of the support function]

The problem for choosing a support measure on Θ is that this measure should always handle a part of ignorance. Indeed, even when it is an *a posteriori* support measure over Θ, there will be hypotheses θ_i with equivalent support from the observed data (through the likelihood function, for instance), and the support measure will have to manage this ignorance between these θ_i. Once again, like in the Bernardo's coin example, this cannot be done by an additive support measure. Let us prove this incompatibility as a consequence of the Invariance $\mathcal{I}$ and Likelihood $\mathcal{L}$ principles.

Suppose that we express our statistical information about θ by means of an additive posterior support measure $S[\theta|E]$. Because of the Invariance $\mathcal{I}$ and Likelihood $\mathcal{L}$, two θ-values, θ_1 and θ_2, having the same relative likelihood cannot be distinguished. To prove that, one has just to consider the function $g(\theta)$ used in $\mathcal{I}$ as the permutation of θ_1 and θ_2. $\mathcal{I}$ implies that $\mu_1 \equiv S[\theta_1|E] = S[\theta_2|E] \equiv \mu_2$. The logic of ignorance requires equivalent inference for $\theta_1 \cup \theta_2$ as for θ_1. Considering $S[\theta|E]$ as additive, one gets : $S[\theta_1 \cup \theta_2|E] = \mu_1 + \mu_2 = S[\theta_1|E] = \mu_1$. Hence, $\mu_1 = \mu_2 = 0$, which is far from convincing. All this reasoning about the consequences of $\mathcal{I}$ and $\mathcal{L}$ was already mentioned, in similar terms, by Hartigan (1967).

As it is clear that additive support measures are incompatible with $\mathcal{I}$ and $\mathcal{L}$, one can think that non-additive support measures, like the ones proposed in the possibility theory, will be the correct choice. The next section shows that support measures built on maximization (or minimization) are also to be questioned.

3.4. The case of the possibility measure and its "Maxitivity" property

The theory of possibility, as well as the theory of plausibility, proposes measures defined in terms of *maximization* or *minimization*. Dubois and Prade (2007) have introduced the pretty terms of *"Maxitivity"* and *"Maxitive measure"* in reference to the additivity property of probability measures. For instance, the *possibility measure* for the state $A \subseteq S$ is denoted by $\Pi(A)$ and defined by :

$$\begin{aligned} \Pi(A) &= \sup_{s \in A} \pi(s) \\ & \quad \text{where } \ \pi : S \to [0,1] \text{ is a possibility distribution for the states } s \in S \end{aligned}$$

A *necessity measure* can be defined for $A \subseteq S$ by $N(A) = \inf_{s \in A}(1 - \pi(s)) = 1 - \Pi(A)$. One get the following *"maxitivity"* properties :

$$\begin{aligned} \Pi(A \cup B) &= \max(\Pi(A), \Pi(B)) \\ N(A \cap B) &= \min(N(A), N(B)) \end{aligned}$$

See Dubois and Prade (2007) and Sigarretta et al. (2007) for detailed explanations about possibility and plausibility measures. Let us note that $\Pi^*(A) \equiv (\Pi(A) + N(A^c))/2$ is still a possibility measure with the additional properties that $\Pi^*(A) = 0$ is equivalent to the impossibility of A, and $\Pi^*(A) = 1$ to the certainty of A. This can be interesting for comparison with *a posteriori* Bayesian probability, but it will not be used here.

3.5. The impossibility of a "maxitive" support measure satisfying $\mathcal{I}, \mathcal{L}$ and $\mathcal{R}$

Let us define our support measure in the framework of the theory of possibility, but in relation to the relative likelihood function $l(\theta; E)$:

$$\begin{aligned} S[\theta \in \Theta_0 | E] &= \sup_{\theta \in \Theta_0} l(\theta; E) \\ &= \frac{\sup_{\theta \in \Theta_0} p_\theta(E)}{\sup_{\theta \in \Theta} p_\theta(E)} \end{aligned}$$

It is clear that such a possibility measure satisfies the Invariance and Likelihood principles. However, the Reduction principle is not satisfied. Indeed, such a support measure based on the sole likelihood function is incompatible with the Reduction principle, as far as partial sufficiency principle is concerned. Let us consider the Bernardo's (1979) coin example again.

Remember that our parameter of interest is $\phi = |\theta - \frac{1}{2}|$ where θ is the probability of observing *"Head"*, and that the coin is known to be either fair ($H_1 : \theta = \frac{1}{2}$), double-headed ($H_2 : \theta = 1$) or double-tailed ($H_3 : \theta = 0$). This time, we observe $x_0 =$ *"Head"*. The likelihood function is $l(\theta | x_0) = \theta \;\; \forall \theta \in \Theta$. So, by definition of our support measure, one gets :

$$\begin{aligned} S[\theta | \textit{“Head”}] = 1 - S[\theta | \textit{“Tail”}] &= \theta \\ S[H_1 | \textit{“Head”}] = S[H_1 | \textit{“Tail”}] &= \tfrac{1}{2} \qquad \text{[The support measure for fairness]} \\ S[H_2 \cup H_3 | \textit{“Head”}] = S[H_2 \cup H_3 | \textit{“Tail”}] &= 1 \qquad \text{[The support measure for unfairness]} \end{aligned}$$

We see that the likelihood, as well as our *"maxitive"* support measure, puts its highest support towards the unfairness of the coin, whatever the first toss gives as a result. We see also that there is an invariant structure in the model, concerning our parameter of interest. Indeed, the minimal G-sufficient statistic with respect to ϕ, see Barnard (1963), is $T(\textit{“Head”}) = T(\textit{“Tail”}) = T(\textit{“Head” or “Tail”})$. Thus, from $\mathcal{R}$ (or $\mathcal{PS}$) and the fact that the *Marginal* likelihood $l(\theta|\,\text{“}Head\text{” } or \text{ “}Tail\text{”}) = 1 \;\; \forall\theta \in \Theta$, we should have :

$$S[H_1|\,\text{“}Head\text{”}] = S[H_1|\,\text{“}Tail\text{”}] \overset{\mathcal{R}}{=} S[H_1|\,\text{“}Head\text{” } or \text{ “}Tail\text{”}] = 1$$

$$S[H_2 \cup H_3|\,\text{“}Head\text{”}] = S[H_2 \cup H_3|\,\text{“}Tail\text{”}] \overset{\mathcal{R}}{=} S[H_2 \cup H_3|\,\text{“}Head\text{” } or \text{ “}Tail\text{”}] = 1$$

This is clearly a better situation in terms of inferential support, as the first toss of a coin gives no information at all about the fairness or unfairness of a coin.

This example shows the impossibility for a *"maxitive"* support measure to satisfy the Reduction principle, as this last one, through structural invariance and partial sufficiency, introduces *Marginal likelihood function* in the scene of inference. And therefore, an *additive* operation in terms of likelihood. Moreover, as we observed in the coin example, *single* and *marginal* likelihood functions can express totally different support with respect to the hypotheses. Which one should we prefer ? Our choice will, *de facto*, contradict either the Reduction or the Likelihood principle.

3.6. Summary

In this section, we have seen that neither the *additivity* nor the *maxitivity* approach is "the" solution for our support measure. The first one cannot handle properly the ignorance present in any statistical problem, while the second one cannot cope with its structural invariance (for instance, a location-scale structure emerging with the asymptotic normal model when the number of observations increases).

- The *additive* Bayesian posterior approach satisfies the Likelihood $\mathcal{L}$ and Reduction $\mathcal{R}$ principles, but not the Invariance $\mathcal{I}$ principle. $\mathcal{R}$ will be valid under the condition that a reference parameterization is chosen as well as a proper prior distribution over the parameter space Θ. This *extra information* is required if paradoxes are to be avoided

(Dawid et al. (1973); Stone (1976)).

- The *maxitive* Maximized or Profile Likelihood approach (Barndorff-Nielsen and Cox (1994)) satisfies the Invariance $\mathcal{I}$ and Likelihood $\mathcal{L}$ principles, but not the Reduction $\mathcal{R}$ principle. The Generalized Likelihood Ratio tests are based on this type of support measure, as well as the Maximum Likelihood point estimation. The *extra information* needed here to avoid paradoxes is the long-run behavior of the model, as well as its structural invariance (Stein (1956); Barnard (1965); Berger and Wolpert (1988)).

- The *mixed additive-maxitive* Marginal or Conditional Inference approaches have not yet been considered in this paper. The Marginal (or Conditional) Likelihood approach is defined in the same way as the Profile Likelihood approach, but using the marginal [respectively conditional] likelihood function $l(\theta; T(x))$ $[l(\theta; x|T(x))]$ instead of the simple likelihood function $l(\theta; x)$, as we did in Bernardo's example. The Invariance $\mathcal{I}$ and Reduction $\mathcal{R}$ principles will be satisfied here, but not the Likelihood $\mathcal{L}$ principle. The problem here is the definition of what is a partial nonformative statistic $T(x)$ (Barndorff-Nielsen (1978); Rémon (1984); Cano Sanchez et al. (1989); Zhu and Reid (1994); Barndorff-Nielsen and Cox (1994)). The use of a marginal or conditional likelihood function requires *extra information* as the knowledge of the stopping rule because the marginal or conditional density can differ from one stopping rule to another. $\mathcal{L}$ is no longer valid for this type of inference.

Bayesian, Profile Likelihood and Marginal/Conditional Likelihood inferences are three major approaches corresponding to the possible *two-by-two* combinations of our general principles. Other inference methods can be classified in the same way, depending on the list of principles they satisfy. But none will be able to satisfy all these principles, as there is an internal incompatibility between them. This incompatibility can be seen as a dilemma between an *additive* or a *maxitive* approach for dealing with the ignorance and the structural information contained in the data.

4. Conclusions : the dilemma between *“Additivity”* and *“Maxitivity”*

Our point of view is that discussion about Statistical Schools of Inference should not focus so much on the kind of principle one keeps or rejects, or even by-passes thanks to some well chosen *extra information*. Indeed, any inference theory seems to miss some information, as extra information is always needed to avoid paradoxes. Statisticians should be more aware of and worried by the mathematical properties of the support measure they wish to use. Here comes the dilemma between the *“additivity”* and the *“maxitivity”* of our support measure.

One can think that most statisticians will prefer an additive approach, by similarity with the probability theory, but this is not so clear. Indeed, the core of the point estimation is done in a *maxitive* environment. And if they have to compare hypotheses, they will normally use likelihood ratios, which are based on *maxitive* support measures. We think that neither the *maxitive* nor the *additive* approach should be promoted as the sole possible approach.

Our point of view is that statisticians should use both perspectives, in a dialogal process, like in the Marginal/Conditional Likelihood approach. EM algorithm (Dempster et al. (1977)) is also a good example of this combined use of *“maxitive”* and *“additive”* operators. This double nature of the support measure is, for us, the *characteristic of statistical inference*, as statisticians should consider themselves as staying in the middle of the road, trying to reconcile the logic of ignorance (related to *"Maxitivity"*) and the logic of information (linked to *"Additivity"*). This is also the source of the efficiency of many statistical *ad hoc* methods.

References

Barnard, G. A. (1963). Some logical aspects of the fiducial argument. *Statist. Soc. Ser. B 25*, 111–115.

Barnard, G. A. (1965). Misleading likelihood functions. Communication to Prof. D.R. Cox.

Barnard, G. A. (1967). The use of likelihood function in statistical practice. In *Porc. Fifth Berkeley Symp.*, Volume 1, pp. 27–40.

Barnard, G. A. (1973). On likelihood (with discussion). In O. Barndorff-Nielsen and al. (Eds.), *Conference on Foundational Questions in Statistical Inference*, Volume 1 of *Aarhus Memoirs*, pp. 355–366.

Barndorff-Nielsen, O. E. (1973). *Exponential Families and Conditioning.* New York: Wiley.

Barndorff-Nielsen, O. E. (1978). *Information and Exponential Families in Statistical Theory.* Chichester: Wiley.

Barndorff-Nielsen, O. E. (nov. 1971). On conditional statistical inference. Technical report, Aarhus report.

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and Asymptotics.* Monographs on Statistics and Applied Probability. London: Chapman and Hall.

Basu, D. (1973). Statistical information and likelihood. In O. E. Barndorff-Nielsen and al.

(Eds.), *Conference on foundational questions in statistical inference*, Volume 1 of *Aarhus Memoirs*, pp. 139–247.

Berger, J. O. (1985). In defense of the likelihood principle, axiomatics and coherency. In L. D. V. Bernardo J., De Groot M. A. and S. A. F. M. (Eds.), *Bayesian Statistics*, Volume 2, pp. 33–66. New York: North Holland.

Berger, J. O. and R. Wolpert (1988). *The Likelihood Principle* (2nd ed.), Volume 9 of *IMS Monograph*. California: Hayward.

Bernardo, J. (1979). Reference posterior distribution for bayesian inference (with discussion). *J. Roy. Statist. Soc. Ser. B 41*, 113–147.

Birnbaum, A. (1962). On the foundations of statistical inference. *J. Amer. Statist. Ass. 57*, 269–326.

Birnbaum, A. (1964). The anomalous concept of statistical evidence. Technical Report IMM NYU 332, Courant Institute of Math. Sciences, New York University.

Cano Sanchez, J. A., A. Hernández Bastida, and E. Moreno Bas (1989). On l-sufficiency concept of partial sufficiency. *Statistica 49*, 519–528.

Cox, D. R. (1958). Some problems connected with statistical inference. *Ann. Math. Statist. 29*, 357–372.

Dale, A. I. (1999). *A History of Inverse Probability from Thomas Bayes to Karl Pearson.* Springer-Verlag.

Dawid, A. P. (1977). Conformity of inference patterns. In Barra and al. (Eds.), *Recent Developments in Statistics*, pp. 245–256. Amsterdam: North Holland.

Dawid, A. P., M. Stone, and J. Zidek (1973). Marginalization paradoxes in bayesian and structural inference (with discussion). *J. Roy. Statist. Soc. Ser. B 35*, 189–233.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data with the em algorithm (with discussion). *J. Roy. Statist. Soc. Ser. B 39*, 1–38.

Dubois, D. and H. Prade (2007). Possibility theory. *Scholarpedia 2*(10), 2074.

Fienberg, S. E. (2006). When did bayesian inference become "bayesian" ? *Bayesian Analysis 1*(1), 1–40.

Fisher, R. A. (1950). Two new properties of mathematical likelihood (1934). In *Contributions to Mathematical Statistics*. New York: Wiley.

Fraser, D. A. S. (1968). *The Structure of Inference*. New York: Wiley.

Friedman, N. and J. Y. Halpern (1995). Plausibility measures : A user's guide. In *Proc. Eleventh Conf. on Uncertainty in Artificial Intelligence (UAI95)*.

Hartigan, J. (1967). The likelihood and invariance principles. *J. Roy. Statist. Soc. Ser. B 29*, 533–539.

Jeffreys, H. (1939). *Theory of Probability*. Oxford University Press.

Jorgensen, B. (1993). The rules of conditional inference : is there a universal definition of nonformation ? Technical report, The Department of Statistics, The University of British Columbia, Vancouver.

Kalbfleisch, J. D. and D. A. S. Sprott (1970). Application of likelihood methods to models involving large number of parameters (with discussion). *J. Roy. Statist. Soc. Ser. B 32*, 175–209.

Laplace, P. S. (1812). *Théorie Analytique des Probabilités.* Courcier.

Pratt, J. W. (1962). In discussion of birmbaum (1962). *J. Amer. Statist. Ass. 57*, 314–315.

Rémon, M. (1984). On a concept of partial sufficiency : L–sufficiency. *International Statistical Review 52*, 127–135.

Shafer, G. (1976). *A Mathematical Theory of Statistical Evidence.* University Press, Princeton.

Sigarretta, J. M., P. Ruesga, and M. Rodriguez (2007). On mathematical foundations of the plausibility theory. *International Mathematical Forum 2*(27), 1319–1328.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *3rd Berkeley Symposium*, Volume 1, pp. 197–206.

Stone, M. (1976). Strong inconstency from uniform priors (with discussion). *J. Amer. Statist. Ass. 71*, 114–125.

Von Mises, R. (1939). *Probability, Statistics and Truth.* Hodge and Co. Press. Translated by Neyman, Sholl and Rabinowitsch.

Zadeh, L. (1978). Fuzzy sets as the basis for a theory of possibility. *Fuzzy Sets and Systems 1*, 3–28.

Zadeh, L. A. (2005). Towards a generalized theory of uncertainty (gtu) - an outline. *Information Sciences 172*, 1–40.

Zhu, Y. and N. Reid (1994). Information, ancillarity and sufficiency in the presence of nuisance of parameters. *Canadian J. Stat. 22*, 111–123.