Scientifica Acta **2**, No. 1, 56 – 76 (2008)

The centenary of Bruno de Finetti (1906–1985)

Eugenio Regazzini and **Federico Bassetti**

Università degli Studi di Pavia, Dipartimento di Matematica, via Ferrata 1, 27100 Pavia, Italy
eugenio.regazzini@unipv.it
federico.bassetti@unipv.it

We gather here two distinct, but complementary, contributions to the understanding of some aspects of de Finetti's work in probability and mathematical statistics. The former emphasizes the relevance of a few unknown papers concerning: Bayesian statistics, empirical approximation of a probability distribution, geometry of correlation. The latter provides a detailed description of the first complete de Finetti's contribution to sequences of exchangeable events, and stresses a few issues susceptible of further new developments.

TABLE OF CONTENTS

1 Introduction

Bruno de Finetti was one of the most influential probabilists of the last century. However, this great scientist achieved fame also for his crucial contributions to other important fields of applied mathematics, to say nothing of the philosophic debate at large. In 2006, on the occasion of de Finetti's centenary, many significant cultural initiatives have been taken in order both to enliven the recollection of his work and to analyse its influence on more recent developments in the theory of probability, in statistics and in the other fields (like mathematical finance) that his interests have covered.

The first author of the present essay had the chance to participate in some of the above-mentioned initiatives as invited speaker. He is now delighted by the invitation of *Scientifica Acta* to publish slightly modified versions of two of his contributions to de Finetti's centenary: the former is drawn from an address given during the XLIII Meeting of the *Società Italiana di Statistica (SIS)* held in Torino. The latter, written by the present authors, is contained in the proceedings of the *Bruno de Finetti Centenary Conference* sponsored by *La Sapienza* and the *Accademia Nazionale dei Lincei*, held in Rome in November 2006. The address to the SIS meeting touched on a few unsung aspects of the scientific work of de Finetti in probability and statistics. The joint paper with Bassetti, which, on the contrary, focuses on one of the most renowned subjects of de Finetti's scientific work, i.e. exchangeable random sequences, provides a detailed critical description of the first complete work de Finetti wrote on exchangeable events, and emphasizes its up-to-dateness with respect to fundamental problems such as that of extending the exchangeability property. During the *La Sapienza/Lincei* conference, there was the presentation of two volumes of de Finetti's selected works (de Finetti, B. (2006) *Opere scelte*, Cremonese, Roma) edited by the *Unione Matematica Italiana (UMI)*.

We conclude these introductory remarks with an essential biography drawn from *de Finetti's contribution to probability and statistics* by D. M. Cifarelli and E. Regazzini in vol. 11 of *Statistical Science* (1996).

De Finetti was born of Italian parents on 13^{th} June 1906 in Innsbruck (Austria). At 17 he enrolled at the *Polytechnic of Milan* with a view to obtaining a degree in engineering. During his third year of study (in Italy engineering is a five-year course), a Faculty of Mathematics was opened in Milan. Following his more theoretical bent, he shifted subject. In 1927 he obtained a degree in applied mathematics. He discussed his dissertation on affine geometry with Giulio Vivanti, a mathematician known for some noteworthy contributions to complex analysis. Soon afterwards, de Finetti accepted a position in Rome, at the *Istituto Centrale di Statistica*, presided over, at that time, by an outstanding Italian statistician: Corrado Gini. De Finetti worked there until 1931. In those years, he laid the foundations for his principal contributions to probability theory and statistics: subjective approach to probability; definition and analysis of sequences of exchangeable events; definition and analysis of processes with stationary independent increments and infinitely decomposable laws; theory of mean values. In 1930 de Finetti qualified in a competition as a university lecturer of Mathematical Analysis.

In 1931, he moved to Trieste, where he accepted a position as actuary at the *Assicurazioni Generali*. During his stay in Trieste, de Finetti developed the research he had started in Rome, but he also obtained significant results in actuarial and financial mathematics as well as in economics. In addition, he was active in the mechanization of some actuarial services. It was probably this operational background which later enabled him to understand the revolutionary impact of the computer on scientific calculus. As a matter of fact, he was one of the first mathematicians in Italy able to solve problems of numerical analysis by means of computer. In spite of his impending actuarial activity, in Trieste de Finetti also started his teaching career (mathematical analysis, actuarial and financial mathematics, probability), with a few years' stint at the renowned *University of Padua*. However, it was only in 1947 that he obtained his chair as full Professor of Financial Mathematics in Trieste. He had actually won it in a nationwide competition back in 1939, but at that time he was unmarried and, by a law in force in those days in Italy, bachelors were not permitted to hold any position in the public service.

In 1954 de Finetti moved to the Faculty of Economics at the University of Rome *La Sapienza*. In 1961, he changed to the Faculty of Sciences in Rome, where he was Professor of Theory of Probability until 1976.

He died in Rome on 20^{th} July 1985.

2 On Some of de Finetti's Unknown Contributions to Statistics

This section links a few remarks on the power of exchangeability in detecting inconsistencies in the conventional Bayesian paradigm with the analysis of a couple of papers belonging to the least known part of de Finetti's impressive scientific corpus. The former shows de Finetti's decisive contribution to the so-called fundamental theorem of mathematical statistics so far known as *Glivenko-Cantelli theorem*. The latter contains a concise treatment of the elementary problem of linear correlation, all the more informative, though as it has a wealth of original cues conducive to further research.

From the very title of this section it is clear that we do not intend it to be a mere review of Bruno de Finetti's life and work. Neither does it claim to be exhaustive in any respect. Somewhat pointedly, we meant it to draw attention to some of his specific contributions to Statistics that might have passed unnoticed although they proved to be excitingly seminal. We hope this might be a proper way both to (implicitly) disclose some hidden aspects of his character, and to acknowledge priority of his work with respect to some key probabilistic and statistical topics.

2.1 Exchangeability and Bayesian reasoning

In 1973, de Finetti was invited to give a lecture at the 39th Session (Vienna) of the *International Statistical Institute*. Under the title "Bayesianism: its unifying role for both the foundations and the applications of statistics", he dealt with a number of issues: statistical inference, decision theory, their frequency and non-frequency interpretations, inductive reasoning and inductive behaviour. Part of his lecture focused on the problem of how to get rid of the framework of "hypotheses" about parameters. In point of fact, along with statistical schemes (or models) where parameter θ is a factual, although unknown or hidden, quantity, there coexists a statistical modeling in which

> θ is a merely fictitious, or mythical, pseudoentity, allowing to perform some reasoning "as if" it existed, but leading to absurdities would it be thought as really existing.
>
> Some old-time demographers, in the age when probabilities where ordinarily conceived only with reference to urn-schemes, used to explain the mortality table with the image of the Parcae drawing each year a ball for each of us, to decide about life or death according to its color – white or black –, and using an urn where the fraction of black balls was increasing with the age.
>
> But to imagine an urn with unknown but constant composition explaining at any drawing from the Pólya urn its outcome as resulting from the "hidden urn" would be even more difficult: much more artificial and preposterous than the plainly mythological picture of the Parcae. The "hidden urn" should in effect have, as its unknown but predetermined composition, the one which corresponds to the limit to which the composition of the Pólya urn should approach through endless addition of new balls to the few put into it till now. (And it is almost sure that not even the Vestals would assure the continuation of such experiment for eternity, what would imply, incidentally, to get sometime more balls than atoms in the world; and, on the other side, there is no reason to expect such limit to exist, since stochastic, even if strong, convergence does not guarantee any conclusion on this point.).
>
> In such a situation, it is obviously only the predictive aspect (concerning the future outcomes – not the parameter!) that matters.

De Finetti pointed out that getting rid of the framework of "hypotheses" is actually the aim as explicitly expressed in [15], by Roberts, who advocated that attention should be focussed on the predictive distribution concerning directly the quantities of interest rather than on the ones concerned with the parameters,which are nothing but auxiliary ingredients. He then mentioned that

> a particular case (but a particularly simple and important one) of Robert's distinction is the one where (as in the example just discussed) we have – in the terminology of Objectivists – independent events E_i (or random quantities X_i) with constant but unknown probability θ (or probability distribution, say F_θ). ...Under such assumption the events are *exchangeable*...
>
> But also the converse holds...That is what some Colleague...called the de Finetti representation theorem. My aim was precisely the same as Roberts', if only restricted to the case of the most usual and important example of inductive reasoning (and, then, of inductive behaviour). The aim was to present induction as a very natural way of reasoning on probabilities of observable facts, avoiding metaphysical pseudoentities and obscurities.

These fragments from the ISI lecture do confirm in the most perfect possible way the thought expressed by their author more than forty years beforehand, and they give a clear-cut picture of the specificity of de Finetti's position in the Bayesian circle.

In fact, the problem he intended to deal with towards the end of the 1920th was, in a sense, of a more general extent, i.e.: "Can the subjective theory of probability provide a conceptually rigorous justification of probability assessments based on observed frequencies of analogous events?". A positive answer would imply that the subjective view, far from being incompatible with empirical approaches to probability, does comprise them as processes to evaluate probabilities when particular circumstances occur. In a complete probabilistic setting, these circumstances must be expressed through probabilistic models which translate the empirical idea of "analogous" events (or more general random elements) into a proper mathematical language. As a matter of fact, the idea of "repeatable events" which recurs both in the so-called frequentistic definitions of probability and in the explanation of objective approaches to statistical methods, boils down to the assumption that sequences of trials are carried out under analogous conditions. In a talk (see [6]) given at the *International Mathematical Congress*, held in Bologna in 1928, de Finetti had introduced *exchangeability* as a probabilistic characterization of a *random phenomenon*, that is, a phenomenon which can be repeatedly observed under analogous environment conditions. He argued that a correct probabilistic translation of such an empirical circumstance leads us to think of the probability of k successes and $(n-k)$ failures in n trials as invariant with respect to which successes and failures alternate, whatever k may be (and, in the case of an infinite sequence of trials, whatever n may be). In other words, events $E_1, E_2, \dots, E_N$ are said to be *exchangeable* if their law is the same as the law of $E_{\pi(1)}, \dots, E_{\pi(N)}$ for any permutation π of $(1, \dots, N)$; analogously, the elements of a sequence of events are called exchangeable when the previous condition obtains for $(E_i)_{1 \le i \le n}$ and for every n. For any sequence of events the problem whether the frequency of success can be used to estimate the probability of success has a precise formulation: For every pair of positive integers n, k, define the conditional probability $P(E_{n+k} \mid E_1, \dots, E_n)$ and the frequency of success

$$f_n(E_1, \dots, E_n) := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\mathbb{E}_k}{}^{1};$$

then, (a) check the deviation $|P(E_{n+k} \mid E_1, \dots, E_n) - f_n(E_1, \dots, E_n)|$, and (in the case of sequences) (b) verify whether it converges (stochastically) to 0 as the number n of trials goes to infinity.

One of the main achievements of de Finetti's research is the following theorem presented in 1928 and proved in [5]:

If $(E_n)_{n \neq 1}$ is a sequence of exchangeable events, then

$$\inf_{k,p \in \mathbb{N}} P\Big\{ \max_{n \le m \le n+p} |P(E_{m+k} \mid f_m(E_1, \dots, E_m)) - f_m(E_1, \dots, E_m)| \le \epsilon \Big\} \ge 1 - \eta \tag{1}$$

holds true for every pair of positive numbers ϵ, η, provided that n is larger than a suitable integer depending only on (ϵ, η).

Roughly speaking, one can say that in the case of sequences of exchangeable events, that is, in the case of analogous trials, one can legitimately use frequency as an estimator of the conditional probability of any future event E_{n+k}, given any outcome of $(E_1, \dots, E_n)$ yielding that particular frequency, provided that n is sufficiently large.

In order to avoid unessential problems of a purely technical nature, we intend to stick to the field of events and thus to try to explain de Finetti remarks already hinted at above about "hypotheses". As a matter of fact, it is easy to devise statistical problems in which attention must be focused on prevision of frequency of N events, under the hypothesis one can get information frequency of n events with $n < N$. To this end, one can note that, for any $(x_1, \dots, x_n)$ in $\{0,1\}^n$ with $s_n = x_1 + \ldots + x_n$, in the case of exchangeable events,

$$\begin{aligned} & P\left\{ \mathbb{1}_{E_1} = x_1, \dots, \mathbb{1}_{E_n} = x_n, \sum_{i=1}^{N} \mathbb{1}_{E_i} = M \right\} \\ = \; & \frac{\binom{N-n}{M-s_n}}{\binom{N}{M}} P\left\{ \sum_{i=1}^{N} \mathbb{1}_{E_i} = M \right\} \qquad (M = s_n, \dots, N-n+s_n) \end{aligned}$$

[1] For any event A, $\mathbb{1}_A$ stands for its indicator, that is, $\mathbb{1}_A = 1$ or 0 depending on A comes true or not.

obtains and, by the Bayes theorem,

$$\begin{aligned} & P\left(\sum_{i=1}^{N} \mathbb{1}_{E_i} = M \mid \sum_{i=1}^{n} \mathbb{1}_{E_i} = s_n\right) \\ = & \frac{\binom{n}{s_n} P\{\mathbb{1}_{E_1} = x_1, \ldots, \mathbb{1}_{E_n} = x_n, \sum_{i=1}^{N} \mathbb{1}_{E_i} = M\}}{P\{\sum_{i=1}^{n} \mathbb{1}_{E_i} = s_n\}} \end{aligned} \tag{2}$$

where

$$P\left\{\sum_{i=1}^{n} \mathbb{1}_{E_i} = s_n\right\} = \sum_{M=s_n}^{N-n+s_n} \frac{\binom{n}{s_n}\binom{N-n}{M-s_n}}{\binom{N}{M}} P\left\{\sum_{i=1}^{N} \mathbb{1}_{E_i=M}\right\}. \tag{3}$$

De Finetti considers with the utmost care the case when $E_1, \ldots, E_N$ represent the initial segment of an infinite sequence of (exchangeable) events and N (e.g., size of statistical population) is immensely large compared with n (size of sample). The study of this realistic situation is based on the noteworthy result that the sequence $(\sum_1^n \mathbb{1}_{E_i}/n)_{n \geq 1}$ *mutually* converges in probability which, in turn, implies that the corresponding sequence of probability distributions $F_n(x) := P\{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{E_i} \leq x\}, x \in R$ and $n = 1, 2, \ldots$, weakly converges to a probability distribution function F, supported by $[0, 1]$. Then, passing to the limit, as $N \to +\infty$, in (3),

$$P\left\{\sum_{i=1}^{n} \mathbb{1}_{E_i} = s_n\right\} = \binom{n}{s_n} \int_{[0,1]} \theta^{s_n}(1-\theta)^{n-s_n} dF(\theta) \tag{4}$$

obtains for every n and $s_n = 0, 1, \ldots, n$. This is the celebrated *de Finetti's representation theorem*, described according to the original way of arguing of its discoverer. Consequently, the main problem of providing previsions for $\sum_{i=1}^{N} \mathbb{1}_{E_i}$ can be solved, besides (2), by

$$\begin{aligned} & P\left(\sum_{i=1}^{N} \mathbb{1}_{E_i} = M \mid \sum_{i=1}^{n} \mathbb{1}_{E_i} = s_n\right) \\ = & \binom{N-n}{M-s_n} \frac{\int_{[0,1]} \theta^{M}(1-\theta)^{N-M} dF(\theta)}{\int_{[0,1]} \theta^{s_n}(1-\theta)^{n-s_n} dF(\theta)}. \end{aligned} \tag{5}$$

Whence, (2) and (3) can be re-written, and simplified to a certain extent, as (5) and (4) respectively, and this shows that it is enough to assign the limiting distribution of the frequency of success, i.e. F, in order to master all the pertinent distributional aspects of an *infinite* sequence of exchangeable events. Then, although F can be considered as a very useful tool to specify those distributional aspects, it is but an "auxiliary" ingredient for the most statistical models. In spite of these incontrovertible remarks about the nature of what is generally called *initial* or a *priori distribution*, Bayesians persevere in formulating inferences on some random element "initially" distributed according to F, even when discovering any real empirical meaning for such a random element is clearly impossible. Under these unfortunately all too common circumstances, unknown parameters are but fictions, or mythical, pseudoentities, as recalled at the beginning of this section, and de Finetti's invitation mentioned therein, to get down to predictive aspects, does warn us against inconsistent arguments. Objectivists are used to study statistical procedures under an "extreme" form of exchangeability, that is, when observations $\xi_1, \xi_2, \ldots$ are supposed to be independent and identically distributed with an unknown distribution. If observations are real valued, it is customary to make reference to their common probability distribution function $x \mapsto F(x)$. In view of the assumption of stochastic independence, the conditional distribution for each future observation, given any number of past observations, must coincide with F. Then, the discussion and the study of relations between prevision of frequencies relative to future observations and observed frequencies (that, in the case of general exchangeable sequences, resort to statements of theorems like (1)) is based on the study of the behaviour of some distance between F and the *empirical distribution* of $\xi_1, \ldots, \xi_n$, that is

$$F_n(x) := \frac{1}{n} \sum_{i=n}^{n} \mathbb{1}_{(-\infty,x]}(\xi_i) \qquad (x \in \mathbb{R}, n = 1, 2, \ldots).$$

For a fixed x, the problem boils down to a simple application of the classical law of large numbers, but its complete solution entails more sophisticated tools from the theory of empirical stochastic processes (random functions whose realizations are frequency distribution functions).

In the first issue dated 1933 of *Giornale dell'Istituto Italiano degli Attuari (GIIA)*, a fundamental paper of Kolmogorov (see [14]) appeared in which the great Russian mathematician gave the limiting distribution, as $n \to +\infty$, of $\sqrt{n}\, ||D_n||_u$ where

$$||D_n||_u := \sup_{x \in \mathbb{R}} |F(x) - F_n(x)|,$$

under the hypothesis that F is *continuous*. This result, besides its ground-breaking value from a pure probabilistic viewpoint, influenced the development of statistical tests of goodness of fit and of the so-called two-sample problem.

In that very same *GIIA* issue of 1933, the paper immediately following Kolmogorov's was a paper by Valeri Glivenko (cf. [11]) in which it is proved, under the hypothesis of continuity for F, that $||D_n||_u$ converges to zero, with probability 1, when $n \to +\infty$.

This specific problem did attract de Finetti's attention. As a result he wrote a short paper, published in the third issue of *GIIA* (dated July 1933). See [7]. As stated in the first section, the content consists of a few simple remarks concerning:

(a) The way of proving the Glivenko theorem.

(b) How to take the problem when the discrepancy between probability distribution functions is measured through a more suitable distance than $||D_n||_u$, i.e. Lévy's distance for arbitrary probability distribution functions F and G on $\mathbb{R}$,

$$\lambda(F, G) := \inf\{\epsilon > 0 : F(x - \epsilon) - \epsilon \le G(x) \le F(x + \epsilon) + \epsilon \text{ for all } x\}$$

which metrizes weak convergence of probability distributions on R.

(c) The study of the Glivenko problem, with respect to $||D_n||_u$, when the size of each jump of F represents the probability concentrated at the corresponding discontinuity point.

Before describing and discussing de Finetti's contribution with reference to Cantelli's better known paper (cf. [2]), let us say a few words about point (c). Indeed, within the usual Kolmogorov axiomatic theory of probability, condition mentioned *sub* (c) is a necessary consequence of the axiom of σ-additivity. On the other hand, when simply additive probabilities are admitted, as is the case with de Finetti's theory based on the coherence principle, then the probability concentrated at any discontinuity point might be smaller than the jump, even vanish. This explains the rationale of (c).

Apropos of (a), de Finetti shows that the Glivenko theorem can be deduced, in an easy and elementary way, from Cantelli's classical strong law of large numbers (cf. [1])

> " which can be considered as source of inspiration for all these studies."

As for (b), in Section 3 of de Finetti's paper it is shown that the restriction concerning the continuity of F is needless when the sup-norm is replaced by the Lévy metric, i.e.:
Whatever F may be, given $\epsilon, \eta > 0$ there is $N = N(\epsilon, \eta)$ such that

$$\sup_p P\left\{ \max_{N \le n \le N+p} \lambda(F_n, F) \ge \epsilon \right\} \le \eta \tag{6}$$

which, within the usual σ-additive setting, is tautamount to asserting

$$\lim_{N\to+\infty} P\left\{\sup_{n\geq N} \lambda(F_n, F) > \epsilon\right\} = 0 \qquad (\epsilon > 0).$$

When the circumstance mentioned sub (c) occurs, it is easy to show that the argument used to prove (6) for arbitrary probability distribution functions holds true in a slightly more restrictive sense, but basically just enough to extend Glivenko's result to discontinuous F.

Whence, de Finetti's remarks about the Glivenko problem highlight the very nature of problem itself and include the part of the theorem that is customarily attributed to Cantelli. Having ascertained that the name of de Finetti ought to appear, with good reason, among the discoverers of the so-called *fundamental theorem of mathematical statistics*, it is worthwhile examining things from Cantelli's vantage point. It should be recalled that Cantelli himself, as editor of *GIIA* (from 1930 to 1957), decided to accept de Finetti's paper. Moreover, it should be mentioned that his extension of the Glivenko theorem appears in the same July issue of *GIIA*, straight after de Finetti's paper. He quotes Glivenko's and de Finetti's papers, and it would pretty tough indeed to pinpoint what actually led to attribute the honor of the discovery to Cantelli and Cantelli alone. The only explanation could stem from Cantelli's hint that the content of the paper had actually be the subject of a few lectures given at the *Institut H. Poincaré* (Paris) in May of 1933. In point of fact, Cantelli's paper did appeared, but two years later, in the *Annales* of that Institute, and covered the convergence of sequences of random elements. Cf. [3]. It is split up into three parts, the first of which contains a detailed proof of the proposition formulated in his 1933 paper.

2.2 Correlation and concordance

The last paper we intend to describe deals with linear correlation between real-valued random variables. It was published in 1937 in the statistical supplement to a magazine of politics, history and economics expression of the so-called left wing of the fascist party. The title of the paper was *A proposito di "correlazione"* (cf. [8]) and can be traced back to a debate on the use and misuse of the Bravais-Galton correlation coefficient which came to the fore during the 23rd Session of the *International Statistical Institute* (London, 1934). Although the main goal of the paper is to seek an effective explanation of the essential properties of the coefficient and of its domain of application, its value is enhanced by a few original hints.

2.2.1 Geometry of correlation

As to the first point under scrutiny, de Finetti – well ahead of his times – suggests considering the following geometric interpretation. One considers the set D of all *random* numbers X, such that $E(X^2) < +\infty$, as a vector space, any vector being thought of as a class of random numbers which differ by a constant (with probability one). Then, denoting mean square deviation by $\sigma(.)$ and linear correlation coefficient between X and Y by $r(X, Y)$, the covariance $\mathrm{Cov}(X, Y) = \sigma(X)\sigma(Y)r(X, Y)$ can be interpreted as *inner product* and, consequently, $\sigma(.)$ defines a norm on D. It turns out that $r(X, Y)$ is the cosine of the angle formed by the vector X with the vector Y, which is determined uniquely by setting $r(X, Y) = \cos[\alpha(X, Y)]\text{with } 0 \leq \alpha(X, Y) \leq \pi$.

By starting from this geometrical interpretation, one easily grasps: the meaning of the extreme situations $(r = -1, \text{or} + 1)$, the meaning of positive (negative, respectively) correlation, and that of absence of correlation (i.e. orthogonality). Moreover, through such a representation, well-known properties of *inner product spaces* may remind us of geometrical images suitable for studying and solving various problems of a probabilistic or statistical nature. The following propositions, for example, can be seen as direct consequences of geometrical properties.

(A) If $X_1, \ldots, X_n$ are linearly independent vectors (in the aforesaid sense), then they can be expressed as linear combinations of n random numbers $Y_1, \ldots, Y_n$ satisfying: $E(Y_i) = 0$ and $\sigma(Y_i) = 1$ for every $i, r(Y_i, Y_j) = 0$, for every $i \neq j$.

(B) If $X_1, X_2, \ldots, X_n$, as in (A), satisfy $r(X_i, X_j) = \rho$ for every $i \neq j$ (pairwise equicorrelation) with $E(X_i) = 0$ and $\sigma(X_i) = \sigma$ for every i, then ρ attains the permitted minimum value if and only if $\sum_{i=1}^{n} X_i = 0$ with probability 1, in which case: $\rho = -1/(n-1)$ (Just think, in particular, to application to exchangeable random numbers.)

2.2.2 Classes of distributions with given marginals

Moving on to further original issues contained in the paper, we come across an amazing *ante litteram* argument involving distributions with fixed marginals, with which Hoeffding and Fréchet are generally credited. Cf. [12] and [9], respectively. Consider events $E_1, E_2, E = E_1 \cap E_2$ and their probabilities p_1, p_2 and p, respectively; set $q_i = 1 - p_i (i = 1, 2), q = 1 - p$. Now, if p_1, and p_2 are *fixed*, then the range of the correlation coefficient r between (the indicator of) E_1 and (the indicator of) E_2 is, in general, a proper subset of $[-1, 1]$. For instance, if $p_1 < p_2$ and $p_1 + p_2 \leq 1$, one has

$$-\left(\frac{p_1 p_2}{q_1 q_2}\right)^{1/2} \leq r \leq \left(\frac{p_1 q_2}{q_1 p_2}\right)^{1/2}.$$

Note that the two bounds coincide with -1 and $+1$, respectively, if and only if $p_1 = p_2 = 1/2$. But, on the other hand, both these bound could be very close to 0 (when p_1 is close to 0). De Finetti extends these conclusions to pairs of random numbers X_1 and X_2 with assigned marginal distribution functions Φ_1 and Φ_2, respectively. Firstly, he argues that $r(X_1, X_2)$, in this case, can attain the value 1 (-1, respectively) if and only if $\Phi_1(x) = \Phi_2(ax + b)$ for every x and some $a > 0$ and b ($\Phi_1(x) = 1 - \Phi(ax + b)$ for every x and some $a < 0$, and b, respectively). Moreover, the range of r coincides with $[-1, 1]$ if and only if $\Phi_1(x) = \Phi_2(ax + b)$ for every x, for some $a > 0$, and Φ_1 is "symmetric".

At this point, de Finetti shows that, generally speaking, the extremes, say r_1 and r_2 of the range of $r(X_1, X_2)$ can be obtained

> by considering the extreme cases in which the values of (X_1, X_2) lie on a decreasing (an increasing, respectively) curve, with probability 1.

We see that this is the very same principle as the one that Fréchet will consider to define the extreme distribution functions of the classes that we routinely call *Fréchet classes*. In point of fact, de Finetti indicates how to determine the equations of the curves which support he extreme distributions and, consequently, obtains the following expression for r_1 and r_2 (when $E(X_i) = 0$ and $\sigma(X_i) = 1$ for $i = 1, 2$)

$$r_1 = \int_0^1 \Phi_1^{-1}(t)\Phi_2^{-1}(1-t)dt, \quad r_2 = \int_0^1 \Phi_1^{-1}(t)\Phi_2^{-1}(t)dt.$$

Finally, de Finetti uses this approach to clarify the nature of one of the problems which had sparked off the controversy mentioned at the beginning of this section, and to solve it. Reference is made to the relationship between stochastic independence and orthogonality (i.e., $r(X_1, X_2) = 0$). He proves that: If Φ_1 and Φ_2 are assigned, then the two concepts turn out to be equivalent if and only if both the support of Φ_1 and of Φ_2 contains two points at the most.

2.2.3 Measures of concordance

With reference to the correlation coefficient, there was another question open to dispute, that is: Can this index be used to measure the *concordance* between X_1 and X_2?

Roughly speaking, two ordered characters are said to be concordant (discordant, respectively) if, being either character large, this indicates that the other character is large (small, respectively). According to this meaning of the term, due to Gini and studied by his School, concordance is more general than correlation, the latter corresponding to a very particular form of concordance or discordance (linear). Then, the measurement of concordance requires appropriate indices like the *indices of homophily* introduced by Gini. Cf. [10]. In the last section of *A proposito di correlazione* de Finetti defines a new index of concordance/discordance, the one he considers as

> the most simple and the most intrinsically revealing.

In order to point out the difference between the concepts at issue, he gives an example to show that the Bravais coefficient and this new index of concordance and discordance may have discordant signs. As to the definition of the new index for a (bivariate) distribution F, he considers two stochastically independent random vectors (X_1, Y_1) and (X_2, Y_2) with the same probability distribution function F. Clearly, the sign of the product $(X_2 - X_1)(Y_2 - Y_1)$ indicates concordance/discordance between two characters X and Y jointly distributed according to F, and an appropriate index can be obtained by taking the expectation of the signum of that product. This yields

$$C(F) = \int_{\mathbb{R}^4} \text{sign}[(x_2 - x_1)(y_2 - y_1)] \, dF(x_1, y_1) \, dF(x_2, y_2).$$

It should be noted that if, instead of expectation of the sign of the product, we take the expectation of the product, then we obtain a concordance/discordance index which is undoubtably more related, than C, to r. Indeed,

$$\int_{\mathbb{R}^4} (x_2 - x_1)(y_2 - y_1) \, dF(x_1, y_1) \, dF(x_2, y_2) = 2\sigma(X_1)\sigma(Y_1)r(X_1, Y_1).$$

It's a plain fact that the sign of this index, unlike C, must be the same as the one of r.

Attentive readers have certainly noticed that C is but the index which, in statistical literature, is well-known as *Kendall's* τ. Cf. [13]. Yet, de Finetti introduced this index one year before Kendall did, so that one might as well redesignate it as *de Finetti-Kendall index*. This had already been proposed in Cifarelli and Regazzini, but in vain. Cf. [4].

References

[1] F. P. Cantelli, Sulla probabilità come limite delle frequenze, Rendiconti Regia Accademia Nazionale dei Lincei **26**, serie V, 39 (1917).

[2] F. P. Cantelli, Sulla determinazione empirica delle leggi di probabilità., Giornale dell'Istituto Italiano degli Attuari **4**, 421 (1933).

[3] F. P. Cantelli, Considérations sur la convergence dans le calcul de probabilités, Annales de l'Institut H. Poincaré **5**, 3 (1935).

[4] D. M. Cifarelli and E. Regazzini, De Finetti's contribution to probability and statistics, Statistical Science **11**, 253 (1996).

[5] B. de Finetti, Funzione caratteristica di un fenomeno aleatorio, Memorie della Regia Accademia Nazionale dei Lincei **VI**, 86 (1930).

[6] B. de Finetti, Funzione caratteristica di un fenomeno aleatorio, Atti del Convegno Internazionale dei Matematici (Bologna, 1928), **VI**, 179 (Zanichelli, Bologna, 1932).

[7] B. de Finetti, Sull'approssimazione empirica di una legge di probabilità, Giornale dell'Istitituto degli Attuari **4**, 415 (1933).

[8] B. de Finetti, A proposito di "correlazione", Supplemento statistico ai Nuovi Problemi di Politica, Storia ed Economia **3**, 41 (1937).

[9] M. Fréchet, Sur les tableaux de corrélation dont les marges sont données, Annales de l'Université de Lyon, Sciences, section A **14**, 53 (1951).

[10] C. Gini, Di una misura della dissomiglianza tra due gruppi di quantità e delle sue applicazioni allo studio delle relazioni statistiche, Atti Regio Istituto Veneto di Scienze Lettere e Arti **74**, 185 (1915). (Altre quattro Note apparse negli stessi Atti nel biennio 1915-1916)

[11] V. Glivenko, Sulla determinazione empirica delle leggi di probabilità., Giornale dell'Istituto Italiano degli Attuari **4** 92 (1933).
[12] W. Hoeffding, Masstabinvariate Korrelations-theorie, Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin **5**, 181 (1940).
[13] M. G. Kendall, A new measure of rank correlation, Biometrika **30**, 81 (1938).
[14] A. N. Kolmogorov, Sulla determinazione empirica di una legge di distribuzione, Giornale dell' Istituto Italiano degli Attuari **4**, 92 (1933).
[15] H. V. Roberts, Probabilistic prediction, Journal of the American Statistical Association **60**, 50 (1965).

3 De Finetti's first paper about exchangeability

3.1 Introduction

Bruno de Finetti is regarded as the founder of the theory of *sequences* of *exchangeable random variables* or *random exchangeable sequences* for short. His first important article about this subject dates back to 1930 [see [6]]. It appears as a *Memoria*, published in the proceedings of the *Accademia dei Lincei*, jointly presented to the *Accademia* by two of the most outstanding Italian scientists of the time: Guido Castelnuovo (1865–1952) and Tullio Levi–Civita (1873–1941). It is clear, from his biography, that de Finetti had examined the problem of finding a probabilistic description and interpretation for *random phenomena* – those which can be repeatedly observed under homogeneous environmental conditions – ever since his early approach to probability, a couple of years before his degree in mathematics, obtained in 1927 at the *University of Milan*. In fact, he presented a summary of the *Memoria* in a pithy communication at the *International Mathematical Congress* held in Bologna, in September 1928. The text of such a communication was published, in 1932, in the sixth volume of the proceedings of the *Congress*. See [11].

The present section of the paper aims at giving a precise idea of the content of the *Memoria* and, especially, of the methods used therein, since they are different from those employed in later de Finetti's contributions to the same subject. This analysis is split into six points which form Section 3.2. Some new developments of de Finetti's original methods are sketched in Section 3.4. The intermediate Section 3.3 reviews a paper by Jules Haag (1882–1953) in which, so far as we know, the concept of exchangeable events had been introduced and studied for the first time. A comparison between this paper and de Finetti's *Memoria* clarifies the complete independence of the two papers, and it convinces of the prominent merits of de Finetti in this field.

Throughout the section, the main propositions will be marked with $(P.)$.

3.2 Characteristic function of a random phenomenon

In view of the homogeneity of the environmental conditions which distinguishes random phenomena (with *equivalent* trials) from other types of phenomena, de Finetti points out that a correct probabilistic translation of such an empirical circumstance leads us to think of the probability of m successes and $(n-m)$ failures, in n trials, as invariant with respect to the order in which successes and failures alternate, whatever n and m may be. Accordingly, he defines a sequence $(E_n)_{n\geq 1}$ of events to be *equivalent* if, for every *finite* permutation π, the probability distribution of $(\mathbb{I}_{E_1}, \mathbb{I}_{E_2}, \dots)$ [1] is the same as the probability distribution of $(\mathbb{I}_{E_{\pi(1)}}, \mathbb{I}_{E_{\pi(2)}}, \dots)$. So, if $\omega_k^{(n)}$ denotes the probability that the random phenomenon, taken into consideration, comes true k times in m trials, one gets

$$\omega_k^{(m)} = \sum_{h=k}^{n-m+k} \omega_h^{(n)} \frac{\binom{h}{k}\binom{n-h}{m-k}}{\binom{n}{m}} \tag{7}$$

whenever $1 \leq m \leq n$ and $k = 0, \dots, m$.

Nowadays, the term *equivalent* is replaced by the more expressive and unambiguous word *exchangeable*, proposed perhaps by Pólya (cf. [13, 14]) or by Fréchet (cf. [16]).

From (7) de Finetti derives a difference–differential equation for the probability generating function and, consequently, for the characteristic function, of the frequency of success in n trials. Such a characteristic function and its limit, as $n \to +\infty$, in the case of an infinite sequence, becomes the center of de Finetti's treatment of exchangeability.

From a methodological viewpoint, the use of characteristic functions joins the *Memoria* to the contemporaneous de Finetti's studies on processes with stationary independent increments, based on the analysis

[1] For any event E, $\mathbb{I}_E$ will stand for its indicator.

of the *derivative law* defined in terms of the characteristic function ψ_λ of the λ–th coordinate of the process; [5]. See also [24].

As already recalled in the first section, the application of the characteristic functions method to the study of exchangeable sequences is a peculiarity of the *Memoria*. In point of fact, on Khinchin's advice, in all subsequent papers on exchangeable random elements, de Finetti uses more direct tools such as probability distribution functions, moments, and so on. See, e. g., Subsection 2.1 of the present paper. We have experimented that the original de Finetti approach has some remarkable merits with respect to some important problems like, for instance, concrete assessment of finitary exchangeable laws and extendibility of exchangeability. So, we believe that an accurate and faithful account of that approach could come in handy to all scholars who are unable to read Italian scientific literature.

3.2.1 Fundamental recurrence relation

Our description starts with Author's remark that (7), for $m = n - 1$, reduces to

$$n\omega_k^{(n-1)} = (n-k)\omega_k^{(n)} + (k+1)\omega_{k+1}^{(n)}$$

with $\omega_0^{(0)} = 1$. Thus, for any sequence of N exchangeable events, he deduces the difference–differential equation

$$n\Omega_{n-1}(z) = n\Omega_n(z) + (1-z)\Omega_n'(z) \tag{8}$$

valid for $n = 1, \dots, N$ and any complex number z, where Ω_n is the probability generating function defined by

$$\Omega_n(z) := \sum_{h=0}^{n} \omega_h^{(n)} z^h \qquad (n = 1, 2, \dots, N;\ z \in \mathbb{C}) \tag{9}$$

with $\Omega_0(z) \equiv 1$.

Firstly, (8) is used to prove the identity

$$\frac{1}{m!}\Big(\frac{d^m\Omega_n}{dz^m}\Big)(1) = \binom{n}{m}\omega_m^{(m)}$$

and, consequently, to write

$$\Omega_n(1+z) = \sum_{h\geq 0} \binom{n}{h}\omega_h^{(h)} z^h. \tag{10}$$

At this stage, de Finetti defines the characteristic function of the frequency [2] $(\sum_{i=1}^{N} \mathbb{I}_{E_k}/N)$

$$t \mapsto \Psi_N(t/N) := \Omega_N(e^{it/N}) \qquad (t \in \mathbb{R})$$

to be the *characteristic function of the* (finite) *class* $\{E_1, \dots, E_N\}$ *of exchangeable events*. Clearly, such a function characterizes the probability distribution of the random vector $(\mathbb{I}_{E_1}, \mathbb{I}_{E_2}, \dots, \mathbb{I}_{E_N})$. Notice that this distribution is also determined by the sole knowledge of the probabilities $\omega_h^{(h)}$, $h = 0, 1, \dots, N$, with $\omega_0^{(0)} = 1$. To see this, combine (9) with (10).

From a practical viewpoint, the following proposition – that the Author states in Section 35 of the *Memoria* – may be useful.

(P_1) *Any sequence* $(\tilde{\omega}_h^{(N)})_{h=0,\dots,N}$, *satisfying*

$$\tilde{\omega}_h^{(N)} \geq 0 \quad (h = 0, \dots, N) \quad \textit{and} \quad \sum_{h=0}^{N} \tilde{\omega}_h^{(N)} = 1,$$

[2] Throughout the paper, the term frequency is used to designate what other authors call relative frequency.

generates a unique exchangeable law, for the class of events $\{E_1, \dots, E_N\}$*, according to which the probability that a random phenomenon comes true* k *times in* n *trials* $(1 \le n \le N, k = 0, \dots, m)$ *is given by*

$$\sum_{h=k}^{N-n+k} \tilde{\omega}_h^{(N)} \frac{\binom{h}{k}\binom{N-h}{n-k}}{\binom{N}{n}}. \tag{11}$$

Indeed, consider the partition defined by

$$A_h := \{\sum_{k=0}^{N} \mathbb{I}_{E_k} = h\} \qquad h = 0, 1, \dots, N$$

in a probability space such that $\tilde{\omega}_h^{(N)}$ is the probability of A_h. If the event A_h occurs, then h white balls along with $(N-h)$ black balls are placed into an urn. Now, consider an individual who just assesses the quantities $\tilde{\omega}_h^{(N)}$ as the probabilities for the events A_h and who randomly draws n balls from the urn $(n \le N)$, without replacement. So, if he sees the $N(N-1)\dots(N-n+1)$ possible outcomes as equally probable, whatever n may be, then the probability that the sample contains exactly k white balls is given by (11). In other words, any N–exchangeable $\{0,1\}$ sequence is a mixture of hypergeometric N–sequences.

After establishing these basic elementary facts, de Finetti moves on to the analysis of infinite sequences of exchangeable events. Such analysis is focused on the study of the pointwise limit of the characteristic function $\Psi_N(t/N)$, as $N \to +\infty$. As a matter of fact, in all later writings on exchangeability, he will consider a different approach, based on a law of large numbers for exchangeable sequences. He adopted this approach following a suggestion of Alexander Khinchin (1984–1969), he met on the occasion of the *Congress* of Bologna. See [12] and [20, 21].

3.2.2 Representation theorem

Given an infinite sequence $(E_n)_{n\ge1}$ of exchangeable events, consistently with the previous notation define $\omega_h^{(h)}$ to be the probability of $E_1 \cap \dots \cap E_h$, for $h = 1, 2, \dots$, and set

$$\Omega(1+z) := \sum_{h\ge0} \omega_h^{(h)} \frac{z^h}{h!} \qquad (z \in \mathbb{C}).$$

In Section 6 of the *Memoria* de Finetti proves the following preliminary proposition:

(P_2) *For any strictly positive* a *and* ϵ *there is an integer* $N_1 = N_1(a, \epsilon)$ *such that*

$$\sup_{|z|\le a} |\Omega(1+z) - \Omega_n(1+z/n)| \le \epsilon \qquad (n \ge N_1).$$

$\Psi_n(t/n)$, as $n \to +\infty$:

(P_3) *For every* $\tau > 0$ *and* $\epsilon > 0$*, there is* $N_2 = N_2(\epsilon, \tau)$ *such that*

$$\sup_{|t|\le\tau} |\Psi(t) - \Psi_n(t/n)| \le \epsilon$$

holds true for every $n \ge N_2$ *and*

$$\Psi(t) := \Omega(1+it) = \sum_{h\ge0} \omega_h^{(h)} \frac{(it)^h}{h!} \qquad (t \in \mathbb{R}).$$

It is important to note that (P_2) and (P_3) are valid uniformly with respect to Ω and Ψ, respectively. In other words, given ϵ, a and τ, N_1 and N_2 do not depend on Ω and Ψ, respectively. See next Subsection 3.2.5 for a different situation apropos of the connection between frequency and predictive distribution.

So, *if one assumes that* Ψ *is a characteristic function* (see the next subsection), then the corresponding random variable must take values in $[0,1]$, with probability one. Moreover, if Φ is the corresponding probability distribution function, since Ψ can be extended as an entire function, one gets

$$\Psi(t) = \int_{[0,1]} e^{it\xi} d\Phi(\xi), \qquad \omega_h^{(h)} = \int_{[0,1]} \xi^h d\Phi(\xi) \quad (h = 0, 1, \dots).$$

This, in turn, combined with (10), gives

$$\Omega_n(1+z) = \int_{[0,1]} (1 + z\xi)^n d\Phi(\xi)$$

and

$$\begin{aligned}\sum_{h \geq 0} \omega_h^{(n)} z^h = \Omega_n(z) &= \int_{[0,1]} (1 - \xi + z\xi)^n d\Phi(\xi) \\ &= \sum_{h \geq 0} \binom{n}{h} z^h \int_{[0,1]} \xi^h (1-\xi)^{n-k} d\Phi(\xi).\end{aligned}$$

(P_4) *The events* $(E_n)_{n\geq 1}$ *are exchangeable if and only if there is a probability distribution function* Φ *supported by* $[0,1]$ *such that the probability of* $\{\mathbb{I}_{E_1} = x_1, \dots, \mathbb{I}_{E_n} = x_n\}$ *is given by*

$$\int_{[0,1]} \xi^{\sigma_n} (1-\xi)^{n-\sigma_n} d\Phi(\xi)$$

for every $(x_1, \dots, x_n)$ *in* $\{0,1\}^n$ *for which* $x_1 + \dots + x_n = \sigma_n$, *and for every* $n = 1, 2, \dots$. *Moreover,* Φ *is the limit* (*in the sense of weak convergence of probability distributions*) *of the probability distribution function* Φ_n *of the frequency of success in the first* n *trials, as* $n \to +\infty$.

3.2.3 Important remark

The previous argument is based on the presumption that the limit, Ψ, of Ψ_n is a characteristic function. Nowadays, the validity of such an assertion is proved in any good probability textbook. On the contrary, the reference books at de Finetti's disposal – [1] and [22] – although they were superb, they contained the form of the continuity theorem according to which "if Ψ_n converges to a *characteristic function*, uniformly on any compact interval, then ...". Clearly, the argument in Subsection 3.2.2, apart from the fact that $t \mapsto \Omega(1 + it)$ is the limit – uniform on any compact interval – of a sequence of characteristic functions, does not give further indications about the fact that the limit is a characteristic function. So, to complete the proof of the representation theorem, de Finetti was obliged to check whether the above–mentioned limiting condition was enough to assert that $t \mapsto \Omega(1 + it)$ was a characteristic function. He deferred the solution of the problem to the Appendix of the *Memoria*, where he proved the desired completion of the continuity theorem – perhaps for the first time – consistently with the fact that he was dealing with *finitely* (i.e., not necessarily *completely*) *additive* distributions of general real–valued random variables. In point of fact, he explicitly assumes that the sequence of distributions corresponding to $(\Psi_n)_{n\geq 1}$ is tight.

3.2.4 Strong law of large numbers

In the following Sections 11 and 12, de Finetti deals with the extension of Cantelli's strong law for frequencies of Bernoulli trials to frequencies of more general exchangeable trials. Define the random frequency

$\bar{f}_n$ of success in the first n trials of a random phenomenon characterized by an infinite sequence $(E_n)_{n\geq 1}$ of exchangeable events,

$$\bar{f}_n := \frac{1}{n}\sum_{k=1}^{n} \mathbb{I}_{E_k},$$

and consider the sequence $(\bar{f}_n)_{n\geq 1}$. The main result de Finetti achieves apropos of the latter sequence is *a mutual* form of the strong law of large numbers for $(\bar{f}_n)_{n\geq 1}$ that, consistently with the admissibility of simply additive probability distributions, he states correctly in the following "finitary" style.

(P_5) *Given strictly positive numbers* ϵ *and* θ*, there is a positive integer* $N := N(\epsilon, \theta)$ *such that the probability of the event*

$$\bigcap_{j=1}^{k}\{|\bar{f}_n - \bar{f}_{n+j}| \leq \epsilon\}$$

turns out to be uniformly (*with respect to* $k = 1, 2, \dots$) *greater that* $1 - \theta$*, whenever* $n \geq N$.

Apparently de Finetti was aware of the fact that, in a framework of completely additive probability distributions on subsets of a sample space, the above proposition holds with $k = +\infty$, and that one can assert the existence of a random number f^*, with probability distribution function Φ, that can be viewed as the almost sure limit of $(\bar{f}_n)_{n\geq 1}$. But he had at least three good reasons, from his viewpoint, to be uninterested in the "strong" formulation of his strong law of large numbers. These reasons, briefly mentioned in many points of the *Memoria*, are discussed in a more systematic way in a few contemporaneous de Finetti's papers such as [7–9]. It is worth recalling them here, in a three-point summary. (i) Logically speaking, it is unjustifiable to speak of an *infinite* sequence of trials of a random phenomenon: The number of the trials could be arbitrarily great but, in any case, finite. (ii) Without the assumption of complete additivity and with no reference to a sample space, there is no possibility of deducing the existence of a limiting random quantity from the sole mutual convergence of a given sequence. (iii) de Finetti deduces the whole theory of probability from a very natural condition having an obvious meaning – the so-called *condition of coherence* – and shows that *complete* additivity is not necessary for a quantitative measure of probability to be coherent. See [10].

The strong law of large numbers for the frequency of success in a sequence of exchangeable events represents the last issue dealt with in Chapter 1 of the *Memoria*. Chapter 2 contains the definitions of some operators on the set of all characteristic functions, with the intention of providing a rigorous, systematic presentation, in Chapter 3, of asymptotic properties of the posterior distribution and of the merging of the predictive distribution with the frequency of success in past trials of a given random phenomenon. In view of the purely instrumental function of Chapter 2, here we jump to the more important Chapter 3.

3.2.5 Probability and experience: posterior and predictive distributions

At the time of the draft of the *Memoria*, de Finetti was unfamiliar with techniques of statistical inference, and it's amazing how he was, nevertheless, able at picking out the essence of the inductive reasoning and the tools to deal with it, from a coherent mathematical standpoint. In his view of these subjects, exchangeability is a means to study and understand the role played by the knowledge of data, gathered from experience, with regards to the evaluation of probability. In particular, he aims at clarifying how exchangeability can be employed to provide with a basis the common belief that prevision of new facts rests on the analogy with past observed facts. In the case of a random phenomenon, this belief leads to assume, although with caution, past frequency as an approximate value for probability. So, in Section 27, de Finetti provides a new rigorous description of the asymptotic behavior of the posterior distribution [3], and makes use of this statement to show the merging of the predictive distribution with frequency in past trials. Apropos of the former, he considers *infinite sequences* of exchangeable trials of a random phenomenon, generating convergent sequences of frequencies. More precisely,

[3] In point of fact, he was unaware of [26], where a strictly related problem is studied.

(P_6) *Let θ_0 be any point in the intersection of $(0,1)$ with the support of the distribution function, Φ, of the random phenomenon. Then the posterior distribution, given $\{\bar{f}_n = \frac{\sigma_n}{n}\}$, converges weakly to the point mass δ_{θ_0}, whenever $\sigma_n/n \to \theta_0$ as $n \to +\infty$, i.e.*

$$\lim_{n\to+\infty} \frac{\int_{\theta_0-\epsilon}^{\theta_0+\epsilon} \theta^{\sigma_n}(1-\theta)^{n-\sigma_n} d\Phi(\theta)}{\int_{[0,1]} \theta^{\sigma_n}(1-\theta)^{n-\sigma_n} d\Phi(\theta)} = 1 \qquad (\epsilon > 0).$$

Whence, as for the conditional probability of $\{E_{n+k}\}$ given $\{\bar{f}_n = \frac{\sigma_n}{n}\}$, viz.

$$\frac{\int_{[0,1]} \theta^{1+\sigma_n}(1-\theta)^{n-\sigma_n} d\Phi(\theta)}{\int_{[0,1]} \theta^{\sigma_n}(1-\theta)^{n-\sigma_n} d\Phi(\theta)},$$

one obtains that, for any $\epsilon > 0$, there is $N = N(\epsilon, \Phi)$, such that

$$\Big| \frac{\int_{[0,1]} \theta^{1+\sigma_n}(1-\theta)^{n-\sigma_n} d\Phi(\theta)}{\int_{[0,1]} \theta^{\sigma_n}(1-\theta)^{n-\sigma_n} d\Phi(\theta)} - \frac{\sigma_n}{n} \Big| \leq \epsilon$$

holds true for every $n \geq N$.

In Section 28, de Finetti explains the "relative" value of this proposition. Indeed, in view of the dependence of N on Φ, it does not allow a quantitative statement about the approximation of frequency to probability, independently of a complete *a priori* knowledge of the characteristic function of the random phenomenon.

Chapter 3 ends with a brief mention of the use of posterior distribution in the problem of hypothesis–testing: the sole explicit hint to a statistical technique, contained in the *Memoria*.

3.2.6 Classes of exchangeable events and extension of exchangeability

The main issue dealt with in the last chapter (Chapter 4, including Sections 31–36) is extendibility of exchangeability, from a finite sequence to a "longer" sequence of events. The problem can be formulated in the following terms: *Given positive integers n and k, establish conditions on the characteristic function of a random phenomenon of n exchangeable events in order that they may constitute the initial n–segment of a random phenomenon of $(n+k)$ exchangeable events.*

To solve this problem, de Finetti starts from (8), viewed as a first–order linear differential equation in the dependent variable Ω_{n+1}. Since the one–parameter family of solutions of this equation is

$$\Omega_{n+1}(z) = (1-z)^{n+1}\{(n+1)\int_0^z \Omega_n(x)(1-x)^{-(n+2)}dx + c\}, \tag{12}$$

then (12) can be combined with (P_1) to obtain

(P_7) $t \mapsto \Omega_n(e^{it/n})$ *is the characteristic function of the initial n-segment of a sequence of $(n+1)$ exchangeable events if and only if the constant c in* (12) *can be determined in such a way that all the coefficients of the polynomial* (*of degree $(n+1)$*)*, defined by the right–hand side of* (12)*, are nonnegative.*

Analogously, to solve the problem for some $k > 1$, one can start from (12) with $(n+2)$ in the place of $(n+1)$, consider it as an equation in the dependent variable Ω_{n+2} and, finally, substitute Ω_{n+1} with its expression in the right–hand side of (12). So, by an obvious recursive argument, de Finetti states that Ω_{n+k} can be written as

$$\Omega_{n+k}(z) = F(z) + C_1(1-z)^{n+1} + \cdots + C_k(1-z)^{n+k} \tag{13}$$

F being a polynomial, whose coefficients are completely determined by Ω_n. Then:

(P_8) $t \mapsto \Omega_n(e^{it/n})$ *is the characteristic function of the initial* n*–segment of a sequence of* $(n+k)$ *exchangeable events if and only if the constants* $C_1, \ldots, C_k$ *can be determined in such a way that all the coefficients of the polynomial* (*of degree* $(n+k)$)*, defined by the right–hand side of* (13)*, are nonnegative.*

Forty years later, de Finetti came back to the problem from a new standpoint, of a geometrical nature (see [15]), followed by some Authors such as [3,4], [17] and [27].

In Section 3.4 of the present paper, we will resume the original analytical de Finetti's argument, by providing an explicit form for F in (13). Our goal is to reformulate the necessary and sufficient condition in (P_8) in the guise of a system of linear inequalities.

De Finetti gives a complete solution of the extendibility problem when $k = +\infty$, i.e.: To establish conditions on $t \mapsto \Omega_n(e^{it/n})$ in order that it can be viewed as characteristic function of the first n trial of a random phenomenon of *infinite* exchangeable events. Resting on the representation (see Subsection 3.2.2) according to which $\omega_h^{(h)}$ is the h–moment of the probability distribution function Φ of the random phenomenon, via the Hamburger solution of the problem of moments (see, e.g., [25]), de Finetti was able to state:

(P_9) *In order that* $t \mapsto \Omega_n(e^{it/n})$ *may be the characteristic function of the initial* n*–segment of an infinite sequence of exchangeable events it is necessary and sufficient that all the roots of a distinguished polynomial, depending on* n*, belong to the closed interval* $[0,1]$*. The polynomial* (*in* ξ) *is*

$$Det \begin{pmatrix} 1 & \xi & \xi^2 & \ldots & \xi^k \\ \omega_0^{(0)} & \omega_1^{(1)} & \omega_2^{(2)} & \ldots & \omega_k^{(k)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \omega_{k-1}^{(k-1)} & \omega_k^{(k)} & \omega_{k+1}^{(k+1)} & \ldots & \omega_{2k-1}^{(2k-1)} \end{pmatrix}$$

if $n = 2k-1$*, while it is*

$$Det \begin{pmatrix} 1 & \xi & \xi^2 & \ldots & \xi^k \\ \omega_1^{(1)} & \omega_2^{(2)} & \omega_3^{(3)} & \ldots & \omega_{k+1}^{(k+1)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \omega_k^{(k)} & \omega_{k+1}^{(k+1)} & \omega_{k+2}^{(k+2)} & \ldots & \omega_{2k}^{(2k)} \end{pmatrix}$$

if $n = 2k$.

3.3 Haag's contribution to exchangeability

To our knowledge, Haag was the first Author to study sequences of exchangeable events. He publicized his conclusions during the *International Mathematical Congress* held in Toronto, August, 1924. His communication appeared in Vol. 1 of the *Proceedings*, published in 1928, the very same year of the already mentioned Bologna *Congress*. See [18]. It is highly likely that de Finetti was in the dark about the Haag contribution until the 1950s, when Edwin Hewitt and Leonard J. Savage mentioned it in a famous paper about exchangeability. See [19].

It is convenient to pause here and consider what Haag really did. In the first six brief sections, he deals with finite sequences of exchangeable events and furnishes a detailed account of the expressions of the $\omega_k^{(n)}$ both in terms of $\omega_h^{(h)}$ and in terms of $\omega_0^{(h)}$, for $h = 1, 2, \ldots, n$. In Section 7, Haag attains an early version of the representation theorem, but via a rather incomplete argument. He considers a sequence of exchangeable trials with a frequency of success σ_n/n converging to x as $n \to +\infty$. By the way, Haag does not hint at any form of law of large numbers, so that the reader is not able to judge whether the convergence assumption expresses an extraordinary or, instead, a common fact. By resorting to the Stirling formula, and assuming that $f(x)dx$ provides, for some continuous function f defined on $(0,1)$, an asymptotic value for the probability that σ_n/n belongs to $(x, x+dx)$, as $n \to +\infty$, he shows that

$$\sqrt{2\pi x(1-x)} \frac{(x^x(1-x)^{1-x})^n}{\sqrt{n}} f(x)$$

represents an approximate value – for great values of n – of $\omega^{(n)}_{\sigma_n}/\binom{n}{\sigma_n}$. At this stage, by means of a heuristic argument based on formal elementary computations, he concludes that

$$\binom{p+q}{p} x^p (1-x)^q \frac{1}{n} f(x) \tag{14}$$

gives an approximate value for the probability of the event "The limiting frequency belongs to $(x, x+1/n)$ and, simultaneously, one gets p successes in $n = p+q$ trials". So, the probability of p successes in $(p+q)$ trials can be represented as limit (as $n \to +\infty$) of a sum of terms like (14), i.e.

$$\binom{p+q}{p} \int_0^1 x^p (1-x)^q f(x) dx.$$

The *assumption* that the frequency converges to a random variable, weakens the validity of the Haag argument, and emphasizes the difference between his standpoint and de Finetti's stance. Indeed, de Finetti reckons that convergence of frequency must be proved, whilst it is evident that Haag is assuming the validity of some type of empirical law which postulates convergence of frequency. So, while de Finetti introduces exchangeability to explain the role of frequency in evaluating probability – within the framework of a rigorously subjectivisitc or, on depending on taste, axiomatic conception, Haag misses out on these fundamental aspects. Moreover, while de Finetti shows to have an extraordinarily advanced command of the right mathematical apparatus to deal with probabilistic problems, Haag does not go beyond the use of the elementary combinatorial calculus. In point of fact, the final part of his paper, intitled *Applications*, includes a review of classical problems solvable by means of elementary combinatorics.

3.4 Some new developments on extendibility

As anticipated in Subsection 3.2.6, in the last part of this paper we follow de Finetti's ideas, explained in that very same subsection, to obtain new necessary and sufficient conditions for extendibility of a given finite–dimensional exchangeable distribution. These conditions are of an algebraic nature, differently from the above–mentioned conditions derived in the frame of a geometrical approach. Taking (13) as a starting point, one first determines an explicit form for F, i.e.

$$\begin{aligned} F(z) = F(z; n, k) = z^k \frac{(n+k)!}{n!} \int_{(0,1)^k} \Big(\prod_{j=1}^{k} t_j^{j-1} \Big) \{1 - z(1 - t_1 \cdots t_k)\}^n \cdot \\ \cdot\, \Omega_n \left(\frac{t_1 \cdots t_k z}{1 - z(1 - t_1 \cdots t_k)} \right) dt_1 \dots dt_k \qquad (k = 1, 2, \dots). \end{aligned} \tag{15}$$

To prove the validity of this representation, first note that (15) with $k = 1$ is consistent with (12). Then, to complete the proof, use (12), with n replaced by $(n+k)$, and proceed by mathematical induction with respect to k.

Now, substitute expression (9) into (15) to obtain

$$\begin{aligned} F(z) &= \frac{(n+k)!}{n!} \sum_{l=0}^{n} \omega_l^{(n)} z^{l+k} \int_{(0,1)^k} (t_1 \cdots t_k)^l (1 - z(1 - t_1 \cdots t_k))^{n-l} t_2 t_3^2 \dots t_k^{k-1} dt_1 \dots dt_k \\ &= \frac{(n+k)!}{n!} \sum_{l=0}^{n} \omega_l^{(n)} z^{l+k} \frac{1}{\Gamma(k)} \int_0^1 x^l (1-x)^{k-1} \{1 - z(1-x)\}^{n-l} dx \\ &\qquad \text{(see, for example, 3.3.5.11 in [23])} \\ &= \frac{(n+k)!}{n!(k-1)!} \sum_{l=0}^{n} \omega_l^{(n)} \sum_{h=0}^{n-l} \binom{n-l}{h} z^{h+l+k} (-1)^h B(l+1, k+h) \end{aligned}$$

with $B(\alpha,\beta) := \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$. Whence, from (12),

$$\Omega_{n+k}(z) = \sum_{i=0}^{k-1}(-1)^i z^i \sum_{j=1\vee(i-1)} \binom{n+j}{i} C_j + \sum_{i=k}^{n+k}(-1)^i z^i \Big\{ \sum_{j=1\vee(i-n)} \binom{n+j}{i} C_j + \frac{(n+k)!}{n!(k-1)!}(-1)^k \sum_{l=0}^{i-k} \omega_l^{(n)}(-1)^l \binom{n-l}{n-i+k} B(l+1, i-l) \Big\}.$$

Then, (P_8) can be restated as:

(P_{10}) $t \mapsto \Omega_n(e^{it/n}) := \sum_{h=0}^n \omega_h^{(n)} e^{iht/n}$ *is the characteristic function of the initial* n*–segment of a sequence of* $(n+k)$ *exchangeable events if and only if the constants* $C_1, \ldots, C_k$ *can be determined in such a way that*

$$\begin{aligned} &0 \le \rho_i := (-1)^i \sum_{j=1\vee(i-n)}^{k} \binom{n+j}{i} C_j \qquad i = 0,\ldots,k-1 \\ &0 \le \rho_i := (-1)^i \Big\{ \sum_{j=1\vee(i-n)} \binom{n+j}{i} C_j \\ &\qquad + \frac{(n+k)!}{n!(k-1)!}(-1)^k \sum_{l=0}^{i-k} \omega_l^{(n)}(-1)^l \binom{n-l}{n-i+k} B(l+1, i-l) \Big\} \\ &\qquad i = k,\ldots,n+k. \end{aligned} \tag{16}$$

Moreover, if this system of linear inequalities is consistent, then for each of the solutions $(C_1, \ldots, C_k)$*, the vector* $(\rho_0, \ldots, \rho_{n+k})$ *represents an exchangeable assessment for* $(\omega_0^{(n+k)}, \ldots, \omega_{n+k}^{(n+k)})$*, consistent with the initial segment* $(\omega_0^{(n)}, \ldots, \omega_n^{(n)})$*.*

The research of conditions for consistency of systems like (16) originated a wealth of literature on the subject. Here, we propose a solution derived from [2]. In matrix form, (16) becomes $Ax \le b$ where

$$A = [a_{ij}]_{1\le i\le n+k+1, 1\le j\le k}, \quad b' = (b_1, \ldots, b_{n+k+1}), \quad x' = (C_1, \ldots, C_k),$$

with

$$\begin{aligned} &a_{ij} := (-1)^i \binom{n+j}{i-1}, \qquad b_1 = 0, \ldots, b_k = 0, \\ &b_i = (-1)^{i+k-1} \frac{(n+k)!}{n!(k-1)!} \sum_{l=0}^{i-1-k} (-1)^l \omega_l^{(n)} \binom{n-l}{n-i+k+1} B(l+1, i-1-l) \\ &\qquad i = k+1, \ldots, n+k+1. \end{aligned}$$

Since, as it is easy to show, the rank of A is k, Theorem 3 in [2] yields

(P_{11}) $\Omega_n(e^{it/n}) := \sum_{h=0}^n \omega_h^{(n)} e^{iht/n}$ *is the characteristic function of the initial* n*–segment of a sequence of* $(n+k)$ *exchangeable events if, and only if, there exist* $1 \le i_1 < i_2 < \cdots < i_k \le n+k+1$ *such that*

$$\begin{pmatrix} a_{i_1 1} & \ldots & a_{i_1 k} \\ \ldots & \ldots & \ldots \\ a_{i_k 1} & \ldots & a_{i_k k} \end{pmatrix}$$

has a nonvanishing determinant Δ*, and*

$$\frac{1}{\Delta}\, det \begin{pmatrix} a_{i_1 1} & \ldots & a_{i_1 k} & b_{i_1} \\ \ldots & \ldots & \ldots & \ldots \\ a_{i_k 1} & \ldots & a_{i_k k} & b_{i_k} \\ a_{j1} & \ldots & a_{jk} & b_j \end{pmatrix}$$

turns out to be nonnegative for every $j = 1, \ldots, n + k + 1$.

In the particular case of $k = 1$*, this necessary and sufficient condition reduces to require that* $\{\omega_h^{(n)} : h = 0, \ldots, n\}$ *satisfies*

$$\max\{\beta_i : \textit{ for any even integer} \leq n + 1\} \leq \min\{\beta_i : \textit{ for any odd integer} \leq n + 1\}$$

where

$$\beta_i = \sum_{l=0}^{i-1} (-1)^l B(l + 1, n - l + 1)\omega_l^{(n)} \qquad (i = 1, \ldots, n + 1).$$

In fact, this result could be obtained by direct inspection of (16) with $k = 1$.

To conclude, let us remark that (P_{11}) is susceptible of interesting geometrical interpretations that one can deduce directly from the above–mentioned Cernikov paper. It would be interesting to compare them with the geometrical arguments developed by [15] and other Authors, already mentioned in Subsection 2.6.

References

[1] G. Castelnuovo, Calcolo delle Probabilità, vol. 1, 2nd ed. (Bologna, Zanichelli, 1925).

[2] S. N. Cernikov, Systems of linear inequalities, American Mathematical Society Translations **26**, 11 (1963).

[3] L. Crisma, Alcune valutazioni quantitative interessanti la proseguibilità di processi aleatori scambiabili, Rendiconti Istituto di Matematica dell'Università di Trieste **3**, 96 (1971).

[4] L. Crisma, Quantitative analysis of exchangeability in alternative processes, 207–216 in Exchangeability in probability and statistics, G. Koch and F. Spizzichino, eds. (Amsterdam, North-Holland, 1982).

[5] B. de Finetti, Sulle funzioni a incremento aleatorio, Atti della Reale Accademia Nazionale dei Lincei, Rendiconti Serie VI **10**, 163 (1929).

[6] B. de Finetti, Funzione caratteristica di un fenomeno aleatorio, Atti Reale Accademia Nazionale dei Lincei, Memorie Serie VI **4**, 86 (1930).

[7] B. de Finetti, Sui passaggi al limite nel calcolo delle probabilità, Rendiconti Reale Istituto Lombardo di Scienze e Lettere **63**, 155 (1930).

[8] B. de Finetti, A proposito dell'estensione del teorema delle probabilità totali alle classi numerabili, Rendiconti Reale Istituto Lombardo di Scienze e Lettere **63**, 901 (1930).

[9] B. de Finetti, Ancora sull'estensione alle classi numerabili del teorema delle probabilità totali, Rendiconti del Reale Istituto Lombardo di Scienze e Lettere **63**, 1063 (1930).

[10] B. de Finetti, Sul significato soggettivo della probabilità, Fundamenta Mathematicae **17** ,298 (1931).

[11] B. de Finetti, Funzione caratteristica di un fenomeno aleatorio, Atti del Congresso Internazionale dei Matematici (Bologna, 3–10 Settembre, 1928) **6**, 179 (Zanichelli, Bologna, 1932).

[12] B. de Finetti, La prévision : ses lois logiques, ses sources subjectives, Annalles de l'Instut Henri Poincaré **7**, 1 (1937).

[13] B. de Finetti, Resoconto critico del colloquio di Ginevra intorno alla teoria delle probabilità Giornale dell'Istituto Italiano degli Attuari **9**, 1 (1938).

[14] B. de Finetti, Compte rendu critique du colloque de Génève sur la théorie des probabilités, in Actualités Scientifiques et Industrielles, vol. 766 (Hermann, Paris 1939).

[15] B. de Finetti, Sulla proseguibilità di processi aleatori scambiabili, Rendiconti dell'Istituto di Matematica dell' Università di Trieste **1**, 53 (1969).

[16] B. de Finetti, La probabilità: guardarsi dalle contraffazioni!, Scientia **111**, 255 (1976).

[17] P. Diaconis, Finite forms of de Finetti's theorem on exchangeability, Synthese **36**, 271 (1977).

[18] J. Haag, Sur un probléme général de probabilités et ses diverses applications, 659–674 in Proceedings of the International Congress of Mathematicians (Toronto, August 11-26, 1924) vol.1 (University of Toronto Press, 1928).

[19] E. Hewitt and L. J. Savage, Symmetric measures on Cartesian products, Transactions of the American Mathematical Society **80**, 470 (1955).

[20] A. Khinchin, Sur les classes d'événements équivalentes, Matematiceskij Sbornik **39**, 40 (1932).

[21] A. Khinchin, Remarques sur les suites d'événements obéssants á la loi des grands nombres, Matematiceskij Sbornik **39**, 115 (1932).
[22] P. Lévy, Calcul des Probabilités (Gauthier–Villars, Paris, 1925).
[23] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, Integrals and series, Vol. 1 (Gordon & Breach Science Publishers, New York, 1986).
[24] E. Regazzini, De Finetti's contribution to the theory of random functions (to appear in International Journal of Approximate Reasoning, 2007).
[25] J. A. Shohat and J. D. Tamarkin, The Problem of Moments (American Mathematical Society, Providence, 1943).
[26] R. von Mises , Fundamentalsaetze der wahrscheinlick Keitsrechnung, Mathematische Zeitschrift **4**, 1 (1919).
[27] G. R. Wood, Binomial mixtures and finite exchangeability, Annals of probability **20**, 1167 (1992).