RT-MAE-8915

EXCHANGEABILITY AND PREDICTIVISM

by

Sergio Wechsler

Palavras Chaves: Exchangeability; intersubjective model;
(Key words) Poissonian measure; predictivism

Classificação AMS: 62A15
(AMS Classification)

- Junho de 1989 -

Exchangeability and Predictivism

Sergio Wechsler

Universidade de São Paulo

ACKNOWLEDGEMENT : We acknowledge the support of CAPES, Brasilia, Brasil.

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 apparently obvious statement can also be understood as de Finetti's description of Bayesian statistical inference and decision-making in general terms. However, nothing seems to be more distant from that description than the consideration of statistical models having arbitrary parameters existing in it. 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 statistics may be seen as a fundamental cause for arbitrary parametrization : the idea of fixed and unknown probabilities which somehow "exist" develops itself in assumptions of so "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 conditioning 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 for inference. Cifarelli and Regazzini (1982) also call it a "predictivistic" or "non-hypothetical" 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 becomes essentially the coherent methodology of making *previsions*, conditional on data. In particular, sufficiency is redefined in terms of (subjective and to a certain extent qualitative) properties of data, exclusively. The earliest definitions of such "predictive sufficiency" seem to go back to Spizzichino (1978). Under the predictivistic point of view , parametric models - once having arbitrariness removed and being properly interpreted - are still 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 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 (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. He suggested that such an interpretation would resolve the antithesis between "the two 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 and 1979) restored to a pure and uncompromising subjective way de Finetti's (1952) idea of coincidence of opinions as an expression of "objectivity": a group of Bayesians having subjective and possibly distinct opinions about observable quantities agrees on a unanimous conditional probability opinion when given the value of a "sufficient" statistic - by now called the *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 neobayesian 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. 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. This is 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 "). In general, exchangeability alone is not enough for a complete representation of the mixture, but additional invariance judgments may determine it.

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 invariance predictive opinions may not be, however, easier than elicitation of priors and models. The prior families obtained by inversion of mixture representations are conjugate neither. 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 of them 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" .

In Section 3.3 the predictivist approach will be illustrated in the solution for an ESS (Environmental Stress Screening) problem. The predictive representation of the Poisson process will be needed. We now obtain the parametric Poisson model from its predictive description.

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

$$P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n \mid \sum_{i=1}^{n} X_i = s) = \binom{s}{x_1\, x_2 \cdots x_n} n^{-s}\, I(\sum_{i=1}^{n} x_i = s) .$$

(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 $\sum_{i=1}^{n} X_i$ 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 X follows from the

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

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

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

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

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

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

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

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

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 theorem 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

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

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

Therefore every Poissonian process is seen to be a mixture of sequences of conditionally independent Poisson quantities. 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 $1 \le n < N$, is given by

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

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

(*) All probabilities are subjective. In particular those denoted by P throughout.

Bibliography

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

Dawid A. P.(1977). Invariant distributions and analysis of variance models. Biometrika, Volume 64, Number 2, pages 291-297.

Dawid A.P.(1979). Conditional Independence in Statistical Theory. Journal of the Royal Statistical Society B, Volume 41, pages 1-31.

Dawid A. P.(1982). Intersubjective Statistical Models. in Exchangeability in Probability and Statistics. Koch G., Spizzichino F. (editors). North-Holland Publishing Company.

de Finetti B.(1937). La Prévision: ses lois logiques, ses sources subjectives. Annales de l'Institut Henri Poincaré, Volume 7, pages 1-68. Translated in Kyburg Jr., H. E.; Smokler H. E.(1964). Studies in Subjective Probability. John Wiley & Sons.

de Finetti B.(1952). La Notion de "distribution d'opinions" comme base d'un essai d'interprétation 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, Volume 20, pages 2-16.

de Finetti B.(1974). Theory of Probability, Volume 1. John Wiley & Sons.

di Bacco M.(1983). Commento all'articolo "Inferenza Predittiva e 'Fatti' " di D.Costantini e P.Monari. Statistica, XLIII, pages 335-340.

Diaconis P.(1988). Recent Progress on de Finetti's Notions of Exchangeability. Technical Report 297, Department of Statistics, Stanford University.

Hill B.M., Lane D., Sudderth W.(1987). Exchangeable Urn Processes. The Annals of Probability, Volume 15, Number 4, pages 1586-1592.

Mendel M.(1989). Development of Bayesian Parametric Theory. Unpublished Draft. Massachusetts Institute of Technology.

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

Spizzichino F.(1978). Statistiche sufficienti per famiglie di numeri aleatori scambiabili. in Atti del Convegno su Induzione, Probabilità e Statistica. Università degli Studi di Venezia.

RELATÓRIO TÉCNICO
DO
DEPARTAMENTO DE ESTATÍSTICA

TÍTULOS PUBLICADOS EM 1987

8701 - ACHCAR, J.A. & BOLFARINE, H.; Constant Hazard Against a Change-Point Alternative: A Bayesian Approach with Censored Data, São Paulo, IME-USP, 1987, 20p.

8702 - RODRIGUES, J.; Some Results on Restricted Bayes Least Squares Predictors for Finite Populations, São Paulo, IME-USP, 1987, 16p.

8703 - LEITE, J.G., BOLFARINE, H. & RODRIGUES, J.; Exact Expression for the Posterior Mode of a Finite Population Size: Capture-Recapture Sequential Sampling, São Paulo, IME-USP, 1987, 14p.

8704 - RODRIGUES, J., BOLFARINE, H. & LEITE, J.G.; A Bayesian Analysis in Closed Animal Populations from Capture Recapture Experiments with Trap Response, São Paulo, IME-USP, 1987, 21p.

8705 - PAULINO, C.D.M.; Analysis of Categorical Data with Full and Partial Classification: A Survey of the Conditional Maximum Likelihood and Weighted Least Squares Approaches, São Paulo, IME-USP, 1987, 52p.

8706 - CORDEIRO, G.M. & BOLFARINE, H.; Prediction in a Finite Population under a Generalized Linear Model, São Paulo, IME-USP, 1987, 21p.

8707 - RODRIGUES, J. & BOLFARINE, H.; Nonlinear Bayesian Least-Squares Theory and the Inverse Linear Regression, São Paulo, IME-USP, 1987, 15p.

8708 - RODRIGUES, J. & BOLFARINE, H.; A Note on Bayesian Least-Squares Estimators of Time-Varying Regression Coefficients, São Paulo, IME-USP, 1987, 11p.

8709 - ACHCAR, J.A., BOLFARINE, H. & RODRIGUES, J.; Inverse Gaussian Distribution: A Bayesian Approach, São Paulo, IME-USP, 1987, 20p.

8710 - CORDEIRO, G.M. & PAULA, G.A.; Improved Likelihood Ratio Statistics for Exponential Family Nonlinear Models, São Paulo, IME-USP, 1987, 26p.

8711 - SINGER, J.M.; PERES, C.A.& HARLE, C.E.; On the Hardy -Weinberg Equilibrium in Generalized ABO Systems, São Paulo, IME-USP, 1987, 16p.

8712 - BOLFARINE, H. & RODRIGUES, J.; A Review and Some Extensions on Distribution Free Bayesian Approaches for Estimation and Prediction, São Paulo, IME-USP, 1987, 19p.

8713 - RODRIGUES, J.; BOLFARINE, H. & LEITE, J.G.; A Simple Nonparametric Bayes Solution to the Estimation of the Size of a Closed Animal Population, São Paulo, IME-USP, 1987, 11p.

8714 - BUENO, V.C.; Generalizing Importance of Components for Multistate Monotone Systems, São Paulo, IME-USP, 1987, 12p.

8801 - PEREIRA, C.A.B. & WECHSLER, S.; On the Concept of P-value, São Paulo, IME-USP, 1988, 22p.

8802 - ZACKS, S., PEREIRA, C.A.B. & LEITE, J.G.; Bayes Sequential Estimation of the Size of a Finite Population, São Paulo,IME-USP, 1988, 23p.

8803 - BOLFARINE, H.; Finite Population Prediction Under Dynamic Generalized Linear Models, São Paulo, IME-USP, 1988, 21p.

8804 - BOLFARINE, H.; Minimax Prediction in Finite Populations, São Paulo, IME-USP, 1988, 18p.

8805 - SINGER, J.M. & ANDRADE, D.F.; On the Choice of Appropriate Error Terms for Testing the General Linear Hypothesis in Profile Analysis, São Paulo, IME-USP, 1988, 23p.

8806 - DACHS, J.N.W. & PAULA, G.A.; Testing for Ordered Rate Rations in Follow-up Studies with Incidence Density Data, São Paulo, IME-USP, 1988, 18p.

8807 - CORDEIRO, G.M. & PAULA, G.A.; Estimation, Significance Tests and Diagnostic Methods for the Non-Exponential Family Nonlinear Models, São Paulo, IME-USP, 1988, 29p.

8808 - RODRIGUES, J. & ELIAN, S.N.; The Coordinate - Free Estimation in Finite Population Sampling, São Paulo, IME-USP, 1988, 5p.

8809 - BUENO, V.C. & CUADRADO, R.Z.B.; On the Importance of Components for Continuous Structures, São Paulo, IME-USP, 1988, 14p.

8810 - ACHCAR, J.A., BOLFARINE, H & PERICCHI, L.R.; Some Applications of Bayesian Methods in Analysis of Life Data, São Paulo, IME-USP, 1988, 30p.

8811 - RODRIGUES, J.; A Bayesian Analysis of Capture-Recapture Experiments for a Closed Animal Population, São Paulo, IME-USP, 1988, 10p.

8812 - FERRARI, P.A.; Ergodicity for Spin Systems, São Paulo, IME-USP, 1988, 25p.

8813 - FERRARI, P.A. & MAURO, E.S.R.; A Method to Combine Pseudo-Random Number Generators Using Xor, São Paulo, IME-USP, 1988, 10p.

8814 - BOLFARINE, H. & RODRIGUES, J.; Finite Population Prediction Under a Linear Functional Superpopulation Model a Bayesian Perspective, São Paulo, IME-USP, 1988, 22p.

8815 - RODRIGUES, J. & BOLFARINE, H.; A Note on Asymptotically Unbiased Designs in Survey Sampling, São Paulo, IME-USP, 1988, 6p.

8816 - BUENO, V.C.; Bounds for the a Availabilities in a Fixed Time Interval for Continuous Structures Functions, São Paulo, IME-USP, 1988, 22 p.

8817 - TOLOI, C.M.C. & MORETTIN, P.A.; Spectral Estimation for Time Series with Amplitude Modulated Observations: A Review, São Paulo, IME-USP, 1988, 16p.

8818 - CHAYES, J.T.; CHAYES, L,; GRIMMETT, G,R,; KESTEN, H. & SCHONMANN, R.H,; The Correlation Length for the High Density Phase of Bernoulli Percolation, São Paulo, IME-USP, 1988, 46p.

8819 - DURRETT, R,; SCHONMANN, R.H. & TANAKA, N.I.: The Contact Process on a Finite Set, III: The Critical Case, São Paulo, IME-USP, 1988, 31p.

8820 - DURRET, R.; SCHONMANN, R.H. & TANAKA, N.I.; Correlation Lengths for Oriented Percolation, São Paulo, IME-USP, 1988, 18p.

8821 - BRICMONT, J.; KESTEN, H.; LEBOWITZ, J.L. & SCHONMANN, R.H.; A Note on the Ising Model in High Dimensions, São Paulo, IME-USP, 1988, 21p.

8822 - KESTEN, H. & SCHONMANN, R.H.; Behavior in Large Dimensions of the Potts and Heisenberg Models, São Paulo, IME-USP, 1988, 61p.

8823 - DURRET, R. & TANAKA, N.I.; Scaling Inequalities for Oriented Percolation, São Paulo, IME-USP, 1988, 21p.

8901 - RODRIGUES, J.; Asymptotically Design - Unbiased Predictors to Two-Stage Sampling, São Paulo, IME-USP, 1989, 9p.

8902 - TOLOI, C.M.C. & MORETTIN, P.A.; Spectral Analysis for Amplitude Modulated Time Series, São Paulo, IME-USP, 1989, 24p.

8903 - PAULA, G.A.; Influence Measures for Generalized Linear Models with Restrictions in Parameters, São Paulo, IME-USP, 1989, 18p.

8904 - MARTIN, M.C. & BUSSAB, W.O.; An Investigation of the Properties of Raking Ratio Estimators for Cell Frequencies with Simple Random Sampling, São Paulo, IME-USP, 1989, 11p.

8905 - WECHSLER, S.; Yet Another Refutation of Allais' Paradox, São Paulo, IME-USP, 1989, 6p.

8906 - BARLOW, R.E. & PEREIRA, C.A.B.; Conditional Independence and Probabilistic Influence Diagrams, São Paulo, IME-USP, 1989, 21p.

8907 - BARLOW, R.E.; PEREIRA, C.A.B. & WECHSLER, S.; The Bayesian Approach to Ess, São Paulo, IME-USP, 1989, 20p.

8908 - PEREIRA, C.A.B. & BARLOW, R.E.; Medical Diagnosis Using Influence Diagrams, São Paulo, IME-USP, 1989, 13p.

8909 - BOLFARINE, H.; A Note on Finite Population Prediction Under Asymmetric Loss Functions, São Paulo, IME-USP, 1989, 8p.

8910 - BOLFARINE, H.; Bayesian Modelling in Finite Populations, São Paulo, IME-USP, 1989, 8p.

8911 - NEVES, M.M.C & MORETTIN, P.A.; A Generalized Cochrane-Orcutt-Type Estimator for Time Series Regression Models, São Paulo, IME-USP, 1989, 29p.

8912 - MORETTIN, P.A., TOLOI, C.M.C., GAIT, N. & MESQUITA, A.R.; Analysis of the Relationships Between Some Natural Phenomena: Atmospheric Precipitation, Mean Sea Level and Sunspots, São Paulo, IME-USP, 1989, 35p.

8913 - BOLFARINE, H.; Population Variance Prediction Under Normal Dynamic Superpopulation Models, São Paulo, IME-USP, 1989, 7p.

8914 - BOLFARINE, H.; Maximum Likelihood Prediction in Two Stage Sampling, São Paulo, IME-USP, 1989, 4p.