Journal of Theoretical Probability (2025) 38:37
https://doi.org/10.1007/s10959-025-01405-0

Exponential Families, Rényi Divergence and the Almost Sure Cauchy Functional Equation

Gérard Letac[1,2] **· Mauro Piccioni**[3]

Received: 15 April 2024 / Revised: 14 December 2024 / Accepted: 16 January 2025
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

Abstract

If $P_1, \dots, P_n$ and $Q_1, \dots, Q_n$ are probability measures on $\mathbb{R}^d$ and $P_1 * \cdots * P_n$ and $Q_1 * \cdots * Q_n$ are their respective convolutions, the Rényi divergence D_λ of order $\lambda \in (0, 1]$ satisfies $D_\lambda(P_1 * \cdots * P_n || Q_1 * \cdots * Q_n) \leq \sum_{i=1}^n D_\lambda(P_i || Q_i)$. When P_i belongs to the natural exponential family generated by Q_i, with the same natural parameter θ for any $i = 1, \dots, n$, the equality sign holds. The present note tackles the inverse problem, namely "does the equality $D_\lambda(P_1 * \cdots * P_n || Q_1 * \cdots * Q_n) = \sum_{i=1}^n D_\lambda(P_i || Q_i)$ imply that P_i belongs to the natural exponential family generated by Q_i for every $i = 1, \dots, n$?" The answer is not always positive and depends on the set of solutions of a generalization of the celebrated Cauchy functional equation. We discuss in particular the case $P_1 = \cdots = P_n = P$ and $Q_1 = \cdots = Q_n = Q$, with $n = 2$ and $n = \infty$, the latter meaning that the equality holds for all n. Our analysis is mainly devoted to P and Q concentrated on non-negative integers, and P and Q with densities with respect to the Lebesgue measure. The results cover the Kullback–Leibler divergence (KL), this being the Rényi divergence for $\lambda = 1$. We also show that the only f-divergences such that $D_f(P^{*2} || Q^{*2}) = 2D_f(P || Q)$, for P and Q in the same exponential family, are mixtures of KL divergence and its dual.

Keywords Rényi divergence · Kullback–Leibler divergence · f-divergence · Natural exponential family · Pexider functional equation · Cauchy functional equation

Mathematics Subject Classification 62E10 · 62B10 · 46N30

✉ Mauro Piccioni
mauro.piccioni@uniroma1.it

[1] Institut de Mathématiques de Toulouse, Université Paul Sabatier, 31009 Toulouse, France

[2] TéSA, 7 Bd de la Gare, 31500 Toulouse, France

[3] Dipartimento di Matematica, Sapienza Università di Roma, 00185 Rome, Italy

Published online: 06 February 2025

1 Introduction

A divergence between two probability measure on the same state space quantifies their dissimilarity. The Rényi divergence of order $\lambda \in (0, 1)$ has nice properties from this point of view; moreover, the divergence that is mostly privileged because of its statistical properties, the Kullback–Leibler divergence, can be obtained as a limit of Rényi divergences as $\lambda \uparrow 1$. For these divergences, the question considered in the paper is the following: when the divergence between two convolutions (i.e. the law of the sum when the summands are independent) on $\mathbb{R}^d$ is equal to the sum of the divergences between the laws of each summand? This always happens when these alternative laws belong to the same natural exponential family, with the same natural parameter for each summand. But since this could happen also in other instances, the object of the paper is to exclude this possibility in a wide range of cases, mainly when the summand are identically distributed, additionally to being mutually independent. Here are the main objects that will be needed along the discussion.

Rényi divergence

Suppose that P and Q are two probabilities on the same measurable space Ω, with P equivalent to Q (i.e. P and Q have the same family of null sets) and $\lambda \in (0, 1)$. The Rényi divergence $D_\lambda(P||Q)$ of order λ is defined by

$$D_\lambda(P||Q) = -\frac{1}{1-\lambda} \log \int_\Omega \left(\frac{dQ}{dP}\right)^{1-\lambda} dP = -\frac{1}{1-\lambda} \log \int_\Omega \left(\frac{dP}{dQ}\right)^{\lambda} dQ. \quad (1)$$

Since for any $x > 0$ and any $\lambda \in (0, 1)$ it holds $x^{1-\lambda} \geq \min(1, x)$, then

$$\begin{aligned} D_\lambda(P||Q) &\leq -\frac{1}{1-\lambda} \log \int_\Omega \min\left(1, \frac{dQ}{dP}\right) dP \\ &= -\frac{1}{1-\lambda} \log \left\{ P\left(\frac{dQ}{dP} > 1\right) + Q\left(\frac{dP}{dQ} \geq 1\right) \right\} < +\infty. \end{aligned}$$

Indeed, the two summands inside the logarithm never vanish simultaneously. Moreover, the function $x \mapsto x^{1-\lambda}$ defined on $(0, +\infty)$ being strictly concave for any $0 < \lambda < 1$, by Jensen's inequality $D_\lambda(P||Q) > 0$ unless $P = Q$, in which case the divergence vanishes. The reader is addressed to [1] for a general proof that

$$\lim_{\lambda \uparrow 1} D_\lambda(P||Q) = -\int_\Omega \log \frac{dQ}{dP} dP =: D_1(P||Q), \quad (2)$$

where D_1 is the Kullback–Leibler (KL) divergence. However, while the Rényi divergence is always finite, for any $\lambda \in (0, 1)$, the KL divergence of P w.r.t. Q can be infinite, even if P and Q are equivalent. A basic reference on KL and Rényi divergences is the book [2]. Another rather exhaustive summary of the properties of the Rényi divergence is the already cited [1].

Natural exponential families

In the sequel $\Omega = \mathbb{R}^d$ is Euclidean with its usual inner product $\langle , \rangle$. Consider a probability measure Q on $\mathbb{R}^d$, and define its Laplace transform

$$\mathcal{L}_Q(\theta) = \int_{\mathbb{R}^d} e^{\langle x,\theta\rangle} Q(dx) \leq +\infty, \quad \theta \in \mathbb{R}^d.$$

Let $\mathcal{D}(Q) = \{\theta \in \mathbb{R}^d : \mathcal{L}_Q(\theta) < +\infty\}$, and for any $\theta \in \mathcal{D}(Q)$ define the probability measure

$$P_Q^\theta(dx) = e^{\langle \theta,x\rangle - \kappa_Q(\theta)} Q(dx), \tag{3}$$

where $\kappa_Q(\theta) = \log \mathcal{L}_Q(\theta)$ is called the cumulant function of Q. The family (3) is called the natural exponential family generated by Q [3]. By Holder's inequality the parameter set $\mathcal{D}(Q)$ is convex and the function κ_Q is convex as well. For any $\theta \in \mathcal{D}(Q)$ it is immediately verified that

$$D_\lambda(P_Q^\theta||Q) = -\frac{1}{1-\lambda} \log \int_{\mathbb{R}^d} e^{\lambda(\langle \theta,x\rangle - \kappa_Q(\theta))} Q(dx) = \frac{\lambda\kappa_Q(\theta) - \kappa_Q(\lambda\theta)}{1-\lambda}, \ \ \lambda \in (0, 1).$$

Furthermore, the function $\kappa(\lambda\theta)$ is differentiable in $\lambda \in (0, 1)$, with derivative $\int \langle \theta, x\rangle P_Q^{\lambda\theta}(dx)$. However, differentiability could be lost in $\lambda = 1$, when θ lies in the relative boundary $\partial\mathcal{D}(Q)$ of $\mathcal{D}(Q)$, the set of $\theta \in \mathcal{D}(Q)$ such that $\lambda\theta \notin \mathcal{D}(Q)$, for all $\lambda > 1$. Conversely, for $\theta \in \mathcal{D}^0(Q) := \mathcal{D}(Q)\backslash\partial\mathcal{D}(Q)$, by applying de l'Hopital's rule and (2)

$$D_1(P_Q^\theta||Q) = \int \langle \theta, x\rangle P_Q^\theta(dx) - \kappa_Q(\theta) < +\infty.$$

Finally notice that the natural exponential family generated by the product probability $Q_1 \otimes \cdots \otimes Q_n$ on $\mathbb{R}^{dn}$ coincides with the family of all product laws with marginals belonging to the natural exponential families generated by Q_i, $i = 1, \ldots, n$, i.e.

$$P_{Q_1\otimes\cdots\otimes Q_n}^{(\theta_1,\ldots,\theta_n)} = P_{Q_1}^{\theta_1} \otimes \cdots \otimes P_{Q_n}^{\theta_n}, \theta_i \in \mathcal{D}(Q_i), \ \ i = 1, \ldots, n, \tag{4}$$

with $\kappa_{Q_1\otimes\cdots\otimes Q_n}(\theta_1, \ldots, \theta_n) = \sum_{i=1}^n \kappa_{Q_i}(\theta_i)$.

Since the Rényi divergence and the KL divergence tensorize, for any $\lambda \in (0, 1]$

$$D_\lambda(P_{Q_1}^{\theta_1} \otimes \cdots \otimes P_{Q_n}^{\theta_n}||Q_1 \otimes \cdots \otimes Q_n) = \sum_{i=1}^n D_\lambda(P_{Q_i}^{\theta_i}||Q_i).$$

Recalling that a convolution product $*$ of n laws is the distribution of the sum of independent random variables having these laws as marginals, we observe that

$$\frac{dP_{Q_1}^\theta * \cdots * P_{Q_n}^\theta}{dQ_1 * \cdots * Q_n}(y) = E^{Q_1\otimes\cdots\otimes Q_n}\left\{\frac{d(P_{Q_1}^\theta \otimes \cdots \otimes P_{Q_n}^\theta)}{d(Q_1 \otimes \cdots \otimes Q_n)}(X_1, .., X_n)|\sum_{i=1}^n X_i = y\right\}$$

$$= E^{Q_1\otimes\cdots\otimes Q_n}\{e^{\langle\theta,\sum_{i=1}^n X_i\rangle-\sum_{i=1}^n k_{Q_i}(\theta)}|\sum_{i=1}^n X_i = y\}$$
$$= e^{\langle\theta,y\rangle-\sum_{i=1}^n k_{Q_i}(\theta)},$$

for any $\theta \in \cap_{i=1}^n \mathcal{D}(Q_i)$. This implies that the family induced by convolution of the marginals from the family $\{P^\theta_{Q_1} \otimes \cdots \otimes P^\theta_{Q_n}, \theta \in \cap_{i=1}^n \mathcal{D}(Q_i)\}$ on $\mathbb{R}^{dn}$ is a natural exponential family on $\mathbb{R}^d$, generated by $Q_1 * \cdots * Q_n$, with $\kappa_{Q_1*\cdots*Q_n}(\theta) = \sum_{i=1}^n \kappa_{Q_i}(\theta)$. Notice that the former is a subfamily of the whole family of product laws with independent components belonging to the natural exponential families generated by $Q_1, \ldots, Q_n$, since the parameter θ is the same for each of the n components.

Proposition 1.1 *For any* $\lambda \in (0, 1]$

$$D_\lambda(P^\theta_{Q_1} * \cdots * P^\theta_{Q_n}||Q_1 * \cdots * Q_n) = \sum_{i=1}^n D_\lambda(P^\theta_{Q_i}||Q_i),\ \ \theta \in \cap_{i=1}^n \mathcal{D}(Q_i), \tag{5}$$
$$D_1(P^\theta_{Q_1} * \cdots * P^\theta_{Q_n}||Q_1 * \cdots * Q_n) = \sum_{i=1}^n D_1(P^\theta_{Q_i}||Q_i),\ \ \theta \in \cap_{i=1}^n \mathcal{D}^0(Q_i). \tag{6}$$

In particular, when $Q_i =: Q$, *then* $P^\theta_{Q_i} = P^\theta_Q =: P$, *for* $i = 1, \ldots, n$, *hence the l.h.s. in* (5) *e* (6) *are proportional to n, that is for* $0 < \lambda \leq 1$ *we have*

$$D_\lambda(P^{*n}||Q^{*n}) = nD_\lambda(P||Q)), \tag{7}$$

where for $\lambda = 1$ *the formula is always valid as long as* $\theta \in \mathcal{D}^0(Q)$.

Proof The result is immediately obtained by translating (5) and (6) in terms of the cumulant functions κ,

$$\frac{1}{1-\lambda}\left(\lambda\kappa_{Q_1*\cdots*Q_n}(\theta) - \kappa_{Q_1*\cdots*Q_n}(\lambda\theta)\right) = \frac{1}{1-\lambda}\sum_{i=1}^n(\lambda\kappa_{Q_i}(\theta) - \kappa_{Q_i}(\lambda\theta)),$$
$$\int \langle\theta, x_1 + \ldots + x_n\rangle P^\theta_{Q_1}(dx_1)\cdots P^\theta_{Q_n}(dx_n) - \kappa_{Q_1\otimes\cdots\otimes Q_n}(\theta)$$
$$= \sum_{i=1}^n\left(\int \langle\theta, x\rangle P^\theta_{Q_i}(dx) - \kappa_{Q_i}(\theta)\right).$$

□

Remark At this point one can raise the following question: are the Rényi and KL divergences the unique divergences such that the analogue of (7) holds for any n, when P and Q belonging to the same natural exponential family? Since the very definition of divergence is quite different from author to author, it is difficult to give a general answer. However, the interesting class of f-divergences deserves consideration [2]:

when f is a convex function defined on $(0, \infty)$ such that $f(1) = 0$, and if P and Q are equivalent probabilities we define the f-divergence by

$$D_f(P||Q) = \int_\Omega f\left(\frac{dP}{dQ}\right) dQ$$

This number is non negative by the Jensen inequality. This class of f-divergences includes KL divergence $D_1(P||Q)$ and its dual $\tilde{D}_1(P||Q) = D_1(Q||P)$ by taking, respectively, $f(x) = x \log x$ and $f(x) = -\log x$. Note that the Rényi divergence with parameter $0 < \lambda < 1$ is not an f-divergence. Section 5 below shows that the only f-divergences satisfying the analogue of (7) have the form $D_f = AD_1 + B\tilde{D}_1$ with $A, B \geq 0$ and $A + B > 0$.

Inequalities between Rényi and KL divergences under convolutions
In the next section, it will be recalled that, in general, if P_i and Q_i are equivalent probabilities on $\mathbb{R}^d$, for $i = 1, \ldots, n$, and $\lambda \in (0, 1]$

$$D_\lambda(P_1 * \cdots * P_n||Q_1 * \cdots * Q_n) \leq \sum_{i=1}^n D_\lambda(P_i||Q_i), \tag{8}$$

where $D_1(P_i||Q_i) < \infty$ is assumed for $i = 1, \ldots, n$, when $\lambda = 1$. Next, setting $f_i = \frac{dP_i}{dQ_i}$, for $i = 1, \ldots, n$, and $g_n = \frac{dP_1 * \cdots * P_n}{dQ_1 * \cdots * Q_n}$, it will be established that the equality holds in (8) if and only if the following equality holds

$$f_1(x_1) + \cdots + f_n(x_n) = g_n(x_1 + \cdots + x_n), \quad Q_1 \otimes \cdots \otimes Q_n - \text{ a.s. in } x_1, \ldots, x_n. \tag{9}$$

Indeed, in the situation considered in Proposition 1.1, $P_i = P^\theta_{Q_i}$, $f_i(x_i) = \langle \theta, x_i \rangle - \kappa_{Q_i}(\theta)$ for $i = 1, \ldots, n$ and $g_n(x) = \langle \theta, x \rangle - \sum_{i=1}^n \kappa_{Q_i}(\theta)$ clearly satisfy (9). If affine functions are the only solutions of the above equation (9), then the choice $P_i = P^\theta_{Q_i}$ for some value of θ, for $i = 1, \ldots, n$, is compulsory to reach the equality in (8). This drives our interest to the study of the general solutions of the functional almost sure equation (9) known in the literature as the Pexider equation [4]. By suitable techniques, adapted to the almost sure nature of the equation, we reduce this equation to the more familiar (almost sure) Cauchy functional equation. Focussing on the i.i.d. case $P_i = P$ and $Q_i = Q$, for $i = 1, \ldots, n$, we will present in Sect. 3 and Sect. 4 some relevant classes of examples, in the discrete and the absolutely continuous case, where it can be proved that the solutions to (9) are only affine functions. In the absolutely continuous case we will stress the relation with the classical Cauchy functional equation [5].

2 Divergences and Convolutions, Pexider and Cauchy Functional Equations

In order to formulate our first lemma, for any pair of probability distributions P_1, P_2 on $\mathbb{R}^d$, denote by $K_s^{P_1,P_2}$ the stick-breaking kernel, that is the law of X_1 given $X_1+X_2=s$, where $X_i \sim P_i$, for $i=1,2$, X_1 and X_2 being independent. Then the following holds:

Lemma 2.1 *Let P_i and Q_i be probability measures on $\mathbb{R}^d$, with P_i equivalent to Q_i, for $i=1,2$, and define*

$$f_i := \log \frac{dP_i}{dQ_i}, \;\; i=1,2, \;\; g_2 := \log \frac{d(P_1 * P_2)}{d(Q_1 * Q_2)}.$$

Then $Q_1 * Q_2$-a.s. w.r.t. s, $K_s^{P_1,P_2}$ is equivalent to $K_s^{Q_1,Q_2}$, with

$$\log \frac{dK_s^{P_1,P_2}}{dK_s^{Q_1,Q_2}}(x) = f_1(x) + f_2(s-x) - g_2(s), \;\; K_s^{Q_1,Q_2} - a.s. \text{ in } x. \tag{10}$$

Proof Denote by m_{P_1,P_2} the joint distributions of (X_1, X_1+X_2) when $X_1 \sim P_1$, $X_2 \sim P_2$, X_1 and X_2 being independent. The function $(x_1,x_2) \mapsto (x_1, x_1+x_2) = (x,s)$ has the inverse $(x,s) \mapsto (x, s-x) = (x_1,x_2)$. As a consequence

$$\frac{dm_{P_1,P_2}}{dm_{Q_1,Q_2}}(x,s) = \frac{d(P_1 \otimes P_2)}{d(Q_1 \otimes Q_2)}(x, s-x) = e^{f_1(x)+f_2(s-x)} \tag{11}$$

Next consider the function $H(x,s) = \exp\{f_1(x) + f_2(s-x) - g_2(s)\}$, and take any Borel sets A and B of $\mathbb{R}^d$

$$\begin{aligned}
&\int_B (P_1 * P_2)(ds) \int_A H(x,s) K_s^{Q_1,Q_2}(dx) \\
&= \int_B e^{g_2(s)} (Q_1 * Q_2)(ds) \int_A H(x,s) K_s^{Q_1,Q_2}(dx) \\
&= \int_B (Q_1 * Q_2)(ds) \int_A e^{f_1(x)+f_2(s-x)} K_s^{Q_1,Q_2}(dx) \\
&= \int_A \int_B e^{f_1(x)+f_2(s-x)} m_{Q_1,Q_2}(dx, ds) \\
&= \int_A \int_B m_{P_1,P_2}(dx, ds),
\end{aligned}$$

proving that $H(x,s) = \left(\frac{dK_s^{P_1,P_2}}{dK_s^{Q_1,Q_2}}\right)(x)$ as promised. □

The next lemma is the chain rule for the Rényi and the KL divergence, specialized for the partial sums we are interested in. The chain rule for the KL divergence is well known, see [2], Theorem 2.14 c. The chain rule for the Rényi divergence, appearing in its basic form at (7.71)-(7.72) of [2], is much less known. In order to formulate it, for any k-tuple of distributions on $\mathbb{R}^d$, say $P_1, \ldots, P_k$, let $m_{P_1,\ldots,P_k}$ be the law of $(X_1, X_1 + X_2, \ldots, X_1 + X_2 + \ldots + X_k)$, where $X_i \sim P_i$ are mutually independent, $i = 1, \ldots, k$.

Lemma 2.2 *For P_k and Q_k equivalent probability measures on $\mathbb{R}^d$, $k = 1, \ldots, n$ let*

$$f_k := \log \frac{dP_k}{dQ_k},\ \ k = 1, \ldots, n,\ \ g_k := \log \frac{d(P_1 * \cdots * P_k)}{d(Q_1 * \cdots * Q_k)},\ \ k = 2, \ldots, n.$$

Next, for any $\lambda \in (0, 1]$, define the random variable

$$Z_k(S_k) = D_\lambda \left(K_{S_k}^{P_1 * \cdots * P_{k-1}, P_k} || K_{S_k}^{Q_1 * \cdots * Q_{k-1}, Q_k} \right),\ \ k = 2, \ldots, n, \tag{12}$$

where S_k is distributed according to

$$m_k^{(\lambda)}(ds_k) = c_{k,\lambda} \int \cdots \int \left(\frac{dm_{P_1 * \ldots * P_k, P_{k+1}, \ldots, P_n}}{dm_{Q_1 * \ldots * Q_k, Q_{k+1}, \ldots, Q_n}}(s_k, s_{k+1}^n) \right)^\lambda m_{Q_1 * \ldots * Q_k, Q_{k+1}, \ldots, Q_n}(ds_k, ds_{k+1}^n),$$

with $s_{k+1}^n = (s_{k+1}, \ldots, s_n)$, and

$$c_{k,\lambda} = \exp\{-(1-\lambda) D_\lambda (m_{P_1 * \ldots * P_k, P_{k+1}, \ldots, P_n} || m_{Q_1 * \ldots * Q_k, Q_{k+1}, \ldots, Q_n})\},$$

for $\lambda \in (0, 1)$, and S_k is distributed as $P_1 * \cdots * P_k$, for $\lambda = 1$.

Then for $\lambda \in (0, 1)$

$$\sum_{i=1}^n D_\lambda(P_i || Q_i) = D_\lambda(P_n^* || Q_n^*) - \frac{1}{1-\lambda} \sum_{k=2}^n \log \mathbb{E}\{e^{-(1-\lambda) Z_k(S_k)}\} \tag{13}$$

Moreover, if $D_1(P_i || Q_i) < +\infty$ for $i = 1, \ldots, n$, then

$$\sum_{i=1}^n D_1(P_i || Q_i) = D_1(P_n^* || Q_n^*) + \sum_{k=2}^n \mathbb{E} Z_k(S_k). \tag{14}$$

Proof From the tensorization property and the invariance of the D_λ divergences under measurable one-to-one transformations, here

$$(x_1, \ldots, x_n) \mapsto (x_1 + \ldots + x_n, x_1 + \ldots + x_{n-1}, \ldots, x_1 + x_2, x_1), \tag{15}$$

it follows that

$$\sum_{k=1}^{n} D_\lambda(P_k||Q_k) = D_\lambda(P_1 \otimes \cdots \otimes P_n||Q_1 \otimes \cdots \otimes Q_n) = D_\lambda(m_{P_1,\dots,P_n}||m_{Q_1,\dots,Q_n}).$$

Therefore the relation (14) is proved once we show that

$$\begin{aligned} &D_\lambda(m_{P_1*\cdots*P_{k-1},P_k,\dots,P_n}||m_{Q_1*\cdots*Q_{k-1},Q_k,\dots,Q_n}) \\ &\quad= D_\lambda(m_{P_1*\cdots*P_k,P_{k+1},\dots,P_n}||m_{Q_1*\cdots*Q_k,Q_{k+1},\dots,*Q_n}) \\ &\quad -\frac{1}{1-\lambda}\log \mathbb{E}\{e^{-(1-\lambda)Z_k(S_k)}\} \end{aligned} \tag{16}$$

for $k = 2, \dots, n$. In order to prove this relation note that, under any product probability measure, the law of $x_1 + \ldots + x_{k-1}$ conditional to $x_1 + \ldots + x_m,\ m = k, k+1, \dots, n$ coincides with the law conditional to $x_1 + \ldots + x_k$ only. In fact the additional information conveyed by the random variables $x_1 + \ldots + x_m,\ m = k+1, \dots, n$ concerns only the variables $x_m, m = k+1, \dots, x_n$, which are clearly independent of any function of $x_1, \dots, x_k$. This ensures that

$$\begin{aligned} &\frac{dm_{P_1*\cdots*P_{k-1},P_k,\dots,P_n}}{dm_{Q_1*\cdots*Q_{k-1},Q_k,\dots,Q_n}}(s_{k-1}, s_k^n) \\ &\quad= \frac{dm_{P_1*\cdots*P_k,P_{k+1},\dots,P_n}}{dm_{Q_1*\cdots*Q_k,Q_{k+1},\dots,Q_n}}(s_k^n)\frac{dK_{s_k}^{P_1*\cdots*P_{k-1},P_k}}{dK_{s_k}^{Q_1*\cdots*Q_{k-1},Q_k}}(s_{k-1}) \end{aligned}$$

$m_{Q_1*\cdots*Q_{k-1},Q_k,\dots,Q_n}$-a.s. in s_{k-1}^n. Raise to λ both sides of (16), integrate w.r.t. $m_{Q_1*\cdots*Q_{k-1},Q_k,\dots,Q_n}$ and divide by $\lambda - 1$: then the l.h.s. clearly agrees with $D_\lambda(m_{P_1*\cdots*P_{k-1},P_k,\dots,P_n}||m_{Q_1*\cdots*Q_{k-1},Q_k,\dots,Q_n})$. As far as the r.h.s. is concerned, it yields

$$\begin{aligned} &\frac{1}{\lambda-1}\log\int\cdots\int e^{-(1-\lambda)Z_k(s_k)}\left(\frac{dm_{P_1*\cdots*P_k,P_{k+1},\dots,P_n}}{dm_{Q_1*\cdots*Q_k,Q_{k+1},\dots,Q_n}}(s_k^n)\right)^\lambda \\ &\quad\times m_{Q_1*\cdots*Q_k,Q_{k+1},\dots,Q_n}(ds_k^n). \end{aligned} \tag{17}$$

Next, multiply and divide under the integral sign by

$$= \exp\{(\lambda-1)D_\lambda(m_{P_1*\cdots*P_k,P_{k+1},\dots,P_n}||m_{Q_1*\cdots*Q_k,Q_{k+1},\dots,Q_n})\},$$

reducing the expression in (17) to

$$\begin{aligned} &D_\lambda(m_{P_1*\cdots*P_k,P_{k+1},\dots,P_n}||m_{Q_1*\cdots*Q_k,Q_{k+1},\dots,*Q_n}) \\ &\quad+\int\cdots\int e^{-(1-\lambda)Z_k(s_k)}m_k^{(\lambda)}(ds_k) \end{aligned} \tag{18}$$

which has the promised form (13). With similar arguments, the statement (14) then follows from the chain rule for the KL divergence. □

Theorem 2.3 *Let $P_1, P_2, \ldots P_n$ and $Q_1, Q_2, \ldots, Q_n$ be probability measures on $\mathbb{R}^d$ with P_i equivalent to Q_i, and let $\lambda \in (0,1]$ (in case $\lambda = 1$ assume also $D_1(P_i||Q_i) < +\infty$), for $i = 1, \ldots, n$. Then*

$$\sum_{i=1}^{n} D_\lambda(P_i||Q_i) \geq D_\lambda(P_1 * \cdots * P_n||Q_1 * \cdots * Q_n). \tag{19}$$

Moreover, the equality in (19) holds if and only if $Q_1 \otimes \ldots \otimes Q_k$ a.s. in $x_1, \ldots, x_k$ we have

$$f_1(x_1) + \ldots + f_k(x_k) = g_k(x_1 + \cdots + x_k), \ \ k = 2, \ldots, n. \tag{20}$$

where f_k and g_k are defined as in Lemma 2.2, for $k = 1, \ldots, n$.

Proof The relation (19) is a consequence of the second term at the r.h.s. in (13) and (14). In the latter case this is obvious, whereas for the former notice that the random variable under the expectation sign is 1 only for $Z_k(S_k) = 0$, otherwise it lies in $(0, 1)$. The properties of the logarithm immediately yield the conclusion. Hence in both cases the r.h.s. vanishes only when $Z_k(S_k) = 0$ a.s. Now notice that in (14) S_k has the law of $P_1 * \cdots P_k$, which is equivalent to $Q_1 * \cdots Q_k$. In (13) S_k is distributed according to $m_k^{(\lambda)}$, which is the first marginal of a law equivalent to $m_{Q_1 * \cdots * Q_k, Q_{k+1}, \ldots, Q_n}$, hence it is equivalent to $Q_1 * \cdots * Q_k$ as well. Consequently

$$D\left(K_s^{P_1 * \cdots * P_{k-1}, P_k}||K_s^{Q_1 * \ldots * Q_{k-1}, Q_k}\right) = 0, \ \ Q_1 * \cdots * Q_k\text{-a.s. in } s$$

if and only if

$$g_{k-1}(s_{k-1}) + f_k(s - s_{k-1}) = g_k(s), \ \ K_s^{Q_1 * \ldots * Q_{k-1}, Q_k}\text{-a.s. in } s_{k-1}$$

by Lemma 2.1. This relation can be rewritten as

$$g_{k-1}(s_{k-1}) + f_k(x_k) = g_k(s_{k-1} + x_k), \tag{21}$$

$Q_1 * \cdots * Q_{k-1} \otimes Q_k$-almost surely in (s_{k-1}, x_k). By induction the relation (20) is obtained. □

The above results are conveniently supplemented by the following corollary.

Corollary 2.4 *Let $P_1, P_2, \ldots P_n$ and $Q_1, Q_2, \ldots, Q_n$ be probability measures on $\mathbb{R}^d$, with P_i equivalent to Q_i, for $i = 1, \ldots, n$, and let*

$$f_i(x) = \log \frac{dP_i}{dQ_i}(x), \ \ i = 1, ..., n, \ \ g_n(s) = \log \frac{d(P_1 * \cdots * P_n)}{d(Q_1 * \cdots * Q_n)}(s).$$

Suppose that

$$f_1(x_1) + \cdots + f_n(x_n) = G_n(x_1 + \cdots + x_n) \tag{22}$$

holds $Q_1(dx_1)\otimes\ldots\otimes Q_n(dx_n)$*-a.s. in* $x_1,\ldots,x_n$*, for some real measurable function* G_n*. Then*

a) $G_n(y)=g_n(y)$*,* $Q_1*\cdots*Q_n$*-a.s. with respect to* y*;*

b) *for any* $k=2,\ldots,n$ *it holds*

$$f_1(x_1)+\cdots+f_k(x_k)=g_k(x_1+\cdots+x_k)$$

$Q_1(dx_1)\otimes\ldots\otimes Q_k(dx_k)$*-a.s. with respect to* $(x_1,\ldots,x_k)$*, where*

$$g_k(s)=\log\frac{d(P_1*\cdots*P_k)}{d(Q_1*\cdots*Q_k)};$$

c) *Finally let* $\lambda\in(0,1]$ *and in case* $\lambda=1$ *assume also that* $D_1(P_i||Q_i)<+\infty$*, for* $i=1,\ldots,n$*. Then for any* $k=2,\ldots,n$ *it holds*

$$\sum_{i=1}^k D_\lambda(P_i||Q_i)=D_\lambda(P_1*\ldots*P_k||Q_1*\ldots*Q_k). \tag{23}$$

Proof Since

$$e^{f_1(x_1)+\cdots+f_n(x_n)}Q_1(dx_1)\ldots Q_n(dx_n)=e^{G_k(x_1+\cdots+x_n)}Q_1(dx_1)\ldots Q_n(dx_n),$$

by multiplying both sides by an arbitrary bounded positive $h(s)$ on $\mathbb{R}^d$ and integrating we get

$$\begin{aligned}
&\int_{\mathbb{R}^d} h(s)e^{g_n(s)}(Q_1*\ldots*Q_n)(ds)\\
&\quad=\int_{\mathbb{R}^d} h(s)(P_1*\ldots*P_n)(ds)\\
&\quad=\int_{\mathbb{R}^{nd}} h(x_1+\cdots+x_n)e^{f_1(x_1)+\cdots+f_n(x_n)}Q_1(dx_1)\ldots Q_n(dx_n)\\
&\quad=\int_{\mathbb{R}^{nd}} h(x_1+\cdots+x_n)e^{G_n(x_1+\cdots+x_n)}Q_1(dx_1)\ldots Q_n(dx_n)\\
&\quad=\int_{\mathbb{R}^d} h(s)e^{G_n(s)}(Q_1*\ldots*Q_n)(ds).
\end{aligned}$$

Since this is true for all such h, we get the result a). Next, for the proof of b), the R.-N. derivative of $P_1\otimes\ldots\otimes P_{n-1}$ w.r.t. $Q_1\otimes\ldots\otimes Q_{n-1}$ is given by

$$\exp\{f_1(x_1)+\ldots+f(x_{n-1})\}=\int_{\mathbb{R}^d}\exp\{g_n(x_1+\ldots+x_{n-1}+x_n)\}Q_n(dx_n)$$

by taking exponentials at both sides of (22) and integrating w.r..t. x_n. Observe that the equality clearly holds $Q_1 \otimes \cdots \otimes Q_{n-1}$. The r.h.s. is clearly a positive function of $x_1 + \ldots + x_{n-1}$, hence by taking logarithms at both sides one obtains b) for $k = n-1$. The general result is then obtained by induction. Finally, the proof of c) immediately follows from b) and Theorem 2.3. □

Note that all the previous results make reference to a completely arbitrary numbering of the variables. The following result allows to turn the Pexider a.s. equation (22) into a simpler form. In order to formulate it, for any vector $x_0 \in \mathbb{R}^d$ and any probability measure Q an $\mathbb{R}^d$ define its translation by x_0 by $Q^{x_0}(\cdot) = Q(x_0 + \cdot)$.

Proposition 2.5 *Suppose f_i are measurable functions on $\mathbb{R}^d$ and Q_i are probability measures on the same space, for $i = 1, \ldots, n$, that satisfy the a.s. Pexider functional equation*

$$f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n) = G_n(x_1 + \ldots + x_n), (Q_1 \otimes \ldots \otimes Q_n) \\ -a.s.\ w.r.t.\ (x_1, \ldots, x_n). \tag{24}$$

Then there exists $x_0^i \in \mathbb{R}^d$, $i = 1, \ldots, n$, such that the function

$$\tilde{G}_n(z) = G_n\left(\sum_{i=1}^n x_i^0 + z\right) - G_n\left(\sum_{i=1}^n x_i^0\right)$$

satisfies the a.s. Cauchy functional equation

$$\tilde{G}_n(z_1) + \ldots + \tilde{G}_n(z_n) = \tilde{G}_n(z_1 + \ldots + z_n), Q_1^{x_1^0} \otimes \ldots \otimes Q_n^{x_n^0} \\ -a.s.\ in\ (z_1, \ldots, z_n). \tag{25}$$

Conversely, any solution $\tilde{G}_n$ of the above equation generates a family of solutions of Eq. (24)

$$f_i(x) = \tilde{G}_n(x - x_i^0) + b_i, i = 1, \ldots, n, \quad G_n(x) \\ = \tilde{G}_n\left(x - \sum_{i=1}^n x_i^0\right) + \sum_{i=1}^n b_i, \tag{26}$$

where b_i are arbitrary real constants, for $i = 1, \ldots, n$.

Proof The first step is to observe that, by Fubini theorem, the set of n-tuples $(x_1^0, \ldots, x_n^0) \in \mathbb{R}^{dn}$ s.t.

$$f_1(x_1^0) + f_2(x_2^0) + \ldots + f_n(x_n^0) = G_n(x_1^0 + \ldots + x_n^0) \tag{27}$$

and simultaneously, for any $i = 1, \ldots, n$

$$f_i(x_i) + \sum_{j \neq i} f_j(x_j^0) = G_n(x_i + \sum_{j \neq i} x_j^0), Q_i - a.s.\ w.r.t.\ x_i, \tag{28}$$

has $Q_1 \otimes \ldots \otimes Q_n$-probability equal to 1, which makes possible to choose $(x_0^1, \ldots, x_0^n)$ with these properties. Next define

$$\tilde{f}_i(z) = f_i(x_i^0 + z) - f_i(x_i^0), i = 1, \ldots, n,$$

and subtract both sides of (27) from (28). Recalling the definition of $\tilde{G}_n$ we get, for any $i = 1, \ldots, n$

$$\tilde{f}_i(z) = \tilde{G}_n(x_i^0 + z), \;\; Q_i^{x_i^0} - a.s. \; w.r.t. \; z.$$

Finally subtract both sides of (27) from (24), getting the required relation (25). The relation (26) is immediately obtained, by observing that a relation holding $Q_1^{x_1^0} \otimes \ldots \otimes Q_n^{x_n^0}$-*a.s. in* $(z_1, \ldots, z_n)$ must hold $Q_1 \otimes \ldots \otimes Q_n$-*a.s. w.r.t.* $(x_1, \ldots, x_n)$ for $x_i = x_i^0 + z_i$, for $i = 1, \ldots, n$. □

It is clear that the correspondence established with the last result between the solutions of the Pexider a.s. functional equation (24) and the Cauchy a.s. functional equation (25) maps affine functions into linear functions. If $\tilde{G}_n(z) = \langle \theta, z \rangle$ solves (25), then the corresponding solutions f_i of equation (24), obtained in (26), with the choice of $b_i = -k_{Q_i^{x_i^0}}(\theta)$, are associated with the probability measures P_i^θ, for $i = 1, \ldots, n$ in the following way

$$\begin{aligned} \frac{dP_i^\theta}{dQ_i}(x) &= e^{f_i(x)} = e^{\tilde{G}_n(x - x_i^0) + b_i} = \exp\{\langle \theta, x - x_i^0 \rangle - k_{Q_i^{x_i^0}}(\theta)\} \\ &= \exp\{\langle \theta, x \rangle - k_{Q_i}.(\theta)\}, i = 1, \ldots, n \end{aligned}$$

The above probabilities are elements of the natural exponential family generated by Q_i, provided $\theta \in \mathcal{D}(Q_i)$, with

$$k_{Q_i^{x_i^0}}(\theta) = \log \int \exp\{\langle \theta, z \rangle\} Q_i^{x_i^0}(dz),$$

and

$$k_{Q_i}(\theta) = \log \int \exp\{\langle \theta, x \rangle\} Q_i(dx) = k_{Q_i^{x_i^0}}(\theta) + \langle \theta, x_i^0 \rangle, \;\; i = 1, \ldots, n.$$

3 The a.s. Pexider and Cauchy Equation in the Discrete Case

In this section, we study the n-Pexider a.s. equation (24) in the identically distributed case

$$f(x_1) + f(x_2) + \ldots + f(x_n) = G_n(x_1 + \ldots + x_n),$$

$$Q^{\otimes n} - a.s.\ w.r.t.\ (x_1, \ldots, x_n). \tag{29}$$

We also allow $n = \infty$, meaning that the above equations holds for any $n = 2, 3, \ldots$. To shorten the notation, for any $d \in \mathbb{N}$ and $n \geq 2$ or $n = \infty$ we introduce the set of probability measures in $\mathbb{R}^d$

$$\mathcal{H}_d^n = \{Q : \textit{ the equation (29) has only solutions Q-a.s. equal to an affine function}\}.$$

It is also plain that we can replace Q with any equivalent probability measure, without changing the set of solutions.

In this section, we will present some classes of $Q \in \mathcal{H}_d^2$ and $Q \in \mathcal{H}_d^\infty$, that are discrete, that is concentrated on a finite or countable set.

3.1 The General Discrete Case

It is clear that when Q is concentrated on a subset of $\mathbb{R}^d$ which is finite or countable, then equation (29) depends on Q only through its (set-theoretical) support $S = \{s \in \mathbb{R}^d : Q(\{s\}) = 0\}$. Note that in the discrete case we are allowed to choose $x_i^0 = x^0$ for $i = 1, \ldots, n$ in Proposition 2.5, which is not guaranteed to work in general. As a result, by translating and centring, we can reduce our interest to the Cauchy functional equation in S (assumed to contain 0 w.l.o.g.)

$$\tilde{G}_n(z_1) + \ldots + \tilde{G}_n(z_n) = \tilde{G}_n(z_1 + \ldots + z_n), \ z_i \in S, \ i = 1, \ldots, n, \tag{30}$$

where $\tilde{G}_n$ is defined in $S_n = \sum_{i=1}^n S$, and it is an extension of G_{n-1} to S_n. It is clear that the sequence of sets $\{S_n\}$ is non-decreasing and its union $S_\infty = \cup_{n=1}^\infty S_n$ is the additive semigroup generated by S (more precisely it is a monoid, because of the assumption $0 \in S$). Notice the equation (30) can be rewritten as

$$\tilde{G}_n(z_1) + \tilde{G}_n(z_1) = \tilde{G}_n(z_1 + z_2), \ z_1 \in S_m, \ z_2 \in S_{n-m}.$$

for any m and $n \in \mathbb{N}$. In case equation (30) holds for any positive integer n, the union of the graphs of the functions $\tilde{G}_n$, defines a function $\tilde{G}$ on S_∞, that satisfies

$$\tilde{G}(z_1) + \tilde{G}(z_2) = \tilde{G}(z_1 + z_2), \ z_1 \in S_\infty, \ z_2 \in S_\infty. \tag{31}$$

Note this equation has the same form of (30) with $n = 2$, and a possibly different domain S_∞. It says that $\tilde{G}$ is a homomorphism of the monoid S_∞ onto its image $\tilde{G}(S_\infty) \subset \mathbb{R}$.

The set $\mathcal{G}_n(S)$ of solutions of (30) is a linear space containing the linear functions, so its dimension is greater or equal than d. It is equal to d if and only if Q belongs to $\mathcal{H}_d^n$. By Corollary 2.3, the sequence $(\mathcal{G}_n(S), n \geq 2)$ is non-increasing w.r.t. inclusion; consequently the set of solutions to (30) for any integer $n \geq 2$ is the set $\mathcal{G}_\infty(S) := \cap_{n\geq 2}\mathcal{G}_n(S)$.

Remark It is easy to produce an S with the corresponding $\mathcal{G}_\infty(S)$ (hence all $\mathcal{G}_n(S)$ for $n \geq 2$) has dimension larger than d. For example, consider $S = \{0, 1, \sqrt{2}\} \subset \mathbb{R}$, with $S_\infty = \{x = n_1 + n_2\sqrt{2}, n_i \in \mathbb{N}_0, i = 1, 2\}$. A function $G \in \mathcal{G}_\infty(S)$, defined on $n_1 + n_2\sqrt{2} \in S_\infty$, has the form

$$G(n_1 + n_2\sqrt{2}) = n_1 G(1) + n_2 G(\sqrt{2}),$$

hence the space $\mathcal{G}_\infty(S)$ has dimension 2. These functions are linear only when $G(\sqrt{2}) = \sqrt{2}G(1)$.

3.2 S Contained in $\mathbb{N}_0$

The present subsection is devoted to the case $S \subset \mathbb{N}_0$. Even in this case, finding necessary and sufficient conditions on S to get a one-dimensional $\mathcal{G}_2(S)$ in general is a complicated combinatorial problem, which seems solvable only by brute-force computations. In the sequel, the notation $S\setminus\{0\} := S_0$ is needed. In the next proposition we consider two simple cases.

Proposition 3.3 *If $S = \{0, c, 2c, \dots, nc\}$, with $c, n \in \mathbb{N}$ or if $S \subset \mathbb{N}_0$ is an additive semigroup, then $\mathcal{G}_2(S)$ is one-dimensional: any Q supported by such an S lies in $\mathcal{H}_1^2$.*

Proof In the first case it is easily seen by induction on $k = 1, \dots, n$ that $G(kc) = kG(c)$, for any $G \in \mathcal{G}_2(S)$. There is nothing to prove for $k = 1$, and if it holds for $k - 1$ then

$$G(kc) = G((k-1)c + c) = G((k-1)c) + G(c) = (k-1)G(c) + G(c) = kG(c).$$

Choosing $a = \frac{G(c)}{c}$, one obtains $G(x) = ax$ for $x \in S$. For the latter assertion, observe that for any $x_0, y_0 \in S_0$, the assumption implies that

$$\{x_0, 2x_0, (y_0 - 1)x_0, y_0 x_0\} \subset S, \ \{y_0, 2y_0, (x_0 - 1)y_0, x_0 y_0\} \subset S.$$

Arguing as in the former case we prove that if $G \in \mathcal{G}_2(S)$, then $G(kx_0) = kG(x_0)$, for $k = 1, 2, \dots, y_0$ and $G(hy_0) = hG(y_0)$, for $h = 1, 2, \dots, x_0$. Setting $k = y_0$ and $h = x_0$, we get

$$G(x_0 y_0) = y_0 G(x_0) = x_0 G(y_0) \Rightarrow \frac{G(x_0)}{x_0} = \frac{G(y_0)}{y_0} =: a$$

and $G(x) = ax$ for $x \in S$ as before.

In the general case $S \subset \mathbb{N}_0$ we offer a general procedure to help to determine $\mathcal{G}_2(S)$. It consists of an iterative procedure producing a sequence of partitions $\{\mathcal{P}_k, k \in \mathbb{N}_0\}$ of $S_0 = S\setminus\{0\}$. These partitions have the property that for any $A \in \mathcal{P}_k$, and $x, y \in A$

$$\frac{G(x)}{x} = \frac{G(y)}{y}, \quad \forall G \in \mathcal{G}_2(S).$$

Here is the procedure.

1. Set $k = 0$ and $\mathcal{P}_0 = \{\{i\}, i \in S_0\}$;
2. For any $A \in \mathcal{P}_k$ set $\Sigma_A = A \cup (A + A)$;
3. For any $A, B \in \mathcal{P}_k$ define $A \leftrightarrow B$ when $\Sigma_A \cap \Sigma_B \neq \emptyset$;
4. Let $L_1, \ldots, L_{H_k}$ be the connected components of the graph $(\mathcal{P}_k, \leftrightarrow)$ and set

$$\mathcal{P}_{k+1} = \{\cup_{A \in L_1} A, \ldots, \cup_{A \in L_{H_k}} A\};$$

5. Set $k \leftarrow k + 1$;
6. Go to 2.

It is clear that $\mathcal{P}_{k+1}$ is coarser than $\mathcal{P}_k$ (in the sense that for any $A \in \mathcal{P}_k$, there exists $B \in \mathcal{P}_{k+1}$ such that $A \subset B$). If for some $k \in \mathbb{N}$ it is $\mathcal{P}_{k+1} = \mathcal{P}_k$ then $\mathcal{P}_\ell = \mathcal{P}_k$ for $\ell > k$ (this always happens when S is finite). The following result explains our interest in the partition $\mathcal{P}_k$. □

Theorem 3.5 *If $A \in \mathcal{P}_k$, then $\frac{G(x)}{x} = \frac{G(y)}{y}$ for $x, y \in A$, for any $G \in \mathcal{G}_2(S)$. Furthermore $H_k = |\mathcal{P}_k|$ is larger or equal than* $\dim \mathcal{G}_2(S)$. *In particular $H_k = 1$ implies that $\mathcal{G}_2(S)$ is one-dimensional.*

Proof The proof works by induction on k. For $k = 0$ the statement is trivial. Next assume that the statement is true for $k - 1$ and suppose that $B_r = \cup_{h \in L_r} A_h \in \mathcal{P}_k$ and $x, y \in B_r$, for some $r \in \{1, \ldots, H_k\}$. Then, either there exists $h \in L_r$ such that $x, y \in A_h$, in which case $\frac{G(x)}{x} = \frac{G(y)}{y}$ by the inductive assumption, or $x \in A_{h_1}$ and $y \in A_{h_t}$, with $A_{h_1}, A_{h_t} \in L_r$ and $h_1 \neq h_t$, with $A_{h_1} \leftrightarrow A_{h_2} \leftrightarrow \ldots \leftrightarrow A_{h_t}$, for some t-tuple $A_{h_i} \in L_r, i = 1, \ldots, t$. Suppose first that $t = 2$, in which case there exists $z \in \Sigma_{A_{h_1}} \cap \Sigma_{A_{h_2}}$. Next choose $G \in \mathcal{G}_2(S)$ arbitrarily. If $z \in A_{h_1}$, by the inductive hypothesis $\frac{G(x)}{x} = \frac{G(z)}{z}$; else $z = x_1 + x_2$ with $x_1, x_2 \in A_{h_1}$. Again by the inductive hypothesis $\frac{G(x_i)}{x_i} = c$, with $i = 1, 2$. Consequently, using the extension of f to S_2, we have

$$G(z) = G(x_1 + x_2) = G(x_1) + G(x_2) = c(x_1 + x_2) = cz$$

which proves that $\frac{G(z)}{z} = \frac{G(x)}{x}$. With the same argument, one proves that also $\frac{G(y)}{y} = \frac{G(z)}{z}$. If $t > 2$ one constructs a sequence $z_i \in \Sigma_{A_i} \cap \Sigma_{A_{i+1}}, i = 0, 1, \ldots, t-1$ such that for any $G \in \mathcal{G}_2(S)$ it holds

$$\frac{G(x)}{x} = \frac{G(z_1)}{z_1} = \ldots = \frac{G(z_{t-1})}{z_{t-1}} = \frac{G(y)}{y},$$

ending the proof of the first statement of the theorem. For the second statement, recall $\mathcal{P}_k = \{B_1, \ldots, B_{H_k}\}$. For $G \in \mathcal{G}_2(S)$ and $r = 1, \ldots, H_k$, let

$$c_r(G) = \frac{G(x)}{x}, \ x \in B_r.$$

The function $\mathbf{c}(G) = (c_1(G), \ldots, c_{H_k}(G))$ from $\mathcal{G}_2(S)$ into $\mathbb{R}^{H_k}$ is linear and injective, as its kernel is trivial. Consequently, the dimension of the domain does not exceed H_k. □

In general, when S is finite, the above procedure ends with a value of k such that $H_k \geq 1$. In Example 1, it happens that $\mathcal{G}_2(S)$ is one-dimensional. When $H_k > 1$, the complete knowledge of $\mathcal{G}_2(S)$ can be obtained by solving a system of linear equations with a number of H_k unknowns smaller than $|S_0|$. As illustrated in the last two forthcoming examples, it is then possible that $\mathcal{G}_2(S)$ is one-dimensional or not.

Example 1 $s = \{0, 2, 3, 4, 5, 10\}$. The partition $\mathcal{P}_0$ consists of the elements of S_0. To get the $\leftrightarrow$ edges observe that

$$\Sigma_2 = \{2, 4\}, \Sigma_3 = \{3, 6\}, \Sigma_4 = \{4, 8\}, \Sigma_5 = \{5, 10\}, \Sigma_{10} = \{10, 20\}$$

whose non-empty intersections are $\Sigma_2 \cap \Sigma_4$ and $\Sigma_5 \cap \Sigma_{10}$. Consequently $\mathcal{P}_1 = \{\{2, 4\}, \{5, 10\}, \{3\}\}$. Now in order to describe $\mathcal{P}_2$ let us compute

$$\Sigma_{\{2,4\}} = \{2, 4, 6, 8\}, \Sigma_{\{5,10\}} = \{5, 10, 15, 20\}, \Sigma_3 = \{3, 6\}.$$

and observe that $\Sigma_{\{2,4\}} \cap \Sigma_3 = \{6\}$ is the only non-empty intersection, implying $\mathcal{P}_2 = \{\{2, 3, 4\}, \{5, 10\}\}$. In order to compute $\mathcal{P}_3$ observe that

$$\Sigma_{\{2,3,4\}} = \{2, 3, 4, 5, 6, 7, 8\}$$

which has a non-empty intersection with $\Sigma_{\{5,10\}}$. Hence $\mathcal{P}_3$ has a single element that coincides with S_0.

Example 2 $S_0 = \{1, 3, 4, 5\}$. Since $\{\Sigma_i, i \in S_0\}$ are pairwise disjoint $\mathcal{P}_1 = \mathcal{P}_0 = \{\{i\}, i \in S_0\}$. However, the system

$$G(1) + G(5) = 2G(3), G(5) + G(3) = 2G(4),$$
$$G(4) = G(1) + G(3), G(5) = G(1) + G(4),$$

has only constant solutions, in other words $\mathcal{G}_2(S)$ is one-dimensional.

Example 3 $S_0 = \{3, 4, 8, 13\}$. Then

$$\Sigma_3 = \{3, 6\}, \Sigma_4 = \{4, 8\}, \Sigma_8 = \{8, 16\}, \Sigma_{13} = \{13, 26\}$$

showing that $\mathcal{P}_1 = \{\{3\}, \{4, 8\}, \{13\}\}$. Since $\Sigma_{\{4,8\}} = \{4, 8, 12, 16\}$, one obtains that $\mathcal{P}_2 = \mathcal{P}_1$. This means that for any $G \in \mathcal{G}_2$, setting $G(i) = ic_i$, we have $c_4 = c_8$. There is only one constraint on the unknowns c_3, c_4, c_{13}, obtained from the relation $3 + 13 = 2 \times 8$ namely

$$3c_3 + 13c_{13} = 16c_4$$

from which it is obtained that $\mathcal{G}_2(S)$ has dimension 2.

3.3 *S* Contained in $\mathbb{Z}$ and $\mathbb{Z}^d$.

The procedure presented above can be easily extended to cover the case $S \subset \mathbb{Z}$ with the following modifications. First, if there is no $i \in S$ such that $-i \in S$ as well, the procedure is performed separately on $S^+ = S \cap \mathbb{N}$ and $S^- = S \cap (-\mathbb{N})$. If $i \in S$ and $-i \in S$, then $\{i, -i\} \in \mathcal{P}_0$ and the procedure is kept unchanged, defining this time $\Sigma_A = A \cup (A + A)\setminus\{0\}$, for any $A \in \mathcal{P}_k$. In fact, with this modification, the procedure can be used for any finite or denumerable $S \subset \mathbb{R}$.

here are a couple of results with $S \subset \mathbb{Z}$ and $S \subset \mathbb{Z}^d$.

Proposition 3.4 *Let $0 \in S \subset \mathbb{Z}$. Then $\mathcal{G}_\infty(S)$ is one-dimensional and any Q supported by such an S lies in $\mathcal{H}_1^\infty$.*

Proof Recall that $\mathcal{G}_\infty(S) = \mathcal{G}_2(S_\infty)$, as a consequence, we consider (30) with $n = 2$. Denote $S^+ = S \cap \mathbb{N}$ and $S^- = (-S) \cap \mathbb{N}$, where $\mathbb{N} = \{1, 2, \ldots\}$. If one of the two is empty, the result follows from Proposition 3.3. So assume that they are both non-empty and let S_∞^+ and S_∞^- be the corresponding additive semigroups. Define $G^+(x) = G(x)$ for $x \in S_\infty^+$ and $G^-(x) = G(-x)$ for $x \in S_\infty^-$. If $x, y \in S_\infty^+$, then $xy \in S_\infty^+$ and since (30) is true, we have $G^+(xy) = xG^+(y) = yG^+(x)$. This implies that $G(x) = a_+x$ for $x \in S_\infty^+$, for some real constant a_+. If $x \in S_\infty^-$ and $y \in S_\infty^+$ then $xy \in S_\infty^-$ from which $G(x) = a_-x$ for $x \in S_\infty^-$, for some real constant a_-. Let c_+ and c_- the greatest common divisors of S_∞^+ and S_∞^-,, respectively, and let $c = \gcd(c_+, c_-)$. Then from the Bézout identity

$$S_\infty = S_\infty^+ - S_\infty^- = c\mathbb{Z}, \tag{32}$$

by consequence $x \in S_\infty$ if and only if $-x \in S_\infty$, and since $G(0) = 0$, from (30) it follows $G(x) = -G(-x)$, i.e. $a_+x = a_-x$, from which $a_+ = a_-$. □

Theorem 3.6 *If $0 \in S \subset \mathbb{Z}^d$ and 0 is also contained in the interior of the convex hull of S, then $\mathcal{G}_\infty(S)$ is d-dimensional.*

Proof From Theorem 2 in [6], S_∞ is a subgroup of $\mathbb{Z}^d$, which means that there exist $w_1, \ldots, w_d \in \mathbb{Z}^d$ linearly independent, such that any element x of S_∞ has the form

$$x = x_1 w_1 + \ldots + x_d w_d$$

where $x_1, \ldots, x_d \in \mathbb{Z}$. Moving each coordinate at a time, it is clear that equation

$$G(x) + G(y) = G(x + y), \;\; x, y \in S_\infty$$

with $G(0) = 0$, implies that $G(x) = \langle a, x \rangle$, for some $a \in \mathbb{R}^d$. □

4 The Absolutely Continuous Case

In this section we consider probability measures Q on $\mathbb{R}^d$ that are absolutely continuous w.r.t. the Lebesgue measure. As before, our goal is to determine suitable properties that ensure that $Q \in \mathcal{G}_d^2$ or $Q \in \mathcal{G}_d^\infty$. For establishing results of this kind, some of the tricks that have been used in the case of the classical Cauchy functional equation are still valid.

Theorem 4.2 *If Q has a positive density w.r.t. the Lebesgue measure in $\mathbb{R}^d$ on an open connected set O, and f satisfies*

$$f(x_1) + f(x_2) = G_2(x_1 + x_2), \quad Q \otimes Q\text{-a.s. in } (x_1, x_2), \tag{33}$$

then $f(x) = \langle a, x\rangle + b$, for some $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$. As a consequence, Q lies in $\mathcal{H}_d^2$.

Proof By assumption each property concerning points of $\mathbb{R}^d$ that holds Q-a.s., holds almost everywhere w.r.t. the Lebesgue measure restricted to O and conversely. The same is true for properties concerning pairs of points in $\mathbb{R}^d$, replacing Q by $Q^{\otimes 2}$ and the set O by $O \times O$. We start by applying Proposition 2.5: in the present situation we are not allowed to choose $x_0^1 = x_0^2$, as in the discrete case, but, due to the properties of the Lebesgue measure, these two vectors can be chosen as close as we wish. As a result the probability measures $Q^{x_0^1}$ and $Q^{x_0^2}$ have positive density in the open sets O_1 and O_2 containing the origin 0, respectively, obtained the one from the other by a translation, whose length can be made as small as we wish. As a consequence, for $\tilde{O} = O_1 \cap O_2$, it holds

$$\tilde{G}_2(u) + \tilde{G}_2(v) = \tilde{G}_2(u + v), \text{ a.s. in } (u, v) \in \tilde{O}^2. \tag{34}$$

where

$$\tilde{G}_2(z) = G_n\left(\sum_{i=1}^n x_i^0 + z\right) - G_n\left(\sum_{i=1}^n x_i^0\right)$$

By Fubini's theorem the equality (34) holds for any fixed $u \in \tilde{O} \setminus M^*$, $M^* \subset \tilde{O}$ being of d-dimensional Lebesgue measure zero, and for $v \in \tilde{O} \setminus N_u^*$, where $N_u^* \subset \tilde{O}$ has d-dimensional Lebesgue measure zero for any $u \in \tilde{O} \setminus M^*$. We aim to replace $\tilde{G}_2$ by a version such that (34) holds for all $u, v \in \tilde{O}$. Here, the technique used by [7] on the whole space $\mathbb{R}$ can be adapted as follows. Consider any $u \in \tilde{O}$. Then $u - M^*$ has Lebesgue measure 0 like M^*, and thus one can choose $u_1 \in \tilde{O} \setminus M^*$, such that $u - u_1 \in \tilde{O} \setminus M^*$; in fact, since $0 \in \tilde{O}$, u_1 can be chosen arbitrarily close to 0. From

$$-\tilde{G}_2(v) = \tilde{G}_2(u_1) - \tilde{G}_2(u_1 + v), \quad v \in \tilde{O} \setminus N_{u_1}^* \tag{35}$$

and

$$\tilde{G}_2(u - u_1 + w) = \tilde{G}_2(u - u_1) + \tilde{G}_2(w), \quad w \in \tilde{O} \setminus N_{u-u_1}^*$$

by choosing $w = u_1 + v$ one obtains from the latter

$$\tilde{G}_2(u + v) = \tilde{G}_2(u - u_1) + \tilde{G}_2(u_1 + v), \quad v \in (\tilde{O} \setminus N_{u-u_1}^*) - u_1, \tag{36}$$

and so, summing both sides of (35) and (36), we obtain

$$\tilde{G}_2(u+v) - \tilde{G}_2(v) = \tilde{G}_2(u-u_1) + \tilde{G}_2(u_1) =: h(u),\\ v \in (\tilde{O} \setminus N^*_{u_1}) \cap \{(\tilde{O} \setminus N^*_{u-u_1}) - u_1\}.$$

From (34), for $u \in \tilde{O} \setminus M^*$ the l.h.s. of the above display is equal to $\tilde{G}_2(u)$, almost everywhere in $v \in \tilde{O}$ w.r.t. the d-dimensional Lebesgue measure: this implies that $h(u) = \tilde{G}_2(u)$, for $u \in \tilde{O} \setminus M^*$, meaning that h is almost everywhere equal to $\tilde{G}_2$ in $\tilde{O}$. It remains to prove that h satisfies

$$h(u) + h(v) = h(u+v), \;\; u \in \tilde{O}, \;\; v \in \tilde{O}. \tag{37}$$

For this is enough to observe that for any choice of $u, v \in \tilde{O}$ it is possible to select $s, t \in \tilde{O}$, as close as we wish to the origin, such that

$$\begin{aligned}&\tilde{G}_2(u+s) - \tilde{G}_2(s) = h(u), \;\; \tilde{G}_2(v+t) - \tilde{G}_2(t) = h(v),\\ &\tilde{G}_2(s+t) - \tilde{G}_2(u+v+s+t) = -h(u+v),\\ &\tilde{G}_2(s) + \tilde{G}_2(t) - \tilde{G}_2(s+t) = 0,\\ &\tilde{G}_2(u+v+s+t) - \tilde{G}_2(u+s) - \tilde{G}_2(v+t) = 0,\end{aligned}$$

which summed from both sides yield exactly (37). In fact, each of these equalities fails in a subset of $\tilde{O}^2$ with a $2d$-dimensional Lebesgue measure zero. Since $0 \in \tilde{O}$, (37) implies also that $h(0) = 0$.

The next step is to prove that h is continuous in $\tilde{O}$. For this, we borrow a trick from [8]. Since $0 \in \tilde{O}$, for $\varepsilon > 0$ sufficiently small, one can define for $t \in \mathbb{R}$

$$\varphi(t) = \int_{B(0,\varepsilon)} e^{ith(y)} dy,$$

where $B(0, \varepsilon)$ is the ball of radius ε around 0. Now observe that, by dominated convergence, $\varphi(t)$ is nonzero if $|t| \leq \eta$, where $\eta > 0$ is small enough. Next, for any $x \in \tilde{O}$

$$\begin{aligned}\int_{B(x,\varepsilon)} e^{ith(s)} ds &= \int_{B(0,\varepsilon)} e^{ith(x+y)} dy\\ &= \int_{B(0,\varepsilon)} e^{it(h(x)+h(y))} dy = e^{ith(x)} \varphi(t).\end{aligned} \tag{38}$$

Again by dominated convergence, for all $|t| \leq \eta$ the function

$$x \in \tilde{O} \mapsto e^{ith(x)} = \frac{1}{\varphi(t)} \int_{B(x,\varepsilon)} e^{ith(s)} ds$$

is continuous. Finally for $\eta > 0$ sufficiently small

$$h(x) = \frac{i(e^{-i\eta h(x)} - e^{i\eta h(x)})}{\int\limits_{-\eta}^{\eta} e^{ith(x)} dt}$$

thus, by dominated convergence, we deduce that h is continuous as well.

Notice that until now we have not yet exploited the connectedness of O. First, since $0 \in \tilde{O}$, again for x inside a ball $B(0, r_0)$ of sufficiently small radius $r_0 > 0$ around 0, it is easily obtained that, h being continuous and satisfying (37), it is $h(x) = \langle a, x\rangle$ for some $a \in \mathbb{R}^d$. This is first established when x is aligned with each of the vectors in $\mathbb{R}^d$, following the standard one-dimensional argument in Chapter 2 in [5], and then using the decomposition of any vector in the ball as a sum of its projections on the vectors. Using (37), we obtain $h(x) - h(x_0) = \langle a, x - x_0\rangle$ for $x \in B(x_0, r_{x_0}) \subset O$, for any $x_0 \in O$ and $r_{x_0} > 0$ suitably small. Next, by connectedness, for any $x \in O$ there exists a continuous path contained in O, starting from the origin and ending in $x \in O$, that can be covered with a finite number of balls $\{B_i := B(x_i, r_i) \subset O, i = 1, \ldots, I\}$, with $r_i > 0$ and $y_i \in B_i \cap B_{i+1}, i = 1, \ldots, I - 1$. Then

$$\begin{aligned} h(x) = h(x) - h(0) &= \sum_{i=0}^{I-1}(h(x_{i+1}) - h(y_i) + h(y_i) - h(x_i)) \\ &= \sum_{i=0}^{I-1}(\langle a, x_{i+1} - y_i\rangle + \langle a, y_i - x_i\rangle) \\ &= \langle a, \sum_{i=0}^{I-1}(x_{i+1} - x_i)\rangle = \langle a, x\rangle, \end{aligned}$$

as desired. □

Remark the assumption of the previous theorem cannot be weakened. For example, with

$$Q(dx) = (1_{(0,1)}(x) + 1_{(3,4)}(x))\frac{dx}{2},$$

any function f equal to d_1 on $(0, 1)$ and to d_2 on $(3, 4)$ is a solution of (33) for $n = 2$ and does not have an affine form, unless $d_1 = d_2$. More generally, the intervals $(0, 1)$ and $(3, 4)$ can be replaced by (a_1, a_2) and (b_1, b_2), with $2a_2 < a_1 + b_1$ and $a_2 + b_2 < 2b_1$.

The disconnectedness of the support, which is crucial in the previous counterexample, can be circumvented by increasing n in (24). In the next result this is achieved by a rather direct argument, without reducing to an equation of the form (31) as we did in the discrete case.

Corollary 4.3 *assume Q is a probability measure on $\mathbb{R}^d$ which is absolutely continuous w.r.t. the Lebesgue measure (on the whole space), and that its density q is positive in some open set O. Then $Q \in \mathcal{H}_d^\infty$.*

Proof By a suitable translation if needed, assume that $0 \in O$, and that O is connected. Moreover suppose $Q(O) < 1$, otherwise Theorem 4.2 is directly applicable. The Borel set $O^* = \{x \notin O : q(x) > 0\}$ is defined up to null sets and has the property $Q(O) + Q(O^*) = 1$. If f satisfies the equation (24) for all values of n, define $f_O = f 1_O$. For any Borel set $A \subset \mathbb{R}^d$ of positive Q-probability denote by Q_A the probability measure Q conditional to A. Then f_O solves equation (33) with Q replaced by Q_O: consequently, by Theorem 4.2, $f_O(x) = \langle a, x \rangle$, for some $a \in \mathbb{R}^d$, almost everywhere in O w.r.t. the d-dimensional Lebesgue measure. Our goal is to prove that the same is true for $f_{O^*} = f 1_{O^*}$, that solves (33) with Q replaced by Q_{O^*}. For this purpose, note that the sequence $\{O_m = O + \ldots + O \ m\text{-times}, m \in \mathbb{N}\}$ increases to $\mathbb{R}^d$. Consequently, the sequence $\{O_m^* = O^* \cap O_m, m \geq m_0\}$ increases to O^*, and $Q(O_m^*) \uparrow Q(O^*)$ as $m \to \infty$, m_0 being the first value of m such that O_m^* has positive d-dimensional Lebesgue measure. As a consequence, by exploiting (24) in the subset of m-tuples with components in O, whose sum belongs to O_m^*, we have

$$f_{O^*}(x_1 + \ldots + x_m) = f_O(x_1) + \ldots + f_O(x_m)$$

a.e. in $x_1, \ldots, x_m$ w.r.t. the md-dimensional Lebesgue measure. Since $f_O(x) = \langle a, x \rangle$ a.e. w.r.t. the d-dimensional Lebesgue measure in O, $f_{O^*}(y) = \langle a, y \rangle$, for $y \in O_m^*$, a.e. w.r.t. the d-dimensional Lebesgue measure in O_m^*. By increasing m the proof is achieved. □

Remark It is worth observing that the assumption of Corollary 4.3 cannot be replaced by the weaker assumption that Q is absolutely continuous w.r.t. the d-dimensional Lebesgue measure. Indeed, already in dimension 1 there exist closed subsets F of $\mathbb{R}$ with positive Lebesgue measure, but without an interior point, excluding the uniform distribution on F from the statement of the result.

5 Exponential Families, *f* and KL Divergences

This section considers the problem raised in the remark at the end of Sect. 1. If f is a convex function on $(0, \infty)$ such that $f(1) = 0$, define the f-divergence between equivalent probabilities P and Q on some set Ω by

$$D_f(P||Q) = \int_\Omega f\left(\frac{dP}{dQ}\right) dQ.$$

The best examples are $f(x) = x\log x$, and $f(x) = -\log x$, leading to the KL divergence

$$D_1(P||Q) = -\int_\Omega \log\left(\frac{dQ}{dP}\right)dP = \int_\Omega \log\left(\frac{dP}{dQ}\right)dP.$$

and to its dual $\tilde{D}_1(P||Q) = D_1(Q||P)$.

We prove in this section that if P and Q belong to the same natural exponential family on $\mathbb{R}^d$ then the property

$$D_f(P^{*n}||Q^{*n}) = nD_f(P||Q) \tag{39}$$

holds essentially only for a linear combination of the KL divergence D_1 and its dual $\tilde{D}_1$. For showing this, we have to exhibit some natural exponential family and a positive integer n, with the property that the relation (39) holds for all P and Q members of the family, only when D_f has the form $AD_1 + B\tilde{D}_1$. After realizing that the normal, Gamma and Poisson families lead to untractable computations, the simplest choice reveals to be a successful one: the Bernoulli family, with $n = 2$, that is P^{*2} and Q^{*2} binomial.

Theorem 5.1 *Let P and Q are two Bernoulli distributions with means p and $q \neq p$, respectively. If $D_f(P^{*2}||Q^{*2}) = 2D_f(P||Q)$ holds for all $0 < q < p < 1$ then there exists three constants $A, B \geq 0$ and C such that*

$$f(x) = -A\log x + Bx\log x + C(x-1). \tag{40}$$

and then

$$D_f(P||Q) = AD_1(P||Q) + BD_1(Q||P).$$

Proof Here $\frac{dP}{dQ}(0) = (1-p)/(1-q)$ and $\frac{dP}{dQ}(1) = p/q$. Similarly

$$\frac{dP^{*2}}{dQ^{*2}}(0) = \left(\frac{1-p}{1-q}\right)^2, \ \frac{dP^{*2}}{dQ^{*2}}(1) = \frac{(1-p)p}{(1-q)q}, \ \frac{dP^{*2}}{dQ^{*2}}(2) = \left(\frac{p}{q}\right)^2$$

We have therefore to prove that

$$\begin{aligned} D_f(P^{*2}||Q^{*2}) &= f\left(\left(\frac{1-p}{1-q}\right)^2\right)(1-q)^2 + 2f\left(\frac{(1-p)p}{(1-q)q}\right)(1-q)q + f\left(\left(\frac{p}{q}\right)^2\right) \\ &= 2D_f(P||Q) = 2f\left(\frac{1-p}{1-q}\right)(1-q) + 2f\left(\frac{p}{q}\right)q \end{aligned} \tag{41}$$

implies that f has the form (40).

For simplification, we write $X = \frac{p}{q}$, $Y = \frac{1-p}{1-q}$. This implies that $q = \frac{1-Y}{X-Y}$, $1-q = \frac{1-X}{Y-X}$. Note that $0 < q < p < 1 \Leftrightarrow 0 < Y < 1 < X$. We do not need to consider the alternative case. for convenience we denote

$$D = \{(X, Y) : 0 < Y < 1 < X\}.$$

The functional equation (41) becomes for $(X, Y) \in D$

$$f(Y^2)\left(\frac{X-1}{X-Y}\right)^2 + 2f(XY)\frac{(X-1)(1-Y)}{(X-Y)^2} + f(X^2)\left(\frac{1-Y}{X-Y}\right)^2 = f(Y)\frac{X-1}{X-Y} + 2f(X)\frac{1-Y}{X-Y} \tag{42}$$

that we rather rewrite as

$$(X-1)^2 f(Y^2) + 2(X-1)(1-Y)f(XY) + (1-Y)^2 f(X^2) = 2(X-Y)((X-1)f(Y) + (1-Y)f(X)) \tag{43}$$

We define now the function $h(X, Y)$ on $(0, \infty)^2$ by

$$h(X, Y) = f(XY) - f(X) - f(Y) \tag{44}$$

For $(X, Y) \in D$ we plug (44) into (43) and the function f disappears: we get

$$(X-1)^2 h(Y, Y) + 2(X-1)(1-Y)h(X, Y) + (1-Y)^2 h(X, X) = 0 \tag{45}$$

Let us introduce $H(X, Y)$ defined on the set

$$S = (0, \infty)^2 \setminus \{(X, Y) : X \neq 1, Y \neq 1\}$$

by

$$H(X, Y) = \frac{h(X, Y)}{(X-1)(1-Y)},$$

and for simplicity denote $h(X, X) = h(X),\ H(X, X) = H(X)$. Plugging H in (45) we get for $(X, Y) \in D$

$$H(X, Y) = \frac{1}{2}(H(X) + H(Y)) \tag{46}$$

At this point suppose that f has four successive derivatives. Later we show that this hypothesis is necessarily fulfilled.

We compute $\frac{\partial^2}{\partial X \partial Y} h(X, Y) = \frac{1}{2} \frac{\partial^2}{\partial X \partial Y} (H(X) + H(Y))(X - 1)(1 - Y)$ in two ways: the first being

$$\begin{aligned} \frac{\partial^2}{\partial X \partial Y} h(X, Y) &= -\frac{H'(X)}{2}(X - 1) + \frac{H'(Y)}{2}(1 - Y) - \frac{1}{2}(H(X) + H(Y)) \\ &= -\frac{1}{2}(H(X)(X - 1))' + \frac{1}{2}(H(Y)(1 - Y))'. \end{aligned} \tag{47}$$

A remarkable consequence of (47) is that

$$\frac{\partial^4}{\partial^2 X \partial^2 Y} h(X, Y) = 0 \tag{48}$$

The second way from (44) gives

$$\frac{\partial^2}{\partial X \partial Y} h(X, Y) = f'(XY) + XY f''(XY). \tag{49}$$

Applying (48) to the result of (49) we get

$$2 f''(XY) + 4XY f'''(XY) + X^2 Y^2 f^{IV}(XY) = 0.$$

With the replacement $t = XY$ and by using the notation $y(t) = f''(t)$ we get the following Euler linear differential equation on $t > 0$

$$t^2 y''(t) + 4t y'(t) + 2y(t) = 0,$$

whose solutions have the form $y(t) = t^\alpha$, with α root of the characteristic equation $\alpha(\alpha - 1) + 4\alpha + 2 = 0$, i.e. $\alpha = -1$ and $\alpha = -2$. As a result there exist two real numbers A and B such that $f''(t) = \frac{A}{t} + \frac{B}{t^2}$. Since f is convex we obtain that $A, B \geq 0$. Furthermore, by using $f(1) = 0$, we finally obtain the existence of a third constant C such that f has the form

$$f(t) = At \log t - B \log t + C(t - 1).$$

Now we remove the differentiability assumption on f. Thus there exists a positive measure ν on $(0, \infty)$ such that its right-derivative

$$f'_+(t) = f'_+(1) - 1_{(0,1)}(t) \int_{(t,1]} \nu(ds) + 1_{(1,+\infty)}(t) \int_{(1,t]} \nu(ds).$$

Since the left-derivative $f'_-(t)$ is $f'_+(t-)$, the derivative f' exists almost everywhere; more precisely, out of the finite or countable set of atoms of ν). Observe that $H'(X)$

has meaning almost everywhere for $X \neq 1$, since from $h(X) = -(X-1)^2 H(X)$ we deduce

$$-H'(X)(X-1)^2 - 2(X-1)H(X) = h'(X) = 2Xf'(X^2) - 2f'(X)$$

and the set of X such that at least one between $f'(X)$ and $f'(X^2)$ fails to exist is at most countable. Therefore H' is locally with bounded variation, for $X \neq 1$.

The key idea of the proof is to give meaning to $\frac{\partial^2}{\partial X \partial Y} h(X, Y)$ by considering $t = XY$ has the new variable. The transformation $(X, Y) \mapsto (X, t)$ maps D onto

$$E = \{(X, t) : X > 1,\ 0 < t < X\}$$

Now, using (47), we consider

$$h_1(X, Y) = \frac{\partial}{\partial X} h(X, Y) = \frac{H'(X)}{2}(X-1)(1-Y) + \frac{H(X) + H(Y)}{2}(1-Y)$$

for $(X, Y) \in D$, and we introduce

$$h_2(X, t) = X h_1(X, \frac{t}{X}) = \frac{H'(X)}{2}(X-1)(X-t) + \frac{H(X) + H(t/X)}{2}(X-t)$$

leading to

$$\frac{\partial}{\partial t} h_2(X, t) = -\frac{H'(X)}{2}(X-1) - \frac{H(X) + H(t/X)}{2} + \frac{H'(t/X)}{2}(1 - \frac{t}{X}) = A_1(X) + A_1\left(\frac{t}{X}\right) \tag{50}$$

for $(X, t) \in E$, where

$$A_1(X) = -\frac{H'(X)}{2}(X-1) - \frac{H(X)}{2}.$$

Note that A_1 and $t \mapsto \frac{\partial}{\partial t} h_2(X, t)$ are locally of bounded variation. The second computation of $\frac{\partial}{\partial t} h_2(X, t)$ starts from $\frac{\partial}{\partial X} h(X, Y) = Y f'(XY) - f'(X)$ leading to $h_2(X, t) = t f'(t) - X f'(X)$ and

$$\frac{\partial}{\partial t} h_2(X, t) = t f''(t) + f'(t), \tag{51}$$

where $f''(x)$ does exists almost everywhere in $t > 0$. Indeed, by comparison of (50) and (51), we obtain that ν is an absolutely continuous measure. For convenience we denote $\nu(dt) = f''(t)dt$, and obtain that

$$A_2(t) := t f''(t) + f'(t) = A_1(X) + A_1(t/X),$$

which is correct $dX \times dt$ almost everywhere on E. Coming back to the (X, Y) notation, we obtain the multiplicative Pexider equation $A_2(XY) = A_1(X) + A_1(Y)$, true on $(X, Y) \in D$ almost everywhere. This is easily turned into additive form with the change of variable

$$X = e^x, \ \ Y = e^y, \ \ F(x) = A_1(e^x), \ \ G(x) = A_2(e^x),$$

leading to

$$F(x) + F(y) = G(x + y), \quad a.e. \ on \ y < 0 < x \tag{52}$$

almost everywhere. Next choose $x_0 > 0$ and $y_0 < 0$ in such a way that

$$\begin{aligned} &F(x_0) + F(y_0) = G(x_0 + y_0), F(x_0) + F(y) \\ &\quad = G(x_0 + y), a.e. \ y < 0, F(x) + F(y_0) = G(x + y_0), \ a.e. \ x > 0 \end{aligned}$$

so that, once defined for $z_1 > -x_0$, $z_2 < y_0$ and $z \in \mathbb{R}$

$$\begin{aligned} &\tilde{F}_1(z_1) = F(x_0 + z_1) - F(x_0), \tilde{F}_2(z_2) \\ &\quad = F(y_0 + z_2) - F(y_0), \tilde{G}(z) = G(x_0 + y_0 + z) - G(x_0 + y_0), \end{aligned}$$

we have

$$\begin{aligned} &\tilde{F}_1(z_1) = \tilde{G}(z_1), a.e. \ z_1 > -x_0, \\ &\tilde{F}_2(z_2) = \tilde{G}(z_2), a.e. \ z_2 < y_0, \\ &\tilde{G}(z_1) + \tilde{G}_2(z_2) = \tilde{G}(z_1 + z_2), \end{aligned}$$

where the latter holds a.e. in $z_1 > -x_0$ and $z_2 < y_0$, hence it certainly holds a.e. in z_1 and z_2 within a suitably small ball around the origin. The rest of the proof is carried away as in Lemma 4.2, proving that $G(x) = Ax + C$ and $A_2(t) = A \log t + C$. Finally the equation

$$tf''(t) = A \log t + C - tf'(t),$$

proves that f'' is continuous: by two successive differentiations, we obtain that f has four derivatives as promised. This concludes the proof. □

6 Conclusion

In the paper we have considered equivalent probability measures P_i and Q_i on $\mathbb{R}^d$, for $i = 1, \ldots, n$, and their convolution products $P_1 * \cdots P_n$ and $Q_1 * \cdots * Q_n$, i.e. the distributions of $X_1 + \ldots + X_n$ and $Y_1 + \ldots + Y_n$, where $X_i \sim P_i$ and $Y_i \sim Q_i$

are mutually independent, for $i = 1, \ldots, n$. We observed that, for any $\lambda \in (0, 1]$, the Rényi divergence (including the Kullback–Leibler for $\lambda = 1$) satisfies the inequality

$$D_\lambda(P^{*n}||Q^{*n}) \leq nD_\lambda(P||Q), \tag{53}$$

and that the equality occurs when P_i belong to the natural exponential family generated by Q_i, with the same natural parameter θ. But in general the equality holds when the log-densities $f_i = \log \frac{P_i}{Q_i}$ satisfy a Pexider-type equation

$$f_1(x_1) + \ldots + f_n(x_n) = G_n(x_1 + \ldots + x_n), \tag{54}$$

$Q_1 \otimes \ldots \otimes Q_n$-almost surely in $(x_1, \ldots, x_n)$. Saying that P_i and Q_i belong to the same natural exponential family is saying that f_i is an affine function. When the natural parameter θ is independent of i, also G_n is affine and (54) is satisfied. But in some cases there could exist other solutions. Focussing on the i.i.d. case $P_i = P$ and $Q_i = Q$, we have established that for some classes of probability measures Q's supported by $\mathbb{Z}$ and $\mathbb{Z}^d$, as well for other classes of Q's absolutely continuous w.r.t. the Lebesgue measure in $\mathbb{R}^d$, these solutions cannot occur. Some counterexamples are provided as well. We have finally proved that an f divergence cannot satisfy $D_f(P^{*n}||Q^{*n}) = nD_f(P||Q)$ for P and Q members of the same exponential family unless this f divergence has the form $D_f(P||Q) = AD_1(P||Q) + BD_1(Q||P)$ where D_1 is the KL divergence, $A, B \geq 0$ and $A + B > 0$.

Author Contributions All the authors have contributed equally to the redaction of the manuscript.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The corresponding author, on behalf of both authors, declares that they have no conflict of interest as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

References

1. van Erven, T., Harremoes, P.: Rényi divergence and Kullback–Leibler divergence. IEEE Trans. IT **60**, 3797–3820 (2014)
2. Polyanskiy, Y., Wu, Y.: Information Theory, from Coding to Learning. Cambridge University Press, in press (2024)
3. Letac, G.: Lectures on natural exponential families and their variance functions. IMPA, Rio de Janeiro (1992)
4. Kannappan, P.: Functional equations and inequalities with applications. Springer, New York (2009)
5. Aczel, J., Dhombres, J.: Functional equations in several variables, with applications to mathematics, information theory, and to the natural and social sciences. Cambridge University Press, Cambridge (1989)
6. Bandelow, C.: On closed additive semigroups in E^n. Proc. Am. Math. Soc. **34**, 87–89 (1972)
7. de Bruijn, N.G.: On almost additive functions, pp. 59–63. Colloq. Math, XV (1966)
8. Kac, M.: Une remarque sur les équations fonctionnelles. Comment Math Helv. **9**, 170–171 (1936-37)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.