Exchangeability and Predictivism
Author(s): Sergio Wechsler
Source: *Erkenntnis (1975-)*, Vol. 38, No. 3 (May, 1993), pp. 343-350
Published by: Springer
Stable URL: http://www.jstor.org/stable/20012482

SERGIO WECHSLER

EXCHANGEABILITY AND PREDICTIVISM

Bruno de Finetti thought of his celebrated Representation Theorem as '... essentially the fruit of a thorough examination of the subject matter, carried out in an unprejudiced manner, with the aim of rooting out nonsense ...' (Preface of his (1974) *Theory of Probability*). This statement could also be understood more generally as de Finetti's description of Bayesian statistical inference and decision-making. However, nothing seems to be more distant from that description than the consideration of statistical models having arbitrary parameters. The arbitrariness lies in the presupposed existence of the parameters, without any invariance judgment – like exchangeability – about the observable data generating them. This attitude, which is shared by some Bayesians as well, is seen by Diaconis (1988) in part as a consequence of the modern availability of large and fast curve-fitting software and hardware. The frequentist concept of probability and its resulting version of statistical inference may be seen as a fundamental cause for arbitrary parametrization: the idea of fixed but unknown probabilities which somehow "exist" leads to assumptions of "existing" parameters. It seems that another important meaning of de Finetti's Representation Theorem is a description of how a subjectivist can include parameters in a model. De Finetti repeatedly pointed out the unreasonableness of considering objective probabilities and "arbitrary" parametric models. The elimination of such arbitrariness – and possibly of parameters – is a basic component of de Finetti's reductionist program. The attitude of restricting the consideration of parameters to invariance judgments is called *predictivism*. The term is used by Piccinato (1986) who – more radically – defines as "completely predictive" the point of view where only strictly observable events are considered relative to inference. Cifarelli and Regazzini (1982) also call it a "predictivistic" or "nonhypothetical" approach – the latter in contrast to the classical point of view where statistical inference is seen as a process of using data in an attempt to specify the "probability mechanism" which generates them. Cifarelli and Regazzini adhere to the completely predictivistic approach to statistical inference: following de Finetti, statistical inference be-

Erkenntnis **38**: 343–350, 1993.

comes essentially the coherent methodology of making *previsions*, conditional on data. In particular, parametric concepts like sufficiency are redefined in terms of (subjective and to a certain extent qualitative) properties of *data*, exclusively. Under the predictivistic point of view, parametric models – once the arbitrariness is removed and is properly interpreted – still remain as auxiliary tools for statistical inference. Representation theorems can make the "extravagant thing" (di Bacco, 1983) of having "an opinion about a probability" a valid *operational* procedure. The choice is clearly a non-mathematical question. A successful example of practical predictivist behavior is given by Dawid (1977) who obtains the Bayesian model of analysis of variance. The parametrization is induced by exchangeability considerations only, but normality of the observations is arbitrary. In a later paper, Dawid (1982) uses "intersubjective" models – unanimous models conditional on a "sufficient" statistic for a set of predictive invariant opinions held by a group of Bayesians – to describe the parametrization *and* functional form of such opinions. The ingenious idea of capturing the "objectivity" of parameters (or of propensities) through a description of unanimity of opinions was somehow presented by de Finetti (1951, 1952) when he suggested that "problems not admitting an initial opinion" are problems where a "coincidence of opinions" exists among all persons dealing with the problem. As in de Finetti (1951), "nonsubjective" problems should be seen, instead, as "multisubjective" problems presenting a coincidence of opinions. He suggested that such an interpretation would resolve the antithesis between Bayesian and frequentist types of problems in statistics! In the light of de Finetti (1937), this was a rather moderate point of view about the notion of probability. De Finetti later made clear his uncompromising approach in several writings, like his (1974) book and, explicitly, in a (1969) paper where he condemned all non-subjective meanings of probability, such as logical, physical or support probability. He acknowledged these as non-objectivistic by calling their use the "middle-of-the-road approach", but nevertheless rejected them – only "truly" subjective probabilities are meaningful. Dawid (1982, 1979) restored to a pure and uncompromising subjective way de Finetti's (1951, 1952) idea of coincidence of opinions as an expression of "objectivity": a group of Bayesians having subjective and possibly distinct opinions about observable quantities agree on a unanimous conditional probability opinion when given the value of a "sufficient" statistic – which Dawid calls an *intersubjective parameter*, while the conditional unanimous model is an *intersubjective model*. The intersubjective parameters are maximal invariants under the group of transformations on the sample space that expresses the invariance characteristics common to all opinions in the group. Exchangeability appears as the particular case of invariance under permutations only. The (individual) predictive opinion is then represented as a mixture (by a "prior") of the intersubjective models ("likelihood"). An example in a low-dimensional case shows how intuitive the representation is: Suppose you (we call "you" a coherent neo-Bayesian person) have an opinion about the random quantities X and Y which is described by a joint probability density on the plane. Your opinion is that X and Y are not only exchangeable, but you also think that the density is invariant under rotations around the origin. It is then clear that your opinion about X and Y is your opinion about their distance from the origin, without any preference for points in a same circle. The intersubjective parameter is a maximal invariant under rotations like the squared distance $X^2 + Y^2$ and the intersubjective model is the uniform density on the appropriate circle. Your opinion is a mixture of conditional uniform densities weighted by a prior density for the squared distance. A parameter is naturally induced exclusively by your original opinion about the observable X and Y. The representation for the one-dimensional marginals is determined and by taking the limit of the general finite representation the mixture of conditionally independent normally distributed quantities with mean zero and common variance is obtained. For example, by letting $\theta = X^2 + Y^2$, the following marginal representation for the density $f(x)$ of X is easily obtained:

$$f(x) = \int_{x^2}^{\infty} \pi^{-1}[\theta - x^2]^{-1/2} p(\theta) d\theta,$$

where $p(\theta)$ is the prior density for the "parameter" $\theta = X^2 + Y^2$ and the other factor of the integrand is the "likelihood" $f(x \mid \theta)$. (Dawid (1982) and Mendel (1989) have the representation above for general higher dimensions). It should be noticed that in the finite sequence case the likelihood is not necessarily a product of conditionally independent quantities.

In the case where an infinite sequence of random quantities is considered, the representation is less immediate. For an exchangeable process, de Finetti's representation form is obtained. If an additional judgment about a 0–1 process states that it behaves as a sequence of indicators in a Pólya urn scheme, the prior necessarily belongs to the beta family. (Actually, there are not many choices: Hill et al. (1987) proved that an exchangeable *urn* process can only be Pólya, Bernoulli or deterministic).

By partitioning the sample space into sets where a "principle of indifference" applies, Mendel (1989) obtains the orbits of the groups of transformations used by Dawid. Mendel points out the possibility of inverting the de Finetti-type mixture in order to obtain the prior as a function of predictive opinions. He sees the likelihood as an "operator" mapping the set of prior probabilities into the set of predictive distributions. The inverse operator is then the posterior distribution.

All this is conceptually very important for predictivists. It means that when the predictive opinion about the observable quantities is sharp enough, it is represented by a mixture of parametric models and both the likelihood and the prior have their functional forms determined. ("You have to think hard only once; parameters and the form of the likelihood and prior will then be ready"). However, in practice we tend to express our predictive opinions in a "hierarchical" way: by starting with an exchangeability judgment, we obtain a representation. Advancing one step further, we obtain the likelihood through stronger invariance judgments. Finally, the opinion about the "parameter" is the prior and completes the representation.

The results of Dawid, Mendel (and of course of de Finetti) are very positive at a philosophical level. They remove the former "metaphysical" character of parameters and in some situations may even determine the coherent functional form of priors and likelihood models. The practical elicitation of invariant predictive opinions may not be, however, easier than elicitation of priors and models. And the prior families obtained by inversion of mixture representations may not be conjugate. Despite these practical difficulties, the arbitrary parametric approach – even when Bayesian – seems to be loosing ground to the predictivist approach. In the former, it is conceptually possible to have a coherent prior and a coherent model for the data, while the combination does not cohere with the predictive opinion. In a sense, predictivism is also more general insofar as some classical models like exponential families are obtained (rather than rejected) as helpful consequences of predictive judgments.

There is another very important contribution from the predictivistic point of view: it removes the asymmetry usually perceived between the "prior function" and the "likelihood" as concepts. For the predictivist, both are equally valued components of a representation that describes his predictive opinion. This weakens the common anti-Bayesian criticism against the "choice of a prior" in two ways: in the first place, by removing the *special* subjectivity a prior function would hold; and also by making clear that the choice of a "model" (performed by non-Bayesians constantly) is an act completely similar, and even inseparable, from the choice of a "prior". As an example, we now obtain the parametric Poisson model from its predictive description.

DEFINITION: A process $\mathbf{X} = (X_1, X_2, \ldots)$ taking values on $\{0, 1, \ldots\}^\infty$ according to a probability measure P is called *Poissonian* when for every $n \geq 1$ and $s \geq 0$,

$$P\left(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \,\middle|\, \sum_{i=1}^{n} X_i = s\right)$$
$$= \binom{s}{x_1 x_2 \ldots x_n} n^{-s} I\left(\sum_{i=1}^{n} x_i = s\right).$$

(Here $I(\cdot)$ denotes the indicator function: $I(\cdot) = 1$ when $\cdot$ is true, $I(\cdot) = 0$ otherwise).

The above characterization is totally predictive: the intersubjective parameter $\Sigma_{i=1}^{n} X_i$ is observable and the intersubjective multinomial model is completely determined. The condition is obviously stronger than exchangeability and describes how a subjectivist understands (or should understand) the Poisson model. The parametric representation of the process $\mathbf{X}$ follows from the

FACT: A process $\mathbf{X}$ is Poissonian if and only if there exists a distribution function F satisfying $F(0-) = 0$ and such that for every $n \geq 1$

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$
$$= [x_1! x_2! \ldots x_n!]^{-1} \int_0^\infty e^{-n\lambda} \lambda^{\Sigma_{i=1}^{n} x_i} \, dF(\lambda)$$

Proof: Suppose that $\mathbf{X}$ is Poissonian. Then for every $n \geq 1$ and every $m > n$,

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

$$= \binom{y}{x_1 x_2 \ldots x_n} n^{-y} \sum_{j=0}^{\infty} P(S_m = j) \binom{j}{y} (n/m)^y (1 - n/m)^{j-y},$$

where $y = \Sigma_{i=1}^{n} x_i$ and $S_m = \Sigma_{i=1}^{m} X_i$. By letting $m^{-1}j = \lambda$, we obtain

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

$$= \binom{y}{x_1 x_2 \ldots x_n} n^{-y} \int_0^{\infty} \binom{m\lambda}{y} (n/m)^y (1 - n/m)^{m\lambda - y} dF_m(\lambda)$$

for the sequence of discrete distribution functions F_m having jumps of value $P(S_m = j)$ at the points $m^{-1}j$.

The sequence F_m admits a subsequence which converges to a limit F, by Helly's theorem; such a limit is a proper distribution function since by de Finetti's (1937) version of the law of large numbers, $m^{-1} S_m$ converges with subjective probability 1. Therefore, by taking the limit as $m \to \infty$ and recalling the (uniform in λ) Poisson approximation to the binomial distribution, we obtain

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

$$= \binom{y}{x_1 x_2 \ldots x_n} n^{-y} [y!]^{-1} \int_0^{\infty} e^{-n\lambda} (n\lambda)^y dF(\lambda).$$

Therefore, every probability measure corresponding to a Poissonian process is seen to be a mixture of sequences of conditionally independent Poisson probabilities. We omit the easy proof of the converse statement.

If the subjectivist thinks a sequence of size N is Poissonian (but not extendible to an infinite Poissonian one), his "parametric" representation, for every $1 \leq n < N$, is given by

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

$$= \binom{y}{x_1 x_2 \ldots x_n} n^{-y} \sum_{j=0}^{\infty} P(S_N = j) \binom{j}{y} (n/N)^y (1 - n/N)^{j-y}.$$

The example shows how a classical statistical model is obtained – rather than imposed – as a consequence of invariance judgments of a predictive opinion. A stronger (than exchangeability) condition of "multinomiality" is needed. This situation is quite general: as de Finetti's representation is essentially an existence result, its use does not provide a functional form for the "likelihood". The well-known Bernoulli likelihood in the {0, 1} situation is an exception caused by the very simple structure of the sample space. The characterization of functional forms for likelihood functions has been established by several authors. Diaconis and Freedman (1981) call it partial exchangeability and obtain predictivistic descriptions in the exponential families. They also explore the connection with sufficiency as also done by Dynkin (1978), Spizzichino (1978) and Dawid (1982) among others. It seems to us that the idea of predictive sufficiency was contained in de Finetti's (1951, 1952) concept of parameters as measures of unanimity – a pure subjectivistic notion of objectivity. The mathematics of such strengthenings of de Finetti's representation has been made extremely general by Dynkin (1978). In a more applied direction, authors like Barlow and Spizzichino (1992) obtain representations of survival functions through Schur-concavity.

This approach to de Finetti's theorem has the philosophical motivation held by subjectivists in the direction of a correct use and understanding of parameters. At the same time, it emphasizes the basic importance of *modelling*, as opposed to arbitrary use of *models*.

References

Barlow, R. E., and F. Spizzichino: 1992, 'Schur-Concave Survival Functions and Survival Analysis', to appear in *Journal of the American Statistical Association*.

Cifarelli, D. M., and E. Regazzini: 1982, 'Some Considerations About Mathematical Statistics Teaching Methodology Suggested by the Concept of Exchangeability', in G. Koch and F. Spizzichino (eds.), *Exchangeability in Probability and Statistics*, North-Holland Publishing Company.

Dawid, A. P.: 1977, 'Invariant Distributions and Analysis of Variance Models', *Biometrika* **64**, 291–297.

Dawid, A. P.: 1979, 'Conditional Independence in Statistical Theory', *Journal of the Royal Statistical Society B* **41**, 1–31.

Dawid, A. P.: 1982, 'Intersubjective Statistical Models', in G. Koch and F. Spizzichino (eds.), *Exchangeability in Probability and Statistics*, North-Holland, Amsterdam.

de Finetti, B.: 1937, 'La Prévision: ses Lois Logiques, ses Sources Subjectives', *Annales de l'Institut Henri Poincaré*, **7**, 1–68, in H. E. Kyburg Jr. and H. E. Smokler (trans.), *Studies in Subjective Probability*. John Wiley, New York.

de Finetti, B.: 1951, 'Recent Suggestions for the Reconciliation of Theories of Probability', in J. Neyman (ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, California.

de Finetti, B.: 1952, 'La Notion de "distribution d'opinions" comme base d'un essai d'interpretation de la Statistique', *Publ. Institut de Statistique Université de Paris* **I**, 2.

de Finetti, B.: 1969, 'Initial Probabilities: A Prerequisite For Any Valid Induction, *Synthese* **20**, 2–16.

de Finetti, B.: 1974, *Theory of Probability*, Vol. 1, John Wiley, New York.

di Bacco, M.: 1983, 'Commento all'articolo "Inferenza Predittiva e 'Fatti' " di D. Costantini e P. Monari', *Statistica* **XLIII**, 335–340.

Diaconis, P.: 1988, 'Recent Progress on de Finetti's Notion of Exchangeability', Technical Report 297, Department of Statistics, Stanford University, California.

Diaconis, P. and D. Freedman,: 1981, 'Partial Exchangeability and Sufficiency', in *Proceedings of the Indian Statistical Institute Golden Jubilee International Conference on Statistics: Applications and New Directions*, pp. 205–236.

Dynkin, E. B.: 1978, 'Sufficient Statistics and Extreme Points', *The Annals of Probability* **6**, 705–730.

Hill, B. M., D. Lane, and W. Sudderth: 1987, 'Exchangeable Urn Processes', *The Annals of Probability* **15**, 1586–1592.

Mendel, M. B.: 1989, 'Development of Bayesian Parametric Theory', Doctoral Dissertation, MIT, Cambridge, Massachusetts.

Piccinato, L.: 1986, 'De Finetti's Logic of Uncertainty and its Impact on Statistical Thinking and Practice', in P. Goel and A. Zellner (eds.), *Bayesian Inference and Decision Techniques*, Elsevier, Amsterdam.

Spizzichino, F.: 1978, 'Statistiche sufficienti per famiglie di numeri aleatori scambiabili', in *Atti del Convegno su Induzione, Probabilità e Statistica*, U. degli Studi di Venezia.

ACKNOWLEDGEMENTS

The author wishes to thank Professor Richard Barlow for many stimulating and influential conversations about de Finetti's ideas. This research was partially supported by CNPQ grants 304692/89-1 and 352925/92-2 (501864/91-1).

Manuscript submitted June 1, 1991
Final version received September 11, 1992

Universidade de São Paulo
IME-USP
Caixa Postal 20570
(Ag. Iguatemi)
São Paulo
01498 Brazil