Journ@l Electronique d'Histoire des Probabilités et de la Statistique

Electronic Journ@l for History of Probability and Statistics

Vol 1, n°1; Mars 2005

Probability and statistics in Italy during the First World War
I: Cantelli and the laws of large numbers

Eugenio REGAZZINI*

Abstract

Some of the most important Italian contributions to probability and statistics came to light in the WW1 years 1915-1918: the Cantelli's theory of stochastic convergence together with the first correct and (rather) general version of the strong law of large numbers, and the Gini theory of statistical association. These studies were influential in determining the characteristics of the subsequent developments of the statistical and probabilistic research in Italy.

The present paper is the first part of an article that aims at describing the spirit of these studies, and helping readers understand their connections both with pre-existent and contemporary literature. It is devoted to the contributions of Cantelli. The (future) second part will deal with statistics and, in particular, with Gini's works on statistical association.

Résumé

Quelques unes des plus importantes contributions italiennes aux probabilités et statistiques apparurent pendant les années de guerre 1915-1918: la théorie de Cantelli sur la convergence stochastique et la première version correcte et assez générale de la loi forte des grands nombres, ainsi que la théorie de Gini d'association statistique. Ces études furent déterminantes pour caractériser les développements ultérieurs des recherches en probabilité et statistique en Italie.

Cet article est la première partie d'un texte qui a pour but de décrire l'esprit de ces études, et d'aider les lecteurs à comprendre leurs connexions avec celles qui les ont précédées. Elle est consacrée aux contributions de Cantelli. La (future) deuxième partie traitera des statistiques, en particulier des travaux de Gini sur l'association statistique.

1 A few biographical notes

As to probability and statistics, the Italian panorama during 1910-1930 was dominated by Francesco Paolo Cantelli and Corrado Gini.

Cantelli was born in Palermo on December 20, 1875 and graduated from the town University in Pure Mathematics in the year 1899 after discussing a thesis on celestial mechanics. In his initial research, at the Astronomic Observatory in Palermo, he tried to verify if the position of the stars mentioned by Dante in the *Divina Commedia* corresponded either to the year 1300 or to the year

*Università degli Studi di Pavia, Dipartimento di Matematica "F. Casorati", Via Ferrata 1, 27100 Pavia (Italy). E-mail: eugenio@dimat.unipv.it

1301. Cantelli's investigation confirmed the hypotesis favouring the year 1301. He went deeper into the study of probability during his stay, from 1903 to 1923, at the Istituti di Previdenza where he had been employed as an actuary. In 1922, he qualified to undertake university teaching of Calculus of Probabilities and its Applications. In the year 1923 he was appointed professor of Actuarial Mathematics at the University of Catania. After a brief permanence in this town and a longer stay in Naples, in 1931 he moved to Rome where he lived till his death on July 21, 1966.

Among the several initiatives in which he took part with the aim of developing the organization of scientific research in the field of applied mathematics, it should be mentioned the institution of the *Istituto Italiano degli Attuari* and of the *Giornale dell'Istituto Italiano degli Attuari* (GIIA) which Cantelli edited from 1930 to 1958. During this period the *Giornale* was one of the most prestigious journals dealing with probability, statistics and actuarial mathematics, involving the collaboration of the most famous Italian and foreign scholars in these fields. Cantelli was also working as an actuary. He has been, among other things, an actuary of the pension board of the Society of Nations in Geneva.

Cantelli was the first "modern" Italian probabilist. His name is definitively linked with some fundamental results about the convergence of sequences of random variables, published during the years of the First World War. In the present paper we will confine ourselves to considering these papers; further information on the Cantelli scientific activity can be found in [Ottaviani 1966, 1967], [Regazzini 1987] and [Benzi 1988].

2 Cantelli and the laws of large numbers (1916-1917)

In this section we describe and discuss the content of four papers which represent the core of Cantelli's contribution to stochastic convergence and, in particular, to the formulation of weak and strong laws of large numbers. Their titles are: *La tendenza ad un limite nel senso del calcolo delle probabilità*[1], *Sulla legge dei grandi numeri* [2], *Sulla probabilità come limite della frequenza* [3], *Su due applicazioni di un teorema di G. Boole alla statistica matematica* [4]. These papers were published between 1916 and 1917, and from now on they will be mentioned by $C_I, C_{II}, C_{III}, C_{IV}$, respectively.

Before getting to the root of the matter, it should be recalled that Cantelli was fully aware of the necessity of a formal definition of probability which could be taken "to attain greater precision in a conceptual discussion at least for those aspects of the theory that are connected with the formal properties" (de Finetti, 1949). Unfortunately, at the time of the above-mentioned papers, he had no construction of this type at his disposal. On this subject he formulated an *abstract theory* of probability shortly before the publication of Kolmogorov's *Grundbegriffe*; see [Cantelli 1932]. So, as we will see soon, he was not in a position to deal with random variables as measurable functions and, moreover, considered, implicity, probability as a completely additive function on a family of

[1]Convergence to a limit in the sense of the calculus of probabilities
[2]On the law of large numbers
[3]On probability seen as a limit of frequencies
[4]On two applications of a Boole's theorem to mathematical statistics

events, without emphasizing the role of such a hypothesis in restricting the class of the admissible probability assessments. In spite of these aspects, the Cantelli treatment of stochastic convergence turns out to be exceptionally transparent and quite satisfactory even with respect to the present expositions of the subject.

2.1 Definition of convergence in probability

As a matter of fact, C_I starts with a definition of random variable. This definition, to be exploited in all its generality, would require the systematic use of the Stieltjes integral. It seems that Cantelli was unaware of this concept, since - in spite of the generality of his definition - he considers discrete and (absolutely) continuous distributions separately. He defines a *random variable* as a quantity which, for any interval $[x, y]$, can take values in it or in its complement $[x, y]^C$[5], according to whether a suitable event $E_{[x,y]}$ comes true or not (i.e. $E_{[x,y]^C}$ comes true). Cantelli completes his definition by requiring an assessment of probability for the entire class of events $\{E_{[x,y]}, E_{[x,y]^C} : -\infty < x < y < +\infty\}$. Any assessment of this type determines the probability distribution of a random variable at issue: denoting this random variable by X, the probability that X belongs to $[x, y]$ must be the same as the probability of $E_{[x,y]}$, and the probability that X belongs to $[x, y]^C$ is the complement to one of the probability of $E_{[x,y]}$. By resorting to the notion of interval of $\mathbb{R}^k$, he mimics the above argument to provide a satisfactory definition of *random vector* (*system of random variables* in the Cantelli language).

Having introduced these basic concepts, Cantelli continues with the definition of a sequence of random variables, say $(X_n)_{n\geq1}$, which converges, *in the sense of the calculus of probabilities*, to a number N. Such a convergence is characterized by the fact that, for every $\eta > 0$, the probability that X_n belongs to $[N - \eta, N + \eta]$ goes to one, as $n \to +\infty$. In the language of measure theory (the same as the language adopted by Cantelli for his 1932 abstract theory), this convergence is the same as *convergence in measure* and, more specifically, as *convergence in probability.* Subsequently, he provides complete proofs of the following facts:

(F_1) The limit of a sequence which converges in probability is unique.

(F_2) Any subsequence of a convergent sequence in probability must converge to the same limit of the sequence.

(F_3) If $(X_n)_{n\geq1}$ converges in probability to N and, for each n, there exists an interval $[N - \eta_n, N + \eta_n]$ such that the limit superior of the probability of $\{X_n \in [N - \eta_n, N + \eta_n]\}$ does not exceed $c < 1$, then $\eta_n \to 0$, as $n \to +\infty$.

(F_4) If f is a real-valued function, defined on a subset of $\mathbb{R}$ which includes the range of every $X_n, (X_n)_{n\geq1}$ being a sequence of random variables which converges in probability to the real number N, then $(f(X_n))_{n\geq1}$ converges in probability to $f(N)$, provided that f is continuous at N.

Apropos of the convergence in probability of two sequences of random variables, say $(X_n)_{n\geq1}$ and $(Y_n)_{n\geq1}$, the concept of random system enabled Cantelli

[5]From now on, given any subset A of a space S, by A^C we will designate the complement of A with respect to S.

to prove that these sequences converge in probability to X and Y respectively, if and only if the probability of the event $\{X - \varepsilon \leq X_n \leq X + \varepsilon\} \bigcap \{Y - \eta \leq Y_m \leq Y + \eta\}$, assessed according to the law of $((X, Y, X_n, Y_m))_{n,m}$, goes to one as n and m diverge to infinity, for every positive ε and η.

2.2 Convergence in probability of random means

Cantelli concludes C_I with an application of the concept of stochastic convergence made precise in that paper. More precisely, he provides a sharp formulation of the classical Bernoulli (weak) law of large numbers. More general forms of weak laws of large numbers are proved in C_{II}. This paper is ideally split into three parts. The first one extends the method of Bienaymé-Tchebycheff, in order to determine bounds for the probability that a random variable belongs to a given interval, and is related to a previous paper of our Author. See [Cantelli 1910] and, for a systematic treatment, chapter IV of [Fréchet 1937]. The second part applies these bounds to sums of random variables, after tackling the problem of computing the second order moment of a sum of random variables. Finally, the last part contains various forms of laws of large numbers and a careful analysis of their innovative contribution with respect to classical works of the Russian School and, in particular, of [Markov 1906]. In fact, Cantelli focuses on "general" sequences of random variables: $X_1, X_2, \dots$ which needn't be stochastically independent. On the other hand, he assumes that they have finite second moments, and writes the variance of $n^{-1}(X_1 + \cdots + X_n)$ as

$$\mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{1}{n^2}\sum_{1}^{n} \mathrm{Var}(X_i) + \left(1 - \frac{1}{n}\right) C(n) \tag{1}$$

with

$$C(n) = \frac{2}{n(n-1)} \sum_{1 \leq j < i \leq n} \mathrm{Cov}(X_i, X_j).$$

At this stage, he is in a position to prove the following form of *weak law of large numbers*:

(WL) If $n^{-1}(E(X_1) + \cdots + E(X_n))$ converges to M as $n \to \infty$, and

$$\mathrm{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) \to 0 \qquad \text{as } n \to \infty,$$

then

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - M\right| > \varepsilon\right\} \to 0 \qquad \text{as } n \to \infty$$

for every $\varepsilon > 0$.

To prove this, Cantelli puts $M_i = E(X_i),\ i = 1, \dots, n$, and starts from the Bienaymé-Tchebycheff inequality to write

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \frac{M_1 + \cdots + M_n}{n}\right| \leq \frac{\lambda}{n}\sqrt{\mathrm{Var}(X_1 + \cdots + X_n)}\right\} \geq 1 - \frac{1}{\lambda^2}.$$

Then, by the triangle inequality,

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - M\right| \leq \frac{\lambda}{n}\sqrt{\mathrm{Var}(X_1 + \cdots + X_n)} + \left|\frac{M_1 + \cdots + M_n}{n} - M\right|\right\} \geq 1 - \frac{1}{\lambda^2}$$

and, in view of the assumptions made, for every $\varepsilon > 0$ there is $n^* = n^*(\varepsilon, \lambda)$ such that $\lambda n^{-1}\sqrt{\mathrm{Var}(X_1 + \cdots + X_n)} + \left|\frac{M_1+\cdots+M_n}{n} - M\right| < \varepsilon$ holds for every $n \geq n^*$. So, for any positive numbers ε, λ and any integer $n \geq n^*$, one has

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - M\right| \leq \varepsilon\right\} \geq 1 - \frac{1}{\lambda^2}.$$

Cantelli points out that the condition for (WL) is satisfied when

$$C(n) \leq 0 \qquad \text{for every } n, \text{ and } \sum_1^n \frac{\mathrm{Var}(X_i)}{n^2} \to 0 \text{ as } n \to \infty \tag{2}$$

is valid, and provides an example of a sequence of nonindependent random variables $X_1, X_2, \ldots$ satisfying $C(n) = 0$ for every n.

At this point, our Author benefits of an inequality, proved in the first part of the paper, to establish the weak law of large numbers through the behavior of moments of order k, about the terms of some suitable sequence of real numbers, of $n^{-1}(X_1 + \cdots + X_n)$, when k is a positive number. Thus, given a positive real number k, he defines

$$\sigma_k[m_k(n)] = E\left(\left|\frac{X_1 + \cdots + X_n}{n} - m_k(n)\right|^k\right)$$

where $(m_k(n))_{n\geq 1}$ is a real-valued sequence. The inequality mentioned above reads as follows

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - m_k(n)\right| \leq \lambda \sqrt[k]{\sigma_k[m_k(n)]}\right\} \geq 1 - \frac{1}{\lambda^k} \qquad (\lambda > 0) \tag{3}$$

and is nowadays explained in any probability text, sometimes referred to Markov inequality. Then arguing as in proof of *(WL)*, Cantelli obtains the following proposition:

(WL)$_1$ If $\sigma_k[m_k(n)] \to 0$ and if $(m_k(n))_{n\geq 1}$ converges, as $n \to \infty$, then

$$P\left\{\left|\frac{X_1 + \cdots + X_n}{n} - \lim_{n\to\infty} m_k(n)\right| > \varepsilon\right\} \to 0 \qquad \text{as } n \to \infty$$

for every $\varepsilon > 0$.

Cantelli designates the assumptions made in *(WL)*$_1$ by $S^{(k)}_{m_k}$, versus the conditions assumed in *(WL)*, that he indicates by S_{M_2} with $M_2 = M_2(n) := n^{-1}(E(X_1) + \cdots + E(X_n))$. The following propositions explain some mutual relations between these groups of assumptions.

(F_5) If $(X_n)_{n\geq 1}$ satisfies $S^{(k)}_{m_k}$, then it meets $S^{(k)}_{m'_k}$ for any sequence $(m'_k(n))_{n\geq 1}$ satisfying $\sigma_k[m'_k(n)] \leq \sigma_k[m_k(n)]$ for every n and, moreover,

$$\lim_{n\to\infty} m_k(n) = \lim_{n\to\infty} m'_k(n).$$

(F_6) If $(X_n)_{n\geq 1}$ satisfies $S^{(k)}_{m_k}$ for some $k > 2$, then it also satisfies S_{M_2}, and

$$\lim_{n\to\infty} m_k(n) = \lim_{n\to\infty} M_2(n). \tag{4}$$

(F_7) If $(X_n)_{n\geq 1}$ meets $S^{(k)}_{m_k}$ for some positive $k > 2$ and there is some $\delta > 0$ for which $\sigma_{2+\delta}[m_k(n)]$ is finite, then $(X_n)_{n\geq 1}$ satisfies S_{M_2} and (4) comes true.

(F_8) If $(X_n)_{n\geq 1}$ obeys S_{M_2}, and if $\sigma_{k+\delta}[M_2(n)]$ is finite for some $k > 2$ and $\delta > 0$, then $(X_n)_{n\geq 1}$ must satisfy $S^{(k)}_{M_2(n)}$.

(F_9) If $0 < k < 2$ and $(X_n)_{n\geq 1}$ agrees with S_{M_2}, then it meets $S^{(k)}_{M_2(n)}$.

As an example, our Author considers a sequence $(X_n)_{n\geq 1}$ which obeys S_{M_2} with $n^{-1}(E(X_1) + \cdots + E(X_n)) \to M$. Then, from ($F_5$), it satisfies $S^{(2)}_{A(n)}$ whenever $A(n)$ is any median of $n^{-1}(X_1 + \cdots + X_n)$, i.e.

$$E\left(\left|\frac{X_1 + \cdots + X_n}{n} - A(n)\right|^2\right) \to 0 \qquad A(n) \to M \qquad (\text{as } n \to \infty)$$

and, by virtue of (F_9),

$$E\left(\left|\frac{X_1 + \cdots + X_n}{n} - A(n)\right|\right) \to 0, \qquad (\text{as } n \to \infty).$$

On the other hand, if $\sigma_{2+\delta}[A(n)]$ is finite for every n and for some $\delta > 0$, and if

$$\begin{aligned} \lim_{n\to\infty} A(n) &= A \\ \lim_{n\to\infty} E\left(\left|\frac{X_1 + \cdots + X_n}{n} - A(n)\right|\right) &= 0, \end{aligned}$$

then, by (F_7), $(X_n)_{n\geq 1}$ satisfies S_{M_2} together with $M_2(n) \to A$. Moreover, by (F_5),

$$E\left(\left|\frac{X_1 + \cdots + X_n}{n} - A(n)\right|^2\right) \to 0 \qquad (\text{as } n \to \infty).$$

As for the proofs of these propositions, we confine ourselves to mentioning that Cantelli provides an independent proof only for (F_5). His argument can be summarized as follows. It is immediate to obtain $\sigma_k[m'_k(n)] \to 0$ (as $n \to \infty$), so that it remains to show that $(m'_k(n))_{n\geq 1}$ converges to the same limit as $(m_k(n))_{n\geq 1}$. Cantelli proves this fact by a reductio ad absurdum argument, through the inequality mentioned immediately before the statement of *(WL)*$_1$. He completes the proofs of $(F_6) - (F_9)$ by resorting to (F_5) and to the following Liapunoff inequality

$$[E(X^{k_2})]^{k_1-k_3} < [E(X^{k_3})]^{k_1-k_2}[E(X^{k_1})]^{k_2-k_3}$$

which holds for $0 \leq k_3 < k_2 < k_1$ and any X such that $P\{X \geq 0\} = 1$ and $P\{X > 0\} > 0$. See [Liapounoff 1901].

2.3 Uniform convergence in probability of random means

C_{III} is one of the most important Cantelli's papers. In point of fact, it contains the formulation and the complete proof of his celebrated *strong law of large numbers*. As the title points out, the work aims at proving that, under suitable conditions, the probability, that the sequence of the frequencies of success take values which converge (in the usual sense) to some specific limit, is one. Cantelli presents this result for sequences of *Bernoulli trials* (sequences of events which are stochastically independent and have a constant probability) as a corollary of a more general statement, proved through the usual inequality (3) - with $k = 4$ - and a sagacious resort to the Boole theorem according to which, if $E_1, E_2, \dots$ are events, then

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) \geq 1 - (P(E_1^C) + \cdots + P(E_n^C)). \tag{5}$$

Cantelli notices that the sequence $(P(E_1 \cap E_2 \cap \cdots \cap E_n))_{n \geq 1}$ is decreasing (and bounded), so that it converges as $n \to \infty$. At this point he assumes that this limit, say ℓ, represents the probability of $E_1 \cap E_2 \cap \cdots$, on the basis of the supposition that the assumption at issue "is not open to objections of a theoretical nature and agrees with the idea we get about probability on the basis of arguments of an empirical nature". As a matter of fact, this is a delicate passage since the Cantelli assumption is tantamount to assuming that complete additivity must be a necessary property of any probability law. For a sharp discussion of this aspect of the Cantelli attitude, explained by means of instructive examples, see [de Finetti 1930a] and the consequent debate between [de Finetti 1930b,c] and [Fréchet 1930a,b].

So, going on with the Cantelli argument, assume that

$$\ell = P(E_1 \cap E_2 \cap \cdots) \geq 1 - (P(E_1^C) + \cdots + P(E_n^C) + \cdots) \tag{6}$$

holds for any sequence of events: $E_1, E_2, \dots$. Now, we are interested in events concerning means of random variables

$$\overline{S}_{(n)} := \frac{X_1 + \cdots + X_n}{n} \qquad (n = 1, 2, \dots)$$

when $X_1, X_2, \dots$ are *stochastically independent*, have finite fourth moment and

$$M_{(n)} := E(\overline{S}_{(n)}) \to M \qquad (\text{as } n \to \infty),$$

is valid for some real number M.

Then,

$$\begin{aligned} \mathrm{Var}(\overline{S}_{(n)}) &= \frac{1}{n^2}\{\mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)\} \\ E[(\overline{S}_{(n)} - M_{(n)})^4] &= \frac{1}{n^4}\left\{\sum_{i=1}^{n} E[(X_i - E(X_i))^4] + 6 \sum_{1 \leq i < j \leq n} \mathrm{Var}(X_i)\mathrm{Var}(X_j)\right\} \end{aligned}$$

and, for any sequence of positive numbers $(\alpha_n)_{n \geq 1}$, one has

$$\begin{aligned} & P\left(\bigcap_{m \geq n} \{|\overline{S}_{(m)} - M| \leq \alpha_m + |M_{(m)} - M|\}\right) \\ \geq\ & P\left(\bigcap_{m \geq n} \{|\overline{S}_{(m)} - M_{(m)}| + |M_{(m)} - M| \leq \alpha_m + |M_{(m)} - M|\}\right) \\ \geq\ & 1 - \sum_{m \geq n} P\left\{\overline{S}_{(m)} - M_{(m)}| > \alpha_m\right\} \\ =\ & 1 - \sum_{m \geq n} \frac{1}{\alpha_m^4} E[|\overline{S}_{(m)} - M_{(m)}|^4]. \end{aligned}$$

Hence, setting

$$\alpha_n := \lambda_n \sqrt[4]{E[|\overline{S}_{(n)} - M_{(n)}|^4]} \qquad (\lambda_n > 0, n = 1, 2, \cdots),$$

we get

$$\mathcal{P}_n := P\left(\bigcap_{m\geq n} \{|\overline{S}_{(m)} - M| \leq \alpha_m + |M_{(m)} - M|\}\right) \geq 1 - \sum_{m\geq n} \frac{1}{\lambda_m^4}$$

with

$$1-\sum_{m\geq n} \frac{1}{\lambda_m^4} = 1-\sum_{m\geq n} \frac{1}{\alpha_m^4}\frac{1}{m^4}\left\{\sum_{i=1}^{m} E[(X_i - E(X_i))^4] + 6 \sum_{1\leq i<j\leq m} \mathrm{Var}(X_i)\mathrm{Var}(X_j)\right\}.$$

So, if there are positive numbers A, B (independent of n) such that

$$\frac{1}{n}\sum_{i=1}^{n} E[(X_i - E(X_i))^4] \leq A \text{ and } \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \mathrm{Var}(X_i)\mathrm{Var}(X_j) \leq B \qquad (7)$$

hold for every n, we obtain

$$\mathcal{P}_n \geq 1 - \sum_{m\geq n} \frac{1}{\alpha_m^4}\left\{\frac{A}{m^3} + 3\frac{m-1}{m^3}B\right\} = 1 - \sum_{k\geq 0} \frac{1}{\alpha_{n+k}^4 (n+k)^3}\{A + 3(n+k-1)B\}.$$

Thus, by taking

$$\frac{1}{\alpha_{n+k}} = \frac{(n+k)^{\frac{1}{4}(1-\xi)}}{\delta}$$

for every n, k, for some $\delta > 0$ and ξ in $(0, 1)$, one can write

$$\begin{aligned} \mathcal{P}_n &\geq 1 - \frac{1}{\delta^4}\sum_{k\geq 0}\left\{\frac{A}{(n+k)^{2+\xi}} + 3\frac{B}{(n+k)^{1+\xi}}\right\} \\ &\geq 1 - \frac{1}{\delta^4}\left\{\frac{A}{(1+\xi)(n-1)^{1+\xi}} + 3\frac{B}{\xi(n-1)^{\xi}}\right\}. \end{aligned}$$

This is an accurate reproduction of the Cantelli argument to prove main result in C_{III}:

(SL) Let $X_1, X_2, \ldots$ be independent random variables with finite 4^{th} moment, and such that:

$$(i) \qquad \frac{1}{n}(E(X_1) + \cdots + E(X_n)) \to M \qquad \text{as } n \to \infty$$

$$(ii) \qquad \frac{1}{n}\sum_{i=1}^{n} E[(X_i - E(X_i))^4] \leq A \qquad \text{and}$$

$$\frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} \mathrm{Var}(X_i)\mathrm{Var}(X_j) \leq B$$

for every n and for some constants A, B independent of n. Then, for every $\delta > 0$ and ξ in $(0, 1)$:

$$P\left(\bigcap_{m\geq n}\left\{\left|\overline{S}_{(m)} - M\right| \leq \frac{\delta}{m^{(1-\xi)/4}} + \left|\frac{E(X_1) + \cdots + E(X_m)}{m} - M\right|\right\}\right)$$

$$\geq 1 - \frac{1}{\delta^4} \cdot O\left(\frac{1}{n^{\xi}}\right) \qquad (n \to \infty).$$

Conditions $(i), (ii)$ are satisfied, in particular, if besides being independent the X_n's are identically distributed. All these conditions are satisfied, for example, by Bernoulli sequences.

(SL)$_1$ Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with finite 4^{th} moment. Then, as $n \to \infty$,

$$P\left(\bigcap_{m\geq n}\left\{\left|\overline{S}_{(m)} - E(X_1)\right| \leq \frac{\delta}{m^{(1-\xi)/4}}\right\}\right) \geq 1 - \frac{1}{\delta^4}\cdot O\left(\frac{1}{n^\xi}\right)$$

holds for every $\delta > 0$ and ξ in $(0, 1)$.

According to Cantelli's terminology, these theorems assert the *uniform convergence in probability* of the means $n^{-1}(X_1 + \cdots + X_n)$ to the limit M of $n^{-1}(E(X_1) + \cdots + E(X_n))$.

For every $\delta, \eta > 0$ and ξ in $(0, 1)$, there is $\overline{n} = \overline{n}(\delta, \eta, \xi)$ such that the probability that

$$\frac{1}{m}\left|\sum_{i=1}^{m}(X_i - E(X_i))\right| \leq \frac{\delta}{m^{(1-\xi)/4}} + \left|M - \frac{E(X_1) + \cdots + E(X_m)}{m}\right|$$

holds for every $m \geq \overline{n}$ turns out to be greater than $(1 - \delta)$.

So, there is a (deterministic) infinitesimal sequence which, with a probability that goes to one as $n \to \infty$, "dominates" the sequence of random means $(m^{-1}\sum_{j=1}^{n}(X_j - M_j))_{m\geq n}$.

The term "strong law of large numbers" was coined later, to designate this very same property of sequences of means of random variables, by Khinchine, and has superseded the term "uniform law of large numbers" preferred by Cantelli.

It is well-known that in his studies about normal numbers, Borel formulated in 1909 a proposition that may remind us of the uniform convergence of Bernoulli frequencies, i.e. of the particular case of *(SL)*$_1$ in which $X_1, X_2, \ldots$ are indicators of *independent* events $E_1, E_2, \ldots$ with constant probability p, namely $P(E_1) = P(E_2) = \cdots = p$. See [Borel 1909]. Under the circumstances, *(SL)*$_1$ states that the sequence of the frequencies of success $(\sum_{k=1}^{n} X_k/n)_{n\geq 1}$ converges uniformly to p in probability. As a matter of fact, the original Borel statement does not cover the case of the Bernoulli sequences either, since it is based on the assumption that frequencies are independent. In spite of this, many Authors, even if they acknowledge that Cantelli was the first to formulate and prove, in a simple and rigorous way, a general strong law of large numbers, are inclined to give the paternity of the uniform convergence of Bernoulli frequencies to Borel. For example, on page 216 of [Fréchet 1937], apropos of the concept of almost sure convergence, one reads:

> "M. Borel n'a pas fait observer explitement une conséquence de ce résultat [6]qui est absolument immediate, mais extrêmement importante. C'est celle qu'on obtient en interprétant p comme la probabilité supposée constante d'un événement fortuit de nature quelconque autre que le choix d'un nombre. Rien ne sera changé aux

[6]the law of normal numbers

raisonnements. Mais le raisonnement obtenu par ce simple changement de mots apporte au théorème de Bernoulli[7] un complément d'importance capitale."

It should also be added that the proof provided by Borel makes use of an asymptotic evaluation of the Gauss' integral which was not secure since the conditions for the validity of this evaluation were not yet known. Apropos of this point, Cantelli points out, in C_{III}, that if one majorized $P\{|\xi_{(m)} - M_{(m)}| > \alpha_m\}$ either by the usual Bienaymé-Tchebycheff bound, or by the Gauss integral combined with the bound provided in 1901 by Liapounoff, the resulting series to be used in the proof of *(SL)* were divergent.

Contemporary Authors seem to be unaware of these distinctions and tend to couple the names of Borel and Cantelli, so that we deem it useful to continue with the words used by Fréchet to conclude his introduction to the proof of the uniform convergence of frequencies:

> "C'est à M. Cantelli [8] que revient le mérite d'avoir plus tard démontré le premier un théorème plus général exprimant ce que M. Khintchine a plus tard appelé la "loi forte" des grands nombres [...]. La démonstration de M. Cantelli est simple, complète et rigoureuse." Cf. [Fréchet, 1937], page 217.

For the sake of completeness, let us recall that Borel mentions, in his 1909 fundamental paper, that his law of normal numbers might be proved in a more rigorous way than the one actually followed by himself, by resorting to arguments of a "geometrical" nature. In fact, his advise was followed by Authors such as [Faber, 1910] and [Hausdorff, 1914] who, by simple measure-theoretical reasoning on the interval $(0, 1)$, provide unexceptionable proofs of the Borel law for the sole dyadic expansion. This circumstance has induced some Authors to raise new doubts about the priority in the formulation of the strong law of large numbers. See, for example, [Barone and Novikoff, 1977, 1977-78] and Historical and Biographical Notes in [Kallenberg, 1997]. In this connection, it must be stressed that:

(a) Neither Faber nor Hausdorff make explicit the necessary relationship between the probabilistic problem associated to sequences of Bernoulli trials and the "geometrical" assumptions made to deal with the problem of normal numbers.

(b) This relationship has been analyzed subsequently by Cantelli himself [Cantelli, 1932].

(c) Even supposing that Faber and Hausdorff had thought it superfluous to point out the above relationship, it is difficult to see how their methods may be used directly to deal with the general case studied by Cantelli.

[7] the weak law of large numbers for indicators of independent events with constant probability

[8] C_{III}

2.4 Further remarks on the application of the Boole theorem

The proof of *(SL)* rests on a combination of inequality (3) - a special case of a more general result proved in C_{II} - with the extended form of the Boole theorem (5)-(6). Apropos of this, it must be recalled that this use of the Boole theorem is at the core of the celebrated *Borel-Cantelli lemma*. In particular, that part of the lemma which deals with general sequences of events is due to Cantelli. Moreover, the Cantelli reasoning, by means of the Boole theorem, allows us to get a more precise (and complete) formulation of the original proposition stated by Borel for sequences of independent events. These aspects are explained in a clear form on pages 26-27 of [Fréchet, 1937].

Apart from the application of the Boole theorem to prove the uniform convergence in probability of random means, Cantelli points out an interesting application - mentioned in C_{IV} - of this very same theorem to the theory of risk, concerning the field of the scientific and professional interests of Cantelli pertaining actuarial techniques. Let $X(0,t)$ represent the random (positive or negative) gain of an insurance company relative to the period $(0,t]$ with $t = 1, 2, \dots, n$, actualized to the age 0. Then, if G stands for the initial capital, it is of interest to evaluate the probability of the event

$$\{X(0,1) \geq -G, X(0,2) \geq -G, \dots, X(0,n) \geq -G\}.$$

In view of the Boole theorem one can get a lower bound of this probability. Cantelli describes such a bound when $X(0,r)$ has, for every r, the Gaussian distribution with parameters m_r and σ_r^2. This assumption is consistent with the fact that each $X(0,r)$ can be viewed as the sum of a great number of random variables with finite variance, namely the gains from each of the policies managed by the insurance company. It is interesting noticing that, starting from this problem and from the Cantelli paper C_{IV}, Ottaviani - a Cantelli pupil - formulated his remarkable inequality. See [Ottaviani, 1940]. Nowadays, this inequality is also referred to as *Ottaviani-Skorohod inequality.*

References

[Barone, J. and Novikoff, A. 1977] The Borel law of normal numbers, the Borel zero-one law and the work of Van Vleck. *Historia Mathematica* **4** 43-65

[Barone, J. and Novikoff, A. 1977-1978] A history of the axiomatic formulation of probability from Borel to Kolmogorov: Part I. *Archive for History of Exact Sciences* **18** 123-190

[Benzi, M. 1988] Un probabilista neoclassico: F. P. Cantelli. *Historia Mathematica* **14** 614-634

[Borel, E. 1909]. Les probabilités dénombrables et leur applications arithmétiques. *Rendiconti del Circolo Matematico di Palermo* **27** 247-271

[Cantelli, F. P. 1910]. Intorno ad un teorema fondamentale della teoria del rischio. *Bollettino dell'Associazioni degli Attuari Italiani* 1-23

[Cantelli, F. P. 1916a]. La tendenza a un limite nel senso del calcolo delle probabilità. *Rendiconti del Circolo Matematico di Palermo* **41** 191-201

[Cantelli, F. P. 1916b]. Sulla legge dei grandi numeri. *Atti Reale Accademia Nazionale Lincei*, Memorie Cl. Sc. Fis. **11** 329-350

[Cantelli, F. P. 1917a] Sulla probabilità come limite della frequenza. *Atti Reale Accademia Nazionale Lincei* **26** 39-45
[Cantelli, F. P. 1917b] Su due applicazioni di un teorema di G. Boole alla statistica matematica. *Atti Reale Accademia Nazionale Lincei* **26** 295-302
[Cantelli, F. P. 1932] Una teoria astratta del calcolo delle probabilità. *Giornale Istituto Italiano degli Attuari* **3** 257-265
[Castellano, V 1965] Corrado Gini: a memoir. *Metron* **24** 1-84
[de Finetti, B. 1930a] Sui passaggi al limite nel calcolo delle probabilità. *Rendiconti R. Istituto Lombardo di Scienze e Lettere* **63** 1063-1069
[de Finetti, B. 1930b] A proposito dell'estensione del teorema delle probabilità totali alle classi numerabili. *Rendiconti R. Istituto Lombardo di Scienze e Lettere* **63** 901-905
[de Finetti, B. 1930c] Ancora sull'estensione alle classi numerabili del teorema delle probabilità totali. *Rendiconti R. Istituto Lombardo di Scienze e Lettere* **63** 1063-1069
[de Finetti, B. 1949] Sull'impostazione assiomatica del calcolo delle probabilità. Aggiunta alla nota sull'assiomatica della probabilità (two articles). *Annali Triestini* **19** sez. II 29-81; **20** sez. II 3-20
[Fréchet, M. 1930a] Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements. *Rendiconti R. Istituto Lombardo di Scienze e Lettere* **63** 899-900
[Fréchet, M. 1930b] Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements. *Rendiconti R. Istituto Lombardo di Scienze e Lettere* **63** 1059-1062
[Fréchet, M. 1937] *Généralités sur les Probabilités. Variables Aléatoires.* Gauthier-Villars, Paris
[Gini, C. 1908]*Il sesso dal punto di vista statistico.* Sandrom, Milano
[Hausdorff, F. 1914] *Grundzüge der Mengenlehre.* Verlag von Veit & Comp., Leipzig
[Kallenberg, O. 1997] *Foundations of Modern Probability.* Springer, New York
[Kolmogorov, A. N. 1933] *Grundbegriffe des Wahrscheinlichteitsrechung*, Ergentisse der Mathematik. Springer, Berlin.
[Liapounoff, A. M. 1901] Nouvelle forme du théorème sur la limite des probabilités. *Mémoires Acad. des Sciences de St. Pétersbourg* **12** 1-24
[Markov, A. A. 1906] Extension de la loi des grands nombres aux événements dépendants les uns des autres. *Bulletin de la Soc. Physico-Mathématique de Kazan* S 2 **15** 135-156
[Ottaviani, G. 1940] La teoria del rischio del Lundberg e il suo legame con la teoria classica del rischio. *Giornale Istituto Italiano degli Attuari* **11** 163-189
[Ottaviani, G. 1966] Francesco Paolo Cantelli. *Giornale Istituto Italiano degli Attuari* **29** 179-190
[Ottaviani, G. 1967] In memoria di Francesco Paolo Cantelli. *Metron* **26** 1-11
[Regazzini, E. 1987] Probability theory in Italy between the two world wars. A brief historical review. *Metron* **45** 5-42