Optimal Concentration of Information Content for Log-Concave Densities

Matthieu Fradelizi, Mokshay Madiman, and Liyao Wang

Abstract An elementary proof is provided of sharp bounds for the varentropy of random vectors with log-concave densities, as well as for deviations of the information content from its mean. These bounds significantly improve on the bounds obtained by Bobkov and Madiman (Ann Probab 39(4):1528–1543, 2011).

Keywords Concentration • Information • Log-concave • Varentropy

Mathematics Subject Classification (2010). Primary 52A40; Secondary 60E15, 94A17

1 Introduction

Consider a random vector Z taking values in $\mathbb{R}^n$, drawn from the standard Gaussian distribution γ, whose density is given by

$$\phi(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{|x|^2}{2}}$$

for each $x \in \mathbb{R}^n$, where $|\cdot|$ denotes the Euclidean norm. It is well known that when the dimension n is large, the distribution of Z is highly concentrated around

M. Fradelizi
Laboratoire d'Analyse et de Mathématiques Appliquées UMR 8050, Université Paris-Est Marne-la-Vallée, 5 Bd Descartes, Champs-sur-Marne 77454, Marne-la-Vallée Cedex 2, France
e-mail: matthieu.fradelizi@univ-mlv.fr

M. Madiman (✉)
Department of Mathematical Sciences, University of Delaware, 501 Ewing Hall, Newark, DE 19716, USA
e-mail: madiman@udel.edu

L. Wang
J.P. Morgan, New York, NY, USA
e-mail: njuwangliyao@gmail.com

the sphere of radius $\sqrt{n}$; that $\sqrt{n}$ is the appropriate radius follows by the trivial observation that $\mathbf{E}|Z|^2 = \sum_{i=1}^{n} \mathbf{E}Z_i^2 = n$. One way to express this concentration property is by computing the variance of $|Z|^2$, which is easy to do using the independence of the coordinates of Z:

$$\text{Var}(|Z|^2) = \text{Var}\left(\sum_{i=1}^{n} Z_i^2\right) = \sum_{i=1}^{n} \text{Var}(Z_i^2) = 2n.$$

In particular, the standard deviation of $|Z|^2$ is $\sqrt{2n}$, which is much smaller than the mean n of $|Z|^2$ when n is large. Another way to express this concentration property is through a deviation inequality:

$$\mathbf{P}\left\{\frac{|Z|^2}{n} - 1 > t\right\} \leq \exp\left\{-\frac{n}{2}[t - \log(1 + t)]\right\} \tag{1.1}$$

for the upper tail, and a corresponding upper bound on the lower tail. These inequalities immediately follow from Chernoff's bound, since $|Z|^2/n$ is just the empirical mean of i.i.d. random variables.

It is natural to wonder if, like so many other facts about Gaussian measures, the above concentration property also has an extension to log-concave measures (or to some subclass of them). There are two ways one may think about extending the above concentration property. One is to ask if there is a universal constant C such that

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

for every random vector X that has an isotropic, log-concave distribution on $\mathbb{R}^n$. Here, we say that a distribution on $\mathbb{R}^n$ is isotropic if its covariance matrix is the identity matrix; this assumption ensures that $\mathbf{E}|X|^2 = n$, and provides the normalization needed to make the question meaningful. This question has been well studied in the literature, and is known as the "thin shell conjecture" in convex geometry. It is closely related to other famous conjectures: it implies the hyperplane conjecture of Bourgain [13, 14], is trivially implied by the Kannan-Lovasz-Simonovits conjecture, and also implies the Kannan-Lovasz-Simonovits conjecture up to logarithmic terms [12]. The best bounds known to date are those of Guédon and Milman [18], and assert that

$$\text{Var}(|X|^2) \leq Cn^{4/3}.$$

The second way that one may try to extend the above concentration property from Gaussians to log-concave measures is to first observe that the quantity that concentrates, namely $|Z|^2$, is essentially the logarithm of the Gaussian density

function. More precisely, since

$$-\log\phi(x) = \frac{n}{2}\log(2\pi) + \frac{|x|^2}{2},$$

the concentration of $|Z|^2$ about its mean is equivalent to the concentration of $-\log\phi(Z)$ about its mean. Thus one can ask if, for every random vector X that has a log-concave density f on $\mathbb{R}^n$,

$$\mathrm{Var}(-\log f(X)) \leq Cn \qquad (1.2)$$

for some absolute constant C. An affirmative answer to this question was provided by Bobkov and Madiman [2]. The approach of [2] can be used to obtain bounds on C, but the bounds so obtained are quite suboptimal (around 1000). Recently V.H. Nguyen [27] (see also [28]) and Wang [32] independently determined, in their respective Ph.D. theses, that the sharp constant C in the bound (1.2) is 1. Soon after this work, simpler proofs of the sharp variance bound were obtained independently by us (presented in the proof of Theorem 2.3 in this paper) and by Bolley et al. [7] (see Remark 4.2 in their paper). An advantage of our proof over the others mentioned is that it is very short and straightforward, and emerges as a consequence of a more basic log-concavity property (namely Theorem 2.9) of L^p-norms of log-concave functions, which may be thought of as an analogue for log-concave functions of a classical inequality of Borell [8] for concave functions.

If we are interested in finer control of the integrability of $-\log f(X)$, we may wish to consider analogues for general log-concave distributions of the inequality (1.1). Our second objective in this note is to provide such an analogue (in Theorem 4.1). A weak version of such a statement was announced in [3] and proved in [2], but the bounds we provide in this note are much stronger. Our approach has two key advantages: first, the proof is transparent and completely avoids the use of the sophisticated Lovasz-Simonovits localization lemma, which is a key ingredient of the approach in [2]; and second, our bounds on the moment generating function are sharp, and are attained for example when the distribution under consideration has i.i.d. exponentially distributed marginals.

While in general exponential deviation inequalities imply variance bounds, the reverse is not true. Nonetheless, our approach in this note is to first prove the variance bound (1.2), and then use a general bootstrapping result (Theorem 3.1) to deduce the exponential deviation inequalities from it. The bootstrapping result is of independent interest; it relies on a technical condition that turns out to be automatically satisfied when the distribution in question is log-concave.

Finally we note that many of the results in this note can be extended to the class of convex measures; partial work in this direction is done by Nguyen [28], and results with sharp constants are obtained in the forthcoming paper [17].

2 Optimal Varentropy Bound for Log-Concave Distributions

Before we proceed, we need to fix some definitions and notation.

Definition 2.1 Let a random vector X taking values in $\mathbb{R}^n$ have probability density function f. The *information content* of X is the random variable $\tilde{h}(X) = -\log f(X)$. The *entropy* of X is defined as $h(X) = \mathbf{E}(\tilde{h}(X))$. The *varentropy* of a random vector X is defined as $V(X) = \text{Var}(\tilde{h}(X))$.

Note that the entropy and varentropy depend not on the realization of X but only on its density f, whereas the information content does indeed depend on the realization of X. For instance, one can write $h(X) = -\int_{\mathbb{R}^n} f \log f$ and

$$V(X) = \text{Var}(\log f(X)) = \int_{\mathbb{R}^n} f(\log f)^2 - \left(\int_{\mathbb{R}^n} f \log f \right)^2.$$

Nonetheless, for reasons of convenience and in keeping with historical convention, we slightly abuse notation as above.

As observed in [2], the distribution of the difference $\tilde{h}(X) - h(X)$ is invariant under any affine transformation of $\mathbb{R}^n$ (i.e., $\tilde{h}(TX) - h(TX) = \tilde{h}(X) - h(X)$ for all invertible affine maps $T : \mathbb{R}^n \to \mathbb{R}^n$); hence the varentropy $V(X)$ is affine-invariant while the entropy $h(X)$ is not.

Another invariance for both $h(X)$ and $V(X)$ follows from the fact that they only depend on the distribution of $\log(f(X))$, so that they are unchanged if f is modified in such a way that its sublevel sets keep the same volume. This implies (see, e.g., [25, Theorem 1.13]) that if $f^\star$ is the spherically symmetric, decreasing rearrangement of f, and $X^\star$ is distributed according to the density $f^\star$, then $h(X) = h(X^\star)$ and $V(X) = V(X^\star)$. The rearrangement-invariance of entropy was a key element in the development of refined entropy power inequalities in [33].

Log-concavity is a natural shape constraint for functions (in particular, probability density functions) because it generalizes the Gaussian distributions. Furthermore, the class of log-concave distributions is infinite-dimensional, and hence, comprises a nonparametric model in statistical terms.

Definition 2.2 A function $f : \mathbb{R}^n \to [0, \infty)$ is *log-concave* if f can be written as

$$f(x) = e^{-U(x)},$$

where $U : \mathbb{R}^n \mapsto (-\infty, +\infty]$ is a convex function, i.e., $U(tx + (1-t)y) \leq tU(x) + (1-t)U(y)$ for any x, y and $0 < t < 1$. When f is a probability density function and is log-concave, we say that f is a *log-concave density*.

We can now state the optimal form of the inequality (1.2), first obtained by Nguyen [27] and Wang [32] as discussed in Sect. 1.

Theorem 2.3 ([27, 32]) *Given a random vector X in $\mathbb{R}^n$ with log-concave density f,*

$$V(X) \leq n$$

Remark 2.4 The probability bound *does not depend* on f– it is universal over the class of log-concave densities.

Remark 2.5 The bound is *sharp*. Indeed, let X have density $f = e^{-\varphi}$, with $\varphi : \mathbb{R}^n \to [0, \infty]$ being positively homogeneous of degree 1, i.e., such that $\varphi(tx) = t\varphi(x)$ for all $t > 0$ and all $x \in \mathbb{R}^n$. Then one can check that the random variable $Y = \varphi(X)$ has a gamma distribution with shape parameter n and scale parameter 1, i.e., it is distributed according to the density given by

$$f_Y(t) = \frac{t^{n-1}e^{-t}}{(n-1)!}.$$

Consequently $\mathbf{E}(Y) = n$ and $\mathbf{E}(Y^2) = n(n+1)$, and therefore $V(X) = \mathrm{Var}(Y) = n$. Particular examples of equality include:

1. The case where $\varphi(x) = \sum_{i=1}^n x_i$ on the cone of points with non-negative coordinates (which corresponds to X having i.i.d. coordinates with the standard exponential distribution), and
2. The case where $\varphi(x) = \inf\{r > 0 : x \in rK\}$ for some compact convex set K containing the origin (which, by taking K to be a symmetric convex body, includes all norms on $\mathbb{R}^n$ suitably normalized so that $e^{-\varphi}$ is a density).

Remark 2.6 Bolley et al. [7] in fact prove a stronger inequality, namely,

$$\frac{1}{V(X)} - \frac{1}{n} \geq \left[\mathbf{E}\{\nabla U(X) \cdot \mathrm{Hess}(U(X))^{-1} \nabla U(X)\} \right]^{-1}.$$

This gives a strict improvement of Theorem 2.3 when the density $f = e^{-U}$ of X is strictly log-concave, in the sense that $\mathrm{Hess}(U(X))$ is, almost surely, strictly positive definite. As noted by Bolley et al. [7], one may give another alternative proof of Theorem 2.3 by applying a result of Hargé [19, Theorem 2].

In order to present our proof of Theorem 2.3, we will need some lemmata. The first one is a straightforward computation that is a special case of a well known fact about exponential families in statistics, but we write out a proof for completeness.

Lemma 2.7 *Let f be any probability density function on $\mathbb{R}^n$ such that $f \in L^\alpha(\mathbb{R}^n)$ for each $\alpha > 0$, and define*

$$F(\alpha) = \log \int_{\mathbb{R}^n} f^\alpha.$$

Let X_α be a random variable with density f_α on $\mathbb{R}^n$, where

$$f_\alpha := \frac{f^\alpha}{\int_{\mathbb{R}^n} f^\alpha}.$$

Then F is infinitely differentiable on $(0, \infty)$, and moreover, for any $\alpha > 0$,

$$F''(\alpha) = \frac{1}{\alpha^2} V(X_\alpha).$$

Proof Note that the assumption that $f \in L^\alpha(\mathbb{R}^n)$ (or equivalently that $F(\alpha) < \infty$) for all $\alpha > 0$ guarantees that $F(\alpha)$ is infinitely differentiable for $\alpha > 0$ and that we can freely change the order of taking expectations and differentiation.

Now observe that

$$F'(\alpha) = \frac{\int f^\alpha \log f}{\int f^\alpha} = \int f_\alpha \log f;$$

if we wish, we may also massage this to write

$$F'(\alpha) = \frac{1}{\alpha}[F(\alpha) - h(X_\alpha)]. \tag{2.1}$$

Differentiating again, we get

$$
\begin{aligned}
F''(\alpha) &= \frac{\int f^\alpha (\log f)^2}{\int f^\alpha} - \left(\frac{\int f^\alpha \log f}{\int f^\alpha}\right)^2 \\
&= \int f_\alpha (\log f)^2 - \left(\int f_\alpha \log f\right)^2 \\
&= \mathrm{Var}[\log f(X_\alpha)] = \mathrm{Var}\left[\frac{1}{\alpha}\{\log f_\alpha(X_\alpha) + F(\alpha)\}\right] \\
&= \frac{1}{\alpha^2}\mathrm{Var}[\log f_\alpha(X_\alpha)] = \frac{V(X_\alpha)}{\alpha^2},
\end{aligned}
$$

as desired.   □

The following lemma is a standard fact about the so-called perspective function in convex analysis. The use of this terminology is due to Hiriart-Urruty and Lemaréchal [20, p. 160] (see [10] for additional discussion), although the notion has been used without a name in convex analysis for a long time (see, e.g., [30, p. 35]). Perspective functions have also seen recent use in convex geometry [6, 11, 17]) and empirical process theory [31]. We give the short proof for completeness.

Lemma 2.8 *If $U : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a convex function, then*

$$w(z, \alpha) := \alpha U(z/\alpha)$$

is a convex function on $\mathbb{R}^n \times (0, +\infty)$.

Proof First note that by definition, $w(az, a\alpha) = aw(z, \alpha)$ for any $a > 0$ and any $(z, \alpha) \in \mathbb{R}^n \times (0, +\infty)$, which implies in particular that

$$\frac{1}{\alpha}w(z, \alpha) = w\left(\frac{z}{\alpha}, 1\right).$$

Hence

$$w(\lambda z_1 + (1 - \lambda)z_2, \lambda\alpha_1 + (1 - \lambda)\alpha_2)$$
$$= [\lambda\alpha_1 + (1 - \lambda)\alpha_2]\, U\left(\frac{\lambda\alpha_1 \frac{z_1}{\alpha_1} + (1 - \lambda)\alpha_2 \frac{z_2}{\alpha_2}}{\lambda\alpha_1 + (1 - \lambda)\alpha_2}\right)$$
$$\leq \lambda\alpha_1 U\left(\frac{z_1}{\alpha_1}\right) + (1 - \lambda)\alpha_2 U\left(\frac{z_2}{\alpha_2}\right)$$
$$= \lambda w(z_1, \alpha_1) + (1 - \lambda)w(z_2, \alpha_2),$$

for any $\lambda \in [0, 1]$, $z_1, z_2 \in \mathbb{R}^n$, and $\alpha_1, \alpha_2 \in (0, \infty)$. □

The key observation is the following theorem.

Theorem 2.9 *If f is log-concave on $\mathbb{R}^n$, then the function*

$$G(\alpha) := \alpha^n \int f(x)^\alpha dx$$

is log-concave on $(0, +\infty)$.

Proof Write $f = e^{-U}$, with U convex. Make the change of variable $x = z/\alpha$ to get

$$G(\alpha) = \int e^{-\alpha U(z/\alpha)} dz.$$

The function $w(z, \alpha) := \alpha U(z/\alpha)$ is convex on $\mathbb{R}^n \times (0, +\infty)$ by Lemma 2.8, which means that the integrand above is log-concave. The log-concavity of G then follows from Prékopa's theorem [29], which implies that marginals of log-concave functions are log-concave. □

Remark 2.10 An old theorem of Borell [8, Theorem 2] states that if f is concave on $\mathbb{R}^n$, then $G_f(p) := (p + 1)\cdots(p + n)\int f^p$ is log-concave as a function of $p \in (0, \infty)$. Using this and the fact that a log-concave function is a limit of α-concave functions with $\alpha \to 0$, one can obtain an alternate, indirect proof of Theorem 2.9.

One can also similarly obtain an indirect proof of Theorem 2.9 by considering a limiting version of [4, Theorem VII.2], which expresses a log-concavity property of $(p-1)\dots(p-n)\int\phi^{-p}$ for any convex function ϕ on $\mathbb{R}^n$, for $p > n+1$ (an improvement of this to the optimal range $p > n$ is described in [6, 17], although this is not required for this alternate proof of Theorem 2.9).

Proof of Theorem 2.3 Since f is a log-concave density, it necessarily holds that $f \in L^\alpha(\mathbb{R}^n)$ for every $\alpha > 0$; in particular, $G(\alpha) := \alpha^n \int f^\alpha$ is finite and infinitely differentiable on the domain $(0, \infty)$. By definition,

$$\log G(\alpha) = n \log \alpha + \log \int f^\alpha = n \log \alpha + F(\alpha).$$

Consequently,

$$\frac{d^2}{d\alpha^2}[\log G(\alpha)] = -\frac{n}{\alpha^2} + F''(\alpha).$$

By Theorem 2.9, $\log G(\alpha)$ is concave, and hence we must have that

$$-\frac{n}{\alpha^2} + F''(\alpha) \leq 0$$

for each $\alpha > 0$. However, Lemma 2.7 implies that $F''(\alpha) = V(X_\alpha)/\alpha^2$, so that we obtain the inequality

$$\frac{V(X_\alpha) - n}{\alpha^2} \leq 0.$$

For $\alpha = 1$, this implies that $V(X) \leq n$.

Notice that if $f = e^{-U}$, where $U : \mathbb{R}^n \to [0, \infty]$ is positively homogeneous of degree 1, then the same change of variable as in the proof of Theorem 2.9 shows that

$$G(\alpha) = \int e^{-\alpha U(z/\alpha)} dz = \int e^{-U(z)} dz = \int f(z) dz = 1.$$

Hence the function G is constant. Then the proof above shows that $V(X) = n$, which establishes the equality case stated in Remark 2.5.

3   A General Bootstrapping Strategy

The purpose of this section is to describe a strategy for obtaining exponential deviation inequalities when one has uniform control on variances of a family of random variables. Log-concavity is not an assumption made anywhere in this section.

Theorem 3.1 *Suppose $X \sim f$, where $f \in L^{\alpha}(\mathbb{R}^n)$ for each $\alpha > 0$. Let $X_{\alpha} \sim f_{\alpha}$, where*

$$f_{\alpha}(x) = \frac{f^{\alpha}(x)}{\int f^{\alpha}}.$$

If $K = K(f) := \sup_{\alpha>0} V(X_{\alpha})$, then

$$\mathbf{E}\left[e^{\beta\{\tilde{h}(X)-h(X)\}}\right] \leq e^{Kr(-\beta)}, \quad \beta \in \mathbb{R},$$

where

$$r(u) = \begin{cases} u - \log(1+u) & \text{for } u > -1 \\ +\infty & \text{for } u \leq -1 \end{cases}.$$

Proof Suppose X is a random vector drawn from a density f on $\mathbb{R}^n$, and define, for each $\alpha > 0$, $F(\alpha) = \log \int f^{\alpha}$. Set

$$K = \sup_{\alpha>0} V(X_{\alpha}) = \sup_{\alpha>0} \alpha^2 F''(\alpha);$$

the second equality follows from Lemma 2.7. Since $f \in L^{\alpha}(\mathbb{R}^n)$ for each $\alpha > 0$, $F(\alpha)$ is finite and moreover, infinitely differentiable for $\alpha > 0$, and we can freely change the order of integration and differentiation when differentiating $F(\alpha)$.

From Taylor-Lagrange formula, for every $\alpha > 0$, one has

$$F(\alpha) = F(1) + (\alpha - 1)F'(1) + \int_1^{\alpha} (\alpha - u)F''(u)du.$$

Using that $F(1) = 0$, $F''(u) \leq K/u^2$ for every $u > 0$ and the fact that for $0 < \alpha < u < 1$, one has $\alpha - u < 0$, we get

$$F(\alpha) \leq (\alpha - 1)F'(1) + K \int_1^{\alpha} \frac{\alpha - u}{u^2} du$$

$$= (\alpha - 1)F'(1) + K \left[-\frac{\alpha}{u} - \log(u)\right]_1^{\alpha}.$$

Thus, for $\alpha > 0$, we have proved that

$$F(\alpha) \leq (\alpha - 1)F'(1) + K(\alpha - 1 - \log \alpha).$$

Setting $\beta = 1 - \alpha$, we have for $\beta < 1$ that

$$e^{F(1-\beta)} \leq e^{-\beta F'(1)} e^{K(-\beta - \log(1-\beta))}. \tag{3.1}$$

Observe that $e^{F(1-\beta)} = \int f^{1-\beta} = \mathbf{E}[f^{-\beta}(X)] = \mathbf{E}[e^{-\beta \log f(X)}] = \mathbf{E}\left[e^{\beta \tilde{h}(X)}\right]$ and $e^{-\beta F'(1)} = e^{\beta h(X)}$; the latter fact follows from the fact that $F'(1) = -h(X)$ as is clear from the identity (2.1). Hence the inequality (3.1) may be rewritten as

$$\mathbf{E}\left[e^{\beta\{\tilde{h}(X)-h(X)\}}\right] \leq e^{Kr(-\beta)}, \quad \beta \in \mathbb{R}. \tag{3.2}$$

$\square$

Remark 3.2 We note that the function $r(t) = t - \log(1 + t)$ for $t > -1$, (or the related function $h(t) = t \log t - t + 1$ for $t > 0$, which satisfies $sh(t/s) = tr_1(s/t)$ for $r_1(u) = r(u - 1)$) appears in many exponential concentration inequalities in the literature, including Bennett's inequality [1] (see also [9]), and empirical process theory [34]. It would be nice to have a clearer understanding of why these functions appear in so many related contexts even though the specific circumstances vary quite a bit.

Remark 3.3 Note that the function r is convex on $\mathbb{R}$ and has a quadratic behavior in the neighborhood of 0 ($r(u) \sim_0 \frac{u^2}{2}$) and a linear behavior at $+\infty$ ($r(u) \sim_\infty u$).

Corollary 3.4 *With the assumptions and notation of Theorem 3.1, we have for any $t > 0$ that*

$$\mathbf{P}\{\tilde{h}(X) - h(X) \geq t\} \leq \exp\left\{-Kr\left(\frac{t}{K}\right)\right\}$$

$$\mathbf{P}\{\tilde{h}(X) - h(X) \leq -t\} \leq \exp\left\{-Kr\left(-\frac{t}{K}\right)\right\}$$

The proof is classical and often called the Cramér-Chernoff method (see for example Sect. 2.2 in [9]). It uses the Legendre transform φ^* of a convex function $\varphi : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ defined for $y \in \mathbb{R}$ by

$$\varphi^*(y) = \sup_x xy - \varphi(x).$$

Notice that if $\min \varphi = \varphi(0)$ then for every $y > 0$, the supremum is reached at a positive x, that is $\varphi^*(y) = \sup_{x>0} xy - \varphi(x)$. Similarly, for $y < 0$, the supremum is reached at a negative x.

Proof The idea is simply to use Markov's inequality in conjunction with Theorem 3.1, and optimize the resulting bound.

For the lower tail, we have for $\beta > 0$ and $t > 0$,

$$\mathbb{P}[\tilde{h}(X) - h(X) \leq -t] \leq \mathbf{E}\left[e^{-\beta\left(\tilde{h}(X)-h(X)\right)}\right]e^{-\beta t}$$

$$\leq \exp\left\{K\left(r(\beta) - \frac{\beta t}{K}\right)\right\}.$$

Thus minimizing on $\beta > 0$, and using the remark before the proof, we get

$$\mathbb{P}[\tilde{h}(X) - h(X) \le -t] \le \exp\left\{ -K \sup_{\beta > 0} \left(\frac{\beta t}{K} - r(\beta) \right) \right\} = e^{-K r^*\left(\frac{t}{K}\right)}. \quad (3.3)$$

Let us compute the Legendre transform r^* of r. For every t, one has

$$r^*(t) = \sup_u tu - r(u) = \sup_{u > -1} (tu - u + \log(1 + u)).$$

One deduces that $r^*(t) = +\infty$ for $t \ge 1$. For $t < 1$, by differentiating, the supremum is reached at $u = t/(1-t)$ and replacing in the definition we get

$$r^*(t) = -t - \log(1 - t) = r(-t).$$

Thus $r^*(t) = r(-t)$ for all $t \in \mathbb{R}$. Replacing, in the inequality (3.3), we get the result for the lower tail.

For the upper tail, we use the same argument: for $\beta > 0$ and $t > 0$,

$$\mathbb{P}[\tilde{h}(X) - h(X) \ge t] \le \mathbf{E}\left[e^{\beta\left(\tilde{h}(X) - h(X)\right)} \right] e^{-\beta t}$$

$$\le \exp\left\{ K\left(r(-\beta) - \frac{\beta t}{K} \right) \right\}.$$

Thus minimizing on $\beta > 0$, we get

$$\mathbb{P}[\tilde{h}(X) - h(X) \ge t] \le \exp\left\{ -K \sup_{\beta > 0} \left(\frac{\beta t}{K} - r(-\beta) \right) \right\}. \quad (3.4)$$

Using the remark before the proof, in the right hand side term appears the Legendre transform of the function $\tilde{r}$ defined by $\tilde{r}(u) = r(-u)$. Using that $r^*(t) = r(-t) = \tilde{r}(t)$, we deduce that $(\tilde{r})^* = (r^*)^* = r$. Thus the inequality (3.4) gives the result for the upper tail.

$\square$

4   Conclusion

The purpose of this section is to combine the results of Sects. 2 and 3 to deduce sharp bounds for the moment generating function of the information content of random vectors with log-concave densities. Naturally these yield good bounds on the deviation probability of the information content $\tilde{h}(X)$ from its mean $h(X) = \mathbf{E}\tilde{h}(X)$. We also take the opportunity to record some other easy consequences.

Theorem 4.1 *Let X be a random vector in $\mathbb{R}^n$ with a log-concave density f. For $\beta < 1$,*

$$\mathbf{E}\left[e^{\beta[\tilde{h}(X)-h(X)]}\right] \leq \mathbf{E}\left[e^{\beta[\tilde{h}(X^*)-h(X^*)]}\right],$$

where X^* has density $f^* = e^{-\sum_{i=1}^n x_i}$, restricted to the positive quadrant.

Proof Taking $K = n$ in Theorem 3.1 (which we can do in the log-concave setting because of Theorem 2.3), we obtain:

$$\mathbf{E}\left[e^{\beta\{\tilde{h}(X)-h(X)\}}\right] \leq e^{nr(-\beta)}, \quad \beta \in \mathbb{R}.$$

Some easy computations will show:

$$\mathbf{E}\left[e^{\beta\{\tilde{h}(X^*)-h(X^*)\}}\right] = e^{nr(-\beta)}, \quad \beta \in \mathbb{R}.$$

This concludes the proof.

$\square$

As for the case of equality of Theorem 2.3, discussed in Remark 2.5, notice that there is a broader class of densities for which one has equality in Theorem 4.1, including all those of the form $e^{-\|x\|_K}$, where K is a symmetric convex body.

Remark 4.2 The assumption $\beta < 1$ in Theorem 4.1 is strictly not required; however, for $\beta \geq 1$, the right side is equal to $+\infty$. Indeed, already for $\beta = 1$, one sees that for any random vector X with density f,

$$\mathbf{E}\left[e^{\tilde{h}(X)-h(X)}\right] = e^{-h(X)}\mathbf{E}\left[\frac{1}{f(X)}\right] = e^{-h(X)}\int_{\text{supp(f)}} dx$$
$$= e^{-h(X)}\text{Vol}_n(\text{supp(f)}),$$

where $\text{supp(f)} = \overline{\{x \in \mathbb{R}^n : f(x) > 0\}}$ is the support of the density f and Vol_n denotes Lebesgue measure on $\mathbb{R}^n$. In particular, this quantity for X^*, whose support has infinite Lebesgue measure, is $+\infty$.

Remark 4.3 Since

$$\lim_{\alpha \to 0} \frac{2}{\alpha^2}\mathbf{E}\left[e^{\alpha(\log f(X)-\mathbf{E}[\log f(X)])}\right] = V(X),$$

we can recover Theorem 2.3 from Theorem 4.1.

Taking $K = n$ in Corollary 3.4 (again because of Theorem 2.3), we obtain:

Corollary 4.4 *Let X be a random vector in $\mathbb{R}^n$ with a log-concave density f. For $t > 0$,*

$$\mathbb{P}[\tilde{h}(X) - h(X) \leq -nt] \leq e^{-nr(-t)},$$
$$\mathbb{P}[\tilde{h}(X) - h(X) \geq nt] \leq e^{-nr(t)},$$

where $r(u)$ is defined in Theorem 3.1.

The original concentration of information bounds obtained in [2] were suboptimal not just in terms of constants but also in the exponent; specifically it was proved there that

$$\mathbf{P}\left\{ \frac{1}{n} \left| \tilde{h}(X) - h(X) \right| \geq t \right\} \leq 2\, e^{-ct\sqrt{n}} \tag{4.1}$$

for a universal constant $c > 1/16$ (and also that a better bound with $ct^2 n$ in the exponent holds on a bounded range, say, for $t \in (0, 2]$). One key advantage of the method presented in this paper, apart from its utter simplicity, is the correct linear dependence of the exponent on dimension. Incidentally, we learnt from a lecture of Klartag [22] that another proof of (4.1) can be given based on the concentration property of the eigenvalues of the Hessian of the Brenier map (corresponding to optimal transportation from one log-concave density to another) that was discovered by Klartag and Kolesnikov [23]; however, the latter proof shares the suboptimal $\sqrt{n}t$ exponent of [2].

The following inequality is an immediate corollary of Corollary 4.4 since it merely expresses a bound on the support of the distribution of the information content.

Corollary 4.5 *Let X have a log-concave probability density function f on $\mathbb{R}^n$. Then:*

$$h(X) \leq -\log \|f\|_\infty + n.$$

Proof By Corollary 4.4, almost surely,

$$\log f(X) \leq \mathbf{E}[\log f(X)] + n,$$

since when $t \geq 1$, $\mathbb{P}[\log f(X) - \mathbf{E}[\log f(X)] \geq nt] = 0$. Taking the supremum over all realizable values of X yields

$$\log \|f\|_\infty \leq \mathbf{E}[\log f(X)] + n,$$

which is equivalent to the desired statement. □

Corollary 4.5 was first explicitly proved in [4], where several applications of it are developed, but it is also implicitly contained in earlier work (see, e.g., the proof of Theorem 7 in [16]).

An immediate consequence of Corollary 4.5, unmentioned in [4], is a result due to [15]:

Corollary 4.6 *Let X be a random vector in $\mathbb{R}^n$ with a log-concave density f. Then*

$$\|f\|_\infty \leq e^n f(\mathbf{E}[X]).$$

Proof By Jensen's inequality,

$$\log f(\mathbf{E}X) \geq \mathbf{E}[\log f(X)].$$

By Corollary 4.5,

$$\mathbf{E}[\log f(X)] \geq \log \|f\|_\infty - n.$$

Hence,

$$\log f(\mathbf{E}X) \geq \log \|f\|_\infty - n.$$

Exponentiating concludes the proof. $\square$

Finally we mention that the main result may also be interpreted as a small ball inequality for the random variable $f(X)$. As an illustration, we record a sharp form of [24, Corollary 2.4] (cf., [21, Corollary 5.1] and [5, Proposition 5.1]).

Corollary 4.7 *Let f be a log-concave density on $\mathbb{R}^n$. Then*

$$\mathbb{P}\{f(X) \geq c^n \|f\|_\infty\} \geq 1 - \left(e \cdot c \cdot \log\left(\frac{1}{c}\right)\right)^n,$$

where $0 < c < \frac{1}{e}$.

Proof Note that

$$\begin{aligned}
\mathbb{P}\{f(X) \leq c^n \|f\|_\infty\} &= \mathbb{P}\{\log f(X) \leq \log \|f\|_\infty + n \log c\} \\
&= \mathbb{P}\{\tilde{h}(X) \geq -\log \|f\|_\infty - n \log c\} \\
&\leq \mathbb{P}\{\tilde{h}(X) \geq h(X) - n(1 + \log c)\}.
\end{aligned}$$

using Corollary 4.5 for the last inequality. Applying Corollary 4.4 with $t = -\log c - 1$ yields

$$\mathbb{P}\{f(X) \leq c^n \|f\|_\infty\} \leq e^{-nr(-1-\log c)}.$$

Elementary algebra concludes the proof. $\square$

Such "effective support" results are useful in convex geometry as they can allow to reduce certain statements about log-concave functions or measures to statements about convex sets; they thus provide an efficient route to proving functional or probabilistic analogues of known results in the geometry of convex sets. Instances where such a strategy is used include [5, 24]. These and other applications of the concentration of information phenomenon are discussed in [26].

References

1. G. Bennett, Probability inequalities for the sum of independent random variables. J. Am. Stat. Assoc. **57**(297), 33–45 (1962)
2. S. Bobkov, M. Madiman, Concentration of the information in data with log-concave distributions. Ann. Probab. **39**(4), 1528–1543 (2011)
3. S. Bobkov, M. Madiman, Dimensional behaviour of entropy and information. C. R. Acad. Sci. Paris Sér. I Math. **349**, 201–204 Février (2011)
4. S. Bobkov, M. Madiman, The entropy per coordinate of a random vector is highly constrained under convexity conditions. IEEE Trans. Inform. Theory **57**(8), 4940–4954 (2011)
5. S. Bobkov, M. Madiman, Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures. J. Funct. Anal. **262**, 3309–3339 (2012)
6. S. Bobkov, M. Fradelizi, J. Li, M. Madiman, When can one invert Hölder's inequality? (and why one may want to). Preprint (2016)
7. F. Bolley, I. Gentil, A. Guillin, Dimensional improvements of the logarithmic Sobolev, Talagrand and Brascamp-Lieb inequalities. Preprint, arXiv:1507:01086 (2015)
8. C. Borell, Complements of Lyapunov's inequality. Math. Ann. **205**, 323–331 (1973)
9. S. Boucheron, G. Lugosi, P. Massart, *Concentration Inequalities* (Oxford University Press, Oxford, 2013). A nonasymptotic theory of independence, With a foreword by Michel Ledoux
10. S. Boyd, L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004)
11. D. Cordero-Erausquin, M. Fradelizi, G. Paouris, P. Pivovarov, Volume of the polar of random sets and shadow systems. Math. Ann. **362**, 1305–1325 (2014)
12. R. Eldan, Thin shell implies spectral gap up to polylog via a stochastic localization scheme. Geom. Funct. Anal. **23**(2), 532–569 (2013)
13. R. Eldan, B. Klartag, Approximately Gaussian marginals and the hyperplane conjecture, in *Concentration, Functional Inequalities and Isoperimetry*, ed. by C. Houdré, M. Ledoux, E. Milman, M. Milman. Contemporary Mathematics, vol. 545 (American Mathematical Society, Providence, RI, 2011), pp. 55–68
14. R. Eldan, J. Lehec, Bounding the norm of a log-concave vector via thin-shell estimates, in *Geometric Aspects of Functional Analysis*, ed. by B. Klartag, E. Milman. Lecture Notes in Mathematics, vol. 2116 (Springer, Berlin, 2014), pp. 107–122

15. M. Fradelizi, Sections of convex bodies through their centroid. Arch. Math. (Basel) **69**(6), 515–522 (1997)
16. M. Fradelizi, M. Meyer, Increasing functions and inverse Santaló inequality for unconditional functions. Positivity **12**(3), 407–420 (2008)
17. J. Li, M. Fradelizi, M. Madiman, Concentration of information content and other functionals under convex measures, in *Proceedings of the 2016 IEEE International Symposium on Information Theory* (Barcelona, 2016) pp. 1128–1132
18. O. Guédon, E. Milman, Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. Geom. Funct. Anal. **21**(5), 1043–1068 (2011)
19. G. Hargé, Reinforcement of an inequality due to Brascamp and Lieb. J. Funct. Anal. **254**(2), 267–300 (2008)
20. J.-B. Hiriart-Urruty, C. Lemaréchal, *Convex Analysis and Minimization Algorithms. I.* Fundamentals. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 305 (Springer-Verlag, Berlin, 1993)
21. B. Klartag, A central limit theorem for convex sets. Invent. Math. **168**(1), 91–131 (2007)
22. B. Klartag, Eigenvalue distribution of optimal transportation. in *Presentation given at the Workshop on Information Theory and Concentration Phenomena held at the Institute for Mathematics and its Applications (IMA)*, Minneapolis, April (2015)
23. B. Klartag, A. Kolesnikov, Eigenvalue distribution of optimal transportation. Anal. PDE **8**(1), 33–55 (2015)
24. B. Klartag, V.D. Milman, Geometry of log-concave functions and measures. Geom. Dedicata **112**, 169–182 (2005)
25. E.H. Lieb, M. Loss, Analysis, 2nd edn. Graduate Studies in Mathematics, vol. 14 (American Mathematical Society, Providence, RI, 2001)
26. M. Madiman, L. Wang, S. Bobkov, Some applications of the nonasymptotic equipartition property of log-concave distributions. Preprint (2016)
27. V.H. Nguyen, Inégalités fonctionnelles et convexité. Ph.D. thesis, Université Pierre et Marie Curie (Paris VI) (2013)
28. V.H. Nguyen, Dimensional variance inequalities of Brascamp-Lieb type and a local approach to dimensional Prékopa's theorem. J. Funct. Anal. **266**(2), 931–955 (2014)
29. A. Prékopa, On logarithmic concave measures and functions. Acta Sci. Math. (Szeged) **34**, 335–343 (1973)
30. R.T. Rockafellar, *Convex Analysis*. Princeton Mathematical Series, vol. 28 (Princeton University Press, Princeton, NJ, 1970)
31. A. van der Vaart, J.A. Wellner, A local maximal inequality under uniform entropy. Electron. J. Stat. **5**, 192–203 (2011)
32. L. Wang, Heat capacity bound, energy fluctuations and convexity. Ph.D. thesis, Yale University (2014)
33. L. Wang, M. Madiman, Beyond the entropy power inequality, via rearrangements. IEEE Trans. Inf. Theory **60**(9), 5116–5137 (2014)
34. J.A. Wellner, Limit theorems for the ratio of the empirical distribution function to the true distribution function. Z. Wahrsch. Verw. Gebiete **45**(1), 73–88 (1978)