Phenomena in high dimensions in geometric analysis,
random matrices, and computational geometry
Roscoff, France, June 25-29, 2012

BOUNDS AND ASYMPTOTICS FOR FISHER INFORMATION IN THE CENTRAL LIMIT THEOREM

Sergey G. Bobkov

University of Minnesota, Minneapolis, USA

joint work with
Gennadiy P. Chistyakov
and Friedrich Götze

Bielefeld University, Bielefeld, Germany

1 Fisher's quantity of information

X a random variable with values in $\mathbf{R}$

Definition. If X has an absolutely continuous density p, its Fisher information is defined by

$$I(X) = I(p) = \int_{-\infty}^{+\infty} \frac{p'(x)^2}{p(x)} \, dx,$$

where p' is a Radon-Nikodym derivative of p.
In all other cases, $I(X) = +\infty$.

Equivalently,

$$I(X) = \mathbf{E} \left[\frac{p'(X)}{p(X)} \right]^2.$$

Remarks.
1) $\mathbf{P}\{p(X) > 0\} = 1$, so the definition makes sense.
Integration is over $\{x : p(x) > 0\}$.

2) Assume $I(X) < +\infty$. Then $p(x) = 0 \Rightarrow p'(x) = 0$.

3) Translation invariance and homogeneity: $I(a+bX) = \frac{1}{b^2} I(X)$ $(a \in \mathbf{R}, b \neq 0)$.

2 When the Fisher information appears naturally

- Statistics: Estimation of the shift parameter in $p(x-\theta)$.
- Probability: Shifts of product measures, distinguishing a sequence of iid random variables from a translate of itself.

μ a probability measure on $\mathbf{R}$, $\mu_\theta^\infty(A) = \mu^\infty(A+\theta)$, $\theta \in \mathbf{R}^\infty$, $A \subset \mathbf{R}^\infty$

Theorem (Feldman 1961, Shepp 1965)

$$(\forall\, \theta \in \ell^2 \;\; \mu_\theta^\infty \ll \mu^\infty) \iff I(\mu) < +\infty \;\text{ and }\; \tfrac{d\mu(x)}{dx} > 0 \text{ a.e.}$$

- Information Theory: de Bruijn's identity
  Differential entropy

$$h(X) = -\int_{-\infty}^{+\infty} p(x) \log p(x)\; dx.$$

Theorem. If a random variable X has finite variance, then for all $\tau > 0$,

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

where $Z \sim N(0,1)$ is independent of X.

3 Distances to normality

X a r.v., with density $p(x)$, and $a = \mathbf{E}X$, $\sigma^2 = \mathrm{Var}(X) < +\infty$
$Z \sim N(a, \sigma^2)$ with density

$$q(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-a)^2/2\sigma^2}.$$

Relative entropy of X with respect to Z
(informational divergence, Kullback-Leibler distance):

$$\begin{aligned} D(X) &= D(X||Z) \\ &= h(Z) - h(X) \;=\; \int p \,\log \frac{p}{q}\, dx. \end{aligned}$$

Relative Fisher information of X with respect to Z

$$I(X||Z) \;=\; I(X) - I(Z) \;=\; \int \left(\frac{p'}{p} - \frac{q'}{q}\right)^2 p\, dx.$$

Properties

- $0 \leq D(X) \leq +\infty$
- $D(a + bX) = D(X)$
- Same for the standardized Fisher information

$$\sigma^2 I(X||Z) = \sigma^2 I(X) - 1$$

- $D(X) = 0 \Longleftrightarrow I(X||Z) = 0 \Longleftrightarrow X$ is normal

4 Relations between distances

• Csiszár-Kullback-Pinsker inequality for total variation (1967): For all random variables X and Z,

$$\frac{1}{2}\,\|P_X - P_Z\|_{\mathrm{TV}}^2 \;\leq\; D(X||Z).$$

• Stam's inequality (1959) $\sim$ Logarithmic Sobolev inequality: If $Z \sim N(0,1)$,

$$D(X||Z) \;\leq\; \frac{1}{2}\,I(X||Z).$$

Sharpening (still equivalent): If $Z \sim N(a,\sigma^2)$, $\mathbf{E}X = \mathbf{E}Z = a$, $\mathrm{Var}(X) = \mathrm{Var}(Z) = \sigma^2$, then

$$D(X) \;\leq\; \frac{1}{2}\,\log\Big[1 + \sigma^2 I(X||Z)\Big] \;=\; \frac{1}{2}\,\log\,[\sigma^2 I(X)].$$

• Let $\mathbf{E}X = 0$, $\mathrm{Var}(X) = 1$, $X \sim p$, $Z \sim N(0,1)$:

$$\|P_X - P_Z\|_{\mathrm{TV}} \;=\; \int_{-\infty}^{+\infty} |p(x) - \varphi(x)|\,dx \;\leq\; \sqrt{2\,I(X||Z)}.$$

• Shimizu (1975):

$$\sup_x |p(x) - \varphi(x)| \;\leq\; C\sqrt{I(X||Z)}.$$

• Sharpening: One can show that

$$\|p - \varphi\|_{\mathrm{TV}} \;=\; \int_{-\infty}^{+\infty} |p'(x) - \varphi'(x)|\,dx \;\leq\; C\sqrt{I(X||Z)}.$$

5 Central limit theorem

$(X_n)_{n\geq 1}$ independent identically distributed random variables,
$\mathbf{E}X_1 = 0$, $\mathrm{Var}(X_1) = 1$

CLT: Weakly in distribution

$$Z_n = \frac{X_1 + \ldots + X_n}{\sqrt{n}} \Rightarrow Z \sim N(0,1) \qquad (n \to \infty)$$

Theorem (Barron-Johnson 2004)

$$I(Z_n||Z) \to 0, \quad \text{as} \;\; n \to \infty \iff$$
$$I(Z_{n_0}||Z) < +\infty \qquad \text{for some} \;\; n_0.$$

Equivalently: $I(Z_{n_0}) < +\infty$ for some n_0.
Sufficient: $I(X_1) < +\infty$.
Necessary: $\forall\; n \geq n_1$, Z_n have bounded densities p_n and

$$\sup_x |p_n(x) - \varphi(x)| \to 0 \qquad (n \to \infty).$$

Problems

1. How to determine in terms of X_1 ? (range of applicability)
2. What is rate for $I(Z_n||Z)$, and under what conditions?

6 Uniform local limit theorem

Theorem (Gnedenko 1950's)
The following properties are equivalent:

$a)$ For all sufficently large n, Z_n have (continuous) bounded densities p_n satisfying

$$\sup_x |p_n(x) - \varphi(x)| \to 0 \qquad (n \to \infty);$$

$b)$ For some n, Z_n has a (continuous) bounded density p_n;

$c)$ The characteristic function $f_1(t) = \mathbf{E}\, e^{itX_1}$ of X_1 satisfies a "smoothness" condition

$$\int_{-\infty}^{+\infty} |f_1(t)|^\nu \, dt < +\infty, \qquad \text{for some } \ \nu > 0.$$

7 CLT for Fisher information distance

$(X_n)_{n\geq 1}$ independent identically distributed random variables, $\mathbf{E}X_1 = 0$, $\mathrm{Var}(X_1) = 1$.

Theorem 1. The following assertions are equivalent:

$a)$ For some n, Z_n has finite Fisher information;

$b)$ For some n, Z_n has density of bounded total variation;

$c)$ For some n, Z_n has a continuously differentiable density p_n such that

$$\int_{-\infty}^{+\infty} |p_n'(x)|\,dx < +\infty;$$

$d)$ For some $\varepsilon > 0$, the characteristic function $f_1(t) = \mathbf{E}\,e^{itX_1}$ satisfies

$$|f_1(t)| = O(t^{-\varepsilon}), \qquad \text{as} \quad t \to +\infty;$$

$e)$ For some $\nu > 0$,

$$\int_{-\infty}^{+\infty} |f_1(t)|^\nu \ |t|\,dt < +\infty.$$

- In this and only in this case,

$$I(Z_n||Z) \to 0 \qquad (n \to \infty).$$

8 $1/n$ – bounds

Barron, Johnson (2004)
Artstein, Ball, Barthe, Naor (2004)

Theorem. Assume that $\mathbf{E}X_1 = 0$, $\mathrm{Var}(X_1) = 1$, and that X_1 satisfies a Poincaré-type inequality

$$\lambda_1 \ \mathrm{Var}(u(X_1)) \ \leq \ \mathbf{E}\, u'(X_1)^2 \qquad\qquad (0 < \lambda_1 \leq 1).$$

Then

$$I(Z_n||Z) \ \leq \ \frac{1}{1 + \frac{\lambda_1}{2}\ (n-1)}\ I(X_1||Z).$$

Thus, $I(Z_n||Z) = O(1/n)$.

Extension to $Z_n = a_1 X_1 + \ldots + a_n X_n$ $(a_1^2 + \ldots + a_n^2 = 1)$
A-B-B-N (2004):

$$I(Z_n||Z) \ \leq \ \frac{L_4}{\frac{\lambda_1}{2} + (1 - \frac{\lambda_1}{2})\ L_4}\ I(X_1||Z),$$

where

$$L_4 = a_1^4 + \ldots + a_n^4.$$

9 Rate of convergence under moment conditions

$(X_n)_{n\geq 1}$ independent identically distributed random variables. Let $\mathbf{E}X_1 = 0$, $\mathrm{Var}(X_1) = 1$, and $I(Z_{n_0}) < +\infty$, for some n_0.

Theorem 2. If $\mathbf{E}\,|X_1|^s < +\infty$, for some $s > 2$, then

$$I(Z_n||Z) \;=\; \sum_{j=1}^{[(s-2)/2)]} \frac{c_j}{n^j} + o\left(\frac{1}{n^{(s-2)/2}\,(\log n)^{(s-3)/2}}\right),$$

where each c_j is a certain polynomial in cumulants $\gamma_3, \dots, \gamma_{2j+1}$ of X_1, or moments $\mathbf{E}X_1^3, \dots, \mathbf{E}X_1^{2j+1}$.

- $s = 4$: $\mathbf{E}X_1^4 < +\infty \;\Rightarrow$

$$I(Z_n||Z) \;=\; \frac{c_1}{n} + o\left(\frac{1}{n\,(\log n)^{1/2}}\right), \qquad c_1 = \frac{1}{2!}\,\gamma_3^2 = \frac{1}{2}\,(\mathbf{E}X_1^3)^2.$$

- $s = 6$: $\mathbf{E}X_1^6 < +\infty$, $\mathbf{E}X_1^3 = 0 \;\Rightarrow$

$$I(Z_n||Z) \;=\; \frac{c_2}{n^2} + o\left(\frac{1}{n^2\,(\log n)^{3/2}}\right), \quad c_2 = \frac{1}{3!}\,\gamma_4^2 = \frac{1}{6}\,(\mathbf{E}X_1^4 - 3)^2.$$

10 Case $2 < s < 4$. Lower bounds

In case $\mathbf{E}\,|X_1|^s < +\infty$ with $2 < s < 4$, Theorem 2 only yields

$$I(Z_n||Z) \quad = \quad o\left(\frac{1}{n^{(s-2)/2}\,(\log n)^{(s-3)/2}}\right).$$

This is worse than $1/n$–rate.

Let $\eta > \frac{s}{2}$, $2 < s < 4$.

Theorem 3. There exists a sequence $(X_n)_{n \geq 1}$ of independent i.i.d. random variables with symmetric distributions, with

$$\mathbf{E}X_1^2 = 1, \quad \mathbf{E}\,|X_1|^s < +\infty, \quad I(X_1) < +\infty,$$

and such that with some constant $c = c(\eta, s)$

$$I(Z_n||Z) \quad \geq \quad \frac{c}{n^{(s-2)/2}\,(\log n)^{\eta}}, \qquad n \geq n_1(X_1).$$

Remark. The distribution of X_1 may be a mixture of mean zero normal laws.

11 When is Fisher information finite?

Question: What should one assume about X with density p to ensure that

$$I(X) \ = \ \int_{-\infty}^{+\infty} \frac{p'(x)^2}{p(x)}\, dx < +\infty \ ?$$

And if so, how to bound $I(X)$ from above?

- Stam's inequality: If X_1 and X_2 are independent, then

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

- Monotonicity: $I(X_1 + X_2) \leq I(X_1)$.

- Example: $X_j \sim$ Uniform on intervals of length $a_j \Rightarrow$

$$\begin{aligned} I(X_1) \ &= \ +\infty \qquad \text{(uniform distribution)} \\ I(X_1 + X_2) \ &= \ +\infty \qquad \text{(triangle distribution)} \\ I(X_1 + X_2 + X_3) \ &< \ +\infty \qquad \text{(like beta with } \alpha = \beta = 2). \end{aligned}$$

12 Necessary conditions

From the definition

$$I(X) \;=\; \mathbf{E}\left[\frac{p'(X)}{p(X)}\right]^2 \;\geq\; \left[\mathbf{E}\,\frac{|p'(X)|}{p(X)}\right]^2 \;=\; \left[\int_{-\infty}^{+\infty} |p'(x)|\,dx\right]^2.$$

Hence, p is a function of bounded variation with

$$\|p\|_{\mathrm{TV}} \;\leq\; \sqrt{I(X)}.$$

In general, the characteristic function $f(t) = \mathbf{E}\,e^{itX}$ satisfies

$$|f(t)| \;\leq\; \frac{1}{|t|}\,\|p\|_{\mathrm{TV}} \qquad (t \in \mathbf{R}).$$

Conclusion.

$$|f(t)| \;\leq\; \frac{1}{|t|}\,\sqrt{I(X)} \qquad (t \in \mathbf{R}).$$

13 Convolution of densities of bounded variation

Let $S = X_1 + X_2 + X_3$ be the sum of three independent random variables with densities p_1, p_2, p_3 having bounded total variation.

Proposition 1. One has

$$2I(S) \le \|p_1\|_{\rm TV} \, \|p_2\|_{\rm TV} + \|p_1\|_{\rm TV} \, \|p_3\|_{\rm TV} + \|p_2\|_{\rm TV} \, \|p_3\|_{\rm TV}.$$

In particular, if $p_1 = p_2 = p_3 = p$,

$$I(X_1 + X_2 + X_3) \le \frac{3}{2} \, \|p\|_{\rm TV}^2.$$

Definition.

$$\|p\|_{\rm TV} = \sup \sum_{k=1}^{n} |p(x_k) - p(x_{k-1})|,$$

where the sup is over all $x_0 < x_1 < \ldots < x_n$, and where we may assume that $p(x)$ is in between $p(x-)$ and $p(x+)$, for all x.

Particular case: If $X_j \sim$ Uniform on intervals of length a_j, then

$$\frac{1}{2} \, I(X_1 + X_2 + X_3) \le \frac{1}{a_1 a_2} + \frac{1}{a_1 a_3} + \frac{1}{a_2 a_3}.$$

However, $I(X_1 + X_2) = +\infty$.

14 Proof of Proposition 1

Let $\mathcal{P}$ denote the collection of all densities of bounded variation. Let $\mathcal{U}$ denote the collection of all uniform densities

$$q(x) = \frac{1}{b-a}, \quad \text{for } \; a < x < b.$$

Note that

$$\|q\|_{\mathrm{TV}} = \frac{2}{b-a}.$$

Proposition 1 follows from the case of uniform densities and the following:

Lemma. Any density $p \in \mathcal{P}$ can be represented as a convex mixture of uniform densities

$$p(x) = \int_{\mathcal{U}} q(x) \; d\pi(q) \quad \text{a.e.}$$

and with the property that

$$\|p\|_{\mathrm{TV}} = \int_{\mathcal{U}} \|q\|_{\mathrm{TV}} \; d\pi(q).$$

Remark. The mixing probability measure π on $\mathcal{U}$ seems to be unique, but no explicit construction is available.
When p is piece-wise constant, the lemma can be proved by induction on the number of supporting intervals.

15 Proof of Theorem 1

Let $S_n = X_1 + \ldots + X_n$ with i.i.d. summands and characteristic function

$$f_n(t) = \mathbf{E}\, e^{itS_n} = f_1(t)^n.$$

- If $I_n = I(S_n) < +\infty$, then, as noted,

$$|f_1(t)|^n = |f_n(t)| \leq \frac{1}{|t|}\,\sqrt{I_n} \quad \Rightarrow \quad |f_1(t)| = O(t^{-1/n}).$$

- Now, assume that, for some (fixed) n,

$$\int_{-\infty}^{+\infty} |f_1(t)|^n\, |t|\, dt < +\infty.$$

Then S_n has density

$$p_n(x) = \frac{1}{2\pi}\, \int_{-\infty}^{+\infty} e^{-itx} f_1(t)^n\, dt,$$

which has a continuous derivative satisfying

$$(1 + x^2)\, p_n'(x) \, = \, \frac{i}{2\pi}\, \int_{-\infty}^{+\infty} e^{-itx}\, (t f_n''(t) + 2 f_n'(t) - t f_n(t))\, dt.$$

Hence, $|p_n'(x)| \leq \frac{C}{1+x^2}$ and $\|p_n\|_{\mathrm{TV}} < +\infty$. By Proposition 1,

$$I_{3n} < +\infty.$$

16 Towards Theorem 2

Let $(X_n)_{n\geq 1}$ be i.i.d., $\mathbf{E}X_1 = 0$, $\mathrm{Var}(X_1) = 1$,

$$Z_n = \frac{X_1 + \ldots + X_n}{\sqrt{n}}, \qquad I(Z_n) < +\infty \ \ (n \geq n_0),$$

with densities p_n so that

$$I(Z_n||Z) \ = \ \int_{-\infty}^{+\infty} \frac{(p_n'(x) + xp_n(x))^2}{p_n(x)}\, dx \ = \ I_0 + I_1,$$

$$I_0 \ = \ \int_{-T_n}^{T_n} \frac{(p_n'(x) + xp_n(x))^2}{p_n(x)}\, dx, \qquad I_1 \ = \ \int_{|x|\geq T_n} \cdots$$

Good choice:

$$T_n = \sqrt{(s-2)\log n + s \log\log n + \rho_n} \qquad (s > 2),$$

where $\rho_n \to +\infty$ sufficiently slowly to guarantee that

$$\sup_{|x|\leq T_n} \left| \frac{p_n(x)}{\varphi(x)} - 1 \right| \to 0.$$

Case $s = 4$: $T_n^2 \ \sim \ 2\log n + 4\log\log n + \rho_n$.

17 Edgeworth-type expansion for densities

Let $\mathbf{E}X_1^s < +\infty$ ($s \geq 3$ integer).
For $|x| \leq T_n$, one may use a suitable approximation of p_n.
Not enough:

$$(1+|x|^s)\ (p_n(x) - \varphi(x)) = O\left(\frac{1}{\sqrt{n}}\right).$$

Edgeworth approximation of p_n:

$$\varphi_s(x) \ = \ \varphi(x) + \sum_{k=1}^{s-2} q_k(x)\, n^{-k/2}$$

with

$$q_k(x) \ = \ \varphi(x) \sum \ \frac{H_{k+2j}(x)}{r_1! \dots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \dots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k}.$$

Here $r_1 + 2r_2 + \ldots + kr_k = k$, $j = r_1 + \ldots + r_k$ and

$$\gamma_r = i^{-r}\, \frac{d^r}{dt^r}\, \log \mathbf{E}\, e^{itX_1}\Big|_{t=0} \qquad (3 \leq r \leq s).$$

Lemma 1. Let $I(Z_{n_0}) < +\infty$, for some n_0. Fix $l = 0, 1, \ldots$ Then, for all sufficiently large n, for all x,

$$|p_n^{(l)}(x) - \varphi_s^{(l)}(x)| \ \leq \ \frac{\psi_{l,n}(x)}{1+|x|^s}\, \frac{\varepsilon_n}{n^{(s-2)/2}},$$

where $\varepsilon_n \to 0$, as $n \to \infty$, and

$$\sup_x\, |\psi_{l,n}(x)| \leq 1, \qquad \int_{-\infty}^{+\infty} \psi_{l,n}(x)^2\, dx \leq 1.$$

18 Moderate deviations

Second step:

$$I_1 \;=\; \int_{|x| \geq T_n} \frac{(p_n'(x) + x p_n(x))^2}{p_n(x)}\, dx \;=\; o\left(\frac{1}{n^{(s-2)/2}\,(\log n)^{(s-3)/2}}\right).$$

We have

$$I_1 \;\leq\; 2I_{1,1} + 2I_{1,2},$$

where

$$I_{1,1} \;=\; \int_{|x| \geq T_n} \frac{p_n'(x)^2}{p_n(x)}\, dx,$$

$$I_{1,2} \;=\; \int_{|x| \geq T_n} x^2\, p_n(x)\, dx \quad \text{easy.}$$

Integration by parts:

$$\begin{aligned} I_{1,1}^{+} \;&=\; \int_{T_n}^{+\infty} \frac{p_n'(x)^2}{p_n(x)}\, dx \\ &=\; -p_n'(T_n) \log p_n(T_n) - \int_{T_n}^{+\infty} p_n''(x) \log p_n(x)\, dx. \end{aligned}$$

Lemma 2. Assume p is representable as convolution of three densities with Fisher information $\leq I$. Then, for all x,

$$|p'(x)| \;\leq\; I^{3/4} \sqrt{p(x)},$$

$$|p''(x)| \;\leq\; I^{5/4} \sqrt{p(x)}.$$