On the convex infimum convolution inequality with optimal cost function

Marta Strzelecka

(based on joint work with Michał Strzelecki and Tomasz Tkocz)

University of Warsaw

Cargèse, May 16, 2018

Convex infimum convolution inequality

X – a random vector in $\mathbb{R}^n$, $\varphi : \mathbb{R}^n \to [0, \infty]$ – a measurable function (a cost function),
$f : \mathbb{R}^n \to \mathbb{R}$ – a bounded measurable function (a test function),

Convex infimum convolution inequality

X – a random vector in $\mathbb{R}^n$, $\varphi : \mathbb{R}^n \to [0, \infty]$ – a measurable function (a cost function),
$f : \mathbb{R}^n \to \mathbb{R}$ – a bounded measurable function (a test function),

$$f \Box \varphi(x) := \inf\{f(y) + \varphi(x - y) : y \in \mathbb{R}^n\} \text{ – infimum convolution.}$$

Convex infimum convolution inequality

X – a random vector in $\mathbb{R}^n$, $\varphi : \mathbb{R}^n \to [0, \infty]$ – a measurable function (a cost function),
$f : \mathbb{R}^n \to \mathbb{R}$ – a bounded measurable function (a test function),

$$f \Box \varphi(x) := \inf\{f(y) + \varphi(x - y) : y \in \mathbb{R}^n\} \quad \text{– infimum convolution.}$$

We say that a pair (X, φ) satisfies the *infimum convolution inequality* (ICI for short) if for every test function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}e^{f \Box \varphi(X)} \mathbb{E}e^{-f(X)} \leq 1. \tag{1}$$

Convex infimum convolution inequality

X – a random vector in $\mathbb{R}^n$, $\varphi : \mathbb{R}^n \to [0, \infty]$ – a measurable function (a cost function),
$f : \mathbb{R}^n \to \mathbb{R}$ – a bounded measurable function (a test function),

$$f \Box \varphi(x) := \inf\{f(y) + \varphi(x-y) : y \in \mathbb{R}^n\} \text{ – infimum convolution.}$$

We say that a pair (X, φ) satisfies the *infimum convolution inequality* (ICI for short) if for every test function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\mathbb{E}e^{f \Box \varphi(X)} \mathbb{E}e^{-f(X)} \leq 1. \tag{1}$$

We also say that a pair (X, φ) satisfies the *convex infimum convolution inequality* (convex ICI for short) if (1) holds for every **convex** function $f : \mathbb{R}^n \to \mathbb{R}$ bounded from below.

What we know

- Gaussian random variable on the real line satisfies the ICI with a quadratic cost function (Maurey);

What we know

- Gaussian random variable on the real line satisfies the ICI with a quadratic cost function (Maurey);
- exponential r.v. on the real line satisfies the ICI with a quadratic-linear cost function ($\Rightarrow$ Talagrand's two-level concentration) (Maurey);

What we know

- Gaussian random variable on the real line satisfies the ICI with a quadratic cost function (Maurey);
- exponential r.v. on the real line satisfies the ICI with a quadratic-linear cost function ($\Rightarrow$ Talagrand's two-level concentration) (Maurey);
- bounded random variables satisfy the convex ICI with a quadratic cost function (Maurey);

What we know

- Gaussian random variable on the real line satisfies the ICI with a quadratic cost function (Maurey);
- exponential r.v. on the real line satisfies the ICI with a quadratic-linear cost function ($\Rightarrow$ Talagrand's two-level concentration) (Maurey);
- bounded random variables satisfy the convex ICI with a quadratic cost function (Maurey);
- convex ICI (on the real line) with a quadratic-linear cost function $\Leftrightarrow \exists \lambda \in [0,1), h > 0$ such that $\mathbb{P}(X \geq x+h) \leq \lambda \mathbb{P}(X \geq x)$ (Feldheim, Marsiglietti, Nayar, Wang, 2018);

What we know

- Gaussian random variable on the real line satisfies the ICI with a quadratic cost function (Maurey);
- exponential r.v. on the real line satisfies the ICI with a quadratic-linear cost function ($\Rightarrow$ Talagrand's two-level concentration) (Maurey);
- bounded random variables satisfy the convex ICI with a quadratic cost function (Maurey);
- convex ICI (on the real line) with a quadratic-linear cost function $\Leftrightarrow \exists \lambda \in [0,1), h > 0$ such that $\mathbb{P}(X \geq x+h) \leq \lambda \mathbb{P}(X \geq x)$ (Feldheim, Marsiglietti, Nayar, Wang, 2018);
- convex ICIs are the dual formulation of weak transport-entropy inequalities introduced by Gozlan, Roberto, Samson, Tetali, 2017;

What we know

- Gaussian random variable on the real line satisfies the ICI with a quadratic cost function (Maurey);
- exponential r.v. on the real line satisfies the ICI with a quadratic-linear cost function ($\Rightarrow$ Talagrand's two-level concentration) (Maurey);
- bounded random variables satisfy the convex ICI with a quadratic cost function (Maurey);
- convex ICI (on the real line) with a quadratic-linear cost function $\Leftrightarrow \exists \lambda \in [0,1), h > 0$ such that $\mathbb{P}(X \geq x + h) \leq \lambda \mathbb{P}(X \geq x)$ (Feldheim, Marsiglietti, Nayar, Wang, 2018);
- convex ICIs are the dual formulation of weak transport-entropy inequalities introduced by Gozlan, Roberto, Samson, Tetali, 2017;
- on the real line: a characterization of convex ICI with an arbitrary convex cost function quadratic near 0 (Gozlan, Roberto, Samson, Shu, Tetali, 2018+);

What we know

- Gaussian random variable on the real line satisfies the ICI with a quadratic cost function (Maurey);
- exponential r.v. on the real line satisfies the ICI with a quadratic-linear cost function ($\Rightarrow$ Talagrand's two-level concentration) (Maurey);
- bounded random variables satisfy the convex ICI with a quadratic cost function (Maurey);
- convex ICI (on the real line) with a quadratic-linear cost function $\Leftrightarrow \exists \lambda \in [0,1), h > 0$ such that $\mathbb{P}(X \geq x + h) \leq \lambda \mathbb{P}(X \geq x)$ (Feldheim, Marsiglietti, Nayar, Wang, 2018);
- convex ICIs are the dual formulation of weak transport-entropy inequalities introduced by Gozlan, Roberto, Samson, Tetali, 2017;
- on the real line: a characterization of convex ICI with an arbitrary convex cost function quadratic near 0 (G., R., S., Shu, T., 2018+);
- ICI with optimal cost function (scaled Legendre transform) for vectors uniformly distributed on ℓ_p^n-balls and for product log-concave vectors (Latała, Wojtaszczyk, 2008).

Optimal cost function

For a random vector X in $\mathbb{R}^n$ let

$$\Lambda_X(x) := \ln \mathbb{E} e^{\langle x, X \rangle}, \qquad x \in \mathbb{R}^n$$

(the cumulant-generating function). We define its Legendre transform

$$\Lambda_X^*(x) := \mathcal{L}\Lambda_X(x) := \sup_{y \in \mathbb{R}^n} \{\langle x, y \rangle - \ln \mathbb{E} e^{\langle y, X \rangle}\}.$$

Optimal cost function

For a random vector X in $\mathbb{R}^n$ let

$$\Lambda_X(x) := \ln \mathbb{E} e^{\langle x, X \rangle}, \qquad x \in \mathbb{R}^n$$

(the cumulant-generating function). We define its Legendre transform

$$\Lambda_X^*(x) := \mathcal{L}\Lambda_X(x) := \sup_{y \in \mathbb{R}^n} \{\langle x, y \rangle - \ln \mathbb{E} e^{\langle y, X \rangle}\}.$$

Proposition

If a symmetric random vector X satisfies the convex ICI with a convex cost function φ, then $\varphi \leq \Lambda_X^*$.

Optimal cost function

For a random vector X in $\mathbb{R}^n$ let

$$\Lambda_X(x) := \ln \mathbb{E} e^{\langle x, X \rangle}, \qquad x \in \mathbb{R}^n$$

(the cumulant-generating function). We define its Legendre transform

$$\Lambda_X^*(x) := \mathcal{L}\Lambda_X(x) := \sup_{y \in \mathbb{R}^n} \{\langle x, y \rangle - \ln \mathbb{E} e^{\langle y, X \rangle}\}.$$

Proposition

If a symmetric random vector X satisfies the convex ICI with a convex cost function φ, then $\varphi \le \Lambda_X^*$.

We say that X satisfies the (convex) $IC(\beta)$ if the pair$(X, \Lambda_X^*(\frac{\cdot}{\beta}))$ satisfies the (convex) ICI.

Characterization of convex IC on the real line

μ – the distribution of a random variable X, ν – the (symmetric) exponential distribution.

$$F_\mu(t) := \mu(-\infty, t], \quad F_\nu(t) := \nu(-\infty, t].$$

Characterization of convex IC on the real line

μ – the distribution of a random variable X, ν – the (symmetric) exponential distribution.

$$F_\mu(t) := \mu(-\infty, t], \quad F_\nu(t) := \nu(-\infty, t].$$

$$U = U_\mu := F_\mu^{-1} \circ F_\nu$$

is the unique non-decreasing, left-continuous function transporting ν on μ.

Characterization of convex IC on the real line

μ – the distribution of a random variable X, ν – the (symmetric) exponential distribution.

$$F_\mu(t) := \mu(-\infty, t], \quad F_\nu(t) := \nu(-\infty, t].$$

$$U = U_\mu := F_\mu^{-1} \circ F_\nu$$

is the unique non-decreasing, left-continuous function transporting ν on μ.

Theorem [Gozlan, Roberto, Samson, Shu, Tetali, 2018+]

φ – convex, symmetric cost function, $\varphi(t) = t^2$ for $|t| \leq t_0$. Then the following are equivalent:

(i) There exists $a > 0$ such that X satisfies IC with a cost function $\varphi(a\,\cdot)$

(ii) There exists $b > 0$ such that for all $x, y \in \mathbb{R}$,

$$\big|U(x) - U(y)\big| \leq \frac{1}{b}\varphi^{-1}\big(1 + |x - y|\big).$$

Convex IC with optimal cost function

Assume that a symmetric random variable X has log-concave tails:

$$\mathbb{P}(|X| \geq t) = e^{-N(t)}, \quad N \text{ is convex.}$$

How to use the characterization to get the optimal convex IC for X?

Convex IC with optimal cost function

Assume that a symmetric random variable X has log-concave tails:

$$\mathbb{P}(|X| \geq t) = e^{-N(t)}, \quad N \text{ is convex.}$$

How to use the characterization to get the optimal convex IC for X?

- Rescale X so that $\mathbb{E}X^2 = (2e)^{-2}$. Then $N(1/2) \geq 2$ and the Chernoff inequality implies that $N(t) + \ln 2 \geq \Lambda_X^*(t)$. (We may assume that μ is nice.)

Convex IC with optimal cost function

Assume that a symmetric random variable X has log-concave tails:

$$\mathbb{P}(|X| \geq t) = e^{-N(t)}, \quad N \text{ is convex.}$$

How to use the characterization to get the optimal convex IC for X?

- Rescale X so that $\mathbb{E}X^2 = (2e)^{-2}$. Then $N(1/2) \geq 2$ and the Chernoff inequality implies that $N(t) + \ln 2 \geq \Lambda_X^*(t)$. (We may assume that μ is nice.)
- Modify the cost function Λ_X^*:

$$\varphi(x) := \left(x^2 \mathbf{1}_{\{|x|<1\}} + (2|x|-1)\mathbf{1}_{\{|x|\geq 1\}}\right) \vee \Lambda_X^*\left(x/(4e)\right).$$

Convex IC with optimal cost function

Assume that a symmetric random variable X has log-concave tails:

$$\mathbb{P}(|X| \geq t) = e^{-N(t)}, \quad N \text{ is convex.}$$

How to use the characterization to get the optimal convex IC for X?

- Rescale X so that $\mathbb{E}X^2 = (2e)^{-2}$. Then $N(1/2) \geq 2$ and the Chernoff inequality implies that $N(t) + \ln 2 \geq \Lambda_X^*(t)$. (We may assume that μ is nice.)
- Modify the cost function Λ_X^*:

$$\varphi(x) := \left(x^2 \mathbf{1}_{\{|x|<1\}} + (2|x|-1)\mathbf{1}_{\{|x|\geq 1\}}\right) \vee \Lambda_X^*\big(x/(4e)\big).$$

- Roughly speaking $U^{-1}(x) = N(|x|)\operatorname{sgn} x$. One may show that it is enough to prove that

$$\varphi\big(|x-y|\big) \leq 1 + \big|N(|x|)\operatorname{sgn} x - N(|y|)\operatorname{sgn} y\big| \qquad \text{for } x, y \in U(\mathbb{R}).$$

Concentration inequalities

- In the log-concave setting: IC with optimal cost function is equivalent to the optimal concentration (we enlarge a given set by $\mathcal{Z}_p(X)$ instead of pB_2^n) (Latała, Wojtaszczyk, 2008);

Concentration inequalities

- In the log-concave setting: IC with optimal cost function is equivalent to the optimal concentration (we enlarge a given set by $\mathcal{Z}_p(X)$ instead of pB_2^n) (Latała, Wojtaszczyk, 2008);
- Convex IC $\Rightarrow$ one-side concentration (under the condition of regularly growing moments);

Concentration inequalities

- In the log-concave setting: IC with optimal cost function is equivalent to the optimal concentration (we enlarge a given set by $\mathcal{Z}_p(X)$ instead of pB_2^n) (Latała, Wojtaszczyk, 2008);
- Convex IC $\Rightarrow$ one-side concentration (under the condition of regularly growing moments);
- However, convex IC(β) and α-regularly growing moments of coordinates imply also that for every norm $\|\cdot\|$ on $\mathbb{R}^n$,

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \leq 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \geq 2e\alpha\beta\sigma(p),$$

where $\sigma(p)$ is the weak p-th moment of X with respect to $\|\cdot\|$:

$$\sigma(p) = \sigma_{\|\cdot\|,X}(p) := \sup_{\|t\|_* \leq 1} \|\langle t, X\rangle\|_p.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \leq 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \geq 2e\alpha\beta\sigma(p).$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

$$\left(\Lambda_X^*(\cdot/\beta)\Box a\|\cdot\|\right)(x) = \inf_y \sup_z \left\{\beta^{-1}\langle y, z\rangle - \Lambda_X(z) + a\|x - y\|\right\}$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \leq 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \geq 2e\alpha\beta\sigma(p).$$

$$\left(\Lambda_X^*(\cdot/\beta)\Box a\|\cdot\|\right)(x) = \inf_y \sup_z \left\{\beta^{-1}\langle y, z\rangle - \Lambda_X(z) + a\|x - y\|\right\}$$

$$= \inf_y \sup_u \left\{(2e\alpha\beta)^{-1} p\langle y, u\rangle - \Lambda_X((2e\alpha)^{-1} pu) + a\|x - y\|\right\}$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

$$\left(\Lambda_X^*(\cdot/\beta)\Box a\|\cdot\|\right)(x) = \inf_y \sup_z \left\{\beta^{-1}\langle y, z\rangle - \Lambda_X(z) + a\|x - y\|\right\}$$

$$\ge \inf_y \sup_{u:\|\langle u,X\rangle\|_p \le 1} \left\{(2e\alpha\beta)^{-1} p\langle y, u\rangle - \Lambda_X((2e\alpha)^{-1}pu) + a\|x - y\|\right\}$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

$$\left(\Lambda_X^*(\cdot/\beta)\square a\|\cdot\|\right)(x) = \inf_y \sup_z \left\{\beta^{-1}\langle y, z\rangle - \Lambda_X(z) + a\|x-y\|\right\}$$

$$\ge \inf_y \sup_{u:\|\langle u,X\rangle\|_p \le 1} \left\{(2e\alpha\beta)^{-1}p\langle y, u\rangle - \Lambda_X((2e\alpha)^{-1}pu) + a\|x-y\|\right\}$$

Lemma

α-regularly growing moments of $\langle t, X\rangle$ and $\|\langle u, X\rangle\|_p \le 1$ imply

$$\Lambda_X((2e\alpha)^{-1}pu) \le p.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

$$\left(\Lambda_X^*(\cdot/\beta)\square a\|\cdot\|\right)(x) = \inf_y \sup_z \left\{\beta^{-1}\langle y, z\rangle - \Lambda_X(z) + a\|x-y\|\right\}$$

$$\ge \inf_y \sup_{u:\|\langle u,X\rangle\|_p \le 1} \left\{(2e\alpha\beta)^{-1} p\langle y,u\rangle - \Lambda_X((2e\alpha)^{-1}pu) + a\|x-y\|\right\}$$

$$\ge \inf_y \sup_{u:\|\langle u,X\rangle\|_p \le 1} \left\{(2e\alpha\beta)^{-1} p\langle y,u\rangle - p + a\|x-y\|\right\}.$$

Lemma

α-regularly growing moments of $\langle t, X\rangle$ and $\|\langle u, X\rangle\|_p \le 1$ imply

$$\Lambda_X((2e\alpha)^{-1}pu) \le p.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \leq 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \geq 2e\alpha\beta\sigma(p).$$

$$\left(\Lambda_X^*(\cdot/\beta)\Box a\|\cdot\|\right)(x) = \inf_y \sup_z \left\{\beta^{-1}\langle y, z\rangle - \Lambda_X(z) + a\|x-y\|\right\}$$

$$\geq \inf_y \sup_{u:\|\langle u,X\rangle\|_p \leq 1} \left\{(2e\alpha\beta)^{-1} p\langle y, u\rangle - \Lambda_X((2e\alpha)^{-1}pu) + a\|x-y\|\right\}$$

$$\geq \inf_y \sup_{u:\|\langle u,X\rangle\|_p \leq 1} \left\{(2e\alpha\beta)^{-1} p\langle y, u\rangle - p + a\|x-y\|\right\}.$$

Choose $u = \sigma_{\|\cdot\|,X}(p)^{-1}v$ with $\|v\|_* \leq 1$ such that $\langle y, v\rangle = \|y\|$.

Reminder

$$\sigma(p) = \sigma_{\|\cdot\|,X}(p) := \sup_{\|t\|_* \leq 1} \|\langle t, X\rangle\|_p.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\left|\|X\| - \mathbb{E}\|X\|\right| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

$$\left(\Lambda_X^*(\cdot/\beta)\Box a\|\cdot\|\right)(x) = \inf_y \sup_z \left\{\beta^{-1}\langle y, z\rangle - \Lambda_X(z) + a\|x-y\|\right\}$$

$$\ge \inf_y \sup_{u:\|\langle u,X\rangle\|_p \le 1} \left\{(2e\alpha\beta)^{-1} p\langle y,u\rangle - \Lambda_X((2e\alpha)^{-1}pu) + a\|x-y\|\right\}$$

$$\ge \inf_y \sup_{u:\|\langle u,X\rangle\|_p \le 1} \left\{(2e\alpha\beta)^{-1} p\langle y,u\rangle - p + a\|x-y\|\right\}.$$

Choose $u = \sigma_{\|\cdot\|,X}(p)^{-1}v$ with $\|v\|_* \le 1$ such that $\langle y, v\rangle = \|y\|$. Then

$$\Lambda_X^*(\cdot/\beta)\Box f(x) \ge \inf_y \left\{\left(2e\alpha\beta\sigma_{\|\cdot\|,X}(p)\right)^{-1} p\|y\| - p + a\|x-y\|\right\}.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

$$\left(\Lambda_X^*(\cdot/\beta)\Box a\|\cdot\|\right)(x) = \inf_y \sup_z \left\{\beta^{-1}\langle y, z\rangle - \Lambda_X(z) + a\|x-y\|\right\}$$

$$\ge \inf_y \sup_{u:\|\langle u,X\rangle\|_p \le 1} \left\{(2e\alpha\beta)^{-1} p\langle y, u\rangle - \Lambda_X((2e\alpha)^{-1}pu) + a\|x-y\|\right\}$$

$$\ge \inf_y \sup_{u:\|\langle u,X\rangle\|_p \le 1} \left\{(2e\alpha\beta)^{-1} p\langle y, u\rangle - p + a\|x-y\|\right\}.$$

Choose $u = \sigma_{\|\cdot\|,X}(p)^{-1} v$ with $\|v\|_* \le 1$ such that $\langle y, v\rangle = \|y\|$. Then

$$\Lambda_X^*(\cdot/\beta)\Box f(x) \ge \inf_y \left\{\left(2e\alpha\beta\sigma_{\|\cdot\|,X}(p)\right)^{-1} p\|y\| - p + a\|x-y\|\right\}.$$

Set $a = p(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}$ to get

$$\left(\Lambda_X^*\left(\cdot/\beta\right)\Box a\|\cdot\|\right)(x) \ge a\|x\| - p.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

For $a = p(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}$ we have

$$\left(\Lambda_X^*\left(\cdot/\beta\right) \square a\|\cdot\|\right)(x) \ge a\|x\| - p.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\left|\|X\| - \mathbb{E}\|X\|\right| > t\right) \leq 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \geq 2e\alpha\beta\sigma(p).$$

For $a = p(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}$ we have

$$\left(\Lambda_X^*\left(\cdot/\beta\right)\square a\|\cdot\|\right)(x) \geq a\|x\| - p.$$

Hence the infimum convolution inequality with a test function $a\|\cdot\|$ implies

$$\mathbb{E}e^{a\|X\|}\mathbb{E}e^{-a\|X\|} \leq e^p.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\left|\|X\| - \mathbb{E}\|X\|\right| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

For $a = p(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}$ we have

$$\left(\Lambda_X^*\left(\cdot/\beta\right)\Box a\|\cdot\|\right)(x) \ge a\|x\| - p.$$

Hence the infimum convolution inequality with a test function $a\|\cdot\|$ implies

$$\mathbb{E}e^{a\|X\|}\mathbb{E}e^{-a\|X\|} \le e^p.$$

Thus Jensen's inequality imply

$$\mathbb{E}e^{a\|X\|-a\mathbb{E}\|X\|} \le e^p.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

For $a = p(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}$ we have

$$\left(\Lambda_X^*(\cdot/\beta)\,\Box a\|\cdot\|\right)(x) \ge a\|x\| - p.$$

Hence the infimum convolution inequality with a test function $a\|\cdot\|$ implies

$$\mathbb{E}e^{a\|X\|}\mathbb{E}e^{-a\|X\|} \le e^p.$$

Thus Jensen's inequality imply

$$\mathbb{E}e^{a\mathbb{E}\|X\|-a\|X\|} \le e^p.$$

Concentration inequalities – a proof

Aim

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p).$$

For $a = p(2e\alpha\beta\sigma_{\|\cdot\|,X}(p))^{-1}$ we have

$$\left(\Lambda_X^*\left(\cdot/\beta\right)\Box a\|\cdot\|\right)(x) \ge a\|x\| - p.$$

Hence the infimum convolution inequality with a test function $a\|\cdot\|$ implies

$$\mathbb{E}e^{a\|X\|}\mathbb{E}e^{-a\|X\|} \le e^p.$$

Thus Jensen's and Markov's inequalities imply

$$\mathbb{P}\left(a\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-t}e^p \le 2e^{-t/2}, \quad \text{for } t \ge 2p.$$

Comparison of weak and strong moments

Convex IC(β) and α-regularly growing moments of coordinates imply that for every norm $\|\cdot\|$ on $\mathbb{R}^n$,

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \le 2e^{-tp/(4e\alpha\beta\sigma(p)}, \quad \text{for } t \ge 2e\alpha\beta\sigma(p),$$

where $\sigma(p)$ is the weak p-th moment of X with respect to $\|\cdot\|$:

$$\sigma(p) = \sigma_{\|\cdot\|,X}(p) := \sup_{\|t\|_* \le 1} \|\langle t, X\rangle\|_p.$$

Comparison of weak and strong moments

Convex IC(β) and α-regularly growing moments of coordinates imply that for every norm $\|\cdot\|$ on $\mathbb{R}^n$,

$$\mathbb{P}\left(\big|\|X\| - \mathbb{E}\|X\|\big| > t\right) \leq 2e^{-tp/(4e\alpha\beta\sigma(p))}, \quad \text{for } t \geq 2e\alpha\beta\sigma(p),$$

where $\sigma(p)$ is the weak p-th moment of X with respect to $\|\cdot\|$:

$$\sigma(p) = \sigma_{\|\cdot\|,X}(p) := \sup_{\|t\|_* \leq 1} \|\langle t, X\rangle\|_p.$$

Integrate it to obtain a comparison of weak and strong moments:

$$\left(\mathbb{E}\|X\|^p\right)^{1/p} \leq \mathbb{E}\|X\| + D\sigma_{\|\cdot\|,X}(p),$$

Note that the constant at $\mathbb{E}\|X\|$ is equal to 1.

Further questions

1. When

$$\left(\mathbb{E}\|X\|^p\right)^{1/p} \le D_1\mathbb{E}\|X\| + D_2\sigma_{\|\cdot\|,X}(p) \qquad (2)$$

holds with $D_1 = 1$?

Further questions

1. When

$$\left(\mathbb{E}\|X\|^p\right)^{1/p} \le D_1\mathbb{E}\|X\| + D_2\sigma_{\|\cdot\|,X}(p) \tag{2}$$

holds with $D_1 = 1$?

   - In the case of independent coordinates: log-concave tails are sufficient, but α-regularly growing moments are not sufficient:

Further questions

1. When
$$\left(\mathbb{E}\|X\|^p\right)^{1/p} \leq D_1\mathbb{E}\|X\| + D_2\sigma_{\|\cdot\|,X}(p) \tag{2}$$
holds with $D_1 = 1$?
   - In the case of independent coordinates: log-concave tails are sufficient, but α-regularly growing moments are not sufficient:

Example

Define X by $\mathbb{P}(|X| > t) = F_X(t)$:

$$F_X(t) := \mathbf{1}[0,2)(t) + \sum_{k=1}^{\infty} e^{-2^k}\mathbf{1}[2^k, 2^{k+1})(t), \quad t \geq 0,$$

or, in other words, let $|X|$ have the distribution
$$(1 - e^{-2})\delta_2 + \sum_{k=2}^{\infty}\left(e^{-2^{k-1}} - e^{-2^k}\right)\delta_{2^k}.$$

Further questions

1. When
$$\left(\mathbb{E}\|X\|^p\right)^{1/p} \leq D_1\mathbb{E}\|X\| + D_2\sigma_{\|\cdot\|,X}(p) \tag{2}$$
holds with $D_1 = 1$?
   - In the case of independent coordinates: log-concave tails are sufficient, but α-regularly growing moments are not sufficient.
2. Is there any other way to obtain (2) with $D_1 = 1$ than by integrating a concentration inequality?

Further questions

1. When
$$\left(\mathbb{E}\|X\|^p\right)^{1/p} \leq D_1\mathbb{E}\|X\| + D_2\sigma_{\|\cdot\|,X}(p) \tag{2}$$
holds with $D_1 = 1$?
    - In the case of independent coordinates: log-concave tails are sufficient, but α-regularly growing moments are not sufficient.
2. Is there any other way to obtain (2) with $D_1 = 1$ than by integrating a concentration inequality?
3. What is the characterization of random variables satisfying convex IC with optimal cost function? (Note that Λ_X^* does not have to be quadratic near 0.)

Further questions

1. When
$$\left(\mathbb{E}\|X\|^p\right)^{1/p} \leq D_1 \mathbb{E}\|X\| + D_2 \sigma_{\|\cdot\|,X}(p) \tag{2}$$
holds with $D_1 = 1$?
   - In the case of independent coordinates: log-concave tails are sufficient, but α-regularly growing moments are not sufficient.
2. Is there any other way to obtain (2) with $D_1 = 1$ than by integrating a concentration inequality?
3. What is the characterization of random variables satisfying convex IC with optimal cost function? (Note that Λ_X^* does not have to be quadratic near 0.)
   - The previous example shows that in the case of independent coordinates: log-concave tails are sufficient, but α-regularly growing moments are not sufficient.