Moments of the scores

Sergey G. Bobkov[1]

***Abstract*—Upper bounds on absolute moments of the scores are derived for sums of independent random variables in terms of the moments of the scores, as well as in terms of the total variation norm of densities of summands.**

***Index Terms*—Score, score function, Fisher information, total variation, convolution, Stam's inequality.**

I. Introduction

If X is a random variable with an absolutely continuous density f, its score function is defined by $\rho(x) = f'(x)/f(x) = (\log f)'(x)$, where the derivatives may be understood in the Radon-Nikodym sense. The score of X is the random variable

$$\rho(X) = \frac{f'(X)}{f(X)}.$$

It is well defined with probability one, and its distribution plays an important role in various problems of Statistics and Information Theory, cf. e.g. [6], [8], [11].

Let us look at the meaning of the absolute moments

$$I_k(X) = \mathbb{E}\,|\rho(X)|^k$$

for positive integer values of k. The first absolute moment

$$I_1(X) = \|f\|_{\rm TV} = \int_{-\infty}^{\infty} |f'(x)|\,{\rm d}x$$

describes the total variation of the density f. In this case, the definition naturally extends to the larger class of probability distributions on the line, whose densities have bounded total variation in the sense of Theory of Functions (including, for example, the uniform distribution on finite intervals). Note that, if $I_1(X)$ is finite, the first moment $\mathbb{E}\rho(X)$ is necessarily vanishing. The second moment

$$I(X) = I_2(X) = \int_{-\infty}^{\infty} \frac{f'(x)^2}{f(x)}\,{\rm d}x$$

represents the Fisher information contained in the distribution of X. Both quantities, I_1 and I_2, are classical objects in different areas.

Higher order moments I_k were introduced by Lions and Toscani [9] in their study of convergence of densities (and of their powers) in the central limit theorem in Sobolev spaces. As was mentioned in [9], the functional I_4 was also considered by Gabetta [7] in the context of the kinetic theory of gases to study the convergence to equilibrium in Kac's model. See also [1], where these moments appear implicitly with the aim to control translates of product probability measures (and where the finiteness of exponential or Gaussian moments of $\rho(X)$ are required).

[1]School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455 USA (bobkov@math.umn.edu).
Partially supported by NSF grants DMS-1612961 and DMS-1855575.

However, in contrast with $I_1(X)$, which can easily be treated by Fourier methods, it is often not clear how to bound or even to verify that $I_k(X)$ is finite for $k > 1$. Let us restrict ourselves to the case where the distribution of X has a convolution structure, that is, when

$$X = S_n = X_1 + \cdots + X_n \tag{I.1}$$

for some independent random variables X_i. The values $(I_k(S_n))^{1/k}$ are non-decreasing in k, but are non-increasing in n (cf. Corollary III.2 below). If $k = 2$, a stronger property is contained in Stam's inequality, which may be written in the linearized form as

$$I(S_n) \,\le\, a_1^2 I(X_1) + \cdots + a_n^2 I(X_n)$$

with arbitrary real numbers $a_i > 0$ such that $a_1 + \cdots + a_n = 1$. Lions and Toscani proposed an extension of this relation to the moments of the scores of even orders $k = 2m$; in case of the two summands, it similarly states that, for all $a, b > 0$, $a + b = 1$,

$$\begin{aligned} I_k(X_1 + X_2) \quad &\le \quad a^k I_k(X_1) + b^k I_k(X_2) + \\ \sum_{j=2}^{k-2} \binom{k}{j} a^{k-j}\, b^j\, (I_{k-2}(X_1))^{\frac{k-j}{k-2}}\, (I_{k-2}(X_2))^{\frac{j}{k-2}}. \end{aligned} \tag{I.2}$$

If X_i's in (I.1) are identically distributed (i.i.d.) with finite $I_k(X_1)$, the inequality (I.2) ensures not only the finiteness of $I_k(S_n)$, but also implies a uniform boundedness of the moments of the scores along rescaled convolutions. In [9], this is shown to hold by induction on n via a bound implying $I_{2m}(S_n/\sqrt{n}) \le c_m I_{2m}(X_1)$ with implicit constants c_m. Here, we first extend the latter inequality to general weighted sums

$$Z_n = \alpha_1 X_1 + \cdots + \alpha_n X_n,$$

where we assume that $\alpha_i \ge 0$ with $\alpha_1^2 + \cdots + \alpha_n^2 = 1$.

Theorem I.1. *If $I_{2m}(X_i) \le I$ for all $i \le n$, then*

$$I_{2m}(Z_n) \le c_m I \tag{I.3}$$

with $c_m \le (2m)!\,({\rm e}/m)^m$.

The right-hand side of (I.3) does not depend on α_i. The sharpness of the constants in this inequality (for large m) may be tested on the example where X_i's have a two-sided exponential distribution. Then $|\rho(X_i)| = 1$ a.e., so that $I_{2m}(X_i) = 1$ for all m. Choosing $\alpha_i = 1/\sqrt{n}$, we then have $Z_n \Rightarrow Z \sim N(0,1)$, and moreover,

$$I_{2m}(Z_n) \,\to\, I_{2m}(Z) \,=\, \mathbb{E}\,|Z|^{2m} \,=\, \frac{(2m)!}{2^m m!} \quad (n \to \infty).$$

Hence, the optimal constant in (I.3) admits a lower bound $c_m \ge (2m)!/(2^m m!)$ meaning that the upper bound is optimal modulo an exponentially growing factor.

Similar statements remain to be valid when dealing with exponential and Gaussian moments.

Theorem I.2. *If* $\mathbb{E}\,\exp\{|\rho(X_i)|/\sigma\} \le 2$ *for all* $i \le n$ $(\sigma > 0)$, *then*

$$\mathbb{E}\,\exp\left\{|\rho(Z_n)|/4\sigma\right\} \le 2. \tag{I.4}$$

Moreover, if $\mathbb{E}\,\exp\{\rho(X_i)^2/\sigma^2\} \le 2$, *then*

$$\mathbb{E}\,\exp\left\{\rho(Z_n)^2/K\sigma^2\right\} \le 2 \tag{I.5}$$

with some absolute constant (e.g. $K = 6$*).*

The inequality (I.5) exhibits a subgaussian behavior of the scores $\rho(Z_n)$ uniformly over all admissible coefficients α_i under a similar hypothesis that the scores of the summands are subgaussian. This is a full analogue of the well-known property of the weighted sums: If $\mathbb{E}\,\exp\{X_i^2/\sigma^2\} \le 2$ and $\mathbb{E}X_i = 0$, then

$$\mathbb{E}\,\exp\left\{Z_n^2/K\sigma^2\right\} \le 2.$$

A similar property takes place in terms of the exponential moments as well.

While the conclusions such as the one in Theorem I.1 are based on the finiteness of $I_k(X_i)$, one may wonder whether or not, the values $I_k(S_n)$ for the sums in (I.1) become finite for a certain $n = n_0$ assuming that just the first moments $I_1(X_k)$ are finite. This is known to be indeed true in the Fisher information case $k = 2$, cf. [3]. The next statement extends this observation to the whole range $k \ge 1$.

Theorem I.3. *If* $b_i = I_1(X_i) < \infty$ *in* (I.1), *then* $I_k(S_n) < \infty$ *whenever* $n \ge k+1$. *Moreover,*

$$I_k(S_{k+1}) \;\le\; c_k\, b_1 \ldots b_{k+1} \left(\frac{1}{b_1} + \cdots + \frac{1}{b_{k+1}}\right) \tag{I.6}$$

with $c_k = k^k/(2^k k!)$

The main point of (I.6) is that its right-hand side is finite and has an explicit form. Since the constants c_k grow exponentially fast, for an effective estimation of $I_k(S_n)$, it is better to combine (I.6) with (I.2) or (I.3) by splitting the sequence $X_1, \ldots, X_n$ into the groups with at least $k+1$ elements in each group. One may apply (I.6) to the group summands and then involve (I.3) or a corresponding variant of (I.2) for the sum of two or more random variables. On this way, one can reach estimates such as (I.7) below dealing with the i.i.d. situation.

But, first let us note that Theorem I.3 is no longer valid for $k \ge 2$ and $n \le k$, as may be seen on the example of the uniform distribution. Indeed, if X_i are uniformly distributed in $(0,1)$, then S_n has a density described as $f(x) = \frac{1}{(n-1)!}\, x^{n-1}$ for $0 < x < 1$. Since the function

$$\left(\frac{f'(x)}{f(x)}\right)^k f(x) \;=\; (n-1)\, x^{n-k-1}$$

is not integrable on $(0,1)$ for $n \le k$, necessarily $I_k(S_n) = \infty$.

The inequality (I.6) has the following application to the characterization of the finiteness of the moments in the scheme of i.i.d. summands. We denote by $v(t) = \mathbb{E}\,\mathrm{e}^{itX_1}$ the common characteristic function.

Theorem I.4. *Let* $(X_i)_{i\ge1}$ *be i.i.d. random variables such that* $\mathbb{E}\,|X_1| < \infty$. *The following properties are equivalent, for any fixed* $k \ge 1$*:*

a) There exists n *such that* $I_k(S_n) < \infty$*;*
b) There exists n *such that* S_n *has a density with bounded total variation;*
c) For some $\varepsilon > 0$, *we have* $v(t) = o(t^{-\varepsilon})$ *as* $t \to \infty$.

Moreover, if X_1 *has a density with bounded total variation, then with some constants* A_k *depending on* k *only,*

$$\sup_{n\ge k+1} I_k(S_n/\sqrt{n}) \le A_k\,(I_1(X_1))^k. \tag{I.7}$$

The property $b)$ is just $a)$ specialized to $k = 1$. Thus, the property $a)$ does not depend on k.

The paper is organized as follows. In Section II, we discuss connections between the scores and the so-called L-functions associated with given probability distributions. Together with the Brunn-Minkowski inequality, this will allow us to derive the inequality (I.6) in the case of uniform distributions on finite intervals, cf. Section VI. Convexity properties of the moments and Stam-type inequalities for the functional I_k are considered in Sections III-IV. Theorems I.1-I.2 and I.3-I.4 are respectively proved in Sections V and VII. In the last two sections, we discuss applications to the decay of densities and give remarks on the relationship between the moments of the scores and the usual moments (via Cramér-Rao-type inequalities).

II. Distribution of the score and the associated L-function

Suppose that the distribution of a random variable X with distribution function $F(x) = \mathbb{P}\{X \le x\}$, $x \in \mathbb{R}$, is supported on some interval (a,b), finite or not, and has an a.e. positive density f on that interval. The inverse function $F^{-1} : (0,1) \to (a,b)$ is then strictly increasing and continuous on $(0,1)$.

As a preliminary step, let us mention one useful general representation for the moments $I_k(X) = \mathbb{E}\,|\rho(X)|^k$ involving the function

$$L(u) = f(F^{-1}(u)), \qquad 0 < u < 1, \tag{II.1}$$

which often appears in many isoperimetric-type inequalities, serving as the so-called isoperimetric profile of the distribution of X. Note that L uniquely determines F up to a shift parameter. For example, it follows from (II.1) that

$$\int_0^1 \frac{1}{L(u)}\,\mathrm{d}u = b - a. \tag{II.2}$$

Indeed, if a random variable U is uniformly distributed on $(0,1)$, then $F^{-1}(U)$ has the same distribution as X, so that

$$\begin{aligned}\int_0^1 \frac{1}{L(u)}\,\mathrm{d}u \;&=\; \mathbb{E}\,\frac{1}{f(F^{-1}(U))} \\ &=\; \mathbb{E}\,\frac{1}{f(X)} \;=\; \int_a^b 1\,\mathrm{d}x \;=\; b-a.\end{aligned}$$

If additionally f is absolutely continuous and has a finite total variation (hence $f(a+) = f(b-) = 0$), then L is absolutely continuous as well, with $L(0+) = L(1-) = 0$. In this case, denote by L' the Radon-Nikodym derivative of L.

Proposition II.1. *Suppose that the distribution of X has an absolutely continuous density f of bounded total variation, supported and a.e. positive on some interval, finite or not. If a random variable U is uniformly distributed on $(0,1)$, then $\rho(X)$ and $L'(U)$ are equidistributed. In particular, for any k,*

$$I_k(X) = \int_0^1 |L'(u)|^k \, \mathrm{d}u. \tag{II.3}$$

Proof. By the assumption, F^{-1} is absolutely continuous and has the Radon-Nikodym derivative

$$(F^{-1})'(u) = \frac{1}{f(F^{-1}(u))} = \frac{1}{L(u)}.$$

Thus, L is absolutely continuous. By the chain rule applied in (II.1), it has the Radon-Nikodym derivative

$$L'(u) = f'(F^{-1}(u))\,(F^{-1})'(u) = \frac{f'(F^{-1}(u))}{f(F^{-1}(u))}.$$

Inserting $u = U$ and using the property that $F^{-1}(U)$ and X are equidistributed, we obtain the equality $L'(U) = \rho(X)$ in the distributional sense. □

We now express some possible convexity properties of the density f in terms of the L-function (cf. also [2] for related issues about the κ-concave probability measures).

Proposition II.2. *Suppose that the distribution of X has a continuous density f, which is supported and positive on some finite interval (a,b). Given $k \geq 1$, the function $f^{1/k}$ is concave on (a,b), if and only if $L^{(k+1)/k}$ is concave on $(0,1)$.*

Proof. Assuming without loss of generality that f has a continuous derivative, we have $L'(F(x)) = f'(x)/f(x)$, and then

$$(L^{\frac{k+1}{k}})'(F(x)) = (k+1)\,(f^{\frac{1}{k}})'(x), \qquad a < x < b.$$

Therefore, the derivative $(f^{1/k})'$ does not increase on (a,b), if and only if $(L^{(k+1)/k})'$ does not increase on $(0,1)$. □

Let us also note that the distribution of $\rho(X)$ does not determine the distribution of X in a unique way. To see this, one may define the functions L_1 and L_2 in the following way. Put $L_1(t) = \min(t, 1-t)$ which is the L-function for the two-sided exponential distribution with density $f(x) = \frac{1}{2}\,e^{-|x|}$, and let L_2 be piecewise linear on $[0,1]$ with $L_2(0) = L_2(1/2) = L_2(1) = 0$ and $L_2(1/4) = L_2(3/4) = 1/4$, which corresponds to a different symmetric probability measure. In both cases, $L_i'(U)$ and therefore $\rho(X_i)$ have a symmetric Bernoulli distribution on $\{-1,1\}$.

III. **Convexity of moments**

Let us now address some general convexity properties of the moments of scores (needed in the proof of Theorem I.3). We write $I_k(f) = I_k(X)$ when a random variable X has density f, with the convention that $I_k(f) = \infty$, if f is not absolutely continuous (in case $k \geq 2$). In case $k = 1$, let us recall that the total variation norm is defined to be

$$\|f\|_{\mathrm{TV}} = I_1(f) = \sup \sum_{k=1}^N |f(x_k) - f(x_{k-1})|,$$

where the supremum is running over all collections of points $x_0 < x_1 < \cdots < x_N$. For the finiteness of this norm, the function f must have at most countably many points of discontinuity, at which it makes sense to require that f be right-continuous.

If f is a convex mixtures of several densities,

$$f = t_1 f_1 + \cdots + t_N f_N \quad (t_i \geq 0, \quad t_1 + \cdots + t_N = 1),$$

then, as was stressed in [9], we have Jensen's inequality

$$I_k(f) \leq t_1 I_k(f_1) + \cdots + t_N I_k(f_N). \tag{III.1}$$

This readily follows from the fact that the homogeneous function $R(u,v) = u^m/v^{m-1}$ is convex on the upper half-plane $u \in \mathbb{R}$, $v > 0$. We need to extend this inequality to arbitrary "continuous" convex mixtures of densities.

The collection $\mathfrak{P}$ of all (probability) densities on the real line represents a closed subset of L^1 for the weak $\sigma(L^1, L^\infty)$ topology. For any Borel set $A \subset \mathbb{R}$, the functional $q \to \int_A q(x)\,dx$ is bounded and continuous on $\mathfrak{P}$. So, given a Borel probability measure π on $\mathfrak{P}$, one may define the probability measure on the real line

$$\mu(A) = \int_{\mathfrak{P}} \left[\int_A q(x)\,\mathrm{d}x \right] \mathrm{d}\pi(q).$$

It is absolutely continuous with respect to the Lebesgue measure and has some density $f(x) = \frac{\mathrm{d}\mu(x)}{\mathrm{d}x}$ called the convex mixture of densities with mixing measure π. For short,

$$f = \int_{\mathfrak{P}} q \,\mathrm{d}\pi(q).$$

A corresponding extension of (III.1) is given in the following:

Proposition III.1. *The functional $f \to I_k(f)$ is lower semi-continuous (hence Borel measurable) on the space $\mathfrak{P}$. Moreover, if f is a convex mixture of densities with a mixing probability measure π, then*

$$I_k(f) \leq \int_{\mathfrak{P}} I_k(q)\,\mathrm{d}\pi(q). \tag{III.2}$$

In particular, (III.2) implies the monotonicity property of the moments with respect to convolutions.

Corollary III.2. *For all independent random variables X, Y,*

$$I_k(X+Y) \leq \min\{I_k(X), I_k(Y)\}. \tag{III.3}$$

Indeed, assuming that $I_k(X) < \infty$, the distribution of X has an absolutely continuous density, say q. In this case, $X+Y$ has a density f representable as a convex mixture of the shifted densities $q_h(x) = q(x-h)$, for which $I_k(q_h) = I_k(q) = I_k(X)$. Hence, $I_k(X+Y) \leq I_k(X)$. $\square$

In the case $k = 2$, the topological properties of the functional I_k together with the inequality (III.2) are discussed in detail in [3], cf. Propositions 3.2-3.3 therein. Their proofs can easily be extended to the general case $k \geq 1$, so we omit the proof of Proposition III.1. What is important, in the case $k = 1$, (III.2) may actually be reversed for a suitable measure π supported on the (two-dimensional) set $U \subset \mathfrak{P}$ of densities of the form $q(x) = \frac{1}{b-a}\, 1_{(a,b)}(x)$ with parameters $a < b$ (cf. [3], Lemma 4.3):

Proposition III.3. *Any density f of bounded total variation can be represented as a convex mixture $f = \int_U q \, \mathrm{d}\pi(q)$ with a mixing probability measure π on U such that*

$$I_1(f) = \int_U I_1(q) \, \mathrm{d}\pi(q). \tag{III.4}$$

Using the transference, the integration in (III.4) may be carried out over the half-plane $\{(a,b) : a < b\}$. For example, if f is supported and non-increasing on $(0,\infty)$, there is a canonical representation

$$f(x) = \int_0^\infty \frac{1}{x_1}\, 1_{\{0<x<x_1\}} \, \mathrm{d}\pi(x_1) \qquad \text{a.e.}$$

with a unique mixing probability measure π on $(0,\infty)$. In this case, $I_1(f) = 2f(0+)$, and (III.4) is obvious. One may write a similar representation for densities of unimodal distributions.

IV. Stam-type inequalities

Given independent random variables $X_1, \dots, X_n$ with finite moments $I_{2m}(X_i)$ of the scores $\rho(X_i)$, consider the weighted sums

$$Z_n = \alpha_1 X_1 + \dots + \alpha_n X_n,$$

assuming as before that $\alpha_i \geq 0$ with $\alpha_1^2 + \dots + \alpha_n^2 = 1$. Here, we also involve non-absolute moments

$$M_k(X_i) = \mathbb{E}\, \rho(X_i)^k, \quad 1 \leq k \leq 2m.$$

Note that $M_1(X_i) = 0$, while $|M_k| \leq I_k$.

We will need the following multinomial bounds with a natural convention $M_0(X_i) = I_0(X_i) = 1$.

Proposition IV.1. *For any integer $m \geq 1$,*

$$\begin{aligned} I_{2m}(Z_n) \;\; &\leq \;\; \sum \binom{2m}{k_1 \dots k_n} \alpha_1^{k_1} \dots \alpha_n^{k_n} \\ &\quad \times \; M_{k_1}(X_1) \dots M_{k_n}(X_n), \end{aligned} \tag{IV.1}$$

In particular,

$$\begin{aligned} I_{2m}(Z_n) \;\; &\leq \;\; \sum \binom{2m}{k_1 \dots k_n} \alpha_1^{k_1} \dots \alpha_n^{k_n} \\ &\quad \times \; I_{k_1}(X_1) \dots I_{k_n}(X_n). \end{aligned} \tag{IV.2}$$

In both cases, the summation is performed over all non-negative integers $k_i \neq 1$ such that $k_1 + \dots + k_n = 2m$.

Being specialized to $m = 1$, (IV.1) and (IV.2) coincide and represent an equivalent form of Stam's inequality, which is stronger than the monotonicity property (III.3) of the Fisher information. However, if $m \geq 2$, the inequality (III.3) cannot be deduced from (IV.1).

In Lions and Toscani [9], the inequalities (IV.1)-(IV.2) were obtained along derivation of (I.2) in the binomial form for the weighted sum of two random variables. Let us recall a simple argument which is slightly different than the standard one used in the proof of the Stam inequality (compare e.g. with [8]). Assuming without loss of generality that the densities f_i of X_i are continuously differentiable and positive ($i = 1, 2$), the density f of $X_1 + X_2$ has a derivative representable as

$$\frac{f'(x)}{f(x)} = \int_{-\infty}^\infty \Big(a_1 \frac{f_1'(x-y)}{f_1(x-y)} + a_2 \frac{f_2'(y)}{f_2(y)} \Big) \, \mathrm{d}\mu_x(y)$$

with arbitrary $a_i > 0$, $a_1 + a_2 = 1$, and where $\mathrm{d}\mu_x(y)/\mathrm{d}y = f_1(x-y) f_2(y)/f(x)$. Since μ_x is a probability measure, one may apply Jensen's inequality, which gives

$$\Big(\frac{f'(x)}{f(x)}\Big)^{2m} \leq \int_{-\infty}^\infty \Big(a_1 \frac{f_1'(x-y)}{f_1(x-y)} + a_2 \frac{f_2'(y)}{f_2(y)} \Big)^{2m} \, \mathrm{d}\mu_x(y).$$

One may now expand the integrand according to the binomial formula, multiply both sides by $f(x)$ and integrate over the variable x. We then arrive at

$$I_{2m}(X_1 + X_2) \;\leq\; \sum \binom{2m}{k_1 k_2} a_1^{k_1} a_2^{k_2} M_{k_1}(X_1) M_{k_2}(X_n)$$

without the terms corresponding to $k_1 = 1$ and $k_2 = 1$. To get (IV.1), it remains to write down this bound for the random variables $\alpha_i X_i$ with $a_i = \alpha_i^2$. As for the general case $n \geq 2$, it is easily obtained by induction on the basis of $n = 2$.

Expanding the cosh-function in a power series and using the property $\mathbb{E}\rho(X_i) = 0$, (IV.2) implies:

Corollary IV.2. *For any $t \in \mathbb{R}$, putting $\xi_i = |\rho(X_i)|$, we have*

$$\mathbb{E}\, \cosh(t\rho(Z_n)) \;\leq\; \prod_{i=1}^n \mathbb{E}\, \big(e^{|t|\alpha_i \xi_i} - |t|\alpha_i \xi_i \big). \tag{IV.3}$$

V. Proof of Theorems I.1-I.2

Proof of Theorem I.1. Since

$$(I_{k_i}(X_i))^{1/k_i} \leq (I_{2m}(X_i))^{1/(2m)},$$

while, by the assumption, $I_{2m}(X_i) \leq I$, the inequality (IV.2) implies $I_{2m}(Z_n) \leq K_m I$, where

$$K_m = \sum \binom{2m}{k_1 \dots k_n} \alpha_1^{k_1} \dots \alpha_n^{k_n} \tag{V.1}$$

with summation as before. Put $K_0 = 1$ and introduce the generating function associated to the sequence $(K_m)_{m \geq 0}$,

$$\psi(z) = \sum_{m=0}^\infty \frac{K_m}{(2m)!}\, z^{2m}, \qquad z \in \mathbb{C},$$

so that $K_m = \psi^{(2m)}(0)$. It follows from (V.1) that

$$\psi(z) \;=\; \prod_{j=1}^{n} \sum_{k_j \geq 0,\, k_j \neq 1} \frac{1}{k_j!}\,(\alpha_j z)^{k_j} \;=\; \prod_{j=1}^{n} \big(\mathrm{e}^{\alpha_j z} - \alpha_j z\big).$$

Since $|e^w - w| \leq e^{|w|} - |w| \leq e^{|w|^2}$ for any complex number w, we get a simple bound

$$|\psi(z)| \;\leq\; \prod_{j=1}^{n} \mathrm{e}^{\alpha_j^2\,|z|^2} = \mathrm{e}^{|z|^2}. \tag{V.2}$$

We now use contour integration and Cauchy's formula

$$K_m = \frac{(2m)!}{2\pi i} \int_{|z|=R} \frac{\psi(z)}{z^{2m+1}}\, dz \qquad (R > 0),$$

which together with the upper bound (V.2) yields

$$K_m \leq \frac{(2m)!}{R^{2m}}\, \mathrm{e}^{R^2}.$$

It remains to choose an optimal value $R = \sqrt{m}$, which leads to

$$I_{2m}(Z_n) \leq \frac{(2m)!\,\mathrm{e}^m}{m^m}\, I, \tag{V.3}$$

that is, (I.3). □

Before turning to the next theorem, let us note that the moment bound (V.3) is insufficient to derive (I.4)-(I.5). Therefore, we choose a slightly different route based on Corollary IV.2.

Proof of Theorem I.2. By homogeneity, we may assume that $\sigma = 1$. Putting $\xi_i = |\rho(X_i)|$, on the axis $t \geq 0$, the function

$$\psi_i(t) = \mathbb{E}\,(\mathrm{e}^{t\xi_i} - t\xi_i), \quad i = 1, \dots, n,$$

is smooth, non-negative, with $\psi_i(0) = 1$, $\psi_i'(0) = 0$. Moreover, for $t \leq 1/2$,

$$\psi_i''(t) \;\leq\; \mathbb{E}\,\xi_i^2\,\mathrm{e}^{\xi_i/2}.$$

Using the bound $x^2\,\mathrm{e}^{x/2} \leq c\,\mathrm{e}^x$ with $c = 16/\mathrm{e}^2$, we get

$$\psi_i''(t) \;\leq\; c\,\mathbb{E}\,\mathrm{e}^{\xi_i} \;\leq\; 2c,$$

which, by Taylor's formula, implies $\psi_j(t) \leq 1 + ct^2 \leq \mathrm{e}^{ct^2}$. Hence, applying the bound (IV.3), we obtain that

$$\begin{aligned} \mathbb{E}\,\exp\{t\,|\rho(Z_n)|\} \;&\leq\; 2\,\mathbb{E}\,\cosh(t\rho(Z_n)) \\ &\leq\; 2\prod_{i=1}^{n} \psi_i(\alpha_i t) \;\leq\; 2\,e^{ct^2}. \end{aligned}$$

Choosing $t = 1/2$, this gives

$$\mathbb{E}\,\mathrm{e}^{|\rho(Z_n)|/4} \;\leq\; \big(\mathbb{E}\,\mathrm{e}^{|\rho(Z_n)|/2}\big)^{1/2} \;\leq\; (2\,\mathrm{e}^{c/4})^{1/2} < 2,$$

which was required.

For the proof of the second claim, we apply the bound $x^2\,\mathrm{e}^{x^2/2} \leq 2\,\mathrm{e}^{x^2-1}$. Since $\mathbb{E}\,\mathrm{e}^{\xi_i^2} \leq 2$, we get

$$\psi_i''(t) \;=\; \mathbb{E}\,\xi_i^2\,\mathrm{e}^{t\xi_i} \;\leq\; \mathbb{E}\,\xi_i^2\,\mathrm{e}^{(t^2+\xi_i^2)/2} \;\leq\; \frac{4}{\mathrm{e}}\,\mathrm{e}^{t^2/2},$$

implying

$$\psi_i(t) \;\leq\; 1 + \frac{2t^2}{e}\,e^{t^2/2} \;\leq\; e^{t^2}.$$

Hence, by (IV.3), for any $t \in \mathbb{R}$,

$$\mathbb{E}\,\cosh(t\rho(Z_n)) \;\leq\; \mathrm{e}^{t^2}.$$

If η is a standard normal random variable, independent of $\rho(Z_n)$, this subgaussian bound on the Laplace transform yields

$$\begin{aligned} \mathbb{E}\,\exp\{\,t^2\rho(Z_n)^2/2\} \;&=\; \mathbb{E}\,\cosh(t\rho(Z_n)\,\eta) \\ &\leq\; \mathbb{E}\,\exp\{\,t^2\eta^2\} \;=\; \frac{1}{\sqrt{1-2t^2}}, \end{aligned}$$

which holds true for all $0 < t < 1/2$. The choice $t^2 = 3/8$ yields (I.5) with constant $K = 16/3$. □

VI. The case of uniform distributions

By virtue of Propositions III.1 and III.3, Theorem I.3 may be reduced to the case of uniformly distributed random variables. Hence, as a next step, here we derive the inequality (I.6) for the class of uniform distributions on finite intervals. Suppose that the independent variables X_i, $1 \leq i \leq k+1$, are uniformly distributed in the intervals of lengths $l_i > 0$, respectively.

Lemma VI.1. *For the sum $S_{k+1} = X_1 + \dots + X_{k+1}$, $k \geq 1$, we have*

$$I_k(S_{k+1}) \;\leq\; \frac{k^k}{k!}\,\frac{l_1 + \dots + l_{k+1}}{l_1 \dots l_{k+1}}. \tag{VI.1}$$

Proof. Let X_i be uniformly distributed in the intervals $(0, l_i)$. Then, the density f of S_{k+1} is supported and positive on $(0, l)$, $l = l_1 + \dots + l_{k+1}$, where it has a piecewise continuous Radon-Nikodym derivative f'. For sufficiently small $x > 0$, the distribution function of S_{k+1} is given by

$$F(x) = \mathbb{P}\{S_{k+1} \leq x\} = \frac{1}{V(k+1)!}\,x^{k+1}, \tag{VI.2}$$

where $V = l_1 \dots l_{k+1}$ denotes the volume of the box Q in $\mathbb{R}^{k+1}$ with sides $[0, l_i]$. Correspondingly, for small $x > 0$,

$$f(x) = \frac{1}{Vk!}\,x^k. \tag{VI.3}$$

Note also that f is symmetric about the point $l/2$, and $f(0+) = f(l-) = 0$.

To explore shape properties of the density on the whole supporting interval, we apply the Brunn-Minkowski inequality in Convex Geometry. It asserts that

$$|uA + (1-u)B|^{1/k} \geq u\,|A|^{1/k} + (1-u)\,|B|^{1/k} \tag{VI.4}$$

for all $0 < u < 1$ and all non-empty Borel sets A, B lying in parallel (k-dimensional) hyperplanes of $\mathbb{R}^{k+1}$. Here

$$uA + (1-u)B = \{\,ua + (1-u)b :\; a \in A,\; b \in B\}$$

stands for the Minkowski sum, and $|C|$ is used to denote the k-dimensional Lebesgue measure of a set C in the hyperplane where it lies (cf. e.g. [5], [10]).

Since the random vector $(X_1, \dots, X_{k+1})$ is uniformly distributed in Q, the density of S_n may be written for $0 \leq x \leq l$ as

$$f(x) = \frac{1}{V}\,\big|\{(x_1, \dots, , x_{k+1}) \in Q : x_1 + \dots + x_{k+1} = x\}\big|.$$

Hence, by (VI.4), the function $f(x)^{1/k}$ is concave on the supporting interval $(0, l)$.

As we know, the latter property may also be formulated in terms of the associated function $L(u) = f(F^{-1}(u))$. Namely, by Proposition II.2, the function $L^{(k+1)/k}$ is concave, so that

$$\frac{k}{k+1}\,(L^{\frac{k+1}{k}})' = L^{\frac{1}{k}}\,L'$$

does not increase on $(0, 1)$. Hence, using also the symmetry of L about the point $1/2$, we get that, for all $0 < u < 1$,

$$|L'(u)|^k\,L(u) \,\leq\, c, \quad c \,=\, \lim_{u\to 0} L'(u)^k\,L(u).$$

With this bound we can now apply Proposition II.1 and the equality (II.2), which give

$$I_k(S_{k+1}) = \int_0^1 |L'(u)|^k\,du \leq \int_0^1 \frac{c}{L(u)}\,\mathrm{d}u = cl. \qquad \text{(VI.5)}$$

It remains to find c. From (VI.2)-(VI.3) it follows that, for all $u > 0$ small enough,

$$\begin{aligned} L(u) &= (k+1)\,(k+1)^{-\frac{1}{k+1}}\,V^{-\frac{1}{k+1}}\,u^{\frac{k}{k+1}}, \\ L'(u) &= k\,(k+1)^{-\frac{1}{k+1}}\,V^{-\frac{1}{k+1}}\,u^{-\frac{1}{k+1}}, \end{aligned}$$

and thus

$$L'(u)^k\,L(u) = \frac{1}{V}\,\frac{k^k}{k!} = c.$$

Hence, we arrive in (VI.5) at $I_k(S_{k+1}) \leq lk^k/(Vk!)$ which is (VI.1). □

VII. Proof of Theorems I.3-I.4

Proof of Theorem I.3. Assuming that the random variables X_i are independent and have densities f_i with finite total variation norms $b_i = I_1(X_i)$, one may apply Proposition III.3 and represent them as convex mixtures $f_i = \int q\,\mathrm{d}\pi_i(q)$ with some mixing probability measures π_i supported on the set U of densities for uniform distributions (on all intervals) and satisfying

$$b_i = \int_U I_1(q_i)\,\mathrm{d}\pi_i(q_i), \qquad i = 1, \dots, k+1. \qquad \text{(VII.1)}$$

Taking the convolution, we then have a similar representation for the density f of the sum $S_{k+1} = X_1 + \dots + X_{k+1}$, namely

$$\begin{aligned} f &= f_1 * \dots * f_{k+1} \\ &= \int_U \dots \int_U q_1 * \dots * q_{k+1}\;\mathrm{d}\pi_1(q_1)\dots\mathrm{d}\pi_{k+1}(q_{k+1}). \end{aligned}$$

One can now apply Jensen's inequality (III.2) to get that

$$I_k(f) \,\leq\, \int_U \dots \int_U I_k(q_1 * \dots * q_{k+1})\,\mathrm{d}\pi_1(q_1)\dots\mathrm{d}\pi_{k+1}(q_{k+1}). \qquad \text{(VII.2)}$$

For the uniform distribution on the interval (a, b), $a < b$, with density $q = \frac{1}{b-a}\,1_{(a,b)}$, we have $I_1(q) = 2/(b-a)$. Equivalently, every q in U is supported on an interval of length $l = 2/I_1(q)$. Hence, by Lemma VI.1, putting $l_i = 2/I_1(q_i)$, we have

$$\begin{aligned} I_k(q_1 * \dots * q_{k+1}) \;&\leq\; \frac{k^k}{k!}\,\frac{l_1 + \dots + l_{k+1}}{l_1 \dots l_{k+1}} \\ &= \; c_k \sum_{i=1}^{k+1} I_1(q_1)\dots I_1(q_{i-1}) I_1(q_{i+1})\dots I_1(q_{k+1}), \end{aligned}$$

where $c_k = k^k/(2^k k!)$ Using this bound in (VII.2) and applying (VII.1), we arrive at

$$I_k(f) \,\leq\, c_k \sum_{i=1}^{k+1} b_1 \dots b_{i-1}\,b_{i+1}\dots b_{k+1},$$

which is the desired inequality (I.6). □

To turn to the next theorem, we employ the following bound on the total variation norm in terms of characteristic functions, cf. [3], Proposition 5.2.

Lemma VII.1. *If the characteristic function $u(t) = \mathbb{E}\,e^{itX}$ of a random variable X has a continuous derivative for $t > 0$, with*

$$\int_{-\infty}^{\infty} t^2\big(|u(t)|^2 + |u'(t)|^2\big)\,dt < \infty, \qquad \text{(VII.3)}$$

then X has an absolutely continuous density q with finite total variation norm satisfying

$$\|q\|_{\mathrm{TV}}^4 \,\leq\, \int_{-\infty}^{\infty} |tu(t)|^2\,\mathrm{d}t \int_{-\infty}^{\infty} |(tu(t))'|^2\,\mathrm{d}t. \qquad \text{(VII.4)}$$

Proof of Theorem I.4. By Theorem I.3, if $I_1(S_n)$ is finite, then so is $I_k(S_{n(k+1)})$. Hence, $a)$ and $b)$ are equivalent.

To see that $b)$ and $c)$ are equivalent as well, note that the sum S_n has characteristic function $v_n(t) = v(t)^n$. Assuming $b)$, S_n has a density q of bounded total variation with $q(-\infty) = q(\infty) = 0$. In this case, one may integrate by parts to write

$$v_n(t) = \int_{-\infty}^{\infty} \mathrm{e}^{itx} q(x)\,dx = -\frac{1}{it}\int_{-\infty}^{\infty} \mathrm{e}^{itx}\,dq(x), \quad t \neq 0.$$

Since the latter integral is bounded in absolute value by $I_1(q) = I_1(S_n)$, it follows that $|v(t)|^n = O(1/t)$ as $t \to \infty$, and we get the property $c)$. The implication $c) \Rightarrow b)$ may be based on Lemma VII.1, applied with $u(t) = v_n(t)$, $X = S_n$, in which case

$$|u'(t)| \leq n\,\mathbb{E}\,|X_1|\,|v(t)|^{n-1}.$$

If $|v(t)| = o(t^{-\varepsilon})$ for some $\varepsilon > 0$, and if n is sufficiently large, then all integrals in (VII.3)-(VII.4) are convergent, so that $\|q\|_{\mathrm{TV}} = I_1(S_n)$ is finite.

It remains to derive the inequality (I.7), assuming that X_1 has density with bounded total variation norm $I_1(X_1)$. First, let $k = 2m$ be even, and consider the sums

$$Y_j \;=\; \sum_{(j-1)(k+1) < i \leq j(k+1)} X_i$$

for $j = 1, \dots, r$, $r = [n/(k+1)]$. Put $N = r(k+1)$, so that $1 \le N \le n$. By Corollary III.2, $I_k(S_n) \le I_k(S_N)$, while, by Theorem I.1,

$$\begin{aligned} I_{2m}(S_N) &= I_{2m}(Y_1 + \dots + Y_r) \\ &\le r^{-m} \frac{\mathrm{e}^m\,(2m)!}{m^m}\, I_{2m}(Y_1). \end{aligned}$$

On the other hand, by Theorem I.3 with $k = 2m$,

$$I_{2m}(Y_1) \le \frac{(2m)^{2m+1}}{2^{2m}\,(2m)!}\, I_1(X_1)^{2m}.$$

The two inequalities yield

$$I_{2m}(S_n) \le 2r^{-m}\,\mathrm{e}^m\, m^{m+1}\, I_1(X_1)^{2m}.$$

Since $n \ge k+1$, necessarily $r \ge \frac{n}{2(k+1)} \ge \frac{n}{6m}$, so that

$$I_{2m}(S_n) \le 2n^{-m}\,(6\mathrm{e})^m\, m^{2m+1}\, I_1(X_1)^{2m}. \qquad \text{(VII.5)}$$

Thus, (I.7) follows with constant $A_k = (3\mathrm{e}/2)^{k/2}\, k^{k+1}$.

If $k = 2m-1$ is odd, and $n \ge k+2$, one may apply the previous step (VII.5), to get

$$\begin{aligned} I_{2m-1}(S_n) &\le (I_{2m}(S_n))^{\frac{2m-1}{2m}} \\ &\le \big(2n^{-m}\,(6\mathrm{e})^m\, m^{2m+1}\big)^{\frac{2m-1}{2m}}\; I_1(X_1)^{2m-1}. \end{aligned}$$

Hence, (I.7) follows with constant $A_k = (3\mathrm{e}/2)^{k/2}\,(k+1)^{k+1}$. Finally, if $k = 2m-1$, $n = k+1$, the inequality (I.6) yields (I.7) with a similar constant. □

VIII. Polynomial decay of densities

Let us now consider the moments of the scores $I_k = I_k(X) = \mathbb{E}\,|\rho(X)|^k$ for real values $k > 1$ (not necessarily integer). Here we show that, by invoking the absolute moments

$$\beta_s = \beta_s(X) = \mathbb{E}\,|X|^s \qquad (s > 0 \ \text{ real}),$$

it is possible to control the behavior of the density of X at infinity. Define the conjugate power $k^* = \frac{k}{k-1}$.

Theorem VIII.1. *If X has density f with finite I_k $(k > 1)$ and β_{sk^*} $(s > 0)$, then, for any $x \in \mathbb{R}$,*

$$f(x) \le \frac{c}{1 + |x|^s} \qquad \text{(VIII.1)}$$

with constant

$$c = \begin{cases} s\beta_{s-1} + (1 + \beta_{sk^*}^{\frac{1}{k^*}})\, I_k^{\frac{1}{k}}, & \text{if } s > 1, \\ s + (3 + \beta_{sk^*}^{\frac{1}{k^*}})\, I_k^{\frac{1}{k}}, & \text{if } s \le 1. \end{cases}$$

Moreover,

$$\lim_{x\to\infty} (1 + |x|^s) f(x) = 0. \qquad \text{(VIII.2)}$$

Proof. Since $k > 1$, the density f is absolutely continuous and has a Radon-Nikodym derivative f'. By Hölder's inequality,

$$\begin{aligned} \int_{-\infty}^{\infty} |x|^s\,|f'(x)|\,\mathrm{d}x &= \int_E |x|^s f(x)^{\frac{k-1}{k}}\, \frac{|f'(x)|}{f(x)^{\frac{k-1}{k}}}\,\mathrm{d}x \\ &\le \Big(\int_E |x|^{sk^*} f(x)\,\mathrm{d}x\Big)^{\frac{k-1}{k}} \Big(\int_E \frac{f'(x)^k}{f(x)^{k-1}}\,\mathrm{d}x\Big)^{\frac{1}{k}}, \end{aligned}$$

where $E = \{x \in \mathbb{R} : f(x) > 0\}$. Hence

$$\int_{-\infty}^{\infty} |x|^s\,|f'(x)|\,\mathrm{d}x \le \beta_{sk^*}^{\frac{1}{k^*}}\, I_k^{\frac{1}{k}}, \qquad s > 0. \qquad \text{(VIII.3)}$$

Now, assuming first that $s > 1$, the function

$$u(x) = (1 + |x|^s) f(x)$$

is (locally) absolutely continuous and has a Radon-Nikodym derivative satisfying

$$|u'(x)| \le s\,|x|^{s-1} f(x) + (1 + |x|^s)\,|f'(x)|. \qquad \text{(VIII.4)}$$

Integrating this inequality, we see that u is a function of bounded total variation. Since u is also integrable, we get

$$\lim_{x\to-\infty} u(x) = \lim_{x\to\infty} u(x) = 0, \qquad \text{(VIII.5)}$$

thus implying (VIII.2). In addition, by (VIII.4) and (VIII.3),

$$\begin{aligned} u(x) &= \int_{-\infty}^{x} u'(y)\,\mathrm{d}y \le \int_{-\infty}^{\infty} |u'(y)|\,\mathrm{d}y \\ &\le s\int_{-\infty}^{\infty} |x|^{s-1} f(x)\,\mathrm{d}x \\ &\quad + \int_{-\infty}^{\infty} (1 + |x|^s)\,|f'(x)|\,\mathrm{d}x \\ &\le s\beta_{s-1} + I_1 + \beta_{sk^*}^{\frac{1}{k^*}}\, I_k^{\frac{1}{k}}. \end{aligned} \qquad \text{(VIII.6)}$$

Since $I_1 \le I_k^{\frac{1}{k}}$, we arrive at (VIII.1).

If $0 < s \le 1$, one may still use (VIII.4). Since $f(x) \le I_1$ for all $x \in \mathbb{R}$, we have

$$\begin{aligned} \int_{-\infty}^{\infty} s\,|x|^{s-1} f(x)\,\mathrm{d}x &\le \int_{|x|>1} s\,|x|^{s-1} f(x)\,\mathrm{d}x \\ &\quad + \int_{-1}^{1} s\,|x|^{s-1} f(x)\,\mathrm{d}x \\ &\le s\int_{|x|>1} f(x)\,\mathrm{d}x + 2I_1 < \infty. \end{aligned}$$

Recalling (VIII.3), we conclude that u is an integrable function of bounded variation, which implies (VIII.5) and thus (VIII.2) again. With a similar argument, (VIII.6) should be modified to

$$u(x) \le s + 3I_1 + \beta_{sk^*}^{\frac{1}{k^*}}\, I_k^{\frac{1}{k}},$$

which yields (VIII.1) in this case as well. □

Remark. Under stronger moment assumptions, one can obtain better bounds for the decay of the density. For example, if for some $\lambda > 0$, the exponential moment

$$\beta = \mathbb{E}\,\mathrm{e}^{\lambda k^*|X|} = \int_{-\infty}^{\infty} \mathrm{e}^{\lambda k^*|x|}\, f(x)\,\mathrm{d}x$$

is finite, then by similar arguments, $f(x) \le c\,\mathrm{e}^{-\lambda|x|}$ for any $x \in \mathbb{R}$ with some constant c depending on λ, β and I_k.

IX. Cramér-Rao-type inequality

Finally, let us relate the moments of the scores to the absolute moments of centered random variables.

Theorem IX.1. *If $I_k(X)$ is finite, then, for any $a \in \mathbb{R}$,*

$$(\mathbb{E}\,|X-a|^{k^*})^{\frac{1}{k^*}}\,(I_k(X))^{\frac{1}{k}} \,\geq\, 1. \qquad \text{(IX.1)}$$

In the case $k = 2$, (IX.1) is reduced to the classical relation

$$\mathrm{Var}(X)\,I_2(X) \geq 1. \qquad \text{(IX.2)}$$

It is a particular case of the Cramér-Rao inequality for the parametric family of densities $f(x;\theta) = f(x-\theta)$ and the unbiased estimator $\hat{\theta}(x) = x$ of the shift parameter $\theta \in \mathbb{R}$, cf. e.g. [4], paragraph 26).

The proof of Theorem IX.1 is based on the following generalization of (IX.2), where one should choose $u(x) = x - a$ to obtain (IX.1).

Lemma IX.2. *For any smooth function $u : \mathbb{R} \to \mathbb{C}$ such that $\mathbb{E}\,|u'(X)| < \infty$ and $u(x)f(x) \to 0$ as $|x| \to \infty$, we have*

$$|\mathbb{E}\,u'(X)| \,\leq\, (I_k(X))^{\frac{1}{k}}\,(\mathbb{E}\,|u(X)|^{k^*})^{\frac{1}{k^*}}. \qquad \text{(IX.3)}$$

Proof. One may integrate by parts to write

$$\begin{aligned}\int_a^b u'(x)f(x)\,\mathrm{d}x \;&=\; u(b)f(b) - u(a)f(a) \\ &\quad - \int_a^b u(x)f'(x)\,\mathrm{d}x, \quad a < b.\end{aligned}$$

Assuming that $\mathbb{E}\,|u'(X)| < \infty$ and letting $a \to -\infty$, $b \to \infty$, we obtain the equality

$$\begin{aligned}\mathbb{E}\,u'(X) \;&=\; -\int_{-\infty}^{\infty} u(x)f'(x)\,\mathrm{d}x \\ &=\; -\int_E u(x)f'(x)\,\mathrm{d}x \\ &=\; -\int_E \frac{f'(x)}{f(x)^{\frac{k-1}{k}}}\;u(x)f(x)^{\frac{k-1}{k}}\,\mathrm{d}x.\end{aligned}$$

where $E = \{x \in \mathbb{R} : f(x) > 0\}$. It remains to apply Holder's inequality (as in the proof of Theorem VIII.1). □

In order to justify an application of (IX.3) in Theorem IX.1, one may assume that $I_k(X) < \infty$ and $\mathbb{E}\,|X|^{k^*} < \infty$. This implies $(1+|x|)f(x) \to 0$ as $|x| \to \infty$, according to Theorem VIII.1 with $s = 1$. Hence, the assumptions of Lemma IX.2 are fulfilled.

Acknowledgment

We would like to thank the referees for valuable remarks (in particular, for the reference to the paper by P.-L. Lions and G. Toscani).

References

[1] Bobkov, S. G. The size of singular component and shift inequalities. Ann. Probab. 27 (1999), no. 1, 416–431.

[2] Bobkov, S. G. Large deviations and isoperimetry over convex probability measures with heavy tails. Electron. J. Probab. 12 (2007), 1072–1100.

[3] Bobkov, S. G.; Chistyakov, G. P.; Götze, F. Fisher information and the central limit theorem. Probab. Theory Related Fields 159 (2014), issue 1-2, 1–59.

[4] Borovkov, A. A. Mathematical statistics. Translated from the Russian by A. Moullagaliev and revised by the author. Gordon and Breach Science Publishers, Amsterdam, 1998. xxii+570 pp.

[5] Burago, Yu. D.; Zalgaller, V. A. Geometric inequalities. Translated from the Russian by A. B. Sosinski. Fundamental Principles of Mathematical Sciences, 285. Springer Series in Soviet Mathematics. Springer-Verlag, Berlin, 1988. xiv+331 pp.

[6] Cover, T. M.; Thomas, J. A. Elements of information theory. Second edition. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2006, xxiv+748 pp.

[7] Gabetta, E. On a conjecture of McKean with application to Kac's model. Quaderno, Univ. degli Studi di Milano, 1993, 13 pp. Also: Transport Theory and Statistical Physics 24 (1995), issue 1-3, 305–317.

[8] Johnson, O. Information theory and the central limit theorem. Imperial College Press, London, 2004, xiv+209 pp.

[9] Lions, P.-L.; Toscani, G. A strengthened central limit theorem for smooth densities. J. Funct. Anal. 129 (1995), no. 1, 148–167.

[10] Schneider, R. Convex bodies: the Brunn-Minkowski theory. Second expanded edition. Encyclopedia of Mathematics and its Applications, 151. Cambridge University Press, Cambridge, 2014. xxii+736 pp.

[11] Toscani, G. Score functions, generalized relative Fisher information and applications. Ric. Mat. 66 (2017), no. 1, 15–26.

Sergey Bobkov graduated from Saint Petersburg University, USSR, in 1983, and then pursued graduate studies at Steklov Mathematical Institute of the Academy of Sciences of USSR (Saint Petersburg branch). He received the Ph.D. degree and the degree of Doctor of Mathematics and Physics from Saint Petersburg University in 1988 and 1997, respectively. From 1987 to 2000 he worked at Syktyvkar University, Russia. In 2000 he joined the School of Mathematics at the University of Minnesota, Minneapolis, USA. From November 1995 to July 1997 he was an Alexander von Humboldt Fellow at Bielefeld University, Germany. He spent the summers of 2001 and 2002 as an EPSRC Fellow at Imperial College London, UK. He was awarded a Simons Fellowship (2012) and Humboldt Research Award (2014).

Prof. Bobkov's current research interests are high-dimensional probability distributions, measure concentration phenomena, isoperimetry, transportation, Sobolev-type inequalities, convex geometry, information theory, and discrete mathematics. He is serving as an Associate Editor for Bernoulli.