Concentration of Empirical Distribution Functions for Dependent Data under Analytic Hypotheses

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Ji Hee Kim

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor
SERGEY BOBKOV

May, 2013

Acknowledgements

This thesis was partially supported by the NSF grant N1106530.

Abstract

The concentration property of empirical distribution functions is studied under the Lévy distance for dependent data whose joint distribution satisfies analytic conditions expressed via Poincaré-type and logarithmic Sobolev inequalities. The concentration results are then applied to the following two general schemes. In the first scheme, the data are obtained as coordinates of a point randomly selected within given convex bodies (and more generally – when the sample obeys a log-concave distribution). In the second scheme, the data represent eigenvalues of symmetric random matrices whose entries satisfy the indicated analytic conditions.

Contents

Chapter 1

Introduction

There are many well known classical results about concentration of empirical distribution functions in the i.i.d. case. However, there are also many natural examples that motivate the study for non-i.i.d. case. These examples include stationary processes, random matrices, and sampling from convex bodies. In order to study concentration for dependent data, one must assume some properties on the joint distribution. Two of such hypothese are Poincaré-type inequalities and logarithmic Sobolev inequalities. These are very useful and natural inequalities in studying concentration phoenomena since the first implies exponential concentration and the latter implies Gaussian concentration.

With no other additional assumption on the individual distributions or moments, we measure concentration under Lévy distance. In general, this classical distance is weaker than the Kolmogorov distance but is more appropriate when speaking about the weak closeness and convergence of probability distributions on the real line to an unknown distribution. Our starting point was a recent work by Bobkov and Götze (cf. [8]), where the concentration problem for distribution functions is considered with respect to Kantorovich-Rubinstein distance. Here we develop different tools based on the study of concentration of empirical characteristic functions and applying Zolotarev's Berry-Esseen's bound for the Lévy distance (cf. Appendix A).

In Chapter 2, we present the problem of concentration of empirical distribution functions including its history and examples. Then in Chapter 3, we introduce the Poincaré-type inequalities and its general properties. Some examples of probability measures that satisfy the inequality and some known characterizations are discussed.

The inequality's relation to isoperimetric inequalities is also outlined. We then apply the Poincaré-type inequalities to linear functionals of empirical measures. These are used in Chapter 4 to prove our main result. We discuss concentration results under Lévy distance and deviations from the mean using Orlicz norms (cf. Appendix B). Then we derive similar results under the log-Sobolev inequalities in Chapter 5. In Chapter 6, we discuss the K-L-S conjecture (cf. [22]) and possible application of our main results to convex bodies. We also obtain some refinements of our results in the case of random matrices. Finally, since Zolotarev bound and Orlicz norms play a crucial role in achieving our bounds, we collect in the Appendix the proof of Zolotarev bound and the relation of Orlicz norms to L^p norms.

Chapter 2

The Problem of Concentration of Empirical Distributions

2.1 Definitions, Notations, and the Statement of the Problem

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and $(S, \mathcal{B})$ be a measurable space. Consider a sequence of random elements $X_i : \Omega \to S$ that are identically distributed with a common law $\mathcal{L}(X_i) = F$. When such F is unknown, the general problem in statistics is to estimate F on the basis of particular values, $X_1 = x_1, \ldots, X_n = x_n$, called observations.

The standard approach to this problem is to consider empirical distributions.

Definition 2.1. The random probability measures

$$F_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i},$$

are called empirical distributions, which can be seen as a random probability measure on S.

The most important and classical case is when S represents the real line $\mathbf{R}$, so that the observations are certain real numbers. In this case, the empirical measures may be associated with the corresponding empirical distribution functions defined by

$$F_n(x) = \frac{1}{n} \operatorname{card}\{i \leq n : X_i \leq x\}.$$

Note that for any measurable B in S, $F_n(B)$ is a random variable with $\mathbf{E}F_n(B) = F(B)$, and it is reasonable to believe that $F_n \approx F$ when n is large with high probability. In fact, it is a well-known classical result that when $X_1, \ldots, X_n$ are independent with common law F, it is indeed true that $F_n \approx F$.

A more general interesting problem would be the case where $X_1, \ldots, X_n$ are dependent, although in general it is not the case that $F_n \approx F$. However, the problem makes sense under various hypotheses about the joint distribution $\mathcal{L}(X_1, \ldots, X_n) = \mu$.

Problem 1. Study the closeness of F_n to F as n grows large when $X_1, \ldots, X_n$ are dependent but something is known about their joint distribution μ.

Then as a further generalization, we can also drop the assumption that X_i are identically distributed. Now, we take the average of marginal distributions, so defining

$$F = \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(X_i)$$

gives $\mathbf{E}F_n = F$ and on the statistical language, F_n is an unbiased estimate for F. Therefore, again, we can expect F_n to be close to F when n is large under certain hypotheses.

Problem 2. Is it still true that $F_n \approx F$ when n is large with the above F on average? In particular, we use the Lévy distance in the weak topology on the space of all Borel probability measures to approximate the closeness of F_n and F. So our question is, can we find a bound for the Lévy distance $L(F_n, F)$?

2.2 Classical Results in the I.i.d. Case

Let X_n, $n \geq 1$, be independent identically distributed random elements in S with common distribution F.

2.2.1 Law of large numbers

For any measurable set B in S,

$$F_n(B) = \frac{1}{n} \sum_{k=1}^{n} \xi_k,$$

where $\xi_k = 1_B(X_k)$. Clearly, ξ_k represent independent random variables, taking the value 1 with probability $p = \mathbf{P}\{X_k \in B\} = F(B)$ and zero with probability $q = 1 - p$. By the classical strong law of large numbers (Kolmogorov's theorem), applied to the sequence ξ_k,

$$\lim_{n \to \infty} F_n(B) = F(B)$$

with probability 1.

More generally, for any F-integrable function $g : S \to \mathbf{R}$, with probability 1, we have

$$\int g \, dF_n = \frac{1}{n} \sum_{k=1}^{n} g(X_k) \to \int g \, dF.$$

In this statement, the law of large numbers is applied to the sequence $\xi_k = g(X_k)$.

However, in general one cannot guarantee that the convergence $F_n(B) \to F(B)$ is uniform over all B. That is, it is not true that with probability 1

$$\frac{1}{2} \, \|F_n - F\|_{\mathrm{TV}} = \sup_{B \in \mathcal{B}} |F_n(B) - F(B)| \to 0.$$

Here, $\|F_n - F\|_{\mathrm{TV}}$ denotes the total variation distance. Indeed, F may be a continuous measure, while F_n is always discrete. In this case F_n and F are orthogonal, which implies that $\|F_n - F\|_{\mathrm{TV}} = 2$.

2.2.2 Weak convergence of empirical measures

Let (S, d) be a metric space. In the space $Z(S)$ of all Borel probability measures on S, one considers the topology of the weak convergence.

Definition 2.2. We say $\nu_n \to \nu$ weakly in $Z(S)$, if and only if

$$\int g\,d\nu_n \to \int g\,d\nu \text{ for any bounded continuous function } g \text{ on } S.$$

This topology is metrizable by many standard metrics, one of which is the Lévy-Prokhorov metric

$$\pi(\mu,\nu) = \min\{\varepsilon \geq 0 : \mu(A) \leq \nu(A^\varepsilon) + \varepsilon, \text{ for all Borel sets } A \subset S\},$$

where $A^\varepsilon = \{x \in S : d(x,a) < \varepsilon, \text{ for some } a \in A\}$ denotes an open ε-neighborhood of the set A. In particular, $\nu_n \to \nu$, if and only if $\pi(\nu_n, \nu) \to 0$.
See more information about this convergence in the books [2] and [15].

As for empirical measures F_n, we have, for any bounded continuous function g on S,

$$\int g\,dF_n \to \int g\,dF \quad \text{with probability 1},$$

as mentioned above. In this statement one cannot automatically change the order of the expressions "for any bounded continuous..." and "with probability 1", so that to strengthen this property. This is a content of a separate theorem, due to Varadarajan, see p. 399 of [15] and [31]:

Theorem 2.3. *(Varadarajan) Assume that the metric space S is separable. Then with probability 1, we have that $F_n \to F$ weakly. Equivalently, $\pi(F_n, F) \to 0$ with probability 1.*

2.2.3 Glivenko-Cantelli's theorem

On the real line $S = \mathbf{R}$, it is well-known that the weak convergence $\nu_n \to \nu$ in $Z(\mathbf{R})$ can be expressed in terms of the associated distribution functions $F_n = \nu_n(-\infty, x]$, $F = \nu(-\infty, x]$. Namely, $\nu_n \to \nu$, if and only if $F_n(x) \to F(x)$, for any point x of continuity of F.

As well as in the general case, the weak convergence can be quantified by many metrics that generate the topology of the weak convergence in the space $Z(\mathbf{R})$. In addition to the Lévy-Prokhorov's metric, one may use the Lévy distance $L(F.G)$. Thus, the above Varadarajan's theorem may also be stated on the real line, that is, when all X_n are i.i.d. random variables, as the property that $L(F_n, F) \to 0$ with probability 1.

It turns out out that this particular case can be strengthened in terms of the Kolmogorov's distance

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

Note that in general $L(F,G) \leq \rho(F,G)$, but ρ cannot be bounded from above in terms of L. Nevertheless, for empirical measures one can obtain such a stronger statement about the convergence. Here and below, we may freely identify probability measures on the real line and their distribution functions.

Theorem 2.4. *(Glivenko-Cantelli [15], p. 400) For empirical measures F_n, constructed for i.i.d. random variables X_n with the common distribution F, we have $\rho(F_n, F) \to 0$ with probability 1.*

2.2.4 Kolmogorov-Smirnov's test

Since $D_n = \rho(F_n, F) \to 0$, one may ask about the rate of convergence. Note that, by the central limit theorem, for any fixed $x \in \mathbf{R}$,

$$\sqrt{n}\,(F_n(x) - F(x)) \to N(0,\sigma^2), \quad \text{weakly in distribution,}$$

where $\sigma^2 = F(x)(1 - F(x))$. Indeed,

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \xi_k,$$

where $\xi_k = 1_{(-\infty,x]}(X_k)$. Clearly, ξ_k represent independent random variables, taking the value 1 with probability $p = \mathbf{P}\{X_k \leq x\} = F(x)$ and zero with probability $q = 1 - p$. Hence, $\sigma^2 = \mathrm{Var}(\xi_k) = pq$, and by the central limit theorem, the sums

$$S_n = \xi_1 + \cdots + \xi_n$$

are asymptotically normal, that is, they satisfy

$$\frac{S_n - \mathbf{E}\,S_n}{\sqrt{n}} \to N(0,\sigma^2), \quad \text{weakly in distribution.}$$

This is exactly what was stated in terms of $F_n(x)$.

Therefore, for any fixed x, the difference $|F_n(x) - F(x)|$ is about $1/\sqrt{n}$. One may naturally expect a similar property when taking the supremum over all x.

Theorem 2.5. *(Kolmogorov) For empirical measures F_n, constructed for i.i.d. random variables X_n with the common continuous distribution function F, for all $x \in \mathbf{R}$,*

$$\mathbf{P}\left\{\sqrt{n}\ \rho(F_n, F) \le x\right\} \to K(x) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}e^{-2i^2x^2}.$$

This function K is called the Kolmogorov distribution function [17].

On the basis of this theorem, one applies the so-called Kolmogorov-Smirnov's test and this theorem also admits a simple explanation in terms of the Brownian motion, and is related to the so-called invariance principle.

2.3 Examples of Non-i.i.d. Observations

2.3.1 Stationary processes

One natural generalization of the i.i.d. case is described by the model where we have a (strongly) stationary sequence $(X_n)_{n \geq 0}$ of random variables. By the definition, this sequence is stationary, if the joint distribution of any finite subsequence of length $h \geq 1$

$$(X_n, X_{n+1}, \dots, X_{n+h-1})$$

does not depend on n. In particular, all X_n must have equal distributions.

There are some quantities, called "mixing", which may be used to quantify the strength of dependence between X_n's.

2.3.2 Random matrices

Let ξ_{jk}, $1 \leq j \leq k \leq n$, be independent identically distributed random variables. Then one may define a random symmetric matrix W of size $n \times n$ with entries $W_{jk} = \frac{1}{\sqrt{n}} \xi_{jk}$ for $j \leq k$ and $W_{jk} = \frac{1}{\sqrt{n}} \xi_{kj}$ for $j > k$. This matrix has n real eigenvalues X_k, which can be arranged in increasing order:

$$X_1 \leq \cdots \leq X_n.$$

The random probability measures

$$F_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}$$

are of a large interest in Probability and Physics. They are called spectral empirical distributions. The measure $F = \mathbf{E}\, F_n$, which also depends on n, is the so-called mean spectral empirical distribution.

2.3.3 Sampling from convex bodies

Let a random vector $X = (X_1, \dots, X_n)$ be uniformly distributed in a convex body $K \subset \mathbf{R}^n$. In this model particular values $X_1 = x_1, \dots X_n = x_n$ may be obtained by picking at random a point from K.

2.4 Some Results for Stationary Sequences

There are some related results for stationary sequences (which may be viewed as a natural extension of the family of i.i.d. sequences). Let us formulate one of the results, which appears in T. Y. Kim [23]. For simplicity, we consider one-dimensional random variables.

Let $(X_n)_{n\in\mathbf{Z}}$ be a (strictly) stationary sequence of random variables. By the stationarity, all random variables have a common distribution function F, which will be assumed to be continuous. Consider a new stationary sequence $Y_n = F(X_n)$ and the associated empirical distribution functions

$$G_n(x) = \frac{1}{n}\,\mathrm{card}\{k = 1, \dots, n : Y_k \le x\}, \qquad x \in \mathbf{R}.$$

By the construction, all Y_k take values in (0,1) and have a uniform distribution $G(x) = x$, for $0 < x < 1$. Like in the i.i.d.-case, introduce the Kolmogorov-Smirnov statistic

$$D_n = \sup_x |G_n(x) - G(x)|.$$

Again, the problem is to determine whether or not D_n is close to zero, and if possible – at rate $1/\sqrt{n}$. This can be done in terms of so-called ϕ-mixing.

Given $n \ge 1$ and $k \in \mathbf{Z}$, denote by $\mathcal{F}_k$ the σ-algebra, generated by the random variables X_i, $i \le k$, and by $\mathcal{F}^{k+n}$ the σ-algebra, generated by the random variables X_j, $j \ge k+n$. Define $\phi_n \ge 0$ as the best constant in the inequality

$$|\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)| \le \phi_n \mathbf{P}(A), \qquad A \in \mathcal{F}_k,\ B \in \mathcal{F}^{k+n}.$$

Note that $\phi_n = 0$, if X_n are independent. So, in general, this quantity may be used to measure the strength of dependence between "the past" and "the future" of the process. If ϕ_n is sufficiently small, one can say that all X_i are almost independent of all X_j, when $j - i \ge n$ and n is large.

Theorem 2.6. *If F is continuous, and $\sum_{n=1}^{\infty} \phi_n < +\infty$, then for all $k \ge 1$,*

$$\mathbf{P}\Big\{\sqrt{n}\ \rho(F_n, F) \ge x\Big\} \le \frac{C}{x^k}, \qquad x > 0,$$

with some constant C, depending on k.

The obtained deviation inequality is a bit weaker than the subgaussian bound of [16] for the i.i.d-case, but it also implies that

$$\mathbf{E}\ \rho(F_n, F) \leq \frac{C}{\sqrt{n}}$$

with some constant C.

The theorem is proved in [23] in a more general setting of $\mathbf{R}^d$-valued stationary random vectors X_n; it has refined a previous result of Sen [29].

Chapter 3

Poincaré-type Inequalities and Related Analytic Hypotheses

In this chapter we introduce and discuss basic properties of Poincare-type inequalities on the Euclidean spaces $\mathbf{R}^n$ including tensorization and stability under Lipschitz transforms (cf. Section 3.1). The probability measures satisfying these inequalities possess a number of useful properties. In particular, one can control the distribution of Lipschitz and other "smooth" functionals under such measures (Section 3.2). In Section 3.3 we recall a Mucheknhoupt-type characterization of probability measures satisfying Poincare-type inequalities on the real line and then give a few examples in Section 3.4. The problem of characterization in spaces of dimension 2 and higher is rather difficult and remains open. Nevertheless, probability measures from several standard classes are known to satifisfy Poincare-type inequalities. One such class is the family of all probability measures whose densities are log-concave. Some results about this class are recalled in Section 3.5. The general relationship between Poincare-type and Cheeger-type isoperimetric inequalities is discussed in Section 3.6.

3.1 Poincaré-type Inequalities. Tensorization.

Poincaré-type inequalities are known to hold with finite σ for many natural families of probability measures μ on $\mathbf{R}^n$. However, the problem on effective bounding of the Poincaré constant σ^2 is not simple.

Definition 3.1. A probability measure μ on $\mathbf{R}^n$ satisfies a Poincaré-type inequality with constant σ^2, $\sigma > 0$, if for any bounded smooth function f on $\mathbf{R}^n$ with gradient ∇f

$$\mathrm{Var}_\mu(f) \le \sigma^2 \int |\nabla f|^2 d\mu.$$

Here, $\mathrm{Var}_\mu(f)$ is the variance of f under the measure μ, and $\sigma \ge 0$ is a constant depending on μ only. Note that the inequality itself is required to hold in the class of all bounded smooth functions $f : \mathbf{R}^n \to \mathbf{R}$. However, the smoothness of f may be relaxed to the property of being "locally Lipschitz", which means that near every point x the function f has a finite Lipschitz semi-norm

$$\sup_{0<|x-y|<r} \frac{|u(x)-u(y)|}{|x-y|}.$$

In this case the generalized modulus of the gradient

$$|\nabla u(x)| = \limsup_{y \to x} \frac{|u(x)-u(y)|}{|x-y|}$$

represents a finite Borel measurable function (see [9] for discussion and a general theory).

Now we state a very useful lemma that extends the Poincaré-type inequality to product measures.

Lemma 3.2. *Let* $(\Omega, \mu) = (\Omega_1, \mu_1) \times \cdots \times (\Omega_n, \mu_n)$ *and* $f : \Omega \mapsto \mathbf{R}$ *be measurable. Denote* f_i *as function on* Ω_i *as defined by* $f_i(x_i) = f(x_1, \ldots, x_n)$ *where* $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ *are fixed. Then*

$$\mathrm{Var}_\mu(f) \le \sum_{i=1}^n \int \mathrm{Var}_{\mu_i}(f_i) d\mu.$$

As an application of the lemma, we get the following well-known theorem (see [25]):

Theorem 3.3. *Let* μ_i *on* $\Omega_i = \mathbf{R}$ *satisfy the Poincaré-type inequality. Then*

$$\mathrm{Var}_{\mu_i}(f_i) \le \sigma^2 \int_{-\infty}^{+\infty} |\nabla f_i|^2 d\mu_i$$

for every bounded smooth function f_i *on* $\mathbf{R}$ *with some constant* σ^2*. Then the product measure* $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ *on the product space* $\mathbf{R}^n$ *satisfies the Poincaré-type inequality. So for every bounded smooth function* f *on* Ω*,*

$$\mathrm{Var}_\mu(f) \leq \sigma^2 \int |\nabla f|^2 \, d\mu.$$

Hence, the product measure also satisfies the Poincaré-type inequality with the same constant σ^2.

The Poincaré-type inequality is also stable under Lipschitz transformation.

Theorem 3.4. *Let* μ *be a probability measure on* $\mathbf{R}^n$ *that satisfies a Poincaré-type inequality with constant* σ^2*. Suppose* $T : \mathbf{R}^n \mapsto \mathbf{R}^k$ *is a Lipschitz transformation so that* $\|Tx - Ty\|_{\mathbf{R}^k} \leq C\|x - y\|_{\mathbf{R}^n}$*. Let* $\nu = T\mu^{-1}$*. Then* ν *also satisfies a Poincaré-type inequality with constant* $(C\sigma)^2$*. So we have*

$$\mathrm{Var}_\nu(f) \leq C^2\sigma^2 \int |\nabla f|^2 d\nu.$$

3.2 Integrability Under Poincaré-type Inequalities

Theorem 3.5. *Suppose μ on $\mathbf{R}^n$ satisfies PI(σ^2). Then any Lipschitz function f on $\mathbf{R}^n$ has a finite exponential moment: If $\int f d\mu = 0$ and $\|f\|_{Lip} \leq 1$, then*

$$\int e^{tf/\sigma} d\mu \leq \frac{2+t}{2-t}, \qquad 0 < t < 2.$$

Moreover, for any locally Lipschitz f on $\mathbf{R}^n$ with μ-mean zero,

$$\|f\|_p \leq \sigma p \|\nabla f\|_p, \qquad p \geq 2.$$

More precisely, if $|\nabla f|$ is in $L^p(\mu)$, then so is f, and the second inequality holds true with the standard notations $\|f\|_p = (\int |f|^p d\mu)^{1/p}$ and $\|\nabla f\|_p = (\int |\nabla f|^p d\mu)^{1/p}$ for $L^p(\mu)$-norms. The property of being loally Lipschitz means that the function f has a finite Lipschitz semi-norm on every compact subset of $\mathbf{R}^n$. The first part of the proposition is proved in [11] in Proposition 4.1 and see Theorem 4.1 in [12] for the second inequality.

Choosing in Theorem 3.5, for example, $t = 1$, we get

$$\int e^{f/\sigma} \leq 3.$$

With a similar bound written for $-f$, this gives

$$\int e^{|f|/\sigma} d\mu \leq 6.$$

We can now apply Chebyshev's inequality to obtain

Corollary 3.6. *Under PI(σ^2), for any $f : \mathbf{R}^n \to \mathbf{R}$ such that $\|f\|_{Lip} \leq 1$ and $\int f d\mu = 0$,*

$$\mu\{|f| \geq r\} \leq 6e^{r/\sigma}, \quad r \geq 0.$$

3.3 Characterization of Poincaré-type Inequalities on R

Let μ and ν be Borel measures on $(0, +\infty)$, not necessarily probability measures. We want to know whether there exists a finite C such that

$$\int_0^\infty f(x)^2 d\mu(x) \leq C \int_0^\infty f'(x)^2 d\nu(x)$$

holds true for all absolutely continuous $f : [0, +\infty) \mapsto \mathbf{R}$ with $f(0) = 0$.

We start with the classical Hardy inequality which was first published in 1920 by Hardy in [19].

For $p > 1$ and f, a nonnegative p-integrable function on $(0, \infty)$,

$$\int_0^\infty \left(\frac{1}{x} f(x)\right)^p dx \leq \left(\frac{p}{p-1}\right)^p \int_0^\infty f'(x)^p dx.$$

This inequality is the standard form of the Hardy inequality one can find in many sources. The constant here is sharp if we take $f(x) = 0$ for $x < 1$, $f(x) = x^{-\frac{1}{p}-\epsilon}$ for $x \geq 1$ and take $\epsilon \to 0$.

The more general form of the above inequality considers the case where $\frac{d\mu(x)}{dx} = x^b$ and $\frac{d\nu(x)}{dx} = x^{b+1}$ for some b to be defined below (see [20], p.245).

For $p > 1$ and b such that $bp < -1$,

$$\int_0^\infty |x^b f(x)|^p dx \leq \left(\frac{-p}{bp+1}\right)^p \int_0^\infty |x^{b+1} f'(x)|^p dx.$$

The constant here is also sharp.

Now for the case of general measures, Muckenhoupt proves the following theorem in [27]. The original theorem by Muckenhoupt holds for $1 \leq p \leq +\infty$ but we note here that the case $p = 2$ was originally proved by Kac and Krein in [21] published in 1958. For our purposes, we take the case $p = 2$ of the theorem.

Theorem 3.7. *Let μ and ν be Borel meausres on $(0, +\infty)$. Then there is a finite C for which*

$$\int_0^\infty |f(x)|^2 d\mu(x) \leq C^2 \int_0^\infty |f'(x)|^2 d\nu(x)$$

if and only if

$$B = \sup_{r>0} \left[\mu([r, +\infty)) \int_0^\infty \frac{1}{p_\nu(t)} dt\right] < \infty$$

where $\dfrac{d\nu(x)}{dx} = p_\nu(x)$. *Furthermore, if* C *is the least constant for which the inequality holds, then* $B \le C \le 2B$.

Now we give the characterization on the whole real line.

Let μ be a probability measure on the line with median m, that is, $\mu(-\infty, m) \le \frac{1}{2}$ and $\mu(m, +\infty) \le \frac{1}{2}$. Define the quantities

$$A_0(\mu) = \sup_{x<m} \left[\mu(-\infty, x) \int_{-\infty}^{x} \frac{dt}{p_\mu(t)} \right],\ A_1(\mu) = \sup_{x>m} \left[\mu(x, +\infty) \int_{x}^{+\infty} \frac{dt}{p_\mu(t)} \right]$$

where p_μ denotes the density of the absolutely continuous component of μ (with respect to Lebesgue measure), and where we set $A_0 = 0$, respectively $A_1 = 0$, if $\mu(-\infty, m) = 0$ or $\mu(m, +\infty) = 0$. Then we have:

Theorem 3.8. *The measure* μ *on* $\mathbf{R}$ *satisfies PI(*σ^2*) with some finite constant, if and only if both* $A_0(\mu)$ *and* $A_1(\mu)$ *are finite. Moreover, the optimal value of* σ^2 *satisfies*

$$c_0(A_0(\mu) + A_1(\mu)) \le \sigma^2 \le c_1(A_0(\mu) + A_1(\mu)),$$

where c_0 *and* c_1 *are positive universal constants.*

The above theorem can be found in [5] or [8]. However, we also note that by setting $\mu = \nu$ and dividing μ into two pieces $\mu\big|_{(-\infty, m]}$ and $\mu\big|_{[m, \infty)}$ Theorem 3.8 can be obtained from Theorem 3.7.

3.4 Examples

In many examples, one can prove that μ satisfies PI(σ^2) using direct arguments without using Theorem 3.8. In [13], Borovkov and Utev give a sufficient condition for μ to satisfy PI(σ^2).

Theorem 3.9. *Let $\frac{d\mu(x)}{dx} = p(x)$ on $\mathbf{R}$ with distribution function $F(x)$. Assume that*

$$C = \operatorname*{ess\,sup}_{x\in\mathbf{R}} \frac{\min(F(x), 1-F(x))}{p(x)} < \infty.$$

Then μ satisfies PI(σ^2) with $\sigma^2 \le 4C$.

For example, if $\mu \sim \mathrm{Exp}(1)$, that is, $\frac{d\mu(x)}{dx} = e^{-x}$, $x > 0$, we have $F(x) = 1 - e^{-x}$, so

$$C = \sup_{x>0} \frac{\min(e^{-x}, 1-e^{-x})}{e^{-x}} = \sup_{x>0} \min(1, e^x - 1) = 1.$$

Therefore, $\sigma^2 \le 4$.

But this can be proved directly.

Indeed, if f is bounded and smooth with $f(0) = 0$, we have by integration by parts,

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x)d\mu(x) &= \int_0^{+\infty} f(x)e^{-x}dx \\ &= \int_0^{+\infty} f'(x)e^{-x}dx \\ &= \int_0^{+\infty} f'd\mu. \end{aligned}$$

Applying this identity to f^2, we then get by Cauchy-Schwarz inequality

$$\int f^2 d\mu = \int 2ff'd\mu \le 2\sqrt{\int f^2 d\mu}\sqrt{\int f'^2 d\mu}.$$

Applying the identity again we get,

$$\int f^2 d\mu \le 4 \int f'^2 d\mu, \quad \text{provided} \quad f(0) = 0.$$

Therefore,

$$\mathrm{Var}(f) = \mathbf{E}_\mu (f - \mathbf{E}_\mu f)^2 \le \mathbf{E}_\mu (f - f(0))^2 \le 4\, \mathbf{E}_\mu f'^2.$$

We remark that $\sigma^2 = 4$ here is optimal. To see this, one may consider $f(x) = e^{\lambda x}$, $0 \leq \lambda < 1/2$ and then let $\lambda \to 1/2$.

Our next important example is the Gaussian measure, where the density is given by,

$$\frac{d\mu(x)}{dx} = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Again, assume that f is bounded smooth function with $f(0) = 0$. For $x > 0$, write $f(x) = \int_0^x f'(y)dy$ so that

$$f^2(x) \leq x \int_0^x f'(y)^2 dy.$$

Now, integrating both sides gives

$$\begin{aligned}
\int_0^\infty f^2(x)\phi(x)dx &\leq \iint\limits_{0<y<x} x f'(y)^2 \phi(x) dx dy \\
&= \int_0^\infty \left[\int_y^{+\infty} x\phi(x)dx \right] f'(y)^2 dy \\
&= \int_0^\infty \phi(y) f'(y)^2 dy.
\end{aligned}$$

Similarly,

$$\int_{-\infty}^0 f(x)^2 \phi(x) dx \leq \int_{-\infty}^0 \phi(y) f'(y)^2 dy.$$

Therefore,

$$\mathrm{Var}_\mu(f) \leq \int f^2 d\mu \leq \int f'^2 d\mu.$$

The constant $\sigma^2 = 1$ here is optimal if one applies $f(x) = x$ or any linear functions.

In both of the examples, we applied the Poincaré-type inequality for all bounded smooth functions f. We remark here that if the inequality holds on $\mathbf{R}^n$ for all bounded smooth functions f, it will still remain to hold for all smooth f. More precisely, if $\int |\nabla f|^2 d\mu < +\infty$, then necessarily $\int f^2 d\mu < +\infty$ and the inequality holds. To see this, one may apply the inequality to truncations

$$f(x) = \begin{cases} c, & \text{if} \quad f(x) > c \\ f(x), & \text{if} \quad -c \leq f(x) \leq c \\ -c, & \text{if} \quad f(x) < -c. \end{cases}$$

Rewrite the Poincaré-type inequality as

$$\mathrm{Var}_\mu(f) = \frac{1}{2} \iint |f(x) - f(y)|^2 d\mu(x) d\mu(y) \leq \sigma^2 \int |\nabla f|^2 d\mu$$

and then let $c \to \infty$.

3.5 Estimation of Poincaré Constant - The Case of Log-Concave Distributions

A probability measure F on the real line, having a log-concave density, always satisfies a Poincaré-type inequality. Moreover, for the best constant σ^2 we have

$$c_0 \ \mathrm{Var}(\xi) \leq \sigma^2 \leq c_1 \ \mathrm{Var}(\xi),$$

where c_1 and c_1 are positive numerical constants, and ξ is a random variable with distribution F. This example can be generalized.

Any probability measure μ on $\mathbf{R}^n$, having a log-concave density, satisfies a Poincaré-type inequality. Moreover, for the best constant σ^2 we have

$$\sigma^2 \leq C \int |x|^2 \, d\mu(x),$$

where C is a positive numerical constant (see [3]). The integral on the right may also be written as $\mathbf{E}\ |X|^2$, since as before X has distribution μ. In fact (see [6]), this bound can be sharpened to

$$\sigma^2 \leq C \ \mathrm{Var}^{1/2}(|X|^2).$$

In many practical examples, the coordinates $X_1, \dots, X_n$ behave like weakly dependent random variables in the sense that

$$\mathrm{Var}(|X|^2) = \mathrm{Var}(X_1^2 + \cdots + X_n^2) \leq Cn,$$

and then we obtain that

$$\sigma^2 \leq C\sqrt{n}$$

with some other constant C.

In fact, in many interesting examples σ^2 is bounded.

Example 1. For standard Gaussian measure, $\sigma^2 = 1$ as shown in Section 3.4. This is consistent with the result for the case of log-concave distributions since the Gaussian measure is log-concave.

Example 2. More generally, let $\mu = \nu^n$ be the product measure, where ν is a log-concave probability measure on the real line. In this case, $\sigma^2(\mu) = \sigma^2(\nu)$, according to the tensorization property.

Example 3. Uniform distribution on the ℓ^p-balls of $\mathbf{R}^n$, more precisely, on the sets $K(p,n) = \{(x_1,\dots,x_n) : |x_1|^p + \cdots + |x_1|^p \le r^p\}$ where the constant r is chosen so that $K(p,n)$ has volume 1. The property that σ^2 is uniformly bounded for all $p \ge 1$ and $n \ge 1$ is proved by S. Sodin in [30].

We will return to this problem in Section 6.1.

3.6 Connection with Isoperimetric Inequalities

Isoperimetric problems are concerned with determining sets of the minimal boundary measure among sets of the same measure. We will introduce one particular type of isoperimetric inequality that is associated with Poincaré-type inequalities in this section.

Let (X, d) be a metric space equipped with Borel probability measure μ.

Definition 3.10. For a Borel set $A \in X$ we can define its boundary measure, or Minkowski content with respect to μ as

$$\mu^+(A) = \liminf_{\epsilon \downarrow 0} \frac{\mu(A^\epsilon) - \mu(A)}{\epsilon}$$

where $A^\epsilon = \{x \in X | \exists y \in A, d(x, y) < \epsilon\}$ is the open ϵ-neighborhood of A.

Hence, by isoperimetric inequality, one means inequalities of the form

$$\mu^+(A) \geq I(\mu(A)), \;\; A \subset X \;\text{ Borel},$$

where $I : [0, 1] \mapsto [0, +\infty)$. We call $I = I_\mu$ the isoperimetric function of μ if it is the largest function such that the inequality holds for every Borel set $A \in X$.

For our purposes, we will consider the function $I(t) = h \min\{t, 1 - t\}$, so our inequality will be

$$\mu^+(A) \geq h \min\{\mu(A), 1 - \mu(A)\}.$$

Inequalities of this type are called isoperimetric inequalities of Cheeger type.

Definition 3.11. The best constant in this inequality $h = h_\mu$, is called the (Cheeger) isoperimetric constant of μ.

We give an equivalent functional formation of the inequality (see [25]):
$\forall f : \mathbf{R}^n \mapsto \mathbf{R}$ locally Lipschtz,

$$h \int |f - m| d\mu \leq \int |\nabla f| d\mu \tag{3.1}$$

where $m = m(f)$ is a median of f under μ.

The connection with Poincaré constants $\sigma^2 = \sigma_\mu^2$ in the Poincaré-type inequality is given in the following two theorems.

Theorem 3.12. *(Cheeger, [14]) If μ satisfies the isoperimetric inequality with Cheeger constant h_μ, then μ satisfies the Poincaré-type inequality with Poincaré constant σ_μ^2 which satisfies*

$$\frac{1}{\sigma_\mu^2} \geq \frac{1}{4} h_\mu^2.$$

Therefore, the Poincaré constant σ_μ^2 can be bounded from above. We give a proof of Theorem 3.12 since the argument is simple, but we first prove the funcitonal equivalent form of the Cheeger isoperimetric inequality.

Proof of equation 3.1. Let $f : \mathbf{R}^n \mapsto \mathbf{R}$ be locally Lipschitz and suppose $m(f) = 0$. Then we have the following inequality due to Bobkov and Houdré from [10]

$$\int |\nabla f| d\mu \geq \int_{-\infty}^{+\infty} \mu^+(f > t) dt.$$

If $f \geq 0$, then by the Cheeger type isoperimetric inequality we have

$$\begin{aligned} \int |\nabla f| d\mu &\geq \int_0^{+\infty} \mu^+(f > t) dt \\ &\geq \int_0^{+\infty} h_\mu \min\{\mu(f > t), 1 - \mu(f > t)\} dt \\ &= \int_0^{+\infty} h_\mu \mu(f > t) dt = h \int f d\mu. \end{aligned} \tag{3.2}$$

In the general case (still with $m(f) = 0$), write $f = f_1 - f_2$ where $f_1 = f\big|_{f \geq 0}$ and $f_2 = -f\big|_{f \leq 0}$. Then we still have $m(f_1) = m(f_2) = 0$ and $f_1, f_2 \geq 0$, so we can apply equation 3.2 to both functions to get

$$\int |\nabla f_1| d\mu \geq h_\mu \int f_1 d\mu \quad \text{and} \int |\nabla f_2| d\mu \geq h_\mu \int -f_2 d\mu.$$

Adding both inequalities gives us

$$\int |\nabla f| d\mu \geq h_\mu \int |f| d\mu.$$

Finally, we can remove the assumption $mf(0) = 0$ since subtracting $m(f)$ from f will not affect ∇f. Therefore, we have proven

$$\int |\nabla f| d\mu \geq h_\mu \int |f - m| d\mu.$$

$\square$

Now we prove Theorem 3.12.

Proof of Theorem 3.12. Assume $f : \mathbf{R}^n \mapsto \mathbf{R}$ is locally Lipschitz, $f \geq 0$ and $m(f) = 0$. Then since $m(f^2) = 0$ applying the functional form, equation 3.2, to f^2 gives

$$h_\mu \int f^2 d\mu \leq \int |\nabla f^2| d\mu = 2 \int f |\nabla f| d\mu.$$

Then by Cauchy-Schwarz, we get

$$h_\mu^2 \int f^2 d\mu \leq 4 \int |\nabla f|^2 d\mu.$$

In the general case (still with $m(f) = 0$), write $f = f_1 - f_2$ as we did in proof of equation 3.1 and the above inequality will hold. Then, we can drop the assumption on the median to get

$$\int |f - m|^2 d\mu \leq \frac{4}{h_\mu^2} \int |\nabla f|^2 d\mu.$$

Finally, since variance is minimized at $\mathbf{E}\, f$,

$$\mathrm{Var}_\mu(f) \leq \frac{4}{h_\mu^2} \int |\nabla f|^2 d\mu.$$

Thus, $\sigma_\mu^2 \leq 4/h_\mu^2$. □

The second theorem gives a lower bound on σ_μ^2.

Theorem 3.13. *(Ledoux, [26]) If μ is log-concave on $\mathbf{R}^n$,*

$$\sigma_\mu^2 \geq \frac{c}{h_\mu^2}$$

for some universal constant $c > 0$.

Therefore, a question one may consider is: Given μ on $\mathbf{R}^n$, when is $h_\mu > 0$?
If $n \geq 2$, this problem, as well as the problem of successfully determining finite σ_μ^2 is still open. If $n = 1$, we have discussed in section 3.3 how to bound σ_μ^2. Concerning h_μ, we have the following description by Bobkov and Houdré in [9].

Theorem 3.14. *Let μ be a probability measure on $\mathbf{R}$ with the distribution function $F(x) = (-\infty, x]$ and density $p(x)$ of the absolutely continous component of μ. Then*

$$h_\mu = \operatorname*{ess\,inf}_{0<F(x)<1} \frac{p(x)}{\min\{F(x), 1 - F(x)\}}.$$

From the theorem we can easily find a lower bound for h_μ.

Corollary 3.15. *If* $\min\{F(x), 1 - F(x)\} \leq Cp(x)$ *for all* $x \in \mathbf{R}$*, then* $h_\mu \geq 1/C$ *and therefore,* $\sigma_\mu^2 \leq 4C$.

Chapter 4

Empirical Poincaré-type Inequalities and Related Inequalities

Now, we are ready to consider Poincaré-type inequalities for empirical measures to study more general situation of an arbitrary random vector $X = (X_1, \dots, X_n)$ in $\mathbf{R}^n$ as in *Problem 2* where $X_1, \dots, X_n$ are not necessary indepent nor identically distributed. We consider F_n, the empirical distribution associated with the observations $X_1, \dots, X_n$ and F, the mean of the empirical distribution.

We want to measure the closeness of F_n to F. Since in general F is not continuous, it is hardly possible to work with the Kolmogorov distance $\rho(F_n, F)$ without additional assumptions on F (such as the existence and boundedness of its density). So, it seems more natural to choose weaker metrics, such as the Lévy distance $L(F_n, F)$ or the Lévy-Prokhorov distance $\pi(F_n, F)$, both of which are responsible for the weak convergence. Under moment assumptions, one may also involve Kantorovich-Rubinstein distance

$$W_1(F_n, F) = \int_{-\infty}^{+\infty} |F_n(x) - F(x)|\, dx.$$

4.1 Definition of Empirical Poincaré-type Inequality

By analytic hypotheses we mean integro-differential inequalities, imposed on the joint distribution μ of X. As the simplest example, one may consider Poincaré-type inequalities

$$\mathrm{Var}_\mu(f) \leq \sigma^2 \int |\nabla f|^2 \, d\mu,$$

where $\mathrm{Var}_\mu(f)$ is the variance of f under the measure μ, $\sigma \geq 0$ is a constant depending on μ, only, and the inequality itself is required to hold in the class of all bounded smooth functions $f : \mathbf{R}^n \to \mathbf{R}$.

Let us apply the Poincaré-type inequality to smooth functions of the form

$$f(x) = \frac{g(x_1) + \cdots + g(x_n)}{n} = \int g \, dF_n, \qquad x = (x_1, \ldots, x_n) \in \mathbf{R}^n,$$

where

$$F_n = \frac{1}{n} \sum_{k=1}^{n} \delta_{x_k}.$$

Then

$$|\nabla f(x)|^2 = \frac{g'(x_1)^2 + \cdots + g'(x_n)^2}{n^2} = \frac{1}{n} \int g'^2 \, dF_n.$$

Since $\int F_n \, d\mu = F$, the Poincaré-type inequality will take the form

$$\mathbf{E} \left| \int g \, dF_n - \int g \, dF \right|^2 \leq \frac{\sigma^2}{n} \int |g'|^2 \, dF.$$

This inequality may be called an "empirical Poincaré-type inequality". Note it remains to hold for complex-valued functions g, as well (by separating the real and imaginary parts of g).

Now we extend the empirical Poincaré-type inequality to all L^p spaces. Since $f = \int g dF_n$, consider for $p \geq 2$

$$|\nabla f|^p = \frac{1}{n^{p/2}} \left(\int g'^2 dF_n \right)^{p/2} \leq \frac{1}{n^{p/2}} \int |g'|^p dF_n.$$

Then we get

$$\mathbf{E}_\mu \, |\nabla f|^p \leq \frac{1}{n^{p/2}} \int |g'|^p dF.$$

And applying the second part of Theorem 3.5 to $f - \mathbf{E}_\mu \, f$ gives

$$\|f - \mathbf{E}_\mu \, f\|_p \leq \sigma p \|\nabla f\|_p.$$

So we can conclude:

Theorem 4.1. *If the distribution of $X = (X_1, \dots, X_n)$ in $\mathbf{R}^n$ satisfies PI(σ^2), then for any smooth F-integrable function $g : \mathbf{R} \to \mathbf{R}$, $p \geq 2$,*

$$\mathbf{E}\left|\int g dF_n - \int g dF\right|^p \leq \frac{(\sigma p)^p}{n^{p/2}} \int |g'|^p dF.$$

4.2 Concentration of Empirical Characteristic Functions

The empirical Poincaré-type inequality implies, for example,

$$\mathbf{E}\,\bigg|\int g\,dF_n - \int g\,dF\bigg| \leq \frac{\sigma}{\sqrt{n}}\,\bigg(\int |g'|^2\,dF\bigg)^{1/2}.$$

Thus, linear functionals of the empirical measures, $\int g\,dF_n$, deviate from their mean $\int g\,dF$ on average at rate $\frac{1}{\sqrt{n}}$ like in the i.i.d.-case – but under the additional assumption that g is smooth, such that the integral $\int |g'|^2\,dF$ is finite.

In particular, one cannot apply it to the indicator functions $g = 1_{(-\infty,x]}$ to get that

$$\mathbf{E}\,|F_n(x) - F(x)| \leq \frac{C}{\sqrt{n}},$$

which is known to be true in the i.i.d.-case. Nevertheless, at the expense of the rate, and properly changing the distance, one can suitably approximate indicator functions $g = 1_{(-\infty,x]}$ by smooth ones. The resulting bound should be weaker – and this is a content of a very intriguing problem.

Problem. Estimate $\mathbf{E}\,\rho(F_n,F)$ in terms of σ^2 and n, where ρ is a given metric, responsible for the weak convergence on the real line.

For example, it is known in [8] that if additionally $\mathbf{E}\,X_i = \mathbf{E}\,X_j$, for all i,j, then

$$\mathbf{E}\,W_1(F_n,F) \leq C\sigma\bigg(\frac{\log(n+1)}{n}\bigg)^{1/3}.$$

4.3 Concentration of Empirical Distributions in Lévy-metric

It is well-known that closeness of characteristic functions in some sense implies closeness of the distributions with repsect to the metrics that generate the topology of the weak convergence.

As an example, we will mention a result of Zolotarev about the Lévy distance. Let F and G be distribution functions on the line with the characteristic functions

$$f(t) = \int_{-\infty}^{+\infty} e^{itx}\, dF(x), \qquad g(t) = \int_{-\infty}^{+\infty} e^{itx}\, dG(x) \qquad (t \in \mathbf{R}),$$

respectively. Recall that the Lévy distance $L(F, G)$ is defined as the minimal value $h \geq 0$, such that

$$F(x-h) - h \leq G(x) \leq F(x+h) + h, \quad \text{for all } x \in \mathbf{R}.$$

Theorem 4.2. *(Zolotarev [32]) For any $T > 0$,*

$$L(F, G) \leq c_1 \int_0^T \frac{|f(t) - g(t)|}{t}\, dt + c_2\, \frac{\log(1+T)}{T}$$

where $c_1, c_2 > 0$ are universal constants.

See Appendix for more details of this theorem. Hence, if f is close to g on a long interval $[0, T]$, then $L(F, G)$ will be small. This theorem can be used in our problem.

Namely, one can apply Zolotarev's theorem to the empirical distribution functions to get

$$L(F_n, F) \leq c_1 \int_0^T \frac{|f_n(t) - f(t)|}{t}\, dt + c_2\, \frac{\log(1+T)}{T},$$

where f_n is the characteristic function of F_n and f is the characteristic function of F. Taking the expectation and using Fubini's theorem, we obtain that

$$\mathbf{E}\ L(F_n, F) \leq c_1 \int_0^T \frac{\mathbf{E}\ |f_n(t) - f(t)|}{t}\, dt + c_2\, \frac{\log(1+T)}{T}. \tag{4.1}$$

Now recall that an empirical Poincaré-type inequality has the form

$$\mathbf{E} \left| \int g dF_n - \int g dF \right|^2 \leq \frac{\sigma^2}{n} \int \left| g'(x) \right|^2 dF(x)$$

and taking $g(x) = e^{itx}$ with parameter t gives

$$\int g dF_n = \frac{1}{n}\sum_{k=1}^{n} e^{itX_k} = f_n(t) \text{ and } \int g dF = \int e^{itX} dF = f(t).$$

So with the above characteristic functions the empirical Poincaré-type inequality implies we get

$$\mathbf{E}\,|f_n(t) - f(t)| \leq \sqrt{\mathbf{E}\,|f_n(t) - f(t)|^2} \leq \frac{\sigma|t|}{\sqrt{n}}.$$

Substituting this back into Equation 4.1, we can obtain the following inequality

$$\mathbf{E}\,L(F_n, F) \leq c_1 \int_0^T \frac{\sigma}{\sqrt{n}} dt + \frac{c_2 \log T}{T} = c_1 \sigma \frac{T}{\sqrt{n}} + \frac{c_2 \log(1+T)}{T} \quad \text{for } T > 0.$$

In Zolotarev's bound, we may take constants $c_1 = 0.4$, and $c_2 = 4$. So we can consider following cases on the inequality

$$\mathbf{E}\,L(F_n, F) \leq C\left(\sigma \frac{T}{\sqrt{n}} + \frac{\log(1+T)}{T}\right).$$

Case I: $0 < \sigma \leq 1$. Then choosing $T = n^{1/4}$ gives

$$\mathbf{E}\,L(F_n, F) \leq C \frac{1 + \log(1 + n^{1/4})}{n^{1/4}}.$$

And since indeed $1 + \log(1 + n^{1/4}) \leq 5 \log(n+1)$, we can get

$$\mathbf{E}\,L(F_n, F) \leq C \frac{\log(n+1)}{n^{1/4}}$$

for some universal constant C.

Case II: $\sigma > 1$. Then we choose $T = \frac{n^{1/4}}{\sqrt{\sigma}}$ to get

$$\begin{aligned} c\,\mathbf{E}\,L(F_n, F) &\leq \frac{\sqrt{\sigma}}{n^{1/4}}\left(1 + \log\left(1 + \frac{n^{1/4}}{\sqrt{\sigma}}\right)\right) \\ &\leq \frac{\sqrt{\sigma}}{n^{1/4}}\left(1 + \log(1 + n^{1/4})\right) \\ &\leq C \frac{\sqrt{\sigma}}{n^{1/4}} \log(n+1), \end{aligned}$$

where C is some universal constant.

One can summarize these results as the following theorem:

Theorem 4.3. *Let $X = (X_1, \ldots, X_n)$ be a random vector in $\mathbf{R}^n$ and F_n be the empirical distribution associated with X. Let $F = \mathbf{E}\, F_n$ and suppose $\mu = \mathcal{L}(X)$ satisfies a Poincaré-type inequality with constant σ^2. Then*

1. *if $0 < \sigma \leq 1$, then*
$$\mathbf{E}\, L(F_n, F) \leq \frac{C \log(n+1)}{n^{1/4}}$$
2. *and if $\sigma \geq 1$, then*
$$\mathbf{E}\, L(F_n, F) \leq C \frac{\sqrt{\sigma} \log(n+1)}{n^{1/4}}.$$

The two cases in the theorem may be united by one inequality such as

$$\mathbf{E}\, L(F_n, F) \leq C \frac{\sqrt{1+\sigma}}{n^{1/4}} \log(n+1),$$

which holds for any σ (with some other constant C).

4.4 Deviation Inequalities for Empirical Characteristic Functions

By similar methods, we can find a bound for higher moments. For $p \geq 2$, the Zolotarev bound gives

$$\begin{aligned} L^p(F_n, F) &\leq \left[c_1 \int_0^T \frac{|f_n(t) - f(t)|}{t} dt + c_2 \frac{\log(1+T)}{T} \right]^p \\ &\leq 2^{p-1} \left[c_1 \int_0^T \frac{|f_n(t) - f(t)|}{t} dt \right]^p + 2^{p-1} \left[c_2 \frac{\log(1+T)}{T} \right]^p \\ &\leq 2^{p-1} c_1^p T^{p-1} \int_0^T \frac{|f_n(t) - f(t)|^p}{t^p} dt + 2^{p-1} \left[c_2 \frac{\log(1+T)}{T} \right]^p. \end{aligned}$$

So by Fubini's theorem, we have

$$\mathbf{E}\, L^p(F_n, F) \leq 2^{p-1} c_1^p T^{p-1} \int_0^T \frac{\mathbf{E}\, |f_n(t) - f(t)|^p}{t^p} dt + 2^{p-1} \left[c_2 \frac{\log(1+T)}{T} \right]^p.$$

Now by Theorem 4.1, we get

$$\mathbf{E}\, |f_n(t) - f(t)|^p \leq \frac{(\sigma p)^p}{n^{p/2}} |t|^p.$$

Therefore, substituting gives

$$c^p\, \mathbf{E}\, L^p(F_n, F) \leq \frac{(\sigma p)}{\sqrt{n}}^p T^p + \left[\frac{\log(1+T)}{T} \right]^p$$

with some absolute constant $c > 0$. To make the optimization easier, consider the L^p-norms

$$\|L(F_n, F)\|_p = \left(\mathbf{E}\, L^p(F_n, F)\right)^{1/p}.$$

With some other constant $c > 0$, we obtain

$$c\|L(F_n, F)\|_p \leq \frac{\sigma p}{\sqrt{n}} T + \frac{\log(1+T)}{T}. \tag{4.2}$$

Now, our inequality is very similar to what we had before Theorem 4.3, and the only difference is the appearance of the parameter p on the right-hand side. We consider two cases as before.

Case I: $\sigma \leq 1$. Then the above gives

$$c\, \|L(F_n, F)\|_p \;\leq\; \frac{p}{\sqrt{n}} T + \frac{\log(1+T)}{T}.$$

One may choose, which is almost optimal, the value

$$T = \left(\frac{\sqrt{n}}{p}\right)^{1/2}.$$

With this value, using also that $p \geq 1$, we get

$$c\,\|L(F_n, F)\|_p \,\leq\, \left(\frac{p}{\sqrt{n}}\right)^{1/2} \left(1 + \log\left(1 + n^{1/4}\right)\right)$$

which easily implies

$$\|L(F_n, F)\|_p \,\leq\, C\left(\frac{p}{\sqrt{n}}\right)^{1/2} \log(n+1)$$

with some absolute constant C.

Case II: $\sigma \geq 1$. Then our choice will be

$$T = \left(\frac{\sqrt{n}}{\sigma p}\right)^{1/2}.$$

Since $\log(1+T) \leq \log\left(1 + n^{1/4}\right)$, we obtain a similar estimate

$$\|L(F_n, F)\|_p \,\leq\, C\left(\frac{\sigma p}{\sqrt{n}}\right)^{1/2} \log(n+1).$$

The two cases may united by one inequality.

Theorem 4.4. *Let $X = (X_1, \ldots, X_n)$ be a random vector in $\mathbf{R}^n$ and F_n be the empirical distribution associated with X. Let $F = \mathbf{E}\, F_n$ and suppose $\mu = \mathcal{L}(X)$ satisfies a Poincaré-type inequality with constant σ^2. Then for $p \geq 1$ we have*

$$\|L(F_n, F)\|_p \,\leq\, C\sqrt{(1+\sigma)p}\;\frac{\log(n+1)}{n^{1/4}},$$

where C is an absolute constant.

When $p = 1$, we return to Theorem 4.3, so Theorem 4.4 is more general.

Now that we have an estimate on the L^p norm of the Lévy distance of the empirical characteristic functions, we can use Ψ_α norms (see Appendix B) to derive the deviation inequalities. It is known and easy to see that up to some absolute constants $c_1, c_2 > 0$,

$$c_1 \sup_{p \geq 1} \frac{\|\xi\|_p}{\sqrt{p}} \leq \|\xi\|_{\Psi_2} \leq c_2 \sup_{p \geq 1} \frac{\|\xi\|_p}{\sqrt{p}}$$

where $\|\cdot\|_{\Psi_2}$ is the Orlicz norm with respect to the Young function $\Psi_2(t) = e^{t^2} - 1$ and ξ is a measurable function. Hence, we can conclude that $\|\xi\|_{\Psi_2} \leq \lambda$ and $\|\xi\|_p \leq C\lambda\sqrt{p}$ are equivalent relations. Now we let

$$\lambda = C\sqrt{1+\sigma}\frac{\log(n+1)}{n^{1/4}}.$$

Then by Theorem 4.4,

$$\|L(F_n, F)\|_p \leq C\lambda\sqrt{p},$$

which implies

$$\|L(F_n, F)\|_{\Psi_2} \leq \lambda.$$

And by the definition of Orlicz norm, $\|L(F_n, F)\|_{\Psi_2} \leq \lambda$ if and only if $\mathbf{E}\, e^{\xi^2/\lambda^2} \leq 2$ where $\xi = L(F_n, F)$. Therefore, by Chebyshev's inequality we can obtain the following corollary.

Corollary 4.5. *For $r > 0$,*

$$\mathbf{P}(L(F_n, F) > r) \leq 2e^{-r^2/\lambda^2}$$

where

$$\lambda = C\sqrt{1+\sigma}\frac{\log(n+1)}{n^{1/4}}$$

and C is an absolute constant.

So we have arrived at a Gaussian type concentration for $L(F_n, F)$. If we fix $r > 0$ and plug in λ, we will get the following bound.

$$\mathbf{P}(L(F_n, F) > r) \leq 2e^{-Cr^2\sqrt{n}/\log^2(n+1)}.$$

So we can see that as $n \to +\infty$ the probability decays to 0 very fast.

Chapter 5

Log-Sobolev Inequalities and Related Analytic Hypotheses

Another type of very useful inequalities in the study of concentration of measures is the logarithmic Sobolev inequalities. They may have some similar properties as the Poincaré-type inequalities, but they yield sharper results for concentration and are particularly well-suited for infinite-dimensional analysis.

5.1 Log-Sobolev Inequalities and Their Properties

Definition 5.1. A probability measure μ on $\mathbf{R}^n$ satisfies a logarithmic Sobolev inequality with constant σ^2 and we write $\mathrm{LSI}(\sigma^2)$ for short if, for all bounded smooth f,

$$\mathrm{Ent}_\mu(f^2) \leq 2\sigma^2 \int |\nabla f|^2 d\mu. \tag{5.1}$$

Recall that $\mathrm{Ent}_\mu(f)$ is the entropy of the function f under measure μ defined as

$$\mathrm{Ent}_\mu(f) = \mathbf{E}\, f \log f - \mathbf{E}\, f \log(\mathbf{E}\, f).$$

Note that $\mathrm{Ent}_\mu(f) \geq 0$ by Jensen's inequality and is defined if $\int f \log(1+f) d\mu < \infty$. Similarly as in the case for Poincaré-type inequalities, the definition makes sense if the smoothness of f is relaxed to the property of being "locally Lipschitz."

The first important result of a logarithmic Sobolev inequality is that it implies Poincaré inequality.

Proposition 5.2. *Assme that μ satisfies a log-Sobolev inequality for all bounded smooth functions f with constant $\sigma^2 > 0$. Then for all bounded smooth functions f*

$$\mathrm{Var}_\mu(f) \leq \sigma^2 \int |\nabla f|^2 d\mu$$

i.e. μ satisfies the Poincaré inequality with the same constant σ^2.

Note the convenient choice of constant $2\sigma^2$ in the definition of logarithmic Sobolev inequality in relations to the constant in the Poincaré-type inequality. The reason for such choice can be seen in the proof below.

Proof of Proposition 5.2. We prove the proposition using Taylor expansion of the logarithmic function. Let f be a bounded smooth function. Without loss of generality, we can assume $\mathbf{E}\, f = 0$ since $\mathrm{Var}(f - \mathbf{E}\, f) = \mathrm{Var}\, f$. Now we choose $\epsilon > 0$ such that $|\epsilon f| < 1$ and apply the log-Sobolev inequality to the function $1 + \epsilon f$. Since $\mathrm{Ent}_\mu(f^2) = \mathbf{E}(f^2 \log f^2) - \mathbf{E}\, f^2 \log \mathbf{E}\, f^2$ we consider the two parts separately. By Taylor expansion, we have

$$\begin{aligned}
\mathbf{E}\left[(1+\epsilon f)^2 \log(1+\epsilon f)^2\right] &= \mathbf{E}\left[(1+2\epsilon f+\epsilon^2 f^2) 2\log(1+\epsilon f)\right] \\
&= \mathbf{E}\left[(1+2\epsilon f+\epsilon^2 f^2)(2\epsilon f-\epsilon^2 f^2+o(\epsilon^2)\right] \\
&= \mathbf{E}\left[2\epsilon f+3\epsilon^2 f^2+o(\epsilon^2)\right] \\
&= 3\epsilon^2\,\mathbf{E}(f^2)+o(\epsilon^2)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}(1+\epsilon f)^2 \log \mathbf{E}(1+\epsilon f)^2 &= (1+\epsilon^2\,\mathbf{E}(f^2))\log(1+\epsilon^2\,\mathbf{E}(f^2)) \\
&= (1+\epsilon^2\,\mathbf{E}(f^2))(\epsilon^2\,\mathbf{E}(f^2)+o(\epsilon^2)) \\
&= \epsilon^2\,\mathbf{E}(f^2)+o(\epsilon^2).
\end{aligned}$$

Therefore, combining the above two expansions gives

$$\mathrm{Ent}_\mu(f^2) = 2\epsilon^2\,\mathbf{E}(f^2)+o(\epsilon^2).$$

Now since $|\nabla(1+\epsilon f)| = \epsilon|\nabla f|$, we get

$$2\epsilon^2\,\mathbf{E}(f^2)+o(\epsilon^2) \leq 2\sigma^2\epsilon^2\,\mathbf{E}\,|\nabla f|^2.$$

Finally, dividing both sides by ϵ^2 and taking the limit $\epsilon \to 0$ gives us the desired result since $\mathrm{Var}(f) = \mathbf{E}(f^2)$. So we have proven

$$\mathrm{Var}(f) \leq \sigma^2\,\mathbf{E}\,|\nabla f|^2.$$

□

Therefore, log-Sobolev inequality is a stronger inequality than a Poincaré-inequality. We can see that there is a natural generalization of this relationship by replacing f with $f+C$ in the definition of log-Sobolev inequality. So under the assumption

$$\mathrm{Ent}(f^2) \leq 2\sigma^2\,\mathbf{E}\,|\nabla f|^2$$

we consider

$$\mathrm{Ent}((f+C)^2) \leq 2\sigma^2\,\mathbf{E}\,|\nabla f|^2.$$

Since adding a constant does not affect the gradient, the right side of the inequality remains the same. Now we try to maximize the left side of the inequality over all values of $C \in \mathbf{R}$. In the limiting case, we obtain the Poincaré-type inequality since as $C \to +\infty$ we have

$$\lim_{C \to +\infty} \operatorname{Ent}((f+C)^2) = 2\operatorname{Var}(f).$$

For all other cases, we consider the function

$$\mathscr{L}(f) = \sup_C \operatorname{Ent}[(f+C)^2] \le 2\sigma^2 \, \mathbf{E}\, |\nabla f|^2$$

as studied by Bobkov and Götze in [7]. It was proved that there is a following bound on $\mathscr{L}(f)$.

$$c_1 \|f - \mathbf{E}\, f\|_\Psi^2 \le \mathscr{L}(f) \le c_2 \|f - \mathbf{E}\, f\|_\Psi^2$$

where $\|\cdot\|_\Psi$ denotes the Orlicz norm under the Young function $\Psi(x) = x^2 \log(1+x^2)$. Thus, the log-Sobolev inequality in Equation 5.1 is equivalent, up to a constant, to an inequality

$$\|f - \mathbf{E}\, f\|_\Psi^2 \le C\, \mathbf{E}\, |\nabla f|^2$$

which belongs to the class of Poincaré-type inequalities. Hence, log-Sobolev inequalities can be viewed as Poincaré-type inequalities in Orlicz spaces.

Logarithmic Sobolev inequalities can also be extended to product measures by a following useful lemma on entropy functionals.

Lemma 5.3. *Let $(\Omega, \mu) = (\Omega_1, \mu_1) \times \cdots \times (\Omega_n, \mu_n)$ and $f : \Omega \mapsto \mathbf{R}$ be measurable. Denote f_i as function on Ω_i as defined by $f_i(x_i) = f(x_1, \ldots, x_n)$ where $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ are fixed. Then,*

$$\operatorname{Ent}_\mu(f) \le \sum_{i=1}^n \int_\Omega \operatorname{Ent}_{\mu_i}(f_i) d\mu.$$

Therefore, we can get the following theorem as stated by Ledoux in [25].

Theorem 5.4. *Let μ_i on Ω_i satisfy the log-Sobolev inequality*

$$\operatorname{Ent}_{\mu_i}(f_i^2) \le 2\sigma_i^2 \int_{\Omega_i} |\nabla f_i|^2 d\mu_i$$

for every bounded smooth function f_i on Ω_i. Then the product measure $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ on the product space Ω satisfies the Log-Sobolev inequality. So for every bounded smooth

f on Ω,

$$\mathrm{Ent}_\mu(f^2) \le 2 \max_{1\le i\le n} \sigma_i^2 \int_\Omega \sum_{i=1}^n |\nabla f_i|^2 d\mu.$$

The logarithmic Sobolev inequality is also stable under Lipschitz transformation.

Theorem 5.5. *Let μ be a probability measure on $\mathbf{R}^n$ that satisfies a log-Sobolev inequality with constant σ^2. Suppose $T : \mathbf{R}^n \mapsto \mathbf{R}^k$ is a Lipschitz transformation so that $\|Tx-Ty\|_{\mathbf{R}^k} \le C\|x-y\|_{\mathbf{R}^n}$. Let $\nu = T\mu^{-1}$. Then ν also satisfies a log-Sobolev inequality with constant $(C\sigma^2)$. So we have*

$$\mathrm{Ent}_\nu(f^2) \le 2C^2\sigma^2 \int |\nabla f|^2 d\nu.$$

Under a log-Sobolev inequality LSI(σ^2) one can also sharpen deviation bounds of Theorem 3.5. Bobkov and Götze in [7] has shown that, for any μ-integrable locally Lipschitz function f on $\mathbf{R}^n$,

$$\int e^{f-\int f\,d\mu}\, d\mu \le \int e^{\sigma^2|\nabla f|^2}\, d\mu.$$

In particular, if $\|f\|_{\mathrm{Lip}} \le 1$, applying the above to λf, we get

$$\int e^{\lambda\left(f-\int f\,d\mu\right)}\, d\mu \le\ e^{\sigma^2\lambda^2}\, d\mu, \qquad \lambda \in \mathbf{R}.$$

Up to a numerical constant $c > 0$, this is equivalent to the family of the moment-type relations

$$\left\| f - \int f\, d\mu \right\|_p \le c\sigma\sqrt{p}, \qquad p \ge 2,$$

which can also be united by just one relation in terms of the Orlicz norm generated by the Young function $\Psi_2(t) = e^{t^2} - 1$, namely, as

$$\left\| f - \int f\, d\mu \right\|_{\Psi_2} \le c\sigma.$$

Please see Appendix B for more details on $\|\cdot\|_{\Psi_\alpha}$.

If f is not necessarily Lipschitz, one can relate the moments of f to the moments of $|\nabla f|$, using the following theorem which improves the second part of Theorem 3.5, cf. [25] and [4].

Theorem 5.6. *Under LSI(σ^2), for any Lipschitz function f on $\mathbf{R}^n$,*

$$\left\| f - \int f\, d\mu \right\|_p \le \sigma\sqrt{p}\, \|f\|_p, \qquad p \ge 2.$$

5.2 Characterization of Log-Sobolev Inequalities on R

Using the same notations as in Section 3.3, we give the characterization of log-Sobolev inequalities on the whole real line. Let μ be a probability measure on the line with median m. Define the quantities

$$B_0(\mu) = \sup_{x<m} \left[\mu(-\infty, x) \log \frac{1}{\mu(-\infty, x)} \int_{-\infty}^{x} \frac{dt}{p_\mu(t)} \right],$$

$$B_1(\mu) = \sup_{x>m} \left[\mu(x, +\infty) \log \frac{1}{\mu(x, +\infty)} \int_{x}^{+\infty} \frac{dt}{p_\mu(t)} \right]$$

where p_μ denotes the density of the absolutely continuous component of μ (with respect to Lebesgue measure), and where we set $B_0 = 0$ and $B_1 = 0$ respectively if $\mu(-\infty, m) = 0$ or $\mu(m, +\infty) = 0$. We then have the following theorem.

Theorem 5.7. *The measure μ on $\mathbf{R}$ satisfies LSI(σ^2) with some finite constant if and only if $B_0(\mu)$ and $B_1(\mu)$ are finite. Moreover, the optimal value of σ^2 satisfies*

$$c_0(B_0(\mu) + B_1(\mu)) \leq \sigma^2 \leq c_1(B_0(\mu) + B_1(\mu)),$$

where c_0 and c_1 are positive universal constants.

In this case, using log-Sobolev inequalities as Poincaré inequalities in Orlicz spaces and applying Theorem 3.7 to the Orlicz space $X = L_{\Psi(\nu)}$ we may obtaion Theorem 5.7. Detailed explanations and proof of the theorem can be found in [7].

Specializing to the log-concave case, this characterization shows that μ satisfies LSI(σ^2) with some finite σ^2 if and only if

$$\int e^{\lambda |x|^2} \, d\mu(x) < \infty, \quad \text{for some } \lambda > 0.$$

This description remains to hold on $\mathbf{R}^n$ for arbitrary log-concave probability measures μ, as proved in [3].

5.3 Empirical Log-Sobolev Inequalities.

As before, let $(X_1, \dots, X_n)$ be a random vector in $\mathbf{R}^n$ with joint distribution μ. By a similar argument as in Section 4.1, assuming that μ satisfies a log-Sobolev inequailty on $\mathbf{R}^n$ with constant σ^2, we get the following.

$$\mathrm{Ent}_\mu \left[\left(\int g dF_n \right)^2 \right] \leq \frac{2\sigma^2}{n} \int g'^2 dF.$$

As suggested in [7], it is called an empirical log-Sobolev inequality. Our next purpose is to sharpen Theorem 4.4 and Corollary 4.5 on the basis of this stronger hypothesis LSI(σ^2). First, Theorem 5.6 yields

Theorem 5.8. *If the distribution of $X = (X_1, \dots, X_n)$ in $\mathbf{R}^n$ satisfies LSI(σ^2), then for any smooth F-integrable function $g : \mathbf{R} \mapsto \mathbf{R}$ and $p \geq 2$,*

$$\mathbf{E} \left| \int g dF_n - \int g dF \right|^p \leq \frac{(\sigma\sqrt{p})^p}{n^{p/2}} \int |g'|^p dF.$$

Now, like in the case of the Poincaré-type inequalities, we apply Theorem 5.8 to the functions $g(x) = e^{itx}$ so as to get corresponding bounds for the deviation of the empirical characteristic functions

$$f_n(t) = \int_{-\infty}^{\infty} e^{itx} \, dF_n(x) = \frac{1}{n} \sum_{k=1}^{n} e^{itX_k}, \qquad t \in \mathbf{R},$$

about the average characteristic function

$$f_n(t) = \int_{-\infty}^{\infty} e^{itx} \, dF(x).$$

Here, as before, $F(x) = \mathbf{E} \, F_n(x)$ denotes the average distribution function of the observations X_k. Namely, we then arrive at

$$\mathbf{E} \, |f_n(t) - f(t)|^p \leq \left(\frac{\sigma\sqrt{p}}{\sqrt{n}} \right)^p \, |t|^p.$$

Next, in order to transform this inequality into the bound for the Lévy distance, we argue similarly to the proof of Theorem 4.4. An application of the Zolotarev's bound gives

$$c^p \, \mathbf{E} \, L^p(F_n, F) \, \leq \, \left(\frac{\sigma\sqrt{p}}{\sqrt{n}} \right)^p T^p + \left(\frac{\log(1+T)}{T} \right)^p$$

with some absolute constant $c > 0$. To make the optimization easier, consider the L^p-norms

$$\|L(F_n, F)\|_p = \Big(\mathbf{E}\ L^p(F_n, F)\Big)^{1/p}.$$

With some other constant $c > 0$, we obtain

$$c\,\|L(F_n, F)\|_p \le \frac{\sigma\sqrt{p}}{\sqrt{n}}\,T + \frac{\log(1+T)}{T}.$$

Now, our inequality is very similar to what we had before in the proof of Theorem 4.4. The only difference is that the parameter p in the ratio on the right-hand side is replaced with $\sqrt{p}$. In order to optimize the right-hand side over the parameter $T > 0$, we distinguish between the two cases.

Case I: $\sigma \le 1$. Then the above gives

$$c\,\|L(F_n, F)\|_p \le \frac{\sqrt{p}}{\sqrt{n}}\,T + \frac{\log(1+T)}{T}.$$

One may choose, which is almost optimal, the value

$$T = \left(\frac{\sqrt{n}}{\sqrt{p}}\right)^{1/2}.$$

With this value, using also that $p \ge 1$, we get

$$c\,\|L(F_n, F)\|_p \le \left(\frac{\sqrt{p}}{\sqrt{n}}\right)^{1/2} \left(1 + \log\left(1 + n^{1/4}\right)\right)$$

which easily implies

$$\|L(F_n, F)\|_p \le C\left(\frac{p}{n}\right)^{1/4} \log(n+1)$$

with some absolute constant C.

Case II: $\sigma \ge 1$. Then our choice will be

$$T = \left(\frac{\sqrt{n}}{\sigma\sqrt{p}}\right)^{1/2}.$$

Since $\log(1+T) \le \log\left(1 + n^{1/4}\right)$, we obtain a similar estimate

$$\|L(F_n, F)\|_p \le C\left(\frac{\sigma^2 p}{n}\right)^{1/4} \log(n+1).$$

The two cases may united by one inequality.

Theorem 5.9. *Under the LSI(σ^2), for any $p \geq 1$, we have*

$$\|L(F_n, F)\|_p \;\leq\; C\sqrt{1+\sigma}\, p^{1/4}\, \frac{\log(n+1)}{n^{1/4}},$$

where C is an absolute constant.

The main difference with Theorem 4.4 is that now $p^{1/4}$ replaces $p^{1/2}$. As we know from Proposition B.3 of the Appendix section, the family of moment inequalities in Theorem 5.9 can be united just by one inequality

$$\|L(F_n, F)\|_{\Psi_4} \;\leq\; C\sqrt{1+\sigma}\, \frac{\log(n+1)}{n^{1/4}}$$

in terms of the Orlicz norm generated by the Young function $\Psi_4 = e^{t^4} - 1$, up to some numerical constant C. That is, by the very definition of the Orlicz norm,

$$\mathbf{E}\ \Psi(L(F_n, F)/\lambda) \leq 1$$

with

$$\lambda = C\sqrt{1+\sigma}\, \frac{\log(n+1)}{n^{1/4}}.$$

Thus,

$$\mathbf{E}\ e^{L(F_n,F)^4/\lambda^4} \leq 2.$$

One can now use the Chebyshev inequality, and we obtain a corresponding sharpening of Corollary 4.5.

Corollary 5.10. *Under the LSI(σ^2), for any $r > 0$, we have*

$$\mathbf{P}\{L(F_n, F) > r\} \leq 2e^{-r^4/\lambda^4},$$

where

$$\lambda = C\sqrt{1+\sigma}\, \frac{\log(n+1)}{n^{1/4}}$$

and C is an absolute constant.

Chapter 6

Applications

In this chapter we apply our results from Chapter 4 in the two popular models. The first model is described by the scheme of the observations having a joint log-concave distribution μ on $\mathbf{R}^n$. One typical example illustrating this model is when μ represents the uniform distribution over a convex body K in $\mathbf{R}^n$. That is, one may speak about sampling from the convex body K. In the second model the observations are obtained as eigenvalues of random matrices whose entries satisfy Poincaré-type inequalities. That is, we will discuss the concentration property of the spectral empirical distributions.

6.1 Sampling from Convex Bodies and Log-Concave Distributions

Assume that $X = (X_1, \dots, X_n) \in \mathbf{R}^n$ has a log-concave distribution μ. We already mentioned in Section 3.5 that $\sigma^2 = \sigma_\mu^2 < +\infty$. So the previous results in Chapter 4 can be applied.

Problem. How to estimate σ^2 from above?

In general (when μ is log-concave on $\mathbf{R}^n$), it is known that the Poincaré constant $\sigma^2 = \sigma_\mu^2$, that is, the optimal constant in the analytic inequality

$$\mathrm{Var}_\mu(g) \leq \sigma^2 \int_{\mathbf{R}^n} |\nabla g|^2 \, d\mu$$

admits an upper bound

$$\sigma^2 \leq C \, \mathbf{E} \, |X|^2 = C \int_{\mathbf{R}^n} |x|^2 \, d\mu(x),$$

where $|X| = (X_1^2 + \cdots + X_n^2)^{1/2}$ denotes the Euclidean norm. In the convex body case, this bound was established in 1995 by Kannan, Lovász and Simonovits [22] and then was extended to the general log-concave case in 1999 by Bobkov [3].

In dimension $n = 1$, just taking $g(x) = x$, and assuming without loss of generality that $\mathbf{E}\, X = 0$, we obtain a similar lower bound

$$\sigma^2 \geq \mathbf{E}\, X^2,$$

so, σ^2 is of order $\mathbf{E}\, X^2$ (which was already mentioned before in Section 3.5. In higher dimensions, by testing the Poincaré-type inequality on linear functions $g(x) = \langle x, \theta \rangle$, we have similarly

$$\sigma^2 \, \geq \, \max_{|\theta|=1} \mathbf{E} \, \langle X, \theta \rangle^2 \qquad (\mathbf{E}\, X = 0).$$

It is believed that this lower bound is actually an upper one:

Conjecture (Kannan, Lovász, and Simonovits). For any log-concave probability measure μ on $\mathbf{R}^n$ with baricenter at the origin,

$$\sigma^2 \, \leq \, C \max_{|\theta|=1} \mathbf{E} \, \langle X, \theta \rangle^2 \,,$$

where X is a random vector with distribution μ and C is an absolute constant.

In other words, this conjecture, called nowadays the K-L-S conjecture, expresses the property that some linear function is almost extremal in the Poincaré-type inequality. The K-L-S conjecture is still open, although it is known to hold in some special examples such as the Gaussian measure or more generally products of one-dimensional log-concave measures.

In many problems about log-concave measures without loss of generality one may assume that in some sense μ is "sufficiently regular". To avoid degenerate situations, usually one assumes that μ is isotropic (which is a matter of normalization, only).

Definition 6.1. A random vector $X = (X_1, \dots, X_n)$ in $\mathbf{R}^n$ with finite second moment ($\mathbf{E}\,|X|^2 < \infty$) is called isotropic, or it is said to have an isotropic distribution, if $\mathbf{E}\,X_i = 0$, for all $i \leq n$, and

$$\mathbf{E}\,X_i X_j = \delta_{ij}, \qquad i,j = 1, \dots, n,$$

where δ_{ij} denotes the Kronecker symbol.

One also says that the distribution μ of X is in isotropic position. Equivalently, the isotropicity means that, for any unit vector θ,

$$\mathbf{E}\,\langle X, \theta \rangle = 0 \quad \text{and} \quad \mathbf{E}\,\langle X, \theta \rangle^2 = 1.$$

If the distribution of X is not supported on any hyperplane (for example, when it has a density), there always exists an invertible affine map $T : \mathbf{R}^n \to \mathbf{R}^n$ such that $Y = T(X)$ is isotropic. Indeed, one should put $T(x) = A(x) - v$, where $v = \mathbf{E}\,X$ is the baricenter of μ, and A is a proper linear operator. If $v = 0$, this operator should be chosen so as to satisfy

$$1 = \mathbf{E}\,\langle Y, \theta \rangle^2 = \mathbf{E}\,\langle AX, \theta \rangle^2 = \mathbf{E}\,\langle X, A'\theta \rangle^2 = \langle RA'\theta, A'\theta \rangle = \langle ARA'\theta, \theta \rangle,$$

where R is the covariance operator for X. Since $\det(R) \neq 0$ (by the assumption), one can choose A such that ARA' is the identity operator.

In the isotropic case, the K-L-S conjecture indicates that for the distribution μ of X the Poincaré constant σ^2 satisfies

$$1 \leq \sigma^2 \leq C$$

with some absolute constant C. Assuming it is true, Theorem 4.3 implies the following.

Corollary 6.2. *Let $X = (X_1, \dots, X_n)$ be an isotropic random vector in $\mathbf{R}^n$ having a log-concave distribution, and let F_n be the corresponding empirical distribution function with mean $F = \mathbf{E}\, F_n$. Given that the K-L-S conjecture is true, we have*

$$\mathbf{E}\, L(F_n, F) \le C\, \frac{\log(n+1)}{n^{1/4}},$$

where C is an absolute constant.

One can also write down a more precise statement in terms of the deviation inequalities.

Regardless of whether the K-L-S conjecture is true or not, one can get somewhat weaker estimates on $\mathbf{E}\, L(F_n, F)$, depending on the exising bounds for the Poincaré constant σ^2. For example, one can try to use the general K-L-S bound $\sigma^2 \le C\, \mathbf{E}\, |X|^2$, mentioned above. In the isotropic case, $\mathbf{E}\, |X|^2 = n$, so $\sigma^2 \le Cn$. Hence, the application of Theorem 4.3 would only give $\mathbf{E}\, L(F_n, F) \le C\, \log(n+1)$. This is useless in view of $L \le 1$.

As a sharpening of the K-L-S bound, it was obtained in [6] that

$$\sigma^2 \le C \sqrt{\mathrm{Var}(|X|^2)},$$

where C is an absolute constant and where there is no isotropicity assumption. In the theory of log-concave measures, it is known that L^p-norms of $|X|$ are equivalent to each other within factors depending on p, only. Hence,

$$\sqrt{\mathrm{Var}(|X|^2)} \le \sqrt{\mathbf{E}\, |X|^4} \le c\, \mathbf{E}\, |X|^2.$$

This shows that we have indeed a sharper estimate. Applying Theorem 4.3, we thus get:

Corollary 6.3. *Let $X = (X_1, \dots, X_n)$ be a random vector in $\mathbf{R}^n$ having a log-concave distribution. Then*

$$\mathbf{E}\, L(F_n, F) \le C\, \mathrm{Var}^{1/8}(|X|^2)\, \frac{\log(n+1)}{n^{1/4}},$$

where C is an absolute constant.

In the isotropic case, typically $\mathrm{Var}(|X|^2)$ is of order n (like in the case when the coordinates X_i are independent), and then we arrive at the bound such as

$$\mathbf{E}\, L(F_n, F) \le C\, \frac{\log(n+1)}{n^{1/8}}.$$

Remark. The so-called thin shell conjecture indicates that, for any isotropic random vector X in $\mathbf{R}^n$ having a log-concave distribution, we indeed have

$$\mathrm{Var}(|X|^2) \leq Cn$$

with some numerical constant C. It is weaker than the K-L-S conjecture: This estimate on the variance is obtained from the Poincaré-type inequality written for the particular function $g(x) = |x|^2$. The thin shell conjecture is known to hold true for a large family of log-concave distribution. For example, it is the case where the distribution of X is additionally symmetric about the coordinate axes. This result is due to Klartag [24].

6.2 Applications to Random Matrices

We now show how the bounds attained can be applied to the case of the spectral empirical distributions. Assume we have a random symmetric matrix

$$M = \frac{1}{\sqrt{n}} \begin{pmatrix} \xi_{11} & \dots & \xi_{1n} \\ \xi_{21} & \dots & \xi_{2n} \\ \vdots & \ddots & \vdots \\ \xi_{n1} & \dots & \xi_{nn} \end{pmatrix}$$

with random entries $\xi_{jk} = \xi_{kj}$. Symmetry condition ensures that M has real eigenvalues $X_1 \le \cdots \le X_n$. Consider $F_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$, the spectral empirical distribution associated with the particular values $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$. The distribution of the eigenvalues when ξ_{jk} are i.i.d. can be formulated as follows.

Theorem 6.4. *(Wigner's Semi-circle Law, [1]) Let* $M = \frac{1}{\sqrt{n}}(\xi_{jk})$, $j \le k$ *be an* $n \times n$ *random symmetric matrix with eigenvalues* $X_1 = x_1 \le \cdots \le X_n = x_n$. *If* (ξ_{jk}) *are i.i.d.,* $\mathbf{E}\,\xi_{jk} = 0$, $\mathbf{E}\,\xi_{jk}^2 = 1$, *and* $F_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_n}$, *then*

$$F = \mathbf{E}\,F_n \Longrightarrow G \quad \textit{weakly}$$

where G *is a distribution function of the semi-circle law with density*

$$g(x) = \begin{cases} \sqrt{4 - x^2}/(2\pi), & \textit{if} \quad |x| \le 2, \\ 0, & \textit{if} \quad |x| > 2. \end{cases}$$

Now, we consider the case when ξ_{jk} may be dependent and not necessarily have same distributions. An important thing to note in this case is that the joint distribution μ of the eigenvalues, as a probability measure on $\mathbf{R}^n$, represents the image of the joint distribution of ξ_{jk}'s under a Lipschitz map T. Namely, we will use the following classical fact from the theory of matrix inequalities [8]:

Proposition 6.5. *Let* $A = (a_{jk})$, $B = (b_{jk})$, $j \le k$ *be* $n \times n$ *symmetric matrices with eigenvalues* $x_1 \le \cdots \le x_n$ *and* $y_1 \le \cdots \le y_n$ *respectively. Then*

$$\sum_{i=1}^n (x_i - y_i)^2 \le \sum_{j,k=1}^n (a_{jk} - b_{jk})^2.$$

Hence, if $M = \frac{1}{\sqrt{n}}(\xi_{jk})_{j\le k}$, in terms of our map

$$T : M \mapsto x = (x_1, \ldots, x_n) \in \mathbf{R}^n$$

we have

$$\begin{aligned}
\|x - x'\|^2_{\mathbf{R}^n} &= \|T(M) - T(M')\|^2_{\mathbf{R}^n} \\
&\le \frac{1}{n} \sum_{j,k=1}^{n} (\xi_{jk} - \xi'_{jk})^2 \\
&= \frac{1}{n} \|T(M) - T(M')\|^2_{HS} \\
&\le \frac{2}{n} \sum_{j\le k} (\xi_{jk} - \xi'_{jk})^2.
\end{aligned}$$

Therefore, we have

Corollary 6.6. $\|T\|_{Lip} \le \sqrt{\frac{2}{n}}$.

Now, assume M is random and symmetric as before. Then $M = \frac{1}{\sqrt{n}}(\xi_{jk})_{j\le k}$ can be viewed as a random vector in $\mathbf{R}^{n(n+1)/2}$ with distribution $\mathbf{P} = \mathbf{P}_\xi$. Assume that $\mathbf{P}_\xi$ satisfies PI(σ^2) on $\mathbf{R}^{n(n+1)/2}$. Then T pushes forward $\mathbf{P}_\xi$ to $\mu = \mu_X$ on $\mathbf{R}^n$ and by Theorem 3.4, $\mu = \mathbf{P}_\xi T^{-1}$ satisfies PI(σ_n^2) on $\mathbf{R}^n$ with $\sigma_n^2 = \frac{2}{n}\sigma^2$. As a result, we obtain

Theorem 6.7. *Let $M = \frac{1}{\sqrt{n}}(\xi_{jk})_{1\le j\le k\le n}$ be an $n \times n$ random symmetric matrix with eigenvalues $X_1 \le \cdots \le X_n$. Suppose that the joint distribution of ξ_{jk} satisfies PI(σ^2), then the empirical spectral distributions $F_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ satisfy the empirical Poincaré-type inequality with $\frac{2\sigma^2}{n}$. So*

$$\mathbf{E}\left|\int g dF_n - \int g dF\right|^2 \le \frac{2\sigma^2}{n} \int g'^2 dF$$

where as before, $F = \mathbf{E}\, F_n$.

Therefore, we can apply previous general observations about the concentration of F_n around F in terms of Lévy distance. The L^p bound of the Lévy distance in this situation using Equation 4.2 will be

$$c\|L(F_n, F)\|_p \le \frac{\sigma p}{n} T + \frac{\log(1+T)}{T} \quad \text{for} \quad T > 0$$

where we have replaced σ with $\sigma\sqrt{2/n}$. We break into two cases again and choose T.

Case I: $0 < \sigma \le 1$. Then choosing $T = \sqrt{\frac{n}{p}}$ gives

$$\begin{aligned} c\|L(F_n, F)\|_p &\le \frac{\sigma p}{n} T + \frac{\log(1+T)}{T} \\ &\le \sqrt{\frac{p}{n}} \left(1 + \log(1+\sqrt{n})\right) \\ &\le C\sqrt{\frac{p}{n}} \log(n+1). \end{aligned}$$

Case II: $\sigma \ge 1$. Then choosing $T = \sqrt{\frac{n}{\sigma p}}$ gives

$$\begin{aligned} c\|L(F_n, F)\|_p &\le \frac{\sigma p}{n} T + \frac{\log(1+T)}{T} \\ &\le \sqrt{\frac{\sigma p}{n}} \left(1 + \log(1+\sqrt{n})\right) \\ &\le C\sqrt{\sigma}\sqrt{\frac{p}{n}} \log(n+1). \end{aligned}$$

Therefore, the two cases can be combined in the following theorem.

Theorem 6.8. *Let $M = \frac{1}{\sqrt{n}}(\xi_{jk})_{1 \le j \le k \le n}$ be an $n \times n$ random symmetric matrix with eigenvalues $X_1 \le \cdots \le X_n$. Suppose that the joint distribution of ξ_{jk} satisfies PI(σ^2), then*

$$\|L(F_n, F)\|_p \le C\sqrt{(1+\sigma)p}\frac{\log(n+1)}{n^{1/2}}$$

for some absolute constant C where F_n is the empirical spectral distribution, $F = \mathbf{E}\, F_n$, and L is the Lévy distance.

Under the same PI(σ^2) assumption, we can see that we now have $n^{1/2}$ in the denominator rather than $n^{1/4}$ as in Theorem 4.4. However, $\sqrt{p}$ is still the same, so derivation of the deviation inequalities is not different from that of Theorem 4.4. We follow the same procedure as before and apply Proposition B.3 of the Appendix section to Theorem 6.8. So we have

$$\|L(F_n, F)\|_{\Psi_2} \le C\sqrt{1+\sigma}\frac{\log(n+1)}{n^{1/2}}$$

in terms of the Orlicz norm generated by the Young function $\Psi_2(t) = e^{t^2} - 1$. Hence, by the definition of the Orlicz norm, we get

$$\mathbf{E}\, e^{L(F_n,F)^2/\lambda^2} \le 2$$

where

$$\lambda = C\sqrt{1+\sigma}\frac{\log(n+1)}{n^{1/2}}.$$

Hence, by Chebyshev's inequality, we obtain the following deviation inequality.

Corollary 6.9. *Under PI(σ^2), for any $r > 0$, we have*

$$\mathbf{P}\{L(F_n, F) > r\} \leq 2e^{-r^2/\lambda^2},$$

where

$$\lambda = C\sqrt{1+\sigma}\frac{\log(n+1)}{n^{1/2}}$$

and C is an absolute constant.

Therefore, in terms of the deviation inequality, we get the same Gaussian type concentration as in Corollary 4.5. However, if we fix $r > 0$ and insert λ we now get a slightly faster decay as $n \to +\infty$ since

$$\mathbf{P}\{\text{Ł}(F_n, F) > r\} \leq 2e^{-Cr^2 n/\log^2(n+1)}$$

where we see n in place of $\sqrt{n}$, which was the case before.

Remark. Under the assumption that the entries $\xi_{jk}, j \leq k$ are i.i.d., the concentration property of spectral empirical distributions were studied by many authors. In particular, assuming that each ξ_{jk} satisfies a log-Sobolev inequality with common constant σ^2, an analog of Corollary 6.9 can be found in a paper by Guionnet and Zeitouni here [18].

References

[1] Arnold, L. On Wigner's semicircle law for the eigenvalues of random matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 19* (1971), 191–198.

[2] Billingsley, P. *Convergence of probability measures.* Springer, 1968.

[3] Bobkov, S. Isoperimetric and analytic inequalities for log-concave probability measures. *Ann. Probab. 27*, 4 (1999), 1903–1921.

[4] Bobkov, S. The growth of L^p norms in presence of logarithmic Sobolev inequalities. *Vestnik Syktyvkar Univ. 1*, 11 (2010), 92–111.

[5] Bobkov, S., and Götze, F. Hardy type inequalities via Riccati and Sturm-Liouville equations. In *Sobolev spaces in mathematics. I*, vol. 8 of *Int. Math. Ser. (N. Y.)*. Springer, New York, 2009, pp. 69–86.

[6] Bobkov, S. G. On isoperimetric constants for log-concave probability distributions. In *Geometric aspects of functional analysis*, vol. 1910 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007, pp. 81–88.

[7] Bobkov, S. G., and Götze, F. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal. 163*, 1 (1999), 1–28.

[8] Bobkov, S. G., and Götze, F. Concentration of empirical distribution functions with applications to non-i.i.d. models. *Bernoulli 16*, 4 (2010), 1385–1414.

[9] Bobkov, S. G., and Houdré, C. Isoperimetric constants for product probability measures. *Ann. Probab. 25*, 1 (1997), 184–205.

[10] Bobkov, S. G., and Houdré, C. Some connections between isoperimetric and Sobolev-type inequalities. *Mem. Amer. Math. Soc. 129*, 616 (1997), viii+111.

[11] Bobkov, S. G., and Ledoux, M. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields 107*, 3 (1997), 383–400.

[12] Bobkov, S. G., and Ledoux, M. Weighted Poincaré-type inequalities for Cauchy and other convex measures. *Ann. Probab. 37*, 2 (2009), 403–427.

[13] Borovkov, A. A., and Utev, S. A. An inequality and a characterization of the normal distribution connected with it. *Teor. Veroyatnost. i Primenen. 28*, 2 (1983), 209–218.

[14] Cheeger, J. A lower bound for the smallest eigenvalue of the Laplacian. In *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*. Princeton Univ. Press, Princeton, N. J., 1970, pp. 195–199.

[15] Dudley, R. M. *Real analysis and probability*, vol. 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[16] Dvoretzky, A., Kiefer, J., and Wolfowitz, J. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist. 27* (1956), 642–669.

[17] Feller, W. On the Kolmogorov-Smirnov limit theorems for empirical distributions. *Ann. Math. Statistics 19* (1948), 177–189.

[18] Guionnet, A., and Zeitouni, O. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab. 5* (2000).

[19] Hardy, G. H. Note on a theorem of Hilbert. *Math. Z. 6*, 3-4 (1920), 314–317.

[20] Hardy, G. H., Littlewood, J. E., and Pólya, G. *Inequalities.* Cambridge, at the University Press, 1952. 2d ed.

[21] Kac, I. S., and Kreĭn, M. G. Criteria for the discreteness of the spectrum of a singular string. *Izv. Vysš. Učebn. Zaved. Matematika 1958*, 2 (3) (1958), 136–153.

[22] Kannan, R., Lovász, L., and Simonovits, M. Isoperimetric problems for convex bodies and a localization lemma. *Discrete Comput. Geom. 13*, 3-4 (1995), 541–559.

[23] Kim, T. Y. On tail probabilities of Kolmogorov-Smirnov statistics based on uniform mixing processes. *Statist. Probab. Lett. 43*, 3 (1999), 217–223.

[24] Klartag, B. A Berry-Esseen type inequality for convex bodies with an unconditional basis. *Probab. Theory Related Fields 145*, 1-2 (2009), 1–33.

[25] Ledoux, M. *The concentration of measure phenomenon*, vol. 89 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2001.

[26] Ledoux, M. Spectral gap, logarithmic Sobolev constant, and geometric bounds. In *Surveys in differential geometry. Vol. IX*, Surv. Differ. Geom., IX. Int. Press, Somerville, MA, 2004, pp. 219–240.

[27] Muckenhoupt, B. Hardy's inequality with weights. *Studia Math. 44* (1972), 31–38. Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I.

[28] Petrov, V. V. *Sums of independent random variables.* Springer-Verlag, New York, 1975. Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82.

[29] Sen, P. K. Weak convergence of multidimensional empirical processes for stationary φ-mixing processes. *Ann. Probability 2*, 1 (1974), 147–154.

[30] Sodin, S. An isoperimetric inequality on the l_p balls. *Ann. Inst. Henri Poincaré Probab. Stat. 44*, 2 (2008), 362–373.

[31] Varadarajan, V. S. On the convergence of sample probability distributions. *Sankhyā 19* (1958), 23–26.

[32] Zolotarev, V. M. Estimates of the difference between distributions in the Lévy metric. *Trudy Mat. Inst. Steklov. 112* (1971), 224–231, 388. Collection of articles dedicated to Academician Ivan Matveevič Vinogradov on his eightieth birthday, I.

Appendix A

Zolotarev's Berry-Esseen-Type Bound on the Lévy Distance

To obtain main results of the concentration of empirical measures, we rely heavily on Zolotarev's bound on the Lévy distance. We include the proof of the bound here since it is not given in standard probability textbooks, but it was explicitly written by Professor Sergey Bobkov.

Let F and G be distribution functions on the real line with characteristic functions

$$f(t) = \int_{-\infty}^{\infty} e^{itx}\,dF(x), \qquad g(t) = \int_{-\infty}^{\infty} e^{itx}\,dG(x) \qquad (t \in \mathbf{R}),$$

respectively. The question is how to effectively bound various distances between F and G in terms of f and g. For example, for the (uniform) Kolmogorov distance

$$\|F - G\| = \sup_x |F(x) - G(x)|$$

there is a famous Berry-Esseen's bound

$$\|F - G\| \leq c_1 \int_0^T \frac{|f(t) - g(t)|}{t}\,dt + c_2\,\frac{\|G\|_{\mathrm{Lip}}}{T}, \tag{A.1}$$

which however requires finiteness of the Lipschitz seminorm of G (i.e., an absolute continuity of G and the boundedness of its density). Here, c_1, c_2 are positive universal constants, and $T > 0$ is arbitrary, cf. e.g. [28].

In order to remove any restriction on G, V. M. Zolotarev proposed to replace the Kolmogorov distance in (A.1) with a generally weaker Lévy distance $L(F,G)$, defined as the minimal value $h \geq 0$, such that

$$F(x-h) - h \leq G(x) \leq F(x+h) + h, \quad \forall x \in \mathbf{R}.$$

Clearly, $L(F,G) \leq \|F-G\|$, while

$$\|F-G\| \leq (1+M)L(F,G), \qquad M = \|G\|_{\mathrm{Lip}}.$$

So, these two distances are in essense equivalent, when one of the distributions is known to have a bounded density.

In this important special case, an answer given by Zolotarev in [32] says the following.

Theorem A.1. *For any $T \geq T_0$,*

$$L(F,G) \leq c_1 \int_0^T \frac{|f(t)-g(t)|}{t}\,dt + c_2\,\frac{\log T}{T}, \tag{A.2}$$

where $c_1, c_2 > 0$ and $T_0 > 1$ are universal constants. One may take $c_1 = \frac{1}{\pi}$, $c_2 = 2e$, and $T_0 = 1.3$.

This bound is almost as good as (A.1), but contains an extra $\log T$ term on the right-hand side. So, for example, when applying (A.2) to the central limit theorem in the usual i.i.d. case with the standard normal distribution function $G = \Phi$ and with T of order n, we will get an additonal $\log n$ factor for the rate of convergence of the distributions of the normalized sums to Φ in Kolmogorov distance. On the other hand, (A.2) is more general, since it has no restriction on G. This appears to be rather useful in various applications when we cannot say anything definite about G. In particular, the problem on the rate of convergence of empirical distributions F_n to the mean distribution $F = \mathbf{E}\,F_n$ may be considered in terms of L. Also, let us mention that the $\log T$ term cannot be removed from (A.2) in general, as was shown by A. Zaitsev.

Since the proof of Theorem A.1 is not given in standard books, we will put it below (although with different constants). Note that in some applications, it is desirable to remove the condition $T \geq T_0$. This can be reached using Theorem 1 by replacing $\frac{\log T}{T}$ with, for example, $\frac{\log(1+T)}{T}$. Indeed, the latter quantity is close to 1, when T is small, while $L(F,G) \le 1$. Hence, choosing a bigger constant c_2 in Theorem A.1, one may do this replacement with any $T > 0$.

But with a straightforward argument one can also prove:

Theorem A.2. *For any $T > 0$,*

$$L(F,G) \le 0.4 \int_0^T \frac{|f(t) - g(t)|}{t}\,dt + 4\,\frac{\log(1+T)}{T}.$$

Before proving, introduce the function

$$\mathcal{L}(x,h) = \max\{F(x-h) - G(x+h),\, G(x-h) - F(x+h)\}, \qquad x \in \mathbf{R},\ h \ge 0.$$

From the definition of the Lévy distance it follows that, given constants $A \ge 0$, $h \ge 0$,

$$\sup_x \mathcal{L}(x,h) \le A + 2h \implies L(F,G) \le A + 2h. \tag{A.3}$$

Proof of Theorems A.1 and A.2. We use smoothing with the help of the standard normal distribution

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2}\,dy \qquad (x \in \mathbf{R}).$$

Given a parameter $\sigma > 0$ to be chosen later on, let $\Phi_\sigma(x) = \Phi(\frac{x}{\sigma})$. Consider convolutions $F_\sigma = F * \Phi_\sigma$ and $G_\sigma = G * \Phi_\sigma$, and define, for any $x \in \mathbf{R}$,

$$I(x) \equiv F_\sigma(x) - G_\sigma(x) \ = \int_{-\infty}^{\infty} (F(x-\sigma y) - G(x - \sigma y))\,d\Phi(y). \tag{A.4}$$

Note that, using the Kolmogorov distance, we have a uniform upper bound on these integrals,

$$|I(x)| \le \|F_\sigma - G_\sigma\| \equiv \sup_x |F_\sigma(x) - G_\sigma(x)|. \tag{A.5}$$

Take another parameter $a > 0$ and split the integral in (A.4) into $I(x) = I_0 + I_1$, where

$$\begin{aligned} I_0 &= \int_{|y| \le a} (F(x-\sigma y) - G(x-\sigma y))\,d\Phi(y), \\ I_1 &= \int_{|y| \ge a} (F(x-\sigma y) - G(x-\sigma y))\,d\Phi(y). \end{aligned}$$

In particular,

$$|I_1| \le 2\,(1 - \Phi(a)) \equiv \gamma, \tag{A.6}$$

while, by the monotonicity of distribution functions,

$$I_0 \geq (F(x-\sigma a) - G(x+\sigma a))\,(1-\gamma).$$

Since also

$$-I_0 \geq (G(x-\sigma a) - F(x+\sigma a))\,(1-\gamma),$$

we have $|I_0| \geq (1-\gamma)\,\mathcal{L}(x,\sigma a)$. Thus,

$$|I_0| \geq (1-\gamma)\,\mathcal{L}(x,\sigma a).$$

Using $|I_0| \leq |I(x)| + |I_1|$, together with (4) we get

$$(1-\gamma)\,\mathcal{L}(x,\sigma a) \leq |I(x)| + \gamma.$$

Taking the sup over all x, we obtain that

$$\sup_x \mathcal{L}(x,\sigma a) \leq \frac{1}{1-\gamma}\,\|F_\sigma - G_\sigma\| + \frac{\gamma}{1-\gamma}. \tag{A.7}$$

Now, let us recall that, by the inversion formula

$$F_\sigma(x) - G_\sigma(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\,\frac{f(t)-g(t)}{-it}\,e^{-\sigma^2 t^2/2}\,dt,$$

one has a general elementary bound for the Kolmogorov distance

$$\|F_\sigma - G_\sigma\| \leq \frac{1}{\pi}\int_0^\infty \frac{|f(t)-g(t)|}{t}\,e^{-\sigma^2 t^2/2}\,dt.$$

Here the integral on the right-hand side may be splitted into the two parts corresponding to the regions $0 < t < T$ and $t > T$. For the second one, integrating by parts, we will use the estimate

$$\begin{aligned}
\int_T^\infty \frac{1}{t}\,e^{-\sigma^2 t^2/2}\,dt \ &= \ \int_{\sigma T}^\infty \frac{1}{s}\,e^{-s^2/2}\,ds \ = \ \int_{\sigma T}^\infty \frac{1}{s^2}\,d\big(-e^{-s^2/2}\big) \\
&= \ \frac{1}{\sigma^2 T^2}\,e^{-\sigma^2 T^2/2} - 2\int_{\sigma T}^\infty \frac{1}{s^3}\,e^{-s^2/2}\,ds \ < \ \frac{1}{\sigma^2 T^2}\,e^{-\sigma^2 T^2/2}.
\end{aligned}$$

Since $|f(t)-g(t)| \leq 2$, it gives

$$\int_T^\infty \frac{|f(t)-g(t)|}{t}\,e^{-\sigma^2 t^2/2}\,dt \ \leq \ \frac{2}{\sigma^2 T^2}\,e^{-\sigma^2 T^2/2},$$

so,

$$\|F_\sigma - G_\sigma\| \le \frac{1}{\pi} \int_0^T \frac{|f(t)-g(t)|}{t}\,dt + \frac{2}{\pi\sigma^2 T^2}\,e^{-\sigma^2T^2/2}.$$

Hence, by (A.7),

$$\sup_x \mathcal{L}(x,\sigma a) \ \le\ \frac{1}{\pi(1-\gamma)} \int_0^T \frac{|f(t)-g(t)|}{t}\,dt + \frac{2}{\pi(1-\gamma)\,\sigma^2T^2}\,e^{-\sigma^2T^2/2} + \frac{\gamma}{1-\gamma}.$$

Thus, according to the remark (A.3) with $h = \sigma a$, if

$$\frac{2}{\pi(1-\gamma)\,\sigma^2T^2}\,e^{-\sigma^2T^2/2} + \frac{\gamma}{1-\gamma} \ \le\ 2\sigma a, \tag{A.8}$$

then, we would have that

$$L(F,G) \ \le\ \frac{1}{\pi(1-\gamma)} \int_0^T \frac{|f(t)-g(t)|}{t}\,dt + 2\sigma a. \tag{A.9}$$

Case of Theorem A.1. For $T \ge e$, choose

$$\sigma = \frac{1}{T}\sqrt{2\log T} \qquad \text{and} \qquad a = \sqrt{2\log T}.$$

Then

$$\gamma = 2\,(1-\Phi(a)) \le e^{-a^2/2} = \frac{1}{T}.$$

Since also $a \ge \sqrt{2}$, we have $\gamma \le 2\left(1-\Phi\big(\sqrt{2}\big)\right) \sim 0.16$. Hence, $\frac{1}{\pi(1-\gamma)} < 0.4$ and $\frac{\gamma}{1-\gamma} < \frac{2}{T}$. Moreover,

$$\begin{aligned}\frac{2}{\pi(1-\gamma)\,\sigma^2T^2}\,e^{-\sigma^2T^2/2} + \frac{\gamma}{1-\gamma} \ &=\ \frac{1}{\pi(1-\gamma)\,T\log T} + \frac{\gamma}{1-\gamma} \\ &<\ \frac{0.4}{T\log T} + \frac{2}{T} \ <\ \frac{4\log T}{T} \ =\ 2\sigma a.\end{aligned}$$

Hence, the condition (A.8) is fulfilled, and we obtain (A.9). Theorem A.1 is thus proved with $T_0 = e$, $c_1 = 0.4$ and $c_2 = 4$.

Case of Theorem A.2. For $T > 0$, choose

$$\sigma = \frac{1}{T}\sqrt{2\log(1+T)} \qquad \text{and} \qquad a = \sqrt{2\log(1+T)}.$$

Then

$$\gamma = 2\,(1-\Phi(a)) \le e^{-a^2/2} = \frac{1}{1+T}.$$

Again $a \geq \sqrt{2}$, so $\gamma < 0.16$, $\frac{1}{\pi(1-\gamma)} < 0.4$ and $\frac{\gamma}{1-\gamma} < \frac{2}{1+T}$. Moreover,

$$\begin{aligned}\frac{2}{\pi(1-\gamma)\,\sigma^2 T^2}\, e^{-\sigma^2 T^2/2} + \frac{\gamma}{1-\gamma} &= \frac{1}{\pi(1-\gamma)\,(1+T)\log(1+T)} + \frac{\gamma}{1-\gamma} \\ &< \frac{0.4}{(1+T)\log(1+T)} + \frac{2}{1+T} \\ &< \frac{4\log(1+T)}{T} \;=\; 2\sigma a.\end{aligned}$$

Hence, the condition (A.8) is fulfilled again, and we obtain (A.9). Theorem A.2 is proved. $\square$

Appendix B

ψ_α Norms

Let $\xi \geq 0$ with $\xi = L(F_n, F)$. Then it is well known that Markov's inequality gives many useful deviation inequalities under certain moment assumptions.

For example, if $\mathbf{E}\,\xi \leq c$, then

$$\mathbf{P}(\xi \geq r) \leq \frac{c}{r}.$$

If $\|\xi\|_2 = \sqrt{\mathbf{E}\,\xi^2} \leq c$, then

$$\mathbf{P}(\xi \geq r) \leq \frac{\mathbf{E}\,\xi^2}{r^2} \leq \frac{c}{r^2}.$$

Also if $\|\xi\|_p \leq c_p$, then

$$\mathbf{P}(\xi \geq r) \leq \frac{c_p}{r^p}.$$

In the case of exponential integrability, if $\mathbf{E}\,e^{\xi/\lambda} \leq 2$, then

$$\mathbf{P}(\xi \geq r) \leq \frac{\mathbf{E}\,e^{\xi/\lambda}}{e^{r/\lambda}} \leq 2e^{-r/\lambda}.$$

In our case of interest, We would like to have that $\mathbf{E}\,e^{\xi^2/\lambda^2} \leq 2$ implies

$$\mathbf{P}(\xi \geq r) \leq 2e^{-r^2/\lambda^2}.$$

In order to further analyze our hypothesis we introduce a more generalized form of L^p norms called Orlicz norm and a more general convex function called a Young function.

Definition B.1. (Young function). A function $\Psi : \mathbf{R} \to [0, \infty)$ is called a Young function if $\Psi(0) = 0$, $\Psi(t) > 0$ for $t > 0$, and Ψ is convex.

Some common examples of Young functions are given below.

- $\Psi(t) = |t|^p$
- $\Psi_1(t) = e^{|t|} - 1$
- $\Psi_2(t) = e^{t^2} - 1$

In the literature, Young functions are sometimes required to be lower semicontinuous or allowed to be $+\infty$-valued. For our purposes, we are not interested in trivial cases, and hence our definition can be more restrictive. Now we define the Orlicz norm.

Definition B.2. (Orlicz norm). Let Ψ be a Young function. For any measurable function ξ on $\mathbf{R}$

$$\|\xi\|_\Psi = \inf\left\{\lambda > 0 : \mathbf{E}\,\Psi\left(\frac{|\xi|}{\lambda}\right) \le 1\right\}.$$

The Young function $\Psi(t) = |t|^p$ in the examples yields the usual norm $\|\xi\|_\Psi = (\mathbf{E}\,|\xi|^p)^{1/p} = \|\xi\|_p$ on L_p. As for the other two examples, we can get the necessary conditions for our deviation inequalities in terms of the Orlicz norm. Hence,

$$\mathbf{E}\,e^{|\xi|/\lambda} \le 2 \quad \Leftrightarrow \quad \|\xi\|_{\Psi_1} \le \lambda$$

and

$$\mathbf{E}\,e^{|\xi^2|/\lambda^2} \le 2 \quad \Leftrightarrow \quad \|\xi\|_{\Psi_2} \le \lambda.$$

Using these two Young functions, Ψ_1 and Ψ_2, we can also relate the Orlicz norms with the L^p norms. In fact, we can define a whole class of these norms as Ψ_α norms generated by the Young function $\Psi_\alpha(t) = e^{|t|^\alpha} - 1$ with an arbitrary parameter $\alpha \ge 1$. It is easy to see the following proposition as expained in the Appendix of [4].

Proposition B.3. *Given $\alpha \ge 1$ and for any measurable function ξ on $\mathbf{R}$,*

$$\frac{1}{2}\sup_{p\ge 1}\frac{\|\xi\|_p}{p^{1/\alpha}} \le \|\xi\|_{\Psi_\alpha} \le 2e\sup_{p\ge 1}\frac{\|\xi\|_p}{p^{1/\alpha}}.$$

Therefore, $\|\cdot\|_{\Psi_1}$ and $\|\cdot\|_{\Psi_2}$ are also equivalent to L^p norms, and we have

$$\|\xi\|_{\Psi_1} \sim \sup_{p\ge 1}\frac{\|\xi\|_p}{p} \quad \text{and} \quad \|\xi\|_{\Psi_2} \sim \sup_{p\ge 1}\frac{\|\xi\|_p}{\sqrt{p}}.$$

We can now restate our hypothesis of the deviation inequality to be that in order to obtain $\mathbf{E}\,e^{\xi^2/\lambda^2} \le 2$, we need to show that $\|\xi\|_p \le C\lambda\sqrt{p}$ with C an absolute constant.