Moments of scores

Sergey Bobkov
University of Minnesota

Workshop on Probability and Information Theory
The University of Hong Kong, 19-22 August 2019

Definitions

X random variable with a (locally) abs. continuous density f.
Definition. The score of X is the random variable

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

Well defined: $\mathbb{P}\{f(X) > 0\} = 1$.

Examples:

$X \sim \mathrm{Exp}(1)$, $f(x) = \frac{1}{2}\, e^{-|x|}$, $\rho(X) = \mathrm{sign}(X) \sim \mathrm{Bern}(1/2)$,

$X \sim N(0,1)$, $f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$, $\rho(X) = -X \sim N(0,1)$.

Absolute moments:

$$I_k(X) = \mathbb{E}\, |\rho(X)|^k, \qquad k = 1, 2, \dots$$

First absolute moment $=$ total variation norm

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

Note: $I_1(X) < \infty \;\Rightarrow\; \mathbb{E}\rho(X) = 0$.

The case $I_1(X) = \|f\|_{\mathrm{TV}}$ involves more distributions (including uniform on intervals).

Second moment

Second moment = Fisher information contained in the distribution of X

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

Cramér-Rao inequality:

$$I(X)\,\mathrm{Var}(X) \geq 1$$

with equality iff X is normal.

De Bruijn's identity: If $\mathbb{E}X^2 < \infty$, then for all $t > 0$

$$\frac{\mathrm{d}}{\mathrm{d}t}\, h(X + \sqrt{t}\, Z) = \frac{1}{2}\, I(X + \sqrt{t}\, Z),$$

where

$$h(X) = -\int_{-\infty}^{\infty} f(x) \log f(x)\ \mathrm{d}x$$

Shannon's entropy, and $Z \sim N(0,1)$ independent of X.

Shifts (translates) of product measures

$(X_n)_{n\geq 1}$ iid copies of X with distribution μ on $\mathbb{R}$.
Product measures:

$$\mu^\infty = \mu \otimes \mu \otimes \dots, \quad \mu_h^\infty = \mu_{h_1} \otimes \mu_{h_2} \otimes \dots$$

for $h = (h_n)_{n\geq 1}$, $\mu_{h_n} = \text{distribution of } X_n + h_n$.

Question: When are the sample paths of (X_n) and $(X_n + h_n)$ distinguishable with errors $h = (h_n)$ of centering X_n from ℓ^2 ?

Kakutani's dichotomy: Any two product measures are either equivalent or singular (orthogonal).

Feldman-Shepp's theorem: $\mu^\infty \sim \mu_h^\infty$ on $\mathbb{R}^\infty$ for any h with

$$\|h\|_2^2 = \sum_n h_n^2 < \infty$$

if and only if X has an a.e. positive absolutely continuous density with $I(X) < \infty$.

Quantification (B '99): Put $\sigma^2 = I(X)$. If $\sigma\|h\|_2 < \pi$, then

$$\|\mu_h^\infty - \mu^\infty\|_{\mathrm{TV}} \leq 2 \sin\Big(\frac{\sigma\|h\|_2}{2}\Big).$$

Logarithmic Sobolev inequality

If $\mathbb{E}X^2 = 1$, $Z \sim N(0,1)$, then

$$h(X) \le h(Z), \qquad I(X) \ge I(Z).$$

Informational divergence (Kullback-Leibler distance):

$$D(X||Z) = h(Z) - h(X) = \int f(x) \log \frac{f(x)}{\varphi(x)} \, \mathrm{d}x.$$

Fisher information distance:

$$I(X||Z) = I(X) - I(Z) = \int \frac{(f(x) - \varphi(x))^2}{\varphi(x)} \, \mathrm{d}x.$$

Log-Sobolev inequality (Stam '59, Gross '75):

$$D(X||Z) \ \le \ \frac{1}{2} I(X||Z).$$

Equivalently in terms of $u = f/\varphi$,

$$\int u \log u \, \mathrm{d}\gamma \ \le \ \frac{1}{2} \int \frac{u'^2}{u} \, \mathrm{d}\gamma$$

with respect to the standard Gaussian measure γ.

Higher order moments of scores

Higher order moments:

$$I_k(f) = I_k(X) = \mathbb{E}\,|\rho(X)|^k = \int_{-\infty}^{\infty} \frac{|f'(x)|^k}{f(x)^{k-1}}\,\mathrm{d}x.$$

• General case $k > 1$ (Lions and Toscani '95):
Convergence of densities and their powers in CLT in Sobolev spaces.

• Case $k = 4$ (Gabetta '93): Convergence to equilibrium in Kac's model (in the context of the kinetic theory of gases).

• Exponential and Gaussian moments of $\rho(X)$ (B '99):
To control translates of product probability measures.

Problems

• How to determine that $I_k(X) < \infty$?
• Behaviour of moments along convolutions, i.e. for

$$X = S_n = X_1 + \cdots + X_n$$

with independent summands.

General properties

• Translation invariance and homogeneity:

$$I_k(a+bX) = \frac{1}{|b|^k}\, I_k(X), \qquad a \in \mathbb{R},\ b \neq 0.$$

• Monotonicity with respect to order:

$$k \le l \ \Rightarrow\ (I_k(X))^{1/k} \le (I_l(X))^{1/l}.$$

• Convexity: If $f = \int f_t \, \mathrm{d}\pi(t)$, then

$$I_k(f) \le \int I(f_t)\, \mathrm{d}\pi(t).$$

• Monotonicity with respect to convolutions:

$$I_k(X+Y) \le \min\{I_k(X), I_k(Y)\}.$$

• Therefore, for $S_n = X_1 + \cdots + X_n$ with independent X_i, the sequence

$$n \to I_k(S_n)$$

is decreasing.

Moments of weighted sums

Lions and Toscani '95: If $S_n = X_1 + \cdots + X_n$ with iid X_i, then

$$I_{2m}(S_n/\sqrt{n}) \;\le\; c_m\, I_{2m}(X_1), \qquad m = 1, 2, \ldots$$

Generalization: weighted sums

$$Z_n = \alpha_1 X_1 + \cdots + \alpha_n X_n \qquad (\alpha_1^2 + \cdots + \alpha_n^2 = 1).$$

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

$$I_{2m}(Z_n) \;\le\; c_m I, \qquad c_m = (2m)!\,(e/m)^m.$$

Example: $X_i \sim f(x) = \frac{1}{2}\, e^{-|x|}$, then

$$|\rho(X_i)| = 1 \quad \text{and} \quad I_{2m}(X_i) = 1.$$

For $\alpha_i = \frac{1}{\sqrt{n}}$, we have $Z_n \Rightarrow Z \sim N(0,1)$ as $n \to \infty$ and

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

Hence, $c_m \ge (2m)!/(2^m m!)$

Gaussian moments

Theorem 2. If for some $\sigma > 0$

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

then

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

with $K = 6$.

Note (well known): If $\mathbb{E}X_i = 0$ and

$$\mathbb{E}\,\exp\{X_i^2/\sigma^2\} \le 2,$$

then similarly

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

Finiteness of moments of scores

Let

$$S_n = X_1 + \dots + X_n$$

with independent X_i.

Question: Is it true that $I_k(S_n) < \infty$ for some $n = n_0$ assuming only that $I_1(X_i) < \infty$?

Case $k = 2$ (B-Chistyakov-Götze '14): $n_0 = 3$.

Application: CLT in Fisher information distance.
Open: Higher dimensions.

Theorem 3. For $k \geq 2$, it is enough to take $n_0 = k + 1$. Moreover, putting $b_i = I_1(X_i)$, we have

$$I_k(S_{k+1}) \,\leq\, c_k \, b_1 \dots b_{k+1} \left(\frac{1}{b_1} + \dots + \frac{1}{b_{k+1}}\right)$$

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

Characterization in the iid case

Let $(X_n)_{n\geq 1}$ be i.i.d. random variables with $\mathbb{E}\,|X_1| < \infty$ and characteristic function

$$v(t) = \mathbb{E}\, e^{itX_1}, \qquad t \in \mathbb{R}.$$

Fix $k \geq 1$.

Theorem 4. The following properties are equivalent:

- $a)$ 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 $t \to \infty$.

If X_1 has density with bounded total variation, then

$$\sup_{n \geq k+1} I_k(S_n/\sqrt{n}) \leq A_k\,(I_1(X_1))^k.$$

Reduction to uniform distributions

For $X \sim f$, write $I_k(f) = I_k(X)$.

Triangle inequality: If $f = \int q \, \mathrm{d}\pi(q)$, then

$$I_1(f) \leq \int I_1(q) \, \mathrm{d}\pi(q).$$

Let U be the collection of uniform densities $q(x) = \frac{1}{b-a} 1_{a<x<b}$.

Lemma 1 (B-C-G). Any probability 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

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

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 measure π. In this case, $I_1(f) = 2f(0+)$.

The case of uniform distributions

Lemma 2. For the sum $X = X_1 + \cdots + X_{k+1}$ with $X_i \sim U(0, l_i)$,

$$I_k(X) \ \leq \ \frac{k^k}{k!} \, \frac{l_1 + \cdots + l_{k+1}}{l_1 \dots l_{k+1}}.$$

Put

$$l = l_1 + \cdots + l_{k+1}, \qquad v = l_1 \dots l_{k+1} = |Q|,$$

where Q is the box in $\mathbb{R}^{k+1}$ with sides $[0, l_i]$. Distribution of X is supported on $(0, l)$ and is symmetric about $l/2$, with density

$$f(x) = \frac{1}{v} \left| \{(x_1, \dots, , x_{k+1}) \in Q : x_1 + \cdots + x_{k+1} = x\} \right|.$$

For small $x > 0$,

$$f(x) \ = \ \frac{1}{vk!} \, x^k.$$

Brunn-Minkowski inequality: For all Borel measurable sets A, B lying in parallel hyperplanes of $\mathbb{R}^{k+1}$ and any $0 < t < 1$,

$$|tA + (1-t)B|^{1/k} \geq t \, |A|^{1/k} + (1-t) \, |B|^{1/k}.$$

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

Proof of Lemma 2

$(f^{1/k})'$ is decreasing, and by the symmetry of f around $l/2$,

$$\begin{aligned}\left|\frac{d}{dx} f(x)^{1/k}\right| &\leq \lim_{x\to 0} \left|\frac{d}{dx} f(x)^{1/k}\right| \\ &= \frac{d}{dx}\left(\frac{1}{vk!}\, x^k\right)^{1/k} = \left(\frac{1}{vk!}\right)^{1/k}.\end{aligned}$$

This gives

$$\begin{aligned}I_k(X) &= \int_0^l \left|\frac{f'(x)}{f(x)}\right|^k f(x)\, \mathrm{d}x \\ &= k^k \int_0^l \left|\frac{d}{dx} f(x)^{1/k}\right|^k \mathrm{d}x \leq \frac{k^k}{k!}\,\frac{l}{v}.\end{aligned}$$

Proof of Theorem 3

X_i independent and have densities f_i with finite total variation norms $b_i = I_1(X_i)$, $1 \le i \le k+1$. By Lemma 1,

$$f_i(x) = \int_U q(x)\,\mathrm{d}\pi_i(q)$$

with some mixing probability measures π_i on the set U of densities for uniform distributions (on all intervals) and satisfying

$$b_i = \int_U I_1(q_i)\,\mathrm{d}\pi_i(q_i).$$

Hence $S_{k+1} = X_1 + \cdots + X_{k+1}$ has density

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

By Jensen's inequality,

$$I_k(f) \le \int_U \dots \int_U I_k(q_1 * \cdots * q_{k+1})\ \mathrm{d}\pi_1(q_1) \dots \mathrm{d}\pi_{k+1}(q_{k+1}).$$

For the uniform distribution with density $q = \frac{1}{b-a} 1_{(a,b)}$, we have

$$I_1(q) = \frac{2}{b-a}.$$

Equivalently, every q in U is supported on an interval of length $l = 2/I_1(q)$. Hence, by Lemma 2, putting $l_i = 2/I_1(q_i)$,

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

where

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

It remains to integrate this inequality over $\pi_1 \otimes \cdots \otimes \pi_{k+1}$ and use $b_i = \int I_1(q_i) \, \mathrm{d}\pi_i(q_i)$ to get

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

Stam's inequality

Theorem (Stam '59). If X_1 and X_2 are independent, then

$$\frac{1}{I(X_1+X_2)} \;\geq\; \frac{1}{I(X_1)} + \frac{1}{I(X_2)}.$$

Linearized form: For all $a_1, a_2 > 0$, $a_1 + a_2 = 1$,

$$I(X_1+X_2) \;\leq\; a_1^2\, I(X_1) + a_2^2\, I(X_2).$$

Weighted sums: For all $\alpha_1, \alpha_2 > 0$ such that $\alpha_1^2 + \alpha_2^2 = 1$,

$$I(\alpha_1 X_1 + \alpha_2 X_2) \;\leq\; \alpha_1^2\, I(X_1) + \alpha_2^2\, I(X_2).$$

Theorem (Lions-Toscani '95). Given $m \geq 1$,

$$I_{2m}(\alpha_1 X_1 + \alpha_2 X_2) \;\leq\; \sum \binom{2m}{k} \alpha_1^k \alpha_2^{2m-k} \;\times\; I_k(X_1) I_{2m-k}(X_2).$$

with summation over all $k \neq 1$, $0 \leq k \leq 2m$.

Multinomial extension for weighted sums

For independent X_i with finite $I_{2m}(X_i) = \mathbb{E}\,\rho(X_i)^{2m}$, consider the weighted sums

$$Z_n = \alpha_1 X_1 + \cdots + \alpha_n X_n \quad (\alpha_1^2 + \cdots + \alpha_n^2 = 1)$$

with $\alpha_i > 0$. Put $I_0(X_i) = 1$.

Lemma 3. For any integer $m \geq 1$,

$$I_{2m}(Z_n) \,\leq\, \sum \binom{2m}{k_1 \dots k_n} \alpha_1^{k_1} \dots \alpha_n^{k_n} \times\, I_{k_1}(X_1) \dots I_{k_n}(X_n).$$

The summation is performed over all non-negative $k_i \neq 1$ such that $k_1 + \cdots + k_n = 2m$.

Proof of Lemma 3

Let $n = 2$. If X_i has densities f_i, the density f of $X_1 + X_2$ is given by

$$f(x) = \int_{-\infty}^{\infty} f_1(x-y) f_2(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} f_2(x-y) f_1(y)\,\mathrm{d}y.$$

It has derivative

$$f'(x) = \int_{-\infty}^{\infty} f_1'(x-y) f_2(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} f_2'(x-y) f_1(y)\,\mathrm{d}y.$$

That is, for any $a_1, a_2 > 0$, $a_1 + a_2 = 1$,

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

Hence

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

with

$$\mathrm{d}\mu_x(y)/\mathrm{d}y = f_1(x-y) f_2(y)/f(x).$$

By Jensen's inequality,

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

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

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

without terms corresponding to $k_1 = 1$ and $k_2 = 1$.
Next, write down this bound for $\alpha_i X_i$ with $a_i = \alpha_i^2$.

Proof of Theorem 2

Theorem 2. Let $\rho_i = |\rho(X_i)|$. If $\|\rho_i\|_{\psi_2} \le 1$, that is, $\mathbb{E}\, e^{\rho_i^2} \le 2$, then the weighted sums

$$Z_n = \alpha_1 X_1 + \dots + \alpha_n X_n, \qquad \alpha_1^2 + \dots + \alpha_n^2 = 1,$$

satisfy

$$\|\rho(Z_n)\|_{\psi_2} \le K.$$

Proof. Assume that $\alpha_i \ge 0$. By Lemma 3,

$$\frac{\mathbb{E}\,\rho(Z_n)^{2m}}{(2m)!} \ \le\ \sum \frac{1}{k_1! \dots k_n!}\, \alpha_1^{k_1} \dots \alpha_n^{k_n}\, \mathbb{E}\,|\rho_1|^{k_1} \dots \mathbb{E}\,|\rho_n|^{k_n}$$

where the summation is performed over all non-negative $k_i \neq 1$ such that $k_1 + \dots + k_n = 2m$. Expanding the $\cosh$-function in a power series, for any $t \ge 0$,

$$\mathbb{E}\, \cosh(t\rho(Z_n)) \ \le\ \prod_{i=1}^n \mathbb{E}\left(e^{t\alpha_i\rho_i} - t\alpha_i\rho_i\right).$$

The non-negative convex function

$$\psi_i(t) = \mathbb{E}\,(e^{t\rho_i} - t\rho_i)$$

satisfies $\psi_i(0) = 1$, $\psi_i'(0) = 0$. Using

$$x^2 e^{x^2/2} \le e^{x^2} - 1,$$

we get

$$\begin{aligned} \psi_i''(t) &= \mathbb{E}\,\rho_i^2\, e^{t\rho_i} \\ &\le \mathbb{E}\,\rho_i^2\, e^{(t^2+\rho_i^2)/2} \\ &\le e^{t^2/2}\,(\mathbb{E}\, e^{\rho_i^2} - 1) \le e^{t^2/2} \end{aligned}$$

so

$$\psi_i(t) \le 1 + t^2 e^{t^2/2} \le e^{t^2}.$$

This gives

$$\mathbb{E}\,\cosh(t\rho(Z_n)) \le \prod_{i=1}^n \psi_i(\alpha_i t) \le \prod_{i=1}^n e^{\alpha_i^2 t^2} = e^{t^2}$$

for any $t \in \mathbb{R}$. If $\eta \sim N(0,1)$,

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

The choice $t^2 = 3/8$ yields the result with $K = 16/3$.

Proof of Theorem 1

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

$$I_{2m}(Z_n) \;\leq\; c_m I, \qquad c_m = (2m)!\,(e/m)^m.$$

Proof. Again, according to Lemma 3,

$$\mathbb{E}\,\rho(Z_n)^{2m} \;\leq\; \sum \frac{(2m)!}{k_1!\dots k_n!}\,\alpha_1^{k_1}\dots\alpha_n^{k_n}\; I_{k_1}(X_1)\dots I_{k_n}(X_n)$$

Since

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

we get

$$I_{2m}(Z_n) \leq K_m I$$

with

$$K_m = \sum \frac{(2m)!}{k_1!\dots k_n!}\,\alpha_1^{k_1}\dots\alpha_n^{k_n}$$

where summation is 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 that

$$\psi(z) \ = \ \prod_{i=1}^n \sum_{k_j \geq 0,\, k_j \neq 1} \frac{1}{k_i!}\,(\alpha_i z)^{k_i} \ = \ \prod_{i=1}^n (e^{\alpha_i z} - \alpha_i z).$$

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

$$|\psi(z)| \ \leq \ \prod_{i=1}^n e^{\alpha_i^2\,|z|^2} = e^{|z|^2}.$$

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 above upper bound yields

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

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

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