Probability Theory and Stochastic Modelling 104

Sergey Bobkov
Gennadiy Chistyakov
Friedrich Götze

Concentration and Gaussian Approximation for Randomized Sums

Springer

Probability Theory and Stochastic Modelling

Volume 104

Probability Theory and Stochastic Modelling publishes cutting-edge research monographs in probability and its applications, as well as postgraduate-level textbooks that either introduce the reader to new developments in the field, or present a fresh perspective on fundamental topics.

Books in this series are expected to follow rigorous mathematical standards, and all titles will be thoroughly peer-reviewed before being considered for publication.

Probability Theory and Stochastic Modelling covers all aspects of modern probability theory including:

- Gaussian processes
- Markov processes
- Random fields, point processes, and random sets
- Random matrices
- Statistical mechanics, and random media
- Stochastic analysis
- High-dimensional probability

as well as applications that include (but are not restricted to):

- Branching processes, and other models of population growth
- Communications, and processing networks
- Computational methods in probability theory and stochastic processes, including simulation
- Genetics and other stochastic models in biology and the life sciences
- Information theory, signal processing, and image synthesis
- Mathematical economics and finance
- Statistical methods (e.g. empirical processes, MCMC)
- Statistics for stochastic processes
- Stochastic control, and stochastic differential games
- Stochastic models in operations research and stochastic optimization
- Stochastic models in the physical sciences

Probability Theory and Stochastic Modelling is a merger and continuation of Springer's Stochastic Modelling and Applied Probability and Probability and Its Applications series.

Sergey Bobkov • Gennadiy Chistyakov
Friedrich Götze

Concentration and Gaussian Approximation for Randomized Sums

Springer

Sergey Bobkov
School of Mathematics
University of Minnesota
Minneapolis, MN, USA

Gennadiy Chistyakov
Fakultät für Mathematik
Universität Bielefeld
Bielefeld, Germany

Friedrich Götze
Fakultät für Mathematik
Universität Bielefeld
Bielefeld, Germany

ISSN 2199-3130 ISSN 2199-3149 (electronic)
Probability Theory and Stochastic Modelling
ISBN 978-3-031-31148-2 ISBN 978-3-031-31149-9 (eBook)
https://doi.org/10.1007/978-3-031-31149-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Given a random vector $X = (X_1, \dots, X_n)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in the Euclidean space $\mathbb{R}^n$, $n \geq 2$, define the weighted sums

$$\langle X, \theta \rangle = \sum_{k=1}^{n} \theta_k X_k,$$

parameterized by points $\theta = (\theta_1, \dots, \theta_n)$ of the unit sphere $\mathbb{S}^{n-1}$ in $\mathbb{R}^n$ (with center at the origin and radius one). In general, the distribution functions of weighted sums $\langle X, \theta \rangle$, say

$$F_\theta(x) = \mathbb{P}\{\langle X, \theta \rangle \leq x\}, \quad x \in \mathbb{R},$$

essentially depend on the parameter θ. On the other hand, a striking observation made by V. N. Sudakov [169] in 1978 indicates that, under mild correlation-type conditions on the distribution of X, and when n is large, most of the F_θ's are concentrated around a certain "typical" distribution function F. Here "most" should be understood in the sense of the normalized Lebesgue measure $\mathfrak{s}_{n-1}$ on $\mathbb{S}^{n-1}$. A more precise statement can be given, for example, under the isotropy condition

$$\mathbb{E}\,\langle X, \theta \rangle^2 = 1, \quad \theta \in \mathbb{S}^{n-1},$$

which frequently appears in many applications. Similar to the classical central limit theorem, Sudakov's result thus represents a rather general principle of convergence, with various interesting aspects. A related phenomenon was discovered later by Diaconis and Freedman [79] in terms of low-dimensional projections of non-random data (cf. also von Weizsäcker [176]).

The phenomenon of concentration of the family $\{F_\theta\}_{\theta \in \mathbb{S}^{n-1}}$ naturally begs the question of closeness of F_θ to F for all θ from a large part of the sphere in terms of standard distances d in the space of probability distributions on the real line. A canonical choice would be the Kolmogorov (uniform) distance

$$\rho(F_\theta, F) = \sup_x |F_\theta(x) - F(x)|.$$

Less sensitive alternatives would be the Lévy distance

$$L(F_\theta, F) = \inf\Big\{h \ge 0 : F(x-h) - h \le F_\theta(x) \le F(x+h) + h \ \text{ for all } x \in \mathbb{R}\Big\},$$

as well as the distances in L^p-norms

$$d_p(F_\theta, F) = \Big(\int_{-\infty}^{\infty} |F_\theta(x) - F(x)|^p \, \mathrm{d}x\Big)^{1/p}, \quad p \ge 1,$$

among which $W = d_1$ and $\omega = d_2$ are most natural. For a given distance d, the behavior of the average value $m = \mathbb{E}_\theta \, d(F_\theta, F)$, as well as the deviation from the mean in spherical probability $\mathfrak{s}_{n-1}\{d(F_\theta, F) \ge m + r\}$, is of interest as a function of n and $r > 0$.

In this context the model of independent random variables X_k has been intensively studied in the literature. When X_k are independent and identically distributed (the i.i.d. case) and have mean zero and variance one, the distribution functions F_θ are known to be close to the standard normal distribution function

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-y^2/2} \, \mathrm{d}y,$$

as long as $\max_k |\theta_k|$ is small. If the 3-rd absolute moment $\beta_3 = \mathbb{E}\,|X_1|^3$ is finite, the Berry–Esseen theorem allows us to quantify this closeness by virtue of the bound

$$\rho(F_\theta, \Phi) \le c\beta_3 \sum_{k=1}^{n} |\theta_k|^3,$$

which holds for some absolute constant $c > 0$. Although the right-hand side is greater than or equal to $c\beta_3/\sqrt{n}$ for any $\theta \in \mathbb{S}^{n-1}$, the bound above implies a similar upper bound on average: $\mathbb{E}_\theta \, \rho(F_\theta, \Phi) \le c'\beta_3/\sqrt{n}$.

The i.i.d. case inspires the idea that, under some natural moment and correlation-type assumptions, most of the F_θ might also be close to the standard normal law. But, in light of Sudakov's theorem, this is equivalent to a similar assertion about the typical distribution – a property which is determined by the distribution of the Euclidean norm

$$|X| = \big(X_1^2 + \cdots + X_n^2\big)^{1/2}.$$

Indeed, in general, the typical distribution can be identified as the spherical average

$$F(x) = \int_{\mathbb{S}^{n-1}} F_\theta(x) \, \mathrm{d}\mathfrak{s}_{n-1}(\theta) \equiv \mathbb{E}_\theta F_\theta(x),$$

which may be alternatively described as the distribution of $|X|\,\theta_1$, where the first coordinate of a point on the sphere is treated as a random variable independent of X. (In the sequel, $\mathbb{E}_\theta$ is always understood as the integral with respect to the measure $\mathfrak{s}_{n-1}$.) Since $\theta_1\sqrt{n}$ is nearly standard normal, F will be close to Φ if and only if the random variable $R^2 = \frac{1}{n}\,|X|^2$ is approximately 1 in the sense of the weak topology. In many situations, this can be verified directly by computing, for example, the variance of R^2, while in some others it represents a non-trivial "thin shell" type concentration problem.

This book aims to describe the current state of the art concerning Sudakov's theorem. In particular, using the metrics d mentioned above, we will focus on the derivation of various bounds for $\mathbb{E}_\theta\, d(F_\theta, F)$ and $\mathbb{E}_\theta\, d(F_\theta, \Phi)$, as well as on large deviation bounds. Our investigations rely on several basic tools. Besides classical techniques of Fourier Analysis (such as Berry–Esseen-type bounds), many arguments rely upon the spherical concentration phenomenon, that is, concentration properties of the measures $\mathfrak{s}_{n-1}$ for growing dimensions n, including the associated Sobolev-type and infimum-convolution inequalities. Concentration tools are also used for various classes of distributions of X when trying to approximate the typical distribution F by the standard normal law.

In order to facilitate the readability of the presentation of the results related to the Sudakov phenomenon, we decided to make the presentation more self-contained by including these auxiliary techniques in the first three chapters. Thus we describe in a separate part (Part II) some general results on concentration in the setting of Euclidean and abstract metric spaces. Most of this material can also be found in other publications, including the excellent survey and monograph by M. Ledoux [129], [130], and the recent book by D. Bakry, I. Gentil, and M. Ledoux [8].

The spherical concentration is discussed separately in Part III. It is a classical well-known fact (whose importance was first emphasized by V. Milman in the early 1970s) that any mean zero smooth (say, Lipschitz) function f on the unit sphere $\mathbb{S}^{n-1}$ has deviations at most of the order $1/\sqrt{n}$ with respect to the growing dimension n. Moreover, as a random variable, $\sqrt{n}\, f$ has Gaussian tails under the measure $\mathfrak{s}_{n-1}$. In addition to this spherical phenomenon, we present recent developments on the so-called second order concentration, which was pushed forward by the authors as an advanced tool in the theory of randomized summation. Roughly speaking, the second order concentration phenomenon indicates that, under proper normalization hypotheses in terms of the Hessian, any smooth f on $\mathbb{S}^{n-1}$ orthogonal to all affine functions in $L^2(\mathfrak{s}_{n-1})$ actually has deviations at most of the order $1/n$. Moreover, as a random variable, nf has exponential tails under the measure $\mathfrak{s}_{n-1}$. Part III also contains various bounds on deviations of elementary polynomials under $\mathfrak{s}_{n-1}$ and collects asymptotic results on special functions related to the distribution of the first coordinate on the sphere.

These tools are needed to quantify Sudakov's theorem in terms of several moment and correlation-type conditions, and for various classes of distributions of X. With this aim, we shall introduce and discuss the following moment type quantities for a parameter $p \geq 1$,

$$M_p = \sup_{\theta \in \mathbb{S}^{n-1}} \big(\mathbb{E}\,|\langle X, \theta\rangle|^p\big)^{1/p}, \quad m_p = \frac{1}{\sqrt{n}}\,\big(\mathbb{E}\,|\langle X, Y\rangle|^p\big)^{1/p},$$

as well as the variance-type functionals

$$\sigma_{2p} = \sqrt{n}\,\bigg(\mathbb{E}\,\Big|\frac{|X|^2}{n} - 1\Big|^p\bigg)^{1/p}, \quad \Lambda = \sup_{\sum a_{ij}^2 = 1} \mathrm{Var}\bigg(\sum_{i,j=1}^n a_{ij} X_i X_j\bigg),$$

where Y is an independent copy of X. For example, $M_2 = m_2 = 1$ in the isotropic case, and $\sigma_4^2 = \frac{1}{n}\,\mathrm{Var}(|X|^2)$, which can often be estimated via evaluation of the covariances of X_i^2 and X_j^2. The relevance of these functionals will be clarified in various examples; they are also connected with analytic properties of the distribution μ of X expressed in terms of isoperimetric or Poincaré-type inequalities. For example, there is a simple bound $\Lambda \le 4/\lambda_1^2$ via the spectral gap λ_1 associated to μ.

We shall now outline several results on upper bounds for $\mathbb{E}_\theta\, d(F_\theta, F)$ and $\mathbb{E}_\theta\, d(F_\theta, \Phi)$ involving these functionals for various distances d. They are discussed in detail in the remaining Parts IV–VI of this monograph.

• *Lévy distance*. Here the moments M_1 and M_2 will control quantitative bounds on fluctuations of F_θ around the typical distribution F in the metric L providing polynomial rates with respect to n. Namely, for some absolute constant $c > 0$ we have

$$\mathbb{E}_\theta\, L(F_\theta, F) \le c\,\frac{M_1 + \log n}{n^{1/4}}, \quad \mathbb{E}_\theta\, L(F_\theta, F) \le c\,\Big(\frac{\log n}{n}\Big)^{1/3} M_2^{2/3}.$$

• *Kantorovich L^1 transport distance*. Here rates can be improved in terms of the moments M_p of higher order. In particular, we have

$$\mathbb{E}_\theta\, W(F_\theta, F) \le c_p M_p\, n^{-\frac{p-1}{2p}} \quad (p > 1),$$

where the constants c_p depend on p only. However, a classical rate of $1/\sqrt{n}$ from other contexts will not be achievable via these bounds.

• *Kolmogorov distance*. Using the variance-type functionals σ_p, it is possible not only to replace the typical distribution F with the normal distribution function Φ, thus proving a law of attraction for F_θ, but also to show a standard rate as well, assuming a finite third moment. Analogously to the classical Berry–Esseen theorem, it is shown that, if $\mathbb{E}\,|X|^2 = n$ and $\mathbb{E}X = a$, then

$$\mathbb{E}_\theta\, \rho(F_\theta, \Phi) \le \frac{A}{\sqrt{n}}$$

with $A = c\,(m_3^{3/2} + \sigma_3^{3/2} + |a|)$ up to some absolute constant c. Here, one may eliminate the parameter a, by using elementary bounds $m_3 \le M_3^2$ and $\sigma_3 \le \sigma_4$ (the latter requires, however, the finiteness of the 4-th moment). A slightly worse estimate can also be derived under less restrictive moment assumptions. For example,

$$\mathbb{E}_\theta\, \rho(F_\theta, \Phi) \le c\,(M_2^2 + \sigma_2)\,\frac{\log n}{\sqrt{n}}.$$

• *Trigonometric and other functional models of random variables*. Modulo a logarithmic factor, the upper bounds such as

$$\mathbb{E}_\theta\, \rho(F_\theta, \Phi) \le c\, \frac{\log n}{\sqrt{n}}$$

turn out to be optimal with respect to n in many examples of orthonormal systems $X = (X_1, \dots, X_n)$ of functions in L^2. These include in particular the trigonometric system of size n with components

$$\begin{aligned} X_{2k-1}(t) &= \sqrt{2}\, \cos(kt), \\ X_{2k}(t) &= \sqrt{2}\, \sin(kt), \quad -\pi < t < \pi, \quad k = 1, \dots, n/2 \ \ (n \text{ even}), \end{aligned}$$

with respect to the normalized Lebesgue measure $\mathbb{P}$ on $\Omega = (-\pi, \pi)$. More precisely, we derive lower bounds such as

$$\mathbb{E}_\theta\, \rho(F_\theta, \Phi) \ge \frac{c}{\sqrt{n}\, (\log n)^s}$$

for some positive c and s independent of n. A similar bound also holds for the sequence of the first n Chebyshev polynomials on the interval $\Omega = (-1, 1)$, for the Walsh system on the Boolean cube $\{-1, 1\}^p$ (with $n = 2^p - 1$), for systems of functions $X_k(t_1, t_2) = f(kt_1 + t_2)$ with 1-periodic f (such functions X_k form a strictly stationary sequence of pairwise independent random variables on the square $\Omega = (0, 1) \times (0, 1)$ under the restricted Lebesgue measure), and some others.

• *L^2 distance*. In order to develop lower bounds as above, similar upper and lower bounds will be needed for the L^2-distance ω, in combination with upper bounds for the Kantorovich-distance W. A number of general results in this direction will be obtained under moment and correlation-type assumptions, as in the case of Kolmogorov distance ρ. In fact, in the case of ω, the correct asymptotic behavior of $\mathbb{E}_\theta\, \omega^2(F_\theta, F)$ will be derived up to the order $1/n^2$. For instance, when the random vector X has an isotropic symmetric distribution and satisfies $|X| = \sqrt{n}$ a.s. (and thus all $\sigma_p = 0$), one has

$$\mathbb{E}_\theta\, \omega^2(F_\theta, F) \sim \frac{1}{n^4}\, \mathbb{E}\, \langle X, Y \rangle^4$$

with an error term of order $1/n^2$, and a similar result holds for the Gaussian limit Φ instead of the typical distribution F. Here, as before, Y denotes an independent copy of X.

• *Improved rates in the i.i.d. case*. Returning to the classical i.i.d. model with $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$, a remarkable result due to Klartag and Sodin [125] which we include in this monograph improves the pointwise Berry–Esseen bound as follows

$$\mathbb{E}_\theta\, \rho(F_\theta, \Phi) \le \frac{c\beta_4}{n}, \quad \beta_4 = \mathbb{E}X_1^4.$$

In fact, this result holds in the non-i.i.d. situation as well with replacement of β_4 with the arithmetic means of the 4-th moments of X_k. This bound can be complemented by corresponding large deviation bounds. Thus, for typical coefficients, the distances $\rho(F_\theta, \Phi)$ are at most of order $1/n$, which is not true in general when the coefficients are equal to each other!

Moreover, we show that, if the distribution of X_1 is symmetric, and the next moment $\beta_5 = \mathbb{E}\,|X_1|^5$ is finite, it is possible to slightly correct the normal distribution Φ to obtain a better approximation such as

$$\mathbb{E}_\theta\, \rho(F_\theta, G) \le \frac{c\beta_5}{n^{3/2}}.$$

Here, G is a certain function of bounded variation which is determined by β_4 and depends on n (but not on θ).

- *The second order correlation condition*. Certainly Sudakov's theorem begs the question whether or not similar results continue to hold for dependent components X_k. This is often the case, although the orthonormal systems mentioned above may serve as counter examples. More precisely, the variance functional $\Lambda = \Lambda(X)$ turns out to be responsible for improved rates of normal approximation for F_θ's on average and actually for most θ's. When X has an isotropic symmetric distribution, it will be shown by virtue of the second order spherical concentration that

$$\mathbb{E}_\theta\, \rho(F_\theta, \Phi) \le \frac{c \log n}{n}\, \Lambda,$$

which thus extends the i.i.d. case modulo a logarithmic factor. The symmetry assumption can be removed at the expense of additional terms reflecting higher order correlations. In particular, in the presence of the Poincaré-type inequality, we have

$$\mathbb{E}_\theta\, \rho(F_\theta, \Phi) \le \frac{c_1 \log n}{n}\, \lambda_1^{-1},$$

which may be complemented by corresponding deviation bounds.

- *Distributions with many symmetries*. Special attention will be devoted to the case where the distribution of X is symmetric about all coordinate axes and isotropic (which reduces to the normalization condition $\mathbb{E}X_k^2 = 1$). The Λ-functional then simplifies, and under the 4-th moment condition, the Berry–Esseen bound "on average" takes the form

$$\mathbb{E}_\theta\, \rho(F_\theta, \Phi) \le \frac{c\, \log n}{n} \Big(\max_{k \le n}\, \mathbb{E}X_k^4 + V_2\Big),$$

where

$$V_2 = \sup_{\theta \in \mathbb{S}^{n-1}} \mathrm{Var}(\theta_1 X_1^2 + \dots + \theta_n X_n^2).$$

If additionally the distribution of X is invariant under permutations of coordinates, it yields a simpler bound

$$\mathbb{E}_\theta\, \rho(F_\theta, \Phi) \le \frac{c\, \log n}{n}\, \big(\mathbb{E}X_1^4 + \sigma_4^2\big).$$

Here the last term σ_4^2 may be removed in some cases, e.g. when $\mathrm{cov}(X_1^2, X_2^2) \le 0$.

These results can be sharpened under some additional assumptions on the shape of the distribution of X. We include the proof of the following important variant of the central limit theorem due to Klartag [123]: If the random vector X has an isotropic, coordinatewise symmetric log-concave distribution, then, for all $\theta \in \mathbb{S}^{n-1}$,

$$\rho(F_\theta, \Phi) \le c \sum_{k=1}^n \theta_k^4$$

up to some absolute constant c. Here, the average value of the right-hand side is of order $1/n$. Although the class of log-concave probability distributions is studied in many investigations, their basic properties are discussed in this text as well. In particular, we include the proof of the Brascamp–Lieb inequality, which serves as a main tool in Klartag's theorem.

Finally, in the last chapter we conclude with brief historical remarks on results about randomized variants of the central limit theorem, in which coefficients have a special structure.

Acknowledgements. This work started in 2015 during the visit of the first author to the Bielefeld University, Germany, and he is grateful for their hospitality. The authors were supported by the SFB 701 and the SFB 1283/2 2021 – 317210226 at Bielefeld University. The work of the first author was also supported by the Humboldt Foundation, the Simons Foundation, and NSF grants DMS-1855575, DMS-2154001.

It is our great pleasure to thank Michel Ledoux for valuable comments on the draft version of the monograph.

In Memoriam. Shortly after completion of this book, Gennadiy Chistyakov passed away after a prolonged illness in December 2022. We mourn the loss of our friend and colleague.

Contents

Part III Analysis on the Sphere

Part IV First Applications to Randomized Sums