LIMIT THEOREMS OF PROBABILITY THEORY

G. Christoph
Department of Mathematics, Otto-von-Guericke-University of Magdeburg, Germany

Keywords: Sequences of random variables, sums of random variables, modes of convergence, laws of large numbers, law of iterated logarithm, central limit theorem, normal limit distribution, Poisson limit distribution, probabilities of large deviation, local limit theorems, limit distributions of extremes.

Contents

Summary

Some basic theory of sums of random variables with increasing number of terms is presented. Different types of convergence are treated.

The laws of large numbers, the law of iterated logarithm, the central limit theorem and the classical summation theory are given, mainly for sums of independent random variables, and also refinements on these theorems.

Local limit theorems, asymptotic expansions, large deviations results and limit distributions of normalized extremes and order statistics are considered, too.

1. Introduction and Preliminaries

Probability theory is motivated by the idea, that the unknown probability *p* of an event *A* is approximately equal to r/n, if *n* trials result in *r* realisation of the event *A*, and the approximation improves with increasing *n*. Limit theorems in probability theory and statistics are regarded as results giving convergence of sequences of random variables or their distribution functions. Since sequences of random variables are sequences of functions with random influences, different modes of convergence are involved. The law of large numbers and the central limit theorem are the most important limit theorems. They are parts of the classical summation theory, investigating the possible limit distributions for the distributions of certain sums of random variables.

1.1. Sequences of Events and Their Probabilities

Let $(\Omega, \mathcal{A}, P)$ be a *probability space*, where Ω is a set of elements ω, $\mathcal{A}$ is a σ-algebra of subsets (here called events) of the set Ω, and *P* is a probability measure defined on $\mathcal{A}$. Let $\{A_n\}_{n\geq 1} \in \mathcal{A}$ be a sequence of events. (See *"Mathematical Foundations and Interpretations of Probability"*).

Proposition 1.1. *Boole's inequality:* For events $\{A_n\}_{n\geq 1} \in \mathcal{A}$,

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} P(A_k) \quad . \tag{1}$$

Since P is countable additive, the equal sign in (1) holds for pairwise disjoint events, i.e. if $A_i \cap A_j = \varnothing$ for $i \neq j$. Define the following events belonging to $\mathcal{A}$:

$$\limsup_n A_n = \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n = \{A_n, \text{i.o.}\} \quad \text{and} \quad \liminf_n A_n = \bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} A_n \quad . \tag{2}$$

The set $\limsup_n A_n$, denoted by $\{A_n, \text{i.o.}\}$ is the set of events ω such that $\omega \in A_n$ for *infinitely many* values of n, where i.o. abbreviates *"infinitely often"*. The set $\liminf_n A_n$ is the set of such events ω, that $\omega \in A_n$ for *all but finitely many* values of n.

Proposition 1.2. For events $\{A_n\}_{n\geq 1} \in \mathcal{A}$ the following hold:

$$P(\liminf_n A_n) \leq \liminf_{n\to\infty} P(A_n) \quad \text{and}$$
$$\limsup_{n\to\infty} P(A_n) \leq P(\limsup_n A_n). \tag{3}$$

The first inequality in (3) is a consequence of *Fatou's lemma* for probabilities.

Proposition 1.3. *Borel-Cantelli lemma:* Suppose $\{A_n\}_{n\geq 1} \in \mathcal{A}$.

a) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(A_n, \text{i.o.}) = 0$.

b) Let $A_1, A_2, \ldots$ be independent events such that $\sum_{n=1}^{\infty} P(A_n) = \infty$. Then $P(A_n, \text{i.o.}) = 1$.

Proposition 1.4. *Borel zero-one criterion:* If the events in the sequence $\{A_n\}_{n \geq 1} \in \boldsymbol{A}$ are independent, then $P(A_n, \text{i.o.}) = 0$ or 1 according as $\sum_{n=1}^{\infty} P(A_n) < \infty$ or $= \infty$.

1.2. Inequalities for Sums of Random Variables

Let $X, X_1, X_2, \ldots$ be random variables on a common probability space $(\Omega, \boldsymbol{A}, P)$. Denote the *mathematical expectation*, the *variance* and the *d-th order absolute moment* of X by $E\,X$, $\operatorname{Var} X$ and $E|X|^d$, respectively, if they exist. The space $L^d = L^d(\Omega, \boldsymbol{A}, P)$, $0 < d < \infty$, denotes the set of random variables X such that $E|X|^d < \infty$. The usual metric in the space L^d is given by $d(X,Y) = \| X - Y \|_d$ with $\| X \|_d = E\,|X|^d$ or $(E\,|X|^d)^{1/d}$ according as $0 < d < 1$ or $d \geq 1$. There is a type of very important inequalities which are collected by the *Markov-inequality:*

$P(|X| \geq \varepsilon) \leq (g(\varepsilon))^{-1} E\,g(X)$ for any even non-degreasing function $g \geq 0$ and every $\varepsilon > 0$. (4)

With $g(x) = x^2$ the Markov inequality implies the *Bienaymé-Chebyshev inequality:*

$$P(|X - E\,X| \geq \varepsilon) \leq \varepsilon^{-2} \operatorname{Var} X \quad \text{for every } \varepsilon > 0, \tag{5}$$

estimating the probability of deviation of a random variable from its expectation by its variance.

Consider the partial sum $S_n = X_1 + \cdots + X_n$ from the sequence $\{X_n\}_{n \geq 1}$.

Example 1.1. *Weak law of large numbers:* Let $X_1, X_2, \cdots, X_n$ be independent and identically distributed (Two or more random variables are identically distributed, if they have the same distribution.) with finite variance $0 < \sigma^2 = \operatorname{Var} X_1 < \infty$. Then $E\,S_n = n\mu$, $\operatorname{Var} S_n = n\sigma^2$ and by (5)

$$P(|S_n/n - EX_1| \geq \varepsilon) \leq \sigma^2 \varepsilon^{-2} n^{-1} \text{ for any } \varepsilon > 0. \tag{6}$$

Hence, the probability of the event, that the arithmetic mean $\overline{X_n} = S_n / n$ differs from the expectation of the summands $E\,X_1$ by more than ε, tends to zero.

Let $\{X_n\}_{n\geq 1} \in L^2$ be a sequence of independent random variables with $E\,X_k = 0$ and $\sigma_k^2 = \mathrm{Var}\,X_k < \infty$, $k = 1,\ldots,n$. An useful tool in probability theory is the *Hàjek-Rényi inequality:* Suppose $0 < c_n \leq c_{n-1} \leq \cdots \leq c_1$. Then, for all x >0 and every integer $0 < m < n$,

$$P(\max_{m\leq k\leq n} |S_k| \geq x) \;\leq\; x^{-2}\,(c_m^2 \sum_{k=1}^m \sigma_k^2 + \sum_{k=m+1}^n c_k^2\,\sigma_k^2)\,. \qquad (7)$$

In case $c_1 = \cdots = c_n = 1$ one find the *Kolmogorov inequality*:

$$P(\max_{1\leq k\leq n} |S_k| \geq x) \;\leq x^{-2} \sum_{k=1}^n \sigma_k^2 \;. \qquad (8)$$

Consider the *Bernstein condition:* There exists a positive constant *H* such that

$$|\,E\,X_k^{\;m}\,| \;\leq \frac{m!}{2}\,\sigma_k^2\;H^{m-2} \;\text{ for all integers }\; m \geq 2 \;\text{ and }\; k = 1,2...,n\,, \qquad (9)$$

bounding the growth of the moments of X_k. Bernstein's condition (9) implies exponential estimates for the partial sum S_n, the *Bernstein inequalities:* Put $b_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$, then

$$\max\{P(S_n \leq -x), P(S_n \geq x)\} \leq \begin{cases} \exp\{-x^2/(4b_n^2)\} & \text{if } \; 0 \leq x \leq b_n^2 / H\,, \\ \exp\{-x/(4H)\} & \text{if } \; \mathrm{x} \geq b_n^2 / H\,. \end{cases} \qquad (10)$$

The Bernstein inequalities are rather powerful, leading to exponentially fast convergence rates as shown in the following. If the random variables $X_1,...,X_n$ with zero mean are uniformly bounded, i.e. if there is a constant *C* such that $P(|\,X_k\,| \leq C) = 1$ for $k = 1,...,n$, then Bernstein's condition (9) is satisfied with *H* = *C*. In case of identically distributed random variables $X_1,...,X_n$, the Bernstein condition (9) is a consequence of the *Cramér condition*: There exists a positive constant *a* such that

$$E\exp\{\,a\,|\,X\,|\,\} < \infty\,, \qquad (11)$$

ensuring the existence of the *moment generating function* $M(h) = E\exp\{hX\}$ for $|h| \le a$, which leads to the existence of all order moments of the underlying random variable *X*.

Example 1.2. *Large deviation estimate for the arithmetic mean:* Consider a sequence $\{X_n\}_{n\ge1}$ of independent and identically distributed random variables with $P(X_1 = 1) = p = 1 - P(X_1 = 0)$. Then the partial sum $S_n = X_1 + \cdots + X_n$ is binomial distributed with success probability *p*, $0 < p < 1$. In order to apply Bernstein's inequalities (10) the conditions (9) have to be proved for the random variable $(X_1 - p)$, which has now zero mean, as well as the random variable $(X_1 - p) + \cdots + (X_n - p) = S_n - np$. Since $|X_1 - p| \le 1$, Bernstein's condition (9) holds with $H = 1$. Let $\overline{X_n} = S_n / n$ denote the arithmetic mean of the first *n* random variables from the given sequence. Using $4p(1-p) \le 1$, Bernstein inequalities (10) with $x = \varepsilon n$ for some $0 < \varepsilon < p(1-p)$ imply an exponential bound for the deviation of sample mean $\overline{X_n}$ from success probability *p*:

$$\max\{P(\overline{X_n} - p \le -\varepsilon),\ P(\overline{X_n} - p \ge \varepsilon)\} \le \exp\{-n\varepsilon^2\}. \qquad (12)$$

Hence, $P(|\overline{X_n} - p| \ge \varepsilon)$ tends to zero exponentially fast as $n \to \infty$. Inequalities like (12) are known as *large deviation estimations* in the law of large numbers.

Let $X_k \in L^r$, i.e. $E|X_k|^r < \infty$ for some $r \ge 2$. Define $M_{r,n} = \sum_{k=1}^n E|X_k|^r$, then by the *Fuk-Nagaev inequality*:

$$P(S_n \ge x) \le (1 + 2/r)^r M_{r,n}\, x^{-r} + \exp\{-2(r-2)e^{-r} x^2 B_n^{-2}\} \quad \text{for} \quad x > 0. \qquad (13)$$

Finally, moments $E|S_n|^r$ may be estimated by the c_r*-inequality*

$$E\,|S_n|^r \le c_r\, M_{r,n} \quad \text{with} \quad c_r = n^{r-1} \quad \text{if} \quad r \ge 1 \quad \text{or} \quad c_r = 1 \quad \text{if} \quad r \le 1. \qquad (14)$$

For independent sequences the *Rosenthal inequalities* give upper and lower bounds:

$$2^{-r} \max\{M_{r,n}, B_n^r\} \le E(|S_n|^r) \le 2^{r \cdot r} \max\{M_{r,n}, B_n^r\} \qquad (15)$$

with $r \ge 2$ in the first and $r > 1$ in the second of the inequalities.

-
-
-

Bibliography

Araujo A. and Giné E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*, 223 pp. New York: Wiley. [The textbook gives a good introduction to central limit theorems in Banach spaces.]

Bhattacharya R.N. and Ranga Rao R. (1976). *Normal Approximation and Asymptotic Expansions*, 274 pp. New York: Wiley. [The book provides an outlook of multidimensional limit theorems.]

Billingsley P. (1999). *Convergence of Probability Measures*. *Second Edition*, 277 pp. Chichester: Wiley. [Advanced textbook on convergence for processes and partial sums]

Embrechts P. and Klüppelberg C. and Mikosch T. (1996). *Modelling Extremal Events for Insurance and Finance*, 645 pp., Berlin: Springer-Verlag. [The textbook looks at asymptotic problems from a practical point of view, particularly in extreme value theory.]

Feller W. (1968/71). *An Introduction to Probability Theory and its Applications I (Third Edition),. and II (Second Edition),* 509 pp and 645 pp. New York: Wiley. [The classical books contain the most important results where limit theorems are involved.]

Hall P. and Heyde C.C. (1980). *Martingale Limit Theorems and its Applications*, 308 pp. New York: Academic Press. [The book focuses mainly on strong law of large numbers, central limit theorem and law of iterated logarithm for discrete parameter martingales as sequences of dependent random variables with many applications.]

Lin Zhengyan and Lu Chuanrong (1996). *Limit Theory for Mixing Dependent Random Variables*, 426 pp. Dordrecht: Kluwer and New York: Science Press [This textbook provides useful tools to study probability and moment inequalities, weak and strong convergence as well as Berry-Esseen bounds for sequences of dependent random variables with different kinds of dependency.]

Loéve M. (1977/78). *Probability Theory I and II, Fourth edition, Graduate Texts in Mathematics, Vol*. 45/46, 425 pp and 413 pp. Berlin: Springer-Verlag. [The widely used books reflect the crucial role of limit theorems in probability theory.]

Petrov V.V. (1995). *Limit Theorems of Probability Theory. Sequences of Independent Random Variables,* 292 pp. Oxford:Clarendon Press. [This is a well written book that contains most of the very significant results in the theory of summation of independent random variables.]

Prokhorov Y.V. and Statulevicius V. (Eds.) (2000). *Limit Theorems of Probability Theory*, 273 pp. Berlin: Springer-Verlag. [The book is a collection of research level surveys on classical limit theorems for independent and dependent random variables, the accuracy of Gaussian approximation in Banach spaces and large deviation theorems.]

Serfling R.J. (1980). *Approximation Theorems of Mathematical Statistics*, 371 pp. New York: Wiley. [The book provides basic results and references on limit theory in statistics.]

Zolotarev V.M. (1997). *Modern Theory of Summation of Random Variables*, 412 pp. Utrecht: VSP. [The book is based on metrics in the set of random variables, leading to the formulation of limit theorems as stability relations.]

Biographical Sketch

Gerd Christoph is professor of probability theory at the Department of Mathematics at the Otto-von-Guericke University of Magdeburg/Germany. He was born October 30, 1948 in Jena/Germany and is married with 3 children. He studied mathematics at the State University in Leningrad (now St. Petersburg)/Russia from 1967 to 1972 and worked as research assistant at the Technical University of Dresden from 1973 to 1987, where he received the degrees Dr. rer. nat. in 1978 and Dr. rer. nat. habil. in 1987. From 1987 to 1989 he was a lecturer at the Department of Mathematics at the Addis Ababa University in Ethiopia, from 1989 to 1990 a docent and since 1990 he has been professor in Magdeburg. Fields of his interest are limit theorems in probability theory, stable and discrete stable limit distributions, characterization, and stability problems for quadratic forms, extreme value theorems, and asymptotic methods in statistics.