A Note on Statistical Distances for Discrete Log-Concave Measures

Arnaud Marsiglietti and Puja Pandey

Abstract

In this note we explore how standard statistical distances are equivalent for discrete log-concave distributions. Distances include total variation distance, Wasserstein distance, and f-divergences.

Keywords: log-concave, total variation distance, Wasserstein distance, f-divergence.

1 Introduction

The study of convergence of probability measures is central in probability and statistics, and may be performed via statistical distances for which the choice has its importance (see, e.g., [10], [19]). The space of probability measures, say over the real numbers, is infinite dimensional, therefore there is a priori no canonical distance, and distances may not be equivalent. Nonetheless, an essential contribution made by Meckes and Meckes in [16] demonstrates that certain statistical distances between continuous log-concave distributions turn out to be equivalent up to constants that may depend on the dimension of the ambient space (see also [7] for improved bounds, and [14] for the extension to the broader class of so-called s-concave distributions).

The goal of this note is to develop quantitative comparisons between distances for discrete log-concave distributions. Let us denote by $\mathbb{N} = \{0, 1, 2, \dots\}$ the set of natural numbers and by $\mathbb{Z}$ the set of integers. Recall that the probability mass function (p.m.f.) associated with an integer valued random variable X is

$$p(k) = \mathbb{P}(X = k), \quad k \in \mathbb{Z}.$$

An integer-valued random variable X is said to be log-concave if its probability mass function p satisfies

$$p(k)^2 \geq p(k-1)p(k+1)$$

for all $k \in \mathbb{Z}$ and the support of X is an integer interval.

Discrete log-concave distributions form an important class. Examples include Bernoulli, discrete uniform, binomial, geometric and Poisson distributions. We refer to [21], [5], [20], [4] for further background on log-concavity.

Let us introduce the main distances we will work with (we refer to [12], [9], [10], [19] for further background on statistical distances). Our setting is the real line $\mathbb{R}$ equipped with its usual Euclidean structure $d(x, y) = |x - y|$, $x, y \in \mathbb{R}$.

1. The bounded Lipschitz distance between two probability measures μ and ν is defined as

$$d_{BL}(\mu, \nu) = \sup_{\|g\|_{BL} \leq 1} \left| \int g \, d\mu - \int g \, d\nu \right|,$$

where for a function $g\colon \mathbb{R} \to \mathbb{R}$,

$$\|g\|_{BL} = \max\left\{ \|g\|_\infty, \sup_{x\neq y} \frac{|g(x)-g(y)|}{|x-y|} \right\}.$$

2. The Lévy-Prokhorov distance between two probability measures μ and ν is defined as

$$d_{LP}(\mu,\nu) = \inf\left\{\epsilon > 0 : \mu(A) \leq \nu(A^\varepsilon) + \epsilon \text{ for all Borel set } A \subset \mathbb{R}\right\},$$

where $A^\varepsilon = \{x \in \mathbb{R} : d(x,A) < \varepsilon\}$.

Using the Ky-Fan distance, which is defined for two random variables X and Y as

$$K(X,Y) = \inf\{\varepsilon > 0 : \mathbb{P}(|X-Y| > \varepsilon) < \varepsilon\},$$

the Lévy-Prokhorov distance admits the following coupling representation,

$$d_{LP}(\mu,\nu) = \inf K(X,Y), \tag{1}$$

where the infimum runs over all random variables X with distribution μ and random variables Y with distribution ν (see, e.g., [19]).

3. The total variation distance between two probability measures μ and ν is defined as

$$d_{TV}(\mu,\nu) = 2\sup_{A\subset\mathbb{R}} |\mu(A) - \nu(A)|.$$

The total variation distance admits the following coupling representation,

$$d_{TV}(\mu,\nu) = \inf \mathbb{P}(X \neq Y), \tag{2}$$

where the infimum runs over all random variables X with distribution μ and random variables Y with distribution ν (see, e.g., [10]). Moreover, for integer valued measures, one has the following identity,

$$d_{TV}(\mu,\nu) = \sum_{k\in\mathbb{Z}} |\mu(\{k\}) - \nu(\{k\})|. \tag{3}$$

4. The p-th Wasserstein distance, $p \geq 1$, between two probability measures μ and ν is defined as

$$W_p(\mu,\nu) = \inf \mathbb{E}[|X-Y|^p]^{\frac{1}{p}},$$

where the infimum runs over all random variables X with distribution μ and random variables Y with distribution ν.

5. Let $f\colon [0,+\infty) \to \mathbb{R}$ be a convex function such that $f(1) = 0$. The f-divergence between two probability measures μ and ν on $\mathbb{Z}$ is defined as

$$d_f(\mu||\nu) = \sum_{k\in\mathbb{Z}} \nu(\{k\}) f\left(\frac{\mu(\{k\})}{\nu(\{k\})}\right).$$

Note that the choice of convex function $f(x) = x\log(x)$, $x \geq 0$, leads to the Kullback-Leibler divergence

$$D(\mu||\nu) = \sum_{k\in\mathbb{Z}} \mu(\{k\}) \log\left(\frac{\mu(\{k\})}{\nu(\{k\})}\right),$$

the function $f(x) = (x-1)^2$ yields the so-called χ^2-divergence, while $f(x) = |x-1|$ returns us to the total variation distance.

Let us review the known relationships between the above distances. It is known [8, Corollaries 2 and 3] that bounded Lipschitz and Lévy-Prokhorov distances are equivalent,

$$\frac{1}{2}d_{BL}(\mu,\nu) \leq d_{LP}(\mu,\nu) \leq \sqrt{\frac{3}{2}d_{BL}(\mu,\nu)}.$$

One also has

$$d_{LP}(\mu,\nu) \leq d_{TV}(\mu,\nu),$$

and, for μ,ν integer valued probability measures,

$$d_{TV}(\mu,\nu) \leq W_1(\mu,\nu),$$

see [10]. By Hölder's inequality, if $p \leq q$, then

$$W_p(\mu,\nu) \leq W_q(\mu,\nu).$$

As for divergences, the Pinsker-Csiszár inequality ([18], [6]) states that

$$d_{TV}(\mu,\nu) \leq \sqrt{2D(\mu||\nu)}.$$

Also, one has

$$D(\mu||\nu) \leq \log(1+\chi^2(\mu||\nu)).$$

The article is organized as follows. In Section 2, we establish properties for log-concave distributions on $\mathbb{Z}$ that are of independent interests. In section 3, we present our results and proofs.

2 Preliminaries

In this section we gather the main tools used throughout the proofs. First, recall that a real-valued random variable X is said to be isotropic if

$$\mathbb{E}[X] = 0, \quad \mathbb{E}[X^2] = 1.$$

We start with an elementary lemma that will allow us to pass results for log-concave distributions on $\mathbb{N}$ to log-concave distributions on $\mathbb{Z}$, however with sub-optimal constants.

Lemma 2.1. *If X is symmetric log-concave on $\mathbb{Z}$, then $|X|$ is log-concave on $\mathbb{N}$.*

Proof. Denote by p (resp. q) the p.m.f. of X (resp. $|X|$). Then, $q(0) = p(0)$ and $q(k) = 2p(k)$ for $k \geq 1$. Therefore,

$$q^2(1) = 4p^2(1) \geq 4p(0)p(2) = 2q(0)q(2),$$

and for all $k \geq 2$,

$$q^2(k) = 4p^2(k) \geq 4p(k+1)p(k-1) = q(k+1)q(k-1).$$

Hence, q is log-concave. □

We note that Lemma 2.1 no longer holds for non-symmetric log-concave random variables, as can be seen by taking X supported on $\{-1,0,1,2,3\}$ with distribution $\mathbb{P}(X=-1) = \mathbb{P}(X=3) = 0.1$, $\mathbb{P}(X=0) = \mathbb{P}(X=2) = 0.2$, and $\mathbb{P}(X=1) = 0.4$. In this case, $\mathbb{P}(|X|=2)^2 < \mathbb{P}(|X|=1)\mathbb{P}(|X|=3)$.

The next lemma provides moments bounds for log-concave distributions on $\mathbb{Z}$.

Lemma 2.2. *If X is log-concave on $\mathbb{Z}$, then for all $\beta \geq 1$,*

$$\mathbb{E}[|X-\mathbb{E}[X]|^{\beta}]^{\frac{1}{\beta}} \leq \Gamma(\beta+1)^{\frac{1}{\beta}}(2\mathbb{E}[|X-\mathbb{E}[X]|]+1).$$

Proof. It has been shown in [13, Corollary 4.5] that for all log-concave random variable X on $\mathbb{N}$, for all $\beta \geq 1$,

$$\mathbb{E}[X^{\beta}]^{\frac{1}{\beta}} \leq \Gamma(\beta+1)^{\frac{1}{\beta}}(\mathbb{E}[X]+1). \tag{4}$$

Let X be a symmetric log-concave random variable on $\mathbb{Z}$, then by Lemma 2.1, $|X|$ is log-concave on $\mathbb{N}$ so one may apply inequality (4) to obtain

$$\mathbb{E}[|X|^{\beta}]^{\frac{1}{\beta}} \leq \Gamma(\beta+1)^{\frac{1}{\beta}}(\mathbb{E}[|X|]+1). \tag{5}$$

Now, let X be a log-concave random variable on $\mathbb{Z}$. Let Y be an independent copy of X, so that $X-Y$ is symmetric log-concave. Applying inequality (5), we deduce that

$$\mathbb{E}[|X-\mathbb{E}[X]|^{\beta}]^{\frac{1}{\beta}} \leq \mathbb{E}[|X-Y|^{\beta}]^{\frac{1}{\beta}} \leq \Gamma(\beta+1)^{\frac{1}{\beta}}(\mathbb{E}[|X-Y|]+1) \leq \Gamma(\beta+1)^{\frac{1}{\beta}}(2\mathbb{E}[|X-\mathbb{E}[X]|]+1),$$

where the first inequality follows from Hölder's inequality and the last inequality from triangle inequality. □

Let us derive concentration inequalities for log-concave distributions on $\mathbb{Z}$.

Lemma 2.3. *For each log-concave random variable X on $\mathbb{Z}$, one has for all $t \geq 0$,*

$$\mathbb{P}(|X-\mathbb{E}[X]| \geq t) \leq 2e^{-\frac{t}{2(2\mathbb{E}[|X-\mathbb{E}[X]|]+1)}}.$$

Proof. The proof is a standard application of the moments bounds obtained in Lemma 2.2 (see, e.g., [22]). For $\lambda > 0$,

$$\begin{aligned}\mathbb{E}[e^{\lambda|X-\mathbb{E}[X]|}] = 1 + \sum_{\beta\geq 1}\frac{\lambda^{\beta}}{\beta!}\mathbb{E}[|X-\mathbb{E}[X]|^{\beta}] &\leq 1 + \sum_{\beta\geq 1}\frac{\lambda^{\beta}}{\beta!}\beta!(2\mathbb{E}[|X-\mathbb{E}[X]|]+1)^{\beta} \\ &= \sum_{\beta\geq 0}[\lambda(2\mathbb{E}[|X-\mathbb{E}[X]|]+1)]^{\beta} \\ &= \frac{1}{1-\lambda(2\mathbb{E}[|X-\mathbb{E}[X]|]+1)},\end{aligned}$$

where the last identity holds for all $0 < \lambda < \frac{1}{2\mathbb{E}[|X-\mathbb{E}[X]|]+1}$. Choosing $\lambda = \frac{1}{2(2\mathbb{E}[|X-\mathbb{E}[X]|]+1)}$ yields

$$\mathbb{E}[e^{\lambda|X-\mathbb{E}[X]|}] \leq 2.$$

Therefore, by Markov's inequality,

$$\mathbb{P}(|X-\mathbb{E}[X]| \geq t) = \mathbb{P}(e^{\lambda|X-\mathbb{E}[X]|} \geq e^{\lambda t}) \leq \mathbb{E}[e^{\lambda|X-\mathbb{E}[X]|}]e^{-\lambda t} \leq 2e^{-\frac{t}{2(2\mathbb{E}[|X-\mathbb{E}[X]|]+1)}}.$$

□

The following lemma, which provides a bound on the variance and maximum of the probability mass function of log-concave distributions on $\mathbb{Z}$, was established in [3] and [1] (see, also, [2], [11]).

Lemma 2.4 ([3], [1])**.** *Let X be a log-concave distribution on $\mathbb{Z}$ with probability mass function p, then*

$$\sqrt{1+\mathrm{Var}(X)} \le \frac{1}{\|p\|_\infty} \le \sqrt{1+12\,\mathrm{Var}(X)}.$$

The following lemma is standard in information theory and provides an upper bound on the entropy of an integer valued random variable (see [15]). Recall that the Shannon entropy of an integer valued random variable X with p.m.f. p is defined as

$$H(X) = \mathbb{E}[-\log(p(X))] = -\sum_{k\in\mathbb{Z}} p(k)\log(p(k)).$$

Lemma 2.5 ([15])**.** *For any integer valued random variable X with finite second moment,*

$$H(X) \le \frac{1}{2}\log\left(2\pi e\left(\mathrm{Var}(X)+\frac{1}{12}\right)\right).$$

The last lemma of this section provides a bound on the second moment of the information content of a log-concave distribution on $\mathbb{Z}$.

Lemma 2.6. *Let X be a discrete log-concave random variable on $\mathbb{Z}$ with probability mass function p. Then,*

$$\mathbb{E}[\log^2(p(X))] \le 4\left(4e^{-2}+1+\frac{H^2(X)}{\|p\|_\infty}\right).$$

Proof. Let X be a log-concave random variable with p.m.f. p. Then p is unimodal, that is, there exists $m\in\mathbb{Z}$ such that for all $k\le m$, $p(k-1)\le p(k)$ and for all $k\ge m$, $p(k)\ge p(k+1)$. Note that $p(m)=\|p\|_\infty$. Define, for $k\in\mathbb{Z}$,

$$p^{\nearrow}(k) = \frac{p(k)}{\sum_{l\le m} p(l)} 1_{\{k\le m\}},$$

and

$$p^{\searrow}(k) = \frac{p(k)}{\sum_{l\ge m} p(l)} 1_{\{k\ge m\}}.$$

Note that both $p^{\nearrow}$ and $p^{\searrow}$ are monotone log-concave probability mass functions. Denote by $X^{\nearrow}$ (resp. $X^{\searrow}$) a random variable with p.m.f. $p^{\nearrow}$ (resp. $p^{\searrow}$). Denote also $a=\sum_{l\le m} p(l)$ and $b=\sum_{l\ge m} p(l)$. On one hand, by a result of Melbourne and Palafox-Castillo [17, Theorem 2.5],

$$\mathrm{Var}(\log(p^{\nearrow}(X^{\nearrow}))) \le 1, \qquad \mathrm{Var}(\log(p^{\searrow}(X^{\searrow}))) \le 1.$$

On the other hand,

$$H(X^{\nearrow}) = \sum_{k\le m}\frac{p(k)}{a}\log\left(\frac{a}{p(k)}\right) \le \frac{1}{a}\sum_{k\le m} p(k)\log\left(\frac{1}{p(k)}\right) \le \frac{1}{a}H(X),$$

and similarly,

$$H(X^{\searrow}) \le \frac{H(X)}{b}.$$

Therefore,

$$\mathbb{E}[\log^2(p^{\nearrow}(X^{\nearrow}))] = \mathrm{Var}(\log(p^{\nearrow}(X^{\nearrow}))) + H^2(X^{\nearrow}) \le 1 + \frac{H^2(X)}{a^2},$$

and similarly,

$$\mathbb{E}[\log^2(p^{\searrow}(X^{\searrow}))] \leq 1 + \frac{H^2(X)}{b^2}.$$

We deduce that

$$\begin{aligned}
\mathbb{E}[\log^2(p(X))] &= \sum_{k\in\mathbb{Z}} p(k)\log^2(p(k)) \\
&\leq \sum_{k\in\mathbb{Z}} ap^{\nearrow}(k)\log^2(ap^{\nearrow}(k)) + \sum_{k\in\mathbb{Z}} bp^{\searrow}(k)\log^2(bp^{\searrow}(k)) \\
&\leq 2\left(a\log^2(a) + a\mathbb{E}[\log^2(p^{\nearrow}(X^{\nearrow}))] + b\log^2(b) + b\mathbb{E}[\log^2(p^{\searrow}(X^{\searrow}))]\right) \\
&\leq 2\left(4e^{-2} + a + \frac{H^2(X)}{a} + 4e^{-2} + b + \frac{H^2(X)}{b}\right) \\
&\leq 4\left(4e^{-2} + 1 + \frac{H^2(X)}{\|p\|_\infty}\right),
\end{aligned}$$

where we used the fact that $a, b \in [\|p\|_\infty, 1]$. $\square$

Remark 2.7. *For an isotropic log-concave random variable X on $\mathbb{Z}$ with probability mass function p, the above bounds reduce to*

$$\begin{aligned}
\mathbb{E}[|X|^\beta]^{\frac{1}{\beta}} &\leq 3\Gamma(\beta+1)^{\frac{1}{\beta}}, \quad \beta \geq 1, && (6)\\
\mathbb{P}(|X| \geq t) &\leq 2e^{-\frac{t}{6}}, \quad t \geq 0, && (7)\\
\sqrt{2} &\leq \frac{1}{\|p\|_\infty} \leq \sqrt{13}, && (8)\\
H(X) &\leq \frac{1}{2}\log\left(2\pi e\left(1 + \frac{1}{12}\right)\right) \leq 3, && (9)
\end{aligned}$$

in particular, we also deduce

$$\mathbb{E}[\log^2(p(X))] \leq 4(4e^{-2} + 1 + 9\sqrt{13}) \leq 136. \quad (10)$$

3 Main results and proofs

This section contains our main results together with the proofs. The first theorem establishes quantitative reversal bounds between 1-Wasserstein distance and Lévy-Prokhorov distance.

Theorem 3.1. *Let μ and ν be isotropic log-concave probability measures on $\mathbb{Z}$, then*

$$W_1(\mu,\nu) \leq 12 d_{LP}(\mu,\nu)\log\left(\frac{4e}{d_{LP}(\mu,\nu)}\right).$$

Proof. Let $R > 0$. Let X (resp. Y) be distributed according to μ (resp. ν). Note that for all $t \geq 0$,

$$\mathbb{P}(|X-Y| > t) = \mathbb{P}(|X-Y| \geq \lfloor t \rfloor + 1) \leq \mathbb{P}(|X-Y| \geq 1) \leq K(X,Y),$$

therefore,

$$\begin{aligned}
\mathbb{E}[|X-Y|] &= \int_0^R \mathbb{P}(|X-Y| > t)dt + \int_R^\infty \mathbb{P}(|X-Y| > t)dt \\
&\leq RK(X,Y) + \int_R^\infty \mathbb{P}\left(|X| > \frac{t}{2}\right)dt + \int_R^\infty \mathbb{P}\left(|Y| > \frac{t}{2}\right)dt.
\end{aligned}$$

Applying inequality (7), we obtain

$$W_1(\mu,\nu) \le \mathbb{E}[|X-Y|] \le RK(X,Y) + 2\int_R^\infty 2e^{-\frac{t}{12}}dt = RK(X,Y) + 48e^{-\frac{R}{12}}.$$

The above inequality being true for any random variable X with distribution μ and any random variable Y with distribution ν, we deduce by taking infimum over all couplings that

$$W_1(\mu,\nu) \le R d_{LP}(\mu,\nu) + 48e^{-\frac{R}{12}}.$$

Choosing $R = 12\log(4/d_{LP}(\mu,\nu))$, which is nonnegative, yields the desired result. □

The next theorem demonstrates that Wasserstein distances are equivalent for discrete log-concave distributions.

Theorem 3.2. *Let μ and ν be isotropic log-concave probability measures on $\mathbb{Z}$, then for all $1 \le p \le q$,*

$$W_q^q(\mu,\nu) \le 24^{q-p} W_p^p(\mu,\nu) \log^{q-p}\left(\frac{6^q\sqrt{\Gamma(2q+1)}}{W_p^p(\mu,\nu)}\right) + 2W_p^p(\mu,\nu),$$

Proof. Let X (resp. Y) be distributed according to μ (resp. ν). Let $R>0$. One has

$$\begin{aligned}\mathbb{E}[|X-Y|^q] &= \mathbb{E}[|X-Y|^{q-p+p}1_{\{|X-Y|<R\}}] + \mathbb{E}[|X-Y|^q 1_{\{|X-Y|\ge R\}}] \\ &\le R^{q-p}\mathbb{E}[|X-Y|^p] + \sqrt{\mathbb{P}(|X-Y|\ge R)\mathbb{E}[|X-Y|^{2q}]},\end{aligned}$$

where we used the Cauchy-Schwarz inequality. Note that by inequality (6),

$$\mathbb{E}[|X-Y|^{2q}]^{\frac{1}{2q}} \le \mathbb{E}[|X|^{2q}]^{\frac{1}{2q}} + \mathbb{E}[|Y|^{2q}]^{\frac{1}{2q}} \le 6\Gamma(2q+1)^{\frac{1}{2q}}.$$

Moreover, by inequality (7),

$$\mathbb{P}(|X-Y|\ge R) \le \mathbb{P}(|X|\ge \frac{R}{2}) + \mathbb{P}(|Y|\ge \frac{R}{2}) \le 4e^{-\frac{R}{12}}.$$

Combining the above and taking infimum over all couplings yield

$$W_q^q(\mu,\nu) \le R^{q-p}W_p^p(\mu,\nu) + 6^q 2\sqrt{\Gamma(2q+1)}e^{-\frac{R}{24}}.$$

The result follows by choosing $R = 24\log\left(\frac{6^q\sqrt{\Gamma(2q+1)}}{W_p^p(\mu,\nu)}\right)$, which is nonnegative since by inequality (6) and log-convexity of the Gamma function,

$$W_p^p(\mu,\nu) \le \left(\mathbb{E}[|X|^p]^{\frac{1}{p}} + \mathbb{E}[|Y|^p]^{\frac{1}{p}}\right)^p \le 6^p\Gamma(p+1) \le 6^q\sqrt{\Gamma(2q+1)}.$$

□

Let us now turn to f-divergences. Considering f-divergences, such as the Kullback-Leibler divergence, the main question lies in figuring out the distribution of the reference measure. In general, if the support of a measure μ is not included in the support of a measure ν, then $D(\mu||\nu) = +\infty$. Our choice of reference measure will therefore be a measure fully supported on $\mathbb{Z}$, but it turns out that it needs not be log-concave. Given $a > 0$ and $c \geq 1$, let us introduce the following class of functions:

$$\mathcal{Q}(a,c) = \{q\colon \mathbb{Z} \to [0,1],\, \forall k \in \mathbb{Z}, q(k) > 0 \text{ and } \log\left(\frac{1}{q(k)}\right) \leq ak^2 + \log(c)\}.$$

Before stating our next result, let us note that important distributions belong to such a class.

Remark 3.3. *The isotropic symmetric Poisson distribution, whose probability mass function is*

$$q(k) = C\frac{\lambda^{|k|}}{|k|!}, \quad k \in \mathbb{Z},$$

with $\lambda > 0$ *such that* $\sum_{k\in\mathbb{Z}} k^2 q(k) = 1$ *and* $C = (2e^\lambda - 1)^{-1}$ *being the normalizing constant, belongs to* $\mathcal{Q}(1 + \log(4), 2e - 1)$*. Indeed, since*

$$1 = \sum_{k\in\mathbb{Z}} k^2 q(k) = \frac{2e^\lambda}{2e^\lambda - 1}\lambda(1+\lambda),$$

then one may choose $\lambda \in [1/4, 1]$*. Therefore, using* $|k|! \leq |k|^{|k|}$*,*

$$0 \leq \log\left(\frac{1}{q(k)}\right) = \log(|k|!) + |k|\log\left(\frac{1}{\lambda}\right) + \log(2e^\lambda - 1) \leq (1+\log(4))k^2 + \log(2e-1).$$

One may also note that the isotropic symmetric geometric distribution and isotropic discretized Gaussian distribution (whose p.m.f. is of the form $q(k) = Ce^{-\lambda k^2}$*) belong to* $\mathcal{Q}(a,c)$ *for some numerical constants* $a, c > 0$*. The above three measures are natural candidates as a reference measure for Kullback-Leibler divergence.*

As for examples of non-log-concave distributions, consider p.m.f. of the form $Ce^{-\lambda k^\alpha}$*, for* $\alpha \in (0,1)$*.*

The next result provides a comparison between total variation distance and f-divergences. The result is general as it holds for arbitrary convex function f, however the statement is not in a closed form formula. We state it as a lemma, and then apply it to two specific convex functions, yielding a comparison with Kullback-Leibler divergence and χ^2-divergence.

Lemma 3.4. *Let* $f\colon [0,+\infty) \to \mathbb{R}$ *be a convex function such that* $f(1) = 0$*. Let* $a > 0$ *and* $c \geq 1$*. Let* ν *be a measure on* $\mathbb{Z}$ *whose p.m.f.* q *belongs to the class* $\mathcal{Q}(a,c)$*. Let* μ *be an isotropic log-concave measure on* $\mathbb{Z}$ *with p.m.f.* p*. Then, denoting by* Y *a random variable with distribution* μ *and* $W = p(Y)/q(Y)$*,*

$$d_f(\mu||\nu) \leq \inf_{R\geq c}\Bigg[\left(\max\{f(0),0\} + \frac{f(R)}{R-1}\right)d_{TV}(\mu,\nu) + \sqrt{\mathbb{E}\left[\left(\frac{f(W)}{W}\right)^2 1_{\{W>1\}}\right]}\sqrt{2}e^{-\frac{1}{12}\sqrt{\frac{1}{a}\log\left(\frac{R}{c}\right)}}\Bigg].$$

Proof. The idea of proof comes from [16] (see also [14]). Denote by p (resp. q) the p.m.f. of μ (resp. ν). Denote by Y a random variable with p.m.f. p, by Z a random variable with p.m.f. q, and denote

$$X = \frac{p(Z)}{q(Z)}, \qquad W = \frac{p(Y)}{q(Y)}.$$

Using identity (3), one has

$$\mathbb{E}[|X-1|] = d_{TV}(\mu,\nu). \tag{11}$$

Let $R \geq 1$ and write

$$d_f(\mu||\nu) = \mathbb{E}[f(X)] = \mathbb{E}[f(X)1_{\{X<1\}}] + \mathbb{E}[f(X)1_{\{1\leq X\leq R\}}] + \mathbb{E}[f(X)1_{\{X>R\}}].$$

Let us bound all three parts. For the first part, since f is convex and $f(1) = 0$, it holds that for all $x \in [0,1]$,

$$f(x) \leq f(0)|x-1| \leq \max\{f(0),0\}|x-1|.$$

Therefore, using (11),

$$\mathbb{E}[f(X)1_{\{X<1\}}] \leq \max\{f(0),0\}\mathbb{E}[|X-1|1_{\{X<1\}}] \leq \max\{f(0),0\}d_{TV}(\mu,\nu). \tag{12}$$

For the second part, since f is convex and $f(1) = 0$, it holds that for all $x \in [1,R]$,

$$f(x) \leq \frac{f(R)}{R-1}(x-1).$$

Hence, using (11),

$$\mathbb{E}[f(X)1_{\{1\leq X\leq R\}}] \leq \frac{f(R)}{R-1}\mathbb{E}[(X-1)1_{\{1\leq X\leq R\}}] \leq \frac{f(R)}{R-1}d_{TV}(\mu,\nu). \tag{13}$$

For the last part, note that

$$\mathbb{E}[f(X)1_{\{X>R\}}] = \mathbb{E}\left[\frac{f(W)}{W}1_{\{W>R\}}\right] \leq \sqrt{\mathbb{E}\left[\left(\frac{f(W)}{W}\right)^2 1_{\{W>1\}}\right]}\sqrt{\mathbb{P}(W>R)}, \tag{14}$$

where we used the Cauchy-Schwarz inequality. It remains to upper bound $\mathbb{P}(W > R)$. Using that $q \in \mathcal{Q}(a,c)$ and $\|p\|_\infty \leq 1$, we have

$$\mathbb{P}(W>R) = \mathbb{P}\left(\frac{p(Y)}{q(Y)} > R\right) \leq \mathbb{P}\left(\log(c) + aY^2 > \log(R)\right).$$

Using (7), we deduce that for all $R \geq c$,

$$\mathbb{P}(W>R) \leq \mathbb{P}\left(|Y| > \sqrt{\frac{1}{a}\log\left(\frac{R}{c}\right)}\right) \leq 2e^{-\frac{1}{6}\sqrt{\frac{1}{a}\log\left(\frac{R}{c}\right)}}. \tag{15}$$

The result follows by combining (12), (13), (14), and (15), and by taking infimum over all $R \geq c$. □

Applying Lemma 3.4 to the convex function $f(x) = x\log(x)$, $x \geq 0$, yields a comparison between total variation distance and Kullback-Leibler divergence.

Theorem 3.5. *Let $a > 0$ and $c \geq 2$. Let ν be a measure on $\mathbb{Z}$ whose p.m.f. belongs to the class $\mathcal{Q}(a,c)$. Let μ be an isotropic log-concave measure on $\mathbb{Z}$. Then,*

$$D(\mu||\nu) \leq d_{TV}(\mu,\nu)\left(288a\log^2\left(\frac{\left(\sqrt{136}+46a+\log(c)\right)\sqrt{2}}{24\sqrt{a}d_{TV}(\mu,\nu)}\right) + 2\log(c) + 24\sqrt{a}\right).$$

Proof. Recall the notation $W = p(Y)/q(Y)$ from Lemma 3.4. With the choice of convex function $f(x) = x\log(x)$, $x \geq 0$, Lemma 3.4 tells us that for all $R \geq c \geq 2$,

$$D(\mu||\nu) \leq 2\log(R) d_{TV}(\mu,\nu) + \sqrt{\mathbb{E}[|\log(W)|^2]}\sqrt{2}e^{-\frac{1}{12}\sqrt{\frac{1}{a}\log(\frac{R}{c})}}.$$

Next, let us upper bound the term $\mathbb{E}[|\log(W)|^2]^{1/2}$. On one hand, since $q \in \mathcal{Q}(a,c)$,

$$\mathbb{E}[|\log(q(Y))|^2] \leq \mathbb{E}[\left(\log(c) + aY^2\right)^2] \leq \log^2(c) + 2a\log(c) + 1944a^2 := C(a,c), \tag{16}$$

where we used (6). On the other hand, by inequality (10),

$$\mathbb{E}[|\log(p(Y))|^2] \leq 136. \tag{17}$$

Therefore, combining (16) and (17),

$$\mathbb{E}[|\log(W)|^2]^{\frac{1}{2}} \leq \mathbb{E}[|\log(p(Y))|^2]^{\frac{1}{2}} + \mathbb{E}[|\log(q(Y))|^2]^{\frac{1}{2}} \leq \sqrt{136} + \sqrt{C(a,c)}, \tag{18}$$

implying

$$D(\mu||\nu) \leq 2\log(R) d_{TV}(\mu,\nu) + (\sqrt{136} + \sqrt{C(a,c)})\sqrt{2}e^{-\frac{1}{12}\sqrt{\frac{1}{a}\log(\frac{R}{c})}}.$$

Putting $t = \sqrt{\log(\frac{R}{c})}$, the above inequality reads

$$D(\mu||\nu) \leq 2\log(c) d_{TV}(\mu,\nu) + 2d_{TV}(\mu,\nu)t^2 + (\sqrt{136} + \sqrt{C(a,c)})\sqrt{2}e^{-\frac{t}{12\sqrt{a}}}.$$

Choosing $t = 12\sqrt{a}\log\left(\frac{\left(\sqrt{136}+\sqrt{C(a,c)}\right)\sqrt{2}}{24\sqrt{a}d_{TV}(\mu,\nu)}\right)$, which is nonnegative, and using the bound

$$\sqrt{C(a,c)} = \sqrt{(a + \log(c))^2 + 1943a^2} \leq 46a + \log(c)$$

yields the desired result. □

As a last illustration, the next result provides a comparison between total variation distance and χ^2-divergence, under an extra moment assumption.

Theorem 3.6. *Let $a > 0$ and $c \geq 1$. Let ν be a measure on $\mathbb{Z}$ whose p.m.f. belongs to the class $\mathcal{Q}(a,c)$. Let μ be an isotropic log-concave measure on $\mathbb{Z}$. Under the moment assumption*

$$\mathbb{E}[e^{2aY^2}] < +\infty,$$

where Y denotes a random variable with distribution μ, one has

$$d_{\chi^2}(\mu||\nu) \leq c\left(d_{TV}(\mu,\nu) + \sqrt{d_{TV}(\mu,\nu)}\right) + c\sqrt{\mathbb{E}[e^{2aY^2}]}\sqrt{2}e^{-\frac{1}{12}\sqrt{\frac{1}{a}\log\left(1+\frac{1}{\sqrt{d_{TV}(\mu,\nu)}}\right)}}.$$

Proof. Recall the notation $W = p(Y)/q(Y)$ from Lemma 3.4. With the choice of convex function $f(x) = (x-1)^2$, $x \geq 0$, Lemma 3.4 tells us that for all $R \geq c$,

$$d_{\chi^2}(\mu||\nu) \leq R d_{TV}(\mu,\nu) + \sqrt{\mathbb{E}\left[\frac{(W-1)^4}{W^2}1_{\{W>1\}}\right]}\sqrt{2}e^{-\frac{1}{12}\sqrt{\frac{1}{a}\log(\frac{R}{c})}}.$$

Note that

$$\mathbb{E}\left[\frac{(W-1)^4}{W^2}1_{\{W>1\}}\right] \leq \mathbb{E}\left[W^2\right] \leq \mathbb{E}\left[\frac{1}{q^2(Y)}\right] \leq c^2\mathbb{E}[e^{2aY^2}],$$

where the last inequality comes from $q \in \mathcal{Q}(a,c)$. Therefore,

$$d_{\chi^2}(\mu||\nu) \leq Rd_{TV}(\mu,\nu) + c\sqrt{\mathbb{E}[e^{2aY^2}]}\sqrt{2}e^{-\frac{1}{12}\sqrt{\frac{1}{a}\log\left(\frac{R}{c}\right)}}.$$

Choosing $R = c\left(1 + \frac{1}{\sqrt{d_{TV}(\mu,\nu)}}\right)$ yields the desired result. □

References

[1] H. Aravinda. Entropy-variance inequalities for discrete log-concave random variables via degree of freedom, to appear in Discrete Mathematics.

[2] S. G. Bobkov, G. P. Chistyakov. On concentration functions of random variables. J. Theor. Probab. 28 (2015), no. 3, 976-988.

[3] S. G. Bobkov, A. Marsiglietti, J. Melbourne. Concentration functions and entropy bounds for discrete log-concave distributions, Combin. Probab. Comput. 31 (2022), no. 1, 54-72.

[4] P. Brändén. Unimodality, log-concavity, real-rootedness and beyond. Handbook of enumerative combinatorics, Discrete Math. Appl., CRC Press:437-483, 2015.

[5] F. Brenti. Log-concave and unimodal sequences in algebra, combinatorics, and geometry: an update. In Jerusalem combinatorics '93, volume 178 of Contemp. Math., pages 71-89. Amer. Math. Soc., Providence, RI, 1994.

[6] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar. 2 (1967), 299-318.

[7] P. Cattiaux, A. Guillin. On the Poincaré constant of log-concave measures. Geometric Aspects of Functional Analysis: Israel Seminar (GAFA). 2017-2019. Volume I. LNM 2256, Springer Verlag, 171–217, 2020.

[8] R. M. Dudley. Distances of probability measures and random variables. Ann. Math. Statist. 39 (1968), 1563-1572.

[9] R. M. Dudley. Real analysis and probability.(English summary) Revised reprint of the 1989 original Cambridge Stud. Adv. Math., 74 Cambridge University Press, Cambridge, 2002. x+555 pp.

[10] A. L. Gibbs, F. E. Su. On choosing and bounding probability metrics. Preprint, arXiv:math/0209021.

[11] J. Jakimiuk, D. Murawski, P. Nayar, S. Słobodianiuk. Log-concavity and discrete degrees of freedom. Preprint, arXiv:2205.04069.

[12] F. Liese, I. Vajda. Convex statistical distances. With German, French and Russian summaries Teubner-Texte Math., 95[Teubner Texts in Mathematics] BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987. 224 pp.

[13] A. Marsiglietti, J. Melbourne. Moments, concentration, and entropy of log-concave distributions. Preprint, arXiv:2205.08293.

[14] A. Marsiglietti, P. Pandey. On the equivalence of statistical distances for isotropic convex measures, Mathematical Inequalities & Applications, Volume 25, Issue 3, 881-901, 2022.

[15] J. L. Massey. On the entropy of integer-valued random variables. In: Proc. 1988 Beijing Int. Workshop on Information Theory, pages C1.1-C1.4, July 1988.

[16] E. Meckes, M. Meckes. On the Equivalence of Modes of Convergence for Log-Concave Measures. In Geometric aspects of functional analysis (2011/2013), volume 2116 of Lecture Notes in Math., pages 385-394. Springer, Berlin, 2014.

[17] J. Melbourne, G. Palafox-Castillo. A discrete complement of Lyapunov's inequality and its information theoretic consequences. Preprint, arXiv:2111.06997.

[18] M. S. Pinsker. Information and information stability of random variables and processes. Translated and edited by Amiel Feinstein Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1964, xii+243 pp.

[19] S. T. Rachev, L. B. Klebanov, S. V. Stoyanov, and F. J. Fabozzi. The methods of distances in the theory of probability and statistics. Springer-Verlag, New York, 2013.

[20] A. Saumard, J. A. Wellner. Log-concavity and strong log-concavity: a review. Stat. Surv. 8, 45-114, 2014.

[21] R. P. Stanley. Log-concave and unimodal sequences in algebra, combinatorics, and geometry. In Graph theory and its applications: East and West (Jinan, 1986), volume 576 of Ann. New York Acad. Sci., pages 500-535. New York Acad. Sci., New York, 1989.

[22] R. Vershynin. High-dimensional probability. An introduction with applications in data science. With a foreword by Sara van de Geer. Cambridge Series in Statistical and Probabilistic Mathematics, 47. Cambridge University Press, Cambridge, 2018. xiv+284 pp.

Arnaud Marsiglietti
Department of Mathematics
University of Florida
Gainesville, FL 32611, USA
a.marsiglietti@ufl.edu

Puja Pandey
Department of Mathematics
University of Florida
Gainesville, FL 32611, USA
puja.pandey@ufl.edu