Information concentration for convex measures

Jiange Li, Matthieu Fradelizi, Mokshay Madiman

HAL Id: hal-01590229

https://hal-upec-upem.archives-ouvertes.fr/hal-01590229

Submitted on 19 Sep 2017

Information concentration for convex measures

Jiange Li[†], Matthieu Fradelizi[*], and Mokshay Madiman[†]

[*] Université Paris-Est, LAMA (UMR 8050), CNRS, UPEM, UPEC, F-77454, Marne-la-Vallée, France.
E-mail: matthieu.fradelizi@u-pem.fr
[†]Department of Mathematical Sciences, University of Delaware, Newark, DE 19716, USA.
E-mail: lijiange@udel.edu,
madiman@udel.edu

Abstract—To be considered for the 2016 IEEE Jack Keil Wolf ISIT Student Paper Award. Sharp exponential deviation estimates of the information content as well as a sharp bound on the varentropy are obtained for convex probability measures on Euclidean spaces. These provide, in a sense, a nonasymptotic equipartition property for convex measures even in the absence of stationarity-type assumptions.

I. INTRODUCTION

Let $\mathbb{X} = (X_1, X_2, \cdots)$ be a stochastic process with each X_i taking values in $\mathbb{R}$. Suppose that the joint distribution of $X^n = (X_1, \cdots, X_n)$ has a density f with respect to the Lebesgue measure on $\mathbb{R}^n$. We are interested in the random variable

$$\widetilde{h}(X^n) = -\log f(X^n), \tag{1}$$

which may be thought of as the (random) information content of X^n. In the discrete case, the quantity $\widetilde{h}(X^n)$ (using f for the probability mass function) is essentially the number of bits needed to represent X^n by a coding scheme that minimizes the average code length [19]. In the continuous case, one may still call $\widetilde{h}(X^n)$ the information content even though this coding interpretation no longer holds. The quantity $\widetilde{h}(X^n)$ is of central interest in information theory and also naturally arises in several other areas such as probability theory, statistical physics and statistics.

The average value of information content is known more commonly as the (differential) entropy. The entropy of X^n is defined by

$$h(X^n) = -\int_{\mathbb{R}^n} f(x) \log f(x) dx = \mathbb{E}\widetilde{h}(X^n). \tag{2}$$

A typical problem is to study the deviation estimate of information content from the entropy. The entropy rate of the stochastic process $\mathbb{X}$ is defined by

$$h(\mathbb{X}) = \lim_{n \to \infty} \frac{h(X^n)}{n}, \tag{3}$$

when the limit exists. The question of whether the information content per coordinate $\frac{\widetilde{h}(X^n)}{n}$ converges to the entropy rate has been extensively studied. In the discrete case, the affirmative answer goes back to Shannon [19], McMillan [16] and Breiman [10] for stationary ergodic process. The theorem particularly implies the existence of a set of roughly $e^{nh(\mathbb{X})}$ typical sequences of length n all having roughly equal probability (a fact that plays a fundamental role in compressing discrete data from ergodic sources). McMillan [16] called this the asymptotic equipartition property (AEP). Extensions to more general (including continuous) settings were obtained independently in [1] and [18].

For processes that are not asymptotically mean stationary, the entropy rate typically does not exist; so there is no convergence question of $\frac{\widetilde{h}(X^n)}{n}$. With a global restriction on the joint distribution of the process, namely log-concavity, [3] proved that $\frac{\widetilde{h}(X^n)}{n}$ is highly concentrated around the entropy rate. It demonstrates that high-dimensional log-concave measures are in a sense close to uniform distributions on the annulus between two nested convex sets. Their argument is non-trivial and depends on the rather heavy machinery of the so-called Lovász-Simonovits localization technique; optimal concentration bounds were recently obtained in [13] using a particularly simple approach.

The purpose of this note is to extend the concentration property of the information content from log-concave measures to the more general class of "convex measures". It should not be surprising that in general we do not have AEP for convex measures, since this property tends to require exponential tails and convex measures generally posses only polynomial decay. But we can still say a lot of interesting things of convex measures in fixed dimensions, and the study is closely related to many interesting problems in convex geometry, information theory and probability theory as well. For example, Corollary IV.7 for log-concave measures was used in [4] to give an information-theoretic formulation of Bourgain's famous hyperplane conjecture [9]; a weaker form of Corollary IV.7 is a key ingredients in [5] on the reverse entropy power inequality for convex measures; and the concentration property of information content for log-concave measures [3] also has applications in random matrix theory as outlined in [15]. Perhaps most importantly from the information theory point of view, we expect our results to have implications for the study of fundamental limits of finite-blocklength performance in contexts involving convexity, e.g., for additive noise channels where the noise is drawn from a convex measure. Let us note that a very special case of our results, namely information concentration for Gaussians (which can be proved by explicit computations), is a key ingredient in the results of Cover and Pombra [11] on feedback capacities of Gaussian channels.

The note is organized as follows. In Section 2, we show

that the exponential deviation of a functional follows from the log-concavity of normalized moments of that functional. The log-concavity of the normalized moments of s-concave functions are studied in Section 3. Optimal concentration and sharp variance bound of the information content of κ-concave random variables are obtained in Section 4 by combining the results of the preceding two sections.

II. A GENERAL PRINCIPLE FOR EXPONENTIAL DEVIATION

Let X be a random variable taking values in $\mathbb{R}^n$. Suppose that it has density f with respect to the Lebesgue measure on $\mathbb{R}^n$. Let $\varphi : \mathbb{R}^n \to \mathbb{R}$ be real-valued function. One natural way to show the exponential deviation of $\varphi(X)$ from its mean is to prove the finiteness of the moment generating function $\mathbb{E}e^{\alpha\varphi(X)}$ for certain α. The logarithmic moment generating function $L(\alpha)$ is defined by

$$L(\alpha) = \log \mathbb{E}e^{\alpha\varphi(X)}. \tag{4}$$

The following observation is a well known fact about exponential families in statistics.

Lemma II.1. *Let $a,b > 0$, and suppose $L(\alpha) < \infty$ for $\alpha \in (-a,b)$. Then we have*

$$L'(\alpha) = \mathbb{E}\varphi(X_\alpha), \quad L''(\alpha) = \mathrm{Var}(\varphi(X_\alpha)), \tag{5}$$

where X_α is a random variable with density

$$f_\alpha(x) = \frac{e^{\alpha\varphi(x)}f(x)}{\int_{\mathbb{R}^n} e^{\alpha\varphi(x)}f(x)dx}. \tag{6}$$

For $\alpha = 0$, one particularly has $L'(0) = \mathbb{E}\varphi(X)$ and $L''(0) = \mathrm{Var}(\varphi(X))$.

Definition II.2. *A function $f : \mathbb{R}^n \to \mathbb{R}_+$ is called log-concave if we have*

$$f((1-\lambda)x + \lambda y) \geq f(x)^{1-\lambda}f(y)^\lambda \tag{7}$$

for all $x,y \in \mathbb{R}^n$ and all $\lambda \in [0,1]$.

The following lemma tells us that the upper bound of $\mathbb{E}e^{\alpha\varphi(X)}$ emerges as a consequence of the log-concavity of $L(\alpha)$ after certain normalization.

Lemma II.3. *Let $c(\alpha)$ be a smooth function such that $e^{-c(\alpha)}\mathbb{E}e^{\alpha\varphi(X)}$ is log-concave for $-a < \alpha < b$. Then we have*

$$\mathbb{E}e^{\alpha(\varphi(X)-\mathbb{E}\varphi(X))} \leq e^{\psi_c(\alpha)}, \tag{8}$$

where $\psi_c(\alpha) = c(\alpha) - c(0) - c'(0)\alpha$.

Proof. Since $e^{-c(s)}\mathbb{E}e^{s\varphi(X)}$ is log-concave, we have $L''(s) \leq c''(s)$. For any $0 < t < \alpha < b$, integrating the inequality over $(0,t)$ we have

$$L'(t) - L'(0) \leq c'(t) - c'(0).$$

Integrating both sides over $(0,\alpha)$, we have

$$L(\alpha) - L(0) - L'(0)\alpha \leq c(\alpha) - c(0) - c'(0)\alpha. \tag{9}$$

Similarly we can show that the estimate also holds for $-a < \alpha < 0$. Notice that $L(0) = 0$ and $L'(0) = \mathbb{E}\varphi(X)$. Then the lemma follows from exponentiating both sides of (9). $\square$

Remark II.4. *From Lemma II.1 and Lemma II.3, we can see that the study of upper bound of $\mathrm{Var}(\varphi(X_\alpha))$ is equivalent to that of the normalizing function for $\mathbb{E}e^{\alpha\varphi(X)}$ to be log-concave. We can get one from the other by differentiating or integrating twice. That is why variance bounds can imply exponential deviation inequalities when moment generating functions exist.*

Let $f : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ be a real-valued function. For $x \in \mathbb{R}$, the Fenchel-Legendre dual function $f^*(x)$ is defined by

$$f^*(x) = \sup_y (xy - f(y)). \tag{10}$$

Let $\psi_{c,+}(\alpha)$ and $\psi_{c,-}(\alpha)$ be the restrictions of $\psi_c(\alpha)$ on the positive and negative half axis, respectively.

Corollary II.5. *Under the assumptions and notations of Lemma II.3, we have for any $t > 0$ that*

$$\mathbb{P}(\varphi(X) - \mathbb{E}\varphi(X) > t) \leq e^{-\psi_{c,+}^*(t)}, \tag{11}$$

$$\mathbb{P}(\varphi(X) - \mathbb{E}\varphi(X) < -t) \leq e^{-\psi_{c,-}^*(-t)}, \tag{12}$$

where $\psi_{c,+}^*$, $\psi_{c,-}^*$ are Fenchel-Legendre dual functions of $\psi_{c,+}$, $\psi_{c,-}$, respectively.

Proof. The proof follows from the so-called Cramér-Chernoff method: using Markov inequality in conjugation with optimization of the resulting bound. For the upper tail, we have for $0 < \alpha < b$ and $t > 0$ that

$$\begin{aligned}
\mathbb{P}(\varphi(X) - \mathbb{E}(\varphi(X)) > t) &= \mathbb{P}(e^{\alpha(\varphi(X)-\mathbb{E}\varphi(X))} > e^{\alpha t}) \\
&\leq e^{-\alpha t} \cdot \mathbb{E}e^{\alpha(\varphi(X)-\mathbb{E}\varphi(X))} \\
&\leq e^{-(\alpha t - \psi_{c,+}(\alpha))}.
\end{aligned}$$

We use Lemma II.3 in the second inequality. Then the upper tail estimate follows by taking the infimum of the right hand side over $0 < \alpha < b$. The lower tail estimate follows from the same argument for $-a < \alpha < 0$. $\square$

III. RÉNYI ENTROPIES OF s-CONCAVE FUNCTIONS

In this section, we study the log-concavity of the (normalized) moments of s-concave functions, which, in conjugation with the results from the previous section, will enable us to obtain optimal concentration of the information content for convex measures.

Definition III.1. *For $s \in \mathbb{R}$, a function $f : \mathbb{R}^n \to \mathbb{R}_+$ is called s-concave if we have*

$$f((1-\lambda)x + \lambda y) \geq ((1-\lambda)f(x)^s + \lambda f(y)^s)^{1/s} \tag{13}$$

for all x,y such that $f(x)f(y) > 0$ and for all $\lambda \in [0,1]$.

For $s = 0$, the right hand side is defined by continuity, which corresponds to log-concave functions defined before. For $s > 0$, the previous definition is equivalent to that f^s is concave on its support; while for $s < 0$, it is equivalent to that f^s is convex on its support.

Recall that for $x > 0$, the gamma function $\Gamma(x)$ is defined by

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

For $x, y > 0$, the beta function $B(x, y)$ is defined by

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt.$$

The following result is proved by Borell [6] for $s > 0$, except that the function φ is assumed to be decreasing. It was then noticed by some people and available for example in Guédon, Nayar and Tkocz [14] that the result remains true without any monotonicity hypothesis. For $s < 0$, it is proved by Fradelizi, Guédon and Pajor [12], and the case $s = 0$ follows by taking the limits (or reproducing the mechanics of the proof).

Proposition III.2. *Let $s \in \mathbb{R}$ and let $\varphi : [0, +\infty) \to [0, +\infty)$ be an s-concave integrable function.*
1) If $s > 0$, then $p \mapsto B(p, s^{-1} + 1)^{-1} \int_0^{+\infty} t^{p-1} \varphi(t) dt$ is log-concave for $p > 0$.
2) If $s = 0$, then $p \mapsto \Gamma(p)^{-1} \int_0^{+\infty} t^{p-1} \varphi(t) dt$ is log-concave for $p > 0$.
3) If $s < 0$, then $p \mapsto B(p, -s^{-1} - p)^{-1} \int_0^{+\infty} t^{p-1} \varphi(t) dt$ is log-concave for $0 < p < -1/s$.

Let us define the function $\varphi_s(t) = (1-st)_+^{1/s} \mathbf{1}_{\mathbb{R}_+}$ for $s \neq 0$, and $\varphi_0(t) = e^{-t} \mathbf{1}_{\mathbb{R}_+}$. Then the preceding proposition may be expressed in the following way: if $\varphi : [0, +\infty) \to [0, +\infty)$ is s-concave, then the function

$$p \mapsto \frac{\int_0^{+\infty} t^{p-1} \varphi(t) dt}{\int_0^{+\infty} t^{p-1} \varphi_s(t) dt}$$

is log-concave for p such that $1/p > \max(0, -s)$. Using the preceding proposition, we can prove the following theorem which unifies and partially extends previous results of [6], [4], and [13]. A weaker log-concavity statement was also obtained in [17].

Theorem III.1. *Let $s \in \mathbb{R}$ and let $f : \mathbb{R}^n \to \mathbb{R}_+$ be an integrable s-concave function. Then the function*

$$p \mapsto (p+s) \cdots (p+ns) \int_{\mathbb{R}^n} f(x)^p dx \qquad (14)$$

is log-concave for $p > \max(0, -ns)$.

Proof. The case $s = 1$ is due to Borell [6] and the case $s > 0$ deduces directly by applying Borell's result to f^s. The case $s = 0$ was proved by [13]. The case $s = -1$ is due to [4][1], except that the range was $p > n+1$. In the same way, the case $s < 0$ deduces from the case $s = -1$ by applying it to $f^{|s|}$. So we only need to prove the extension of the range for $s = -1$. Let us assume that $s = -1$. Thus f is -1-concave, which means that $g = f^{-1}$ is convex on its support. As done by [4], we write

$$\int_{\mathbb{R}^n} f(x)^p dx = \int_{\mathbb{R}^n} g(x)^{-p} dx = \int_0^{+\infty} p t^{p-1} \psi(1/t) dt,$$

---

[1]The details of this proof were omitted from [4] because of space considerations, and are being presented here. A complete presentation will appear in [2].

where $\psi(t) = |\{x \in \mathbb{R}^n : g(x) \leq t\}|_n$ is the Lebesgue measure of the sub-level set $\{x \in \mathbb{R}^n : g(x) \leq t\}$. Using Brunn-Minkowski theorem, we can see that ψ is a $1/n$-concave function. Using the properties of the perspective function, we can deduce that the function $\varphi(t) = t^n \psi(1/t)$ is also a $1/n$-concave function. Thus it follows that

$$\int_{\mathbb{R}^n} f(x)^p dx = p \int_0^{+\infty} t^{p-n-1} \varphi(t) dt.$$

Applying Proposition III.2 to $s = 1/n$ and p replaced by $p - n$ we get that

$$B(p-n, n+1)^{-1} \int_0^{+\infty} t^{p-1-n} \varphi(t) dt$$

is log-concave on $(n, +\infty)$. Then we can conclude the proof using the following identity

$$B(p-n, n+1)^{-1} = \frac{p(p-1) \cdots (p-n)}{\Gamma(n+1)}.$$

$\square$

The fact that Theorem III.1 is optimal can be seen from the following example. Let $U : \mathbb{R}^n \to [0, \infty]$ be a positively homogeneous convex function of degree 1, i.e. that $U(tx) = tU(x)$ for all $x \in \mathbb{R}^n$ and all $t > 0$. We define $f_{s,U} = (1 - sU)_+^{1/s}$ for $s \neq 0$ and $f_{0,U} = e^{-U}$ for $s = 0$. Then we have

$$\int_{\mathbb{R}^n} f_{s,U}(x)^p dx = \frac{C_U n!}{(p+s) \cdots (p+ns)},$$

where C_U is the Lebesgue measure of the sub-level set $\{x \in \mathbb{R}^n : U(x) \leq 1\}$. We only check the identity for $s > 0$, and the other two cases can be proved similarly.

$$
\begin{aligned}
\int_{\mathbb{R}^n} f_{s,U}(x)^p dx &= p \int_0^1 t^{p-1} |\{x \in \mathbb{R}^n : (1 - sU(x))_+^{1/s} > t\}| dt \\
&= p \int_0^1 t^{p-1} |\{x \in \mathbb{R}^n : U(x) < (1 - t^s)/s\}| dt \\
&= C_U p \int_0^1 t^{p-1} ((1 - t^s)/s)^n dt \\
&= C_U s^{-n-1} p B(p/s, n+1)
\end{aligned}
$$

In the third equation, we use the homogeneity of U and the property of Lebesgue measure. Then we can prove the identity using the following fact

$$B(p/s, n+1) = \frac{n!}{(p/s + n) \cdots p/s}.$$

Thus the preceding theorem can be written in the following way: if $f : \mathbb{R}^n \to \mathbb{R}_+$ is an integrable s-concave function, then

$$p \mapsto \frac{\int_{\mathbb{R}^n} f(x)^p dx}{\int_{\mathbb{R}^n} f_{s,U}(x)^p dx}$$

is log-concave for $p > \max(0, -ns)$.

IV. CONCENTRATION OF INFORMATION CONTENT

Now we are ready to study the concentration property of information content for convex measures introduced and studied by Borell [7], [8].

Definition IV.1. *Let $-\infty \leq \kappa \leq \infty$. A finite Borel measure μ on $\mathbb{R}^n$ is called κ-concave if we have*

$$\mu((1-\lambda)A + \lambda B) \geq ((1-\lambda)\mu(A)^\kappa + \lambda\mu(B)^\kappa)^{1/\kappa} \quad (15)$$

for all $\lambda \in [0,1]$ and all Borel sets $A, B \subseteq \mathbb{R}^n$ such that $\mu(A)\mu(B) > 0$.

Here $(1-\lambda)A + \lambda B = \{(1-\lambda)x + \lambda y : x \in A, y \in B\}$ stands for the Minkowski sum of two sets. The limit cases are interpreted by continuity. Thus the right hand side of (15) is equal to $\min(\mu(A), \mu(B))$ for $\kappa = -\infty$; $\mu(A)^{1-\lambda}\mu(B)^\lambda$ for $\kappa = 0$; and $\max(\mu(A), \mu(B))$ for $\kappa = \infty$. Note that the inequality (15) becomes stronger as κ increases. For $\kappa = -\infty$, we obtain the largest class, whose members are called convex or hyperbolic measures. The case $\kappa = 0$ describes log-concave measures. If μ is a convex measure on $\mathbb{R}^n$ then it is absolutely continuous with respect to the Lebesgue measure on the subspace generated by its support and its density has a concavity property. For example, if μ is κ-concave and has a density f on $\mathbb{R}^n$ then $\kappa \leq 1/n$ and f is $-1/\beta$-concave with $\beta = n - 1/\kappa$.

We say that a $\mathbb{R}^n$-valued random variable X is κ-concave if the probability measure induced by X is κ-concave. In this section, we let X be a κ-concave random variable with density f and $\kappa < 0$. Then Borell's characterization implies that there is a convex function V such that $f = V^{-\beta}$. In the following, we will study the deviation of $\widetilde{h}(X)$ from its mean $h(X)$, that is corresponding to taking $\varphi = -\log f$ in Section 2. Then the moment generating function is

$$\mathbb{E}f^{-\alpha}(X) = \int_{\mathbb{R}^n} f(x)^{1-\alpha} dx.$$

The integral is finite as long as $(1-\alpha)\beta > n$, i.e. that $\alpha < 1 - n/\beta$.

Proposition IV.2. *Let $\beta > n$ and let X be a random variable in $\mathbb{R}^n$ with density f being $-1/\beta$-concave. Then the function*

$$\alpha \mapsto \prod_{i=1}^{n}((1-\alpha)\beta - i)\mathbb{E}f^{-\alpha}(X) \quad (16)$$

is log-concave for $\alpha < 1 - n/\beta$.

Proof. It easily follows from Theorem III.1 with p replaced by $1 - \alpha$ and s replaced by $-1/\beta$. $\qquad\square$

Following Lemma II.3, we can set

$$c(\alpha) = -\sum_{i=1}^{n} \log((1-\alpha)\beta - i). \quad (17)$$

Corollary IV.3. *Under the conditions and notations of Proposition IV.2, we have*

$$\mathrm{Var}(\widetilde{h}(X)) \leq \beta^2 \sum_{i=1}^{n} (\beta - i)^{-2}. \quad (18)$$

Proof. By Lemma II.1, we know that $\mathrm{Var}(\widetilde{h}(X)) = L''(\alpha)$, where X_α is a random variable with density proportional to $f^{1-\alpha}$ and $L(\alpha) = \log \mathbb{E}f^{-\alpha}(X)$ is the logarithmic moment generating function. By Proposition IV.2, we know that $L''(\alpha) \leq c''(\alpha)$, where $c(\alpha)$ is defined in (17). Then the variance bound (18) follows by differentiating $c(\alpha)$ twice and setting $\alpha = 0$. $\qquad\square$

Remark IV.4. *The variance bound is sharp. Suppose X has density $f = (1 + U/\beta)_+^{-\beta}$ with U being a positively homogeneous convex function of degree 1. In this case, the function in Proposition IV.2 is log-affine, i.e. $L''(\alpha) = c''(\alpha)$. Then we have equality in the above variance bound. In particular, it includes the Pareto distribution with density*

$$f(x) = \frac{1}{Z_n(a,\beta)}(a + x_1 + \cdots + x_n)^{-\beta}, \; x_i > 0, \quad (19)$$

where $a > 0$ and $Z_n(a,\beta)$ is a normalizing constant.

Let $\beta > n + 2$ and let X be a random variable in $\mathbb{R}^n$ with density f being $-1/\beta$-concave. In this case, we have $\mathbb{E}|X|^2 < \infty$ and the covariance matrix Σ is defined by

$$\Sigma = \mathbb{E}(X - \mathbb{E}X) \otimes (X - \mathbb{E}X). \quad (20)$$

Then we have

$$
\begin{aligned}
n &= \int_{\mathbb{R}^n} \langle x - \mathbb{E}X, -\nabla\log f(x)\rangle f(x)dx \\
&\leq \left(\int_{\mathbb{R}^n}|x - \mathbb{E}X|^2 f(x)dx \cdot \int_{\mathbb{R}^n}|\nabla\log f(x)|^2 f(x)dx\right)^{1/2} \\
&= \sqrt{\mathrm{tr}(\Sigma)J(X)},
\end{aligned}
$$

where $\mathrm{tr}(\Sigma)$ is the trace of Σ and $J(X)$ is the Fisher information defined by

$$J(X) = \int_{\mathbb{R}^n} \frac{|\nabla f|^2}{f} dx. \quad (21)$$

Combining with Corollary IV.3 we have the following result.

Corollary IV.5. *Let $\beta > n + 2$ and let X be a random variable in $\mathbb{R}^n$ with density f being $-1/\beta$-concave. Then we have*

$$\mathrm{Var}(\widetilde{h}(X)) \leq \frac{tr(\Sigma)\beta^2}{n^2} \sum_{i=1}^{n}(\beta - i)^{-2} J(X). \quad (22)$$

In particular, if X is isotropic, i.e. that $\mathbb{E}X = 0$ and Σ is the identity matrix, we have

$$\mathrm{Var}(\widetilde{h}(X)) \leq \frac{\beta^2}{n} \sum_{i=1}^{n}(\beta - i)^{-2} J(X). \quad (23)$$

Taking $\beta \to \infty$ yields the analogue for log-concave random variables, namely

$$\mathrm{Var}(\widetilde{h}(X)) \leq J(X), \quad (24)$$

which was observed by [17].

Theorem IV.1. *Let $\beta > n$ and let X be a random variable in $\mathbb{R}^n$ with density f being $-1/\beta$-concave. Then we have*

$$\mathbb{E}e^{\alpha(\widetilde{h}(X) - h(X))} \leq e^{\psi_c(\alpha)} \tag{25}$$

for $\alpha < 1 - n/\beta$, where

$$\psi_c(\alpha) = -\alpha\beta \sum_{i=1}^{n}(\beta - i)^{-1} - \sum_{i=1}^{n}\log\frac{(1-\alpha)\beta - i}{\beta - i}. \tag{26}$$

Particularly, we have equality for Pareto distributions.

Proof. The moment generating function bound (25) easily follows from Lemma II.3 and Proposition IV.2. Some easy calculations will show the equality case for Pareto distributions. Essentially that is due to the identity $L''(\alpha) = c''(\alpha)$, where $c(\alpha)$ is defined in (17). $\square$

Corollary IV.6. *Under the conditions and notations of Theorem IV.1, we have for $t > 0$ that*

$$\mathbb{P}(\widetilde{h}(X) - h(X) > t) \leq e^{-\psi_{c,+}^*(t)}, \tag{27}$$

$$\mathbb{P}(\widetilde{h}(X) - h(X) < -t) \leq e^{-\psi_{c,-}^*(-t)}, \tag{28}$$

where $\psi_{c,+}^*$ and $\psi_{c,-}^*$ are Fenchel-Legendre dual functions of $\psi_{c,+}$ and $\psi_{c,-}$, respectively.

In general we do not have explicit expressions for $\psi_{c,+}^*$ or $\psi_{c,-}^*$.

The following result (obtained in [4] when $\beta \geq n+1$) says that the entropy of a κ-concave distribution can not exceed that of the Pareto distribution with the same maximal density value.

Corollary IV.7. *Under the conditions and notations of Theorem IV.1, we have*

$$h(X) \leq -\log\|f\|_\infty + \beta\sum_{i=1}^{n}(\beta - i)^{-1}, \tag{29}$$

where we denote by $\|f\|_\infty$ the essential supremum. We have equality for Pareto distributions.

Proof. As a function of α, we have

$$(-\alpha t - \psi_c(\alpha))' = -t + \beta\sum_{i=1}^{n}(\beta - i)^{-1} - \beta\sum_{i=1}^{n}((1-\alpha)\beta - i)^{-1}.$$

For any $t > \beta\sum_{i=1}^{n}(\beta - i)^{-1}$, we can see that $-\alpha t - \psi_c(\alpha)$ is a decreasing function of $\alpha < 1 - n/\beta$. It is clear that $\lim_{\alpha\to-\infty}(-\alpha t - \psi_c(\alpha)) = \infty$. Therefore we have $\psi_{c,-}^*(-t) = \infty$ for $t > \beta\sum_{i=1}^{n}(\beta - i)^{-1}$. Using the lower tail estimate in Corollary IV.6, almost surely we have

$$\widetilde{h}(X) - h(X) \geq -\beta\sum_{i=1}^{n}(\beta - i)^{-1}.$$

Taking the supremum over all realizable values of X yields

$$-\log\|f\|_\infty - h(X) \geq -\beta\sum_{i=1}^{n}(\beta - i)^{-1}.$$

That is equivalent to the desired statement. $\square$

Remark IV.8. *We can get corresponding estimates for log-concave random variables (see [4], [13]) by taking the limit $\beta \to \infty$.*

REFERENCES

[1] A. Barron, "The strong ergodic theorem for densities: Generalized Shannon-Mcmillan-Breiman theorem," *Ann. Probab.*, vol. 13, pp. 1292–1303, 1985.

[2] S. Bobkov, M. Fradelizi, J. Li, and M. Madiman, "When can one invert Hölder's inequality? (and why one may want to)," *Preprint*, 2015.

[3] S. Bobkov and M. Madiman, "Concentration of the information in data with log-concave distributions," *Ann. Probab.*, vol. 39, no. 4, pp. 1528–1543, 2011.

[4] ——, "The entropy per coordinate of a random vector is highly constrained under convexity conditions," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 4940–4954, August 2011.

[5] ——, "Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures," *J. Funct. Anal.*, vol. 262, pp. 3309–3339, 2012.

[6] C. Borell, "Complements of Lyapunov's inequality," *Math. Ann.*, vol. 205, pp. 323–331, 1973.

[7] ——, "Convex measures on locally convex spaces," *Ark. Mat.*, vol. 12, pp. 239–252, 1974.

[8] ——, "Convex set functions in d-space," *Period. Math. Hungar.*, vol. 6, no. 2, pp. 111–136, 1975.

[9] J. Bourgain, "On high-dimensional maximal functions associated to convex bodies," *Amer. J. Math.*, vol. 108, no. 6, pp. 1467–1476, 1986.

[10] L. Breiman, "The individual ergodic theorem for information theory," *Ann. Math. Stat.*, vol. 28, pp. 809–811, 1957.

[11] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inform. Theory*, vol. 35, no. 1, pp. 37–43, 1989.

[12] M. Fradelizi, O. Guédon, and A. Pajor, "Thin-shell concentration for convex measures," *Studia Math.*, vol. 223, no. 2, pp. 123–148, 2014.

[13] M. Fradelizi, M. Madiman, and L. Wang, "Optimal concentration of information content for log-concave densities," *To appear in High Dimensional Probability VII*, 2015, available online at arXiv:1508.04093.

[14] O. Guédon, P. Nayar, and T. Tkocz, "Concentration inequalities and geometry of convex bodies," in *Analytical and Probabilistic Methods in the Geometry of Convex Bodies*, ser. IM PAN Lecture Notes. Warsaw: Polish Acad. Sci., 2014, vol. 2, pp. 9–86.

[15] M. Madiman, L. Wang, and S. Bobkov, "Some applications of the nonasymptotic equipartition property of log-concave distributions," *Preprint*, 2016.

[16] B. McMillan, "The basic theorems of information theory," *Ann. Math. Stat.*, vol. 24, pp. 196–219, 1953.

[17] V. H. Nguyen, "Dimensional variance inequalities of Brascamp-Lieb type and a local approach to dimensional Prékopa's theorem," *J. Funct. Anal.*, vol. 266, no. 2, pp. 931–955, 2014.

[18] S. Orey, "On the Shannon-Perez-Moy theorem," in *Particle systems, random media and large deviations (Brunswick, Maine, 1984)*. Providence, R.I.: Amer. Math. Soc., 1985, pp. 319–327.

[19] C. Shannon, "A mathematical theory of communication," *Bell System Tech. J.*, vol. 27, pp. 379–423, 623–656, 1948.