Strong Approximations in Probability and Statistics

by M. CSÖRGŐ and P. RÉVÉSZ

AKADÉMIAI KIADÓ, BUDAPEST

Strong Approximations in Probability and Statistics

M. Csörgő and P. Révész

Akadémiai Kiadó · Budapest 1981

1989 -04- 1 0 1999 -05- 3 1
1992 -06- 2 2

Copyright © 1981 by Akadémiai Kiadó, Budapest
All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the publishers.

Joint edition published by Akadémiai Kiadó, Budapest and Academic Press, New York—San Francisco—London

Budapesti Műszaki Egyetem
Közlekedésmérnöki Kar
Mechanika T.
KÖNYVTÁRA

B M E
BG 34
KÖNYVTÁRA

Szaktári száma:
Leltári száma: 2066

ISBN 963 05 2441 4

Printed in Hungary

Preface

Without knowing that both of us were there, the authors of this volume were random-walking on the streets of London in 1966 when, due to a theorem of Pólya, they met. Obviously this occasion called for a beer and a chat about mathematics. The beer turned out to be good enough to suggest that we should work together, and the idea of writing this book was born then. We are deeply indebted to the inkeeper for his hospitality on this occasion.

At that time we did not really know each other, though we had a common root in that both of us were students of Alfréd Rényi. The first named author actually studied mathematics at McGill University in Montreal and never took any courses from him. It was the papers and book of Rényi at that time, however, which influenced him most, and moulded his interest in doing research in probability-statistics. This also led to meeting him several times personally, thus directly benefiting from his most stimulating and unique way of thinking about mathematics. The second named author was a student of Rényi, indeed taking his courses in Budapest, and learning the secrets of doing research in probability directly from him. Rényi's great enthusiasm for the beauty of doing mathematics has inspired him to also try his hands at it. Both of us are deeply convinced that, without his lasting influence and help while we were young, we could have never written this book.

Our real collaboration began in 1972. During these past years we were fortunate enough to be able to visit each other several times, working in Ottawa where M. Csörgő is located and in Budapest where P. Révész is. This intensive collaboration would have been impossible without the understanding and support of our respective home institutions, the Department of Mathematics at Carleton University and the Mathematical Institute of the Hungarian Academy of Sciences.

Introduction

Let $X_1, X_2, \ldots$ be i.i.d.r.v. with $EX_1=0$, $EX_1^2=1$ and let F be their distribution function. Let $Y_1, Y_2, \ldots$ be i.i.d. normal r.v. with mean zero and variance one $(Y_1 \in \mathcal{N}(0, 1))$ and put $S_n = \sum_{i=1}^{n} X_i$, $T_n = \sum_{i=1}^{n} Y_i$ with $S_0 = T_0 = 0$. The classical central limit theorem states

$$P\{n^{-1/2} S_n \leqq y\} \to \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-u^2/2}\, du \tag{0.1}$$

for any real y as $n \to \infty$. Since

$$P\{n^{-1/2} T_n \leqq y\} = \Phi(y) \quad (n = 1, 2, \ldots),$$

the central limit theorem can also be stated as follows:

$$P\{n^{-1/2} S_n \leqq y\} - P\{n^{-1/2} T_n \leqq y\} \to 0, \tag{0.2}$$

which, roughly speaking, means that the limiting behaviour of S_n and T_n is the same. In other words, as time goes on, S_n forgets about the distribution function F where it has come from. However, it is also true that observing the sequence $S_1, S_2, \ldots$ (or, only $S_n, S_{n+1}, \ldots$ from any fixed n on), one can determine F with probability one via the Glivenko–Cantelli theorem.

Thus one can say that each individual S_n forgets about F but the complete sequence $\{S_n;\, n=1, 2 \ldots\}$ (or a tail of it) remembers F. One of the main goals of this book is to investigate to what extent can S_n remember F and to what extent can it forget about it.

The first questions of this type were formulated by Erdős and Kac (1946) (cf. also Kac (1946)). They wanted to evaluate the limit distributions

(i) $$G_1(y) = \lim_{n\to\infty} P\left(n^{-1/2} \max_{1 \leqq k \leqq n} S_k \leqq y\right),$$

(ii) $$G_2(y) = \lim_{n\to\infty} P\left(n^{-1/2} \max_{1 \leqq k \leqq n} |S_k| \leqq y\right),$$

butional properties of $\{S_n(t);\ 0 \leqq t \leqq 1\}$ should coincide[1] with those of $\{W(t);\ 0 \leqq t \leqq 1\}$ as $n \to \infty$. One possible way of saying this precisely is:

Theorem 0.1. (Donsker 1951). *We have*

$$h(S_n(t)) \xrightarrow{\mathcal{D}} h(W(t)) \tag{0.5}$$

for every continuous functional $h\colon C(0, 1) \to R^1$.

We note here that (0.4) only suggests that (0.5) should also be true and a precise proof of it was not at all easy to produce. Indeed, if $\{X_n(t)\}_{n=0}^{\infty}$ is a sequence of stochastic processes taking values from a function space M endowed with a metric ϱ, and

$$(X_n(t_1), X_n(t_2), \ldots, X_n(t_k)) \xrightarrow{\mathcal{D}} (X_0(t_1), X_0(t_2), \ldots, X_0(t_k)) \tag{0.6}$$

for any fixed sequence $t_1 < t_2 < \ldots < t_k$, then the statement that

$$h(X_n(t)) \xrightarrow{\mathcal{D}} h(X_0(t)) \tag{0.7}$$

should hold for every continuous functional $h\colon M \to R^1$, is not necessarily true. A complete methodology for proving (0.7), assuming that (0.6) is true, was worked out by Prohorov (1956) and Skorohod (1956).

In fact they proved a stronger statement to the effect that, under some conditions, the sequence of probability measures generated by $\{X_n(t)\}$ converges (in the so-called weak topology) to the measure generated by $X_0(t)$. An excellent summary and further development of these ideas and techniques can be found in the books of Billingsley (1968) and Parthasarathy (1967).

Replacing the functional h in Theorem 0.1 by

$$h_1(f) = \sup_{0 \leqq t \leqq 1} f(t), \quad h_2(f) = \sup_{0 \leqq t \leqq 1} |f(t)|,$$

$$h_3(f) = \int_0^1 f^2(t)\,dt, \quad h_4(f) = \int_0^1 |f(t)|\,dt,$$

and taking into account that these functionals are continuous with respect to the topology of $C(0, 1)$, Theorem 0.1, in particular, also implies that

[1] In this connection we should also mention that Kolmogorov (1931, 1933a) and Khinchine (1933) investigated the problem of evaluating the asymptotic probability of the event $f_1(t) < S_n(t) \leqq f_2(t)$ for two functions $f_1(t) < 0 < f_2(t)$, and proved that under some conditions on these functions the latter probability is equal to $P\{f_1(t) < W(t) \leqq \leqq f_2(t)\}$. Their approach is based on the heat equation.

$G_i(x)$ $(i=1, 2, 3, 4)$ of (i)–(iv) do not depend on F. That is to say the invariance principle of Erdős and Kac follows from Donsker's theorem and, at the same time, the latter can also be applied for any other continuous functional.

After the development of the theory of weak convergence of probability measures on metric spaces, a completely new form of the invariance principle was introduced by Strassen (1964). He proposed to construct a Wiener process $W(t)$ on the very same probability space where the r.v. $\{X_i\}$ live in such a way that $|S_n - W(n)|$ would be small in the sense that the relation

$$\frac{|S_n - W(n)|}{g(n)} \xrightarrow{\text{a.s.}} 0 \tag{0.8}$$

should hold for a suitably increasing function g. In fact the possibility of such a construction depends not only on the distribution F but also on the structure of the basic space. Hence the question in a more adequate form is the following:

Given a distribution function F with $\int x dF = 0$, $\int x^2 dF = 1$, can we construct a probability space $\{\Omega, \mathscr{A}, P\}$, a sequence $\{X_i\}$ of i.i.d.r.v. with $P(X_1 \leqq y) = F(y)$ living on Ω, and a Wiener process $W(t)$ also defined on Ω, such that (0.8) should hold?

Answering this question Strassen (1964) proved the following

Theorem 0.2.

$$\frac{|S_n - W(n)|}{\sqrt{n \log \log n}} \xrightarrow{\text{a.s.}} 0. \tag{0.9}$$

That is to say for any F with $\int x dF = 0$, $\int x^2 dF = 1$, one can construct a probability space where the i.i.d. sequence $\{X_i\}$ and a Wiener process $W(t)$ can be realized such that (0.9) holds.

In order to get a form of Theorem 0.2 resembling that of Theorem 0.1, we give the following reformulation of the former.

Theorem 0.2*.

$$\sup_{0 \leqq t \leqq 1} \frac{|S_n(t) - n^{-1/2} W(nt)|}{\sqrt{\log \log n}} \xrightarrow{\text{a.s.}} 0. \tag{0.9*}$$

Comparing Theorems 0.1 and 0.2 (or 0.2*), a great advantage of the latter is that is speaks about almost sure convergence instead of convergence in distribution.

Strassen used his strong invariance principle (Theorem 0.2) to prove the law of iterated logarithm for i.i.d.r.v. with finite second moment (the Hartman–Wintner theorem (1941)) via first proving such a theorem for the Wiener process. In fact studying the sequence $\left\{\frac{n^{-1/2}W(nt)}{\sqrt{2\log\log n}};\, 0\leqq t\leqq 1\right\}$ of stochastic processes, Strassen also obtained a deeper insight into the properties of the sequence $\left\{\frac{S_n(t)}{\sqrt{2\log\log n}};\, 0\leqq t\leqq 1\right\}$ (cf. Theorem 1.3.2).

In this spirit then Theorem 0.2 is like Theorem 0.1, the latter being applicable to prove weak convergence theorems for i.i.d.r.v. using distributional properties of the Wiener process, while the former is useful for proving strong theorems via similar properties of the Wiener process.

Theorem 0.2, however, does not imply Theorem 0.1, and this is because the rate of convergence in (0.9) is not strong enough. Should one be able to prove (0.8) with $g(n)=o(n^{1/2})$, then clearly we could also get (0.5) as a consequence of such a strong invariance principle. Chapter 2 of this book is mainly devoted to the question of the best possible rate in (0.8).

The precise connection between weak and strong invariance principles was established by Strassen (1965a) (cf. also Dudley (1968) and Wichura (1970)) via the so-called Prohorov distance of probability measures. In fact these results state a kind of equivalence between these two forms of invariance.

Our book is mainly devoted to the overall question of strong invariance theorems.

Our reason for concentrating on strong invariance methodology (instead of the weak one) can, perhaps, be justified by the fact that this approach has developed so much in recent years that it was capable of producing a number of results in probability and statistics which, in spite of the above mentioned equivalence of the two principles, would have been quite difficult to produce by the usual weak convergence methodology.

When talking about the origin of the invariance principle, another, independent source should be also mentioned besides the 1946 paper of Erdős and Kac. It is the paper of Doob (1949), entitled "Heuristic approach to the Kolmogorov–Smirnov theorems". The idea of this paper is the following: Let $U_1, U_2, \ldots$ be a sequence of i.i.d.r.v., coming from the uniform $U(0, 1)$ law. Let

$$E_n(x) = n^{-1} \sum_{k=1}^{n} I_{(0,x]}(U_k)$$

be the empirical distribution function, and let

$$\alpha_n(x) = n^{1/2}(E_n(x)-x)$$

be the empirical process. Observe that the limit of the joint distribution of $\alpha_n(x_1), \alpha_n(x_2), \ldots, \alpha_n(x_k)$ $(0 \leqq x_1 < x_2 < \ldots < x_k \leqq 1;\ k=1, 2, \ldots)$ is the corresponding finite dimensional distribution of a Brownian bridge; that is to say

$$\{\alpha_n(x_1), \alpha_n(x_2), \ldots, \alpha_n(x_k)\} \xrightarrow{\mathscr{D}} \{B(x_1), B(x_2), \ldots, B(x_k)\} \tag{0.10}$$

as $n \to \infty$, where $B(x)$ is a Brownian bridge. This then suggests that the limit properties of the empirical process $\alpha_n(x)$ should agree with the corresponding properties of a Brownian bridge. For example, the limit distribution of $\sup_x \alpha_n(x)$ (resp. $\sup_x |\alpha_n(x)|$) should agree with the distribution of $\sup_x B(x)$ (resp. $\sup_x B|(x)|$). Since the direct evaluation of the limit distribution of $\sup_x \alpha_n(x)$ (resp. $\sup_x |\alpha_n(x)|$) is rather complicated, while the evaluation of the distribution of $\sup_x B(x)$ (resp. $\sup_x |B(x)|$) is easier, the above sketched approach is obviously useful. Indeed, besides posing the above invariance argument, Doob (1949) proceeded to evaluate the distribution of these latter functionals of $B(x)$, leaving the problem of justification of his approach open. Donsker (1952) was the first one again who attacked this latter problem and succeeded in justifying and extending Doob's heuristic approach.

Comparing this problem to that of Theorem 0.1, we can see that a difficulty is coming from the fact that the sample functions of $\alpha_n(x)$ do not belong to $C(0, 1)$. This difficulty was again solved by Prohorov (1956) and Skorohod (1956), while working on the so-called $D(0, 1)$ function space. Naturally, an analogue of Theorem 0.1 is also true for a continuous approximation of $\alpha_n(x)$ on $C(0, 1)$.

In the light of Strassen's strong invariance principle, it was only natural to look for analogous approximations also for the empirical process $\alpha_n(x)$. This task turned out to be quite difficult and it took a bit of time to get results. The first one of them is due to Brillinger (1969), and reads as follows:

Theorem 0.3. *Given independent* $U(0, 1)$ *r.v.* $U_1, U_2, \ldots,$ *there exists a probability space with sequences of Brownian bridges* $\{B_n(x);\ 0 \leqq x \leqq 1\}$ *and empirical processes* $\{\tilde{\alpha}_n(x);\ 0 \leqq x \leqq 1\}$ *such that*

$$\{\tilde{\alpha}_n(x);\ 0 \leqq x \leqq 1\} \stackrel{\mathscr{D}}{=} \{\alpha_n(x);\ 0 \leqq x \leqq 1\} \quad \textit{for each} \quad n = 1, 2, \ldots, \tag{0.11}$$

and

$$\sup_{0 \leq x \leq 1} |\tilde{\alpha}_n(x) - B_n(x)| \stackrel{\text{a.s.}}{=} O\big(n^{-1/4}(\log n)^{1/2}(\log\log n)^{1/4}\big). \tag{0.12}$$

This theorem immediately implies the above mentioned analogue of Theorem 0.1. Namely, in terms of weak convergence, we have

$$\alpha_n(\cdot) \xrightarrow{\mathcal{D}} B(\cdot). \tag{0.13}$$

On the other hand, in spite of the indicated a.s. convergence in (0.12), Theorem 0.3 is not really a strong approximation theorem like Theorem 0.2 is. The reason for this is that in (0.12) we only have an approximation for each n, and only for a version $\tilde{\alpha}_n(x)$ of $\alpha_n(x)$. More precisely then, while Theorem 0.3 is a good first step in the right direction, it does not succeed in bringing together the stochastic processes $\{\alpha_n(x);\ 0 \leq x \leq 1, n=1, 2, \ldots)$ and $\{B_n(x);\ 0 \leq x \leq 1,\ n=1, 2, \ldots\}$. Consequently, no strong law type behaviour of the process $\alpha_n(x)$, say like the law of iterated logarithm, can be deduced from (0.12).

Kiefer (1969b) was the first one to call attention to the desirability of viewing the empirical process $\alpha_n(x)$ as a two parameter process and that a strong approximation theorem for $\alpha_n(x)$ should be given in terms of an appropriate two dimensional Gaussian process. He also succeeded in giving a first solution to this problem (Kiefer 1972; cf. Theorem 4.3.1). Preceding this work, Müller (1970) proved a corresponding two dimensional weak convergence of $\alpha_n(x)$, using Rényi's (1953) exponential representation of the empirical process.

In the present book we intend to summarize and elaborate on a number of recent strong invariance type results for partial sums and empirical processes of i.i.d.r.v., putting an emphasis on the applicability of strong approximation methodology to a variety of problems in probability and statistics. This is why, in the title, we use the expression "strong approximations" instead of "strong invariance principles".

In Chapter 1 we study the Wiener process together with some further Gaussian processes derived from it. In fact, in this Chapter we have intended to collect mostly those theorems for Gaussian processes which can be extended to partial sums and empirical processes of i.i.d.r.v. via strong approximation methods.

Chapter 2 is addressed to the problem of best possible strong approximations of partial sums of i.i.d.r.v. by a Wiener process, and it contains those theorems which tell us a complete story of this problem.

The content of Chapter 3 can be summarized in one sentence: Take "almost" any theorem of Chapter 1 concerning the one-time parameter Wiener process, then it can be extended to partial sums of i.i.d.r.v. via the results of Chapter 2. In most of the cases when the approximation methods do not work we can also conclude that the corresponding results cannot be extended at all. This Chapter does not intend to give a full systematic treatment of the asymptotic behaviour of partial sum processes and we concentrate only on those properties which can be deduced from invariance principles. For a detailed discussion of sums of random variables we refer to Petrov (1975) and Stout (1974).

Chapter 4 contains strong approximation theorems (in terms of suitable Gaussian processes) for the empirical and quantile processes based on i.i.d.r.v.

The role of Chapter 5 in the theory of empirical and quantile processes is similar to that of Chapter 3 in the theory of partial sums of i.i.d.r.v. Namely, in this Chapter we show that by applying the results of Chapter 4, the theorems of Chapter 1 concerning Brownian bridges and the so-called Kiefer process are also valid for empirical and quantile processes. This phenomenon of inheriting properties from appropriate Gaussian processes is not so complete here as in the case of partial sums of i.i.d.r.v. and, to some extent, we also touch upon the problem of similar and non-similar behaviour beyond invariance (cf. Remark 5.1.1). For a recent and more detailed discussion of this topic we refer to the survey paper of Gaenssler and Stute (1979).

In Chapter 6 we show that suitably defined sequences of empirical density, regression and characteristic functions can be approximated by appropriate Gaussian processes. Here it will be seen that some results on Gaussian processes can be extended also to these by strong approximation methods.

The aim of Chapter 7 is to demonstrate that strong approximation methodology can also be applied to study weak and strong convergence properties of random size partial sum and empirical processes.

A common property of Chapters 3, 5, 6 and 7 is that their respective topics are treated only so far as one can see them via strong approximation methods, and we did not aim at completeness at all in treating them.

The subject of this book is restricted to i.i.d.r.v. when the time and state parameters belong to the real line. There is an exception in Chapter 1, when we also study certain properties of two-time parameter Wiener and Kiefer processes. Our reason for this is due to the fact that certain properties of the empirical process $\alpha_n(x)$ can only be described and handled via viewing it as a two-time parameter process in x and n.

We intend to study the problems of strong approximation of multi-time parameter partial sum and empirical processes by appropriate multi-time parameter Gaussian processes in the second volume of this book.

The case when the state space is also a higher dimensional Euclidean space (or a Banach space) has been investigated by several authors (cf. e.g. J. Kuelbs 1973, J. Hoffman–Jørgensen–G. Pisier 1976, Garling 1976) and it should be the subject of a third volume. The subject of a fourth volume should be the case of non-independent and/or non-identically distributed r.v. (for a preliminary version we refer to W. Philipp and W. Stout (1975), an excellent survey of the present situation of this topic). However, the authors have realized that the lifetime of a human being is not only a one-dimensional but also a strictly bounded r.v. Hence, they do not intend to write the mentioned third and fourth volumes, though they would be glad to live long enough to read these by someone else.