



# Entropy Theory for Random Walks on Lie Groups

Samuel Kittle<sup>1</sup> · Constantin Kogler<sup>2</sup>

Received: 13 May 2025 / Revised: 15 September 2025 / Accepted: 23 November 2025

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

## Abstract

We develop entropy and variance results for the product of independent identically distributed random variables on Lie groups. Our results apply to the study of stationary measures in various contexts.

**Keywords** Entropy · Random Walks on Groups · Lie Groups · Self-similar measures

## Contents

1	Introduction	...
	Notation	...
2	Basic Properties of Entropy on Lie Groups	...
2.1	Entropy and Basic Properties	...
2.2	Kullback–Leibler Divergence	...
2.3	Regular Conditional Distributions	...
2.4	Conditional Entropy	...
3	Entropy Growth for Stopped Random Walks	...
4	Entropy and Trace on Lie groups	...
4.1	Entropy and Trace	...
4.2	Conditional Trace and Entropy Between Scales	...
4.3	Proof of Theorem 1.4	...
5	From Entropy Gap to Trace Sum	...
	References	...

---

✉ Constantin Kogler  
kogler@ias.edu

Samuel Kittle  
s.kittle@ucl.ac.uk

<sup>1</sup> Department of Mathematics, University College London, 25 Gordon Street, London WC1H 0AY, UK

<sup>2</sup> Institute for Advanced Study, 1 Einstein Dr, Princeton, NJ 08540, USA

## 1 Introduction

Entropy is a central tool in the study of random walks. For example, as exposed in the book by Johnson [6], entropic methods can be used to prove the central limit theorem on  $\mathbb{R}^d$  as well as equidistribution of random walks on compact groups. More recently, entropy results have been applied in the study of stationary measures as in the work of Hochman [4, 5] on the dimension of self-similar measures, of Bárány-Hochman-Rapaport [1] for self-affine measures, or of Varjú [10] to construct absolutely continuous Bernoulli convolutions.

The aim of this paper is to establish entropy and variance bounds for smoothings of random walks on arbitrary Lie groups. We strive to be as general as feasible to ensure applicability to various settings. Our results generalise the entropy and variance bounds previously established in [7] for  $\mathrm{SL}_2(\mathbb{R})$ , which were developed to construct absolutely continuous Furstenberg measures. Furthermore, the results presented here have been applied in [8, 9] to study self-similar measures. We anticipate that this paper will contribute to the investigation of random walks of Lie group actions on manifolds in various contexts.

Throughout this paper, let  $G$  be a real Lie group of dimension  $\ell$  and denote by  $\mathfrak{g}$  the Lie algebra of  $G$ . We fix an inner product on  $\mathfrak{g}$ , inducing an associated norm  $|\circ|$  and a left-invariant metric  $d$  on  $G$ . Also, denote by  $\log : G \rightarrow \mathfrak{g}$  the logarithm on  $G$ , which is defined in a small neighbourhood around the identity. For  $g \in G$  we write  $B_\varepsilon(g)$  for the open  $\varepsilon$ -ball around  $g$  and abbreviate  $B_\varepsilon = B_\varepsilon(\mathrm{Id})$  for  $\mathrm{Id}$  the identity in  $G$ . Finally,  $m_G$  is the Haar measure on  $G$  that is normalised such that  $m_G(B_\varepsilon)/m_{\mathfrak{g}}(\log B_\varepsilon) \rightarrow 1$  as  $\varepsilon \rightarrow 0$ , where  $m_{\mathfrak{g}}$  is the volume measure induced by the fixed inner product on  $\mathfrak{g}$ .

Given a  $G$ -valued random variable  $g$ , we denote by

$$H(g)$$

the Shannon entropy of  $g$  when  $g$  is discrete and the differential entropy when  $g$  is absolutely continuous. Precise definitions and some basic results are given in Sects. 2.1 and 2.2.

Similarly to [1, 4, 5] or [10], we study the entropy of a smoothing of  $g$ . If  $s$  is a smoothing distribution independent of  $g$ , we can abstractly define the entropy of  $g$  with respect to  $s$  as

$$H(g; s) = H(gs) - H(s).$$

Concretely, we will choose the following smoothing functions: for given  $r > 0$  and  $a \geq 1$ , denote by  $\beta_{a,r}$  a random variable on  $\mathfrak{g}$  with density function  $f_{a,r} : \mathfrak{g} \rightarrow \mathbb{R}$  given by

$$f_{a,r}(x) = \begin{cases} C_{a,r} e^{-\frac{|x|^2}{2r^2}} & \text{if } |x| \leq ar, \\ 0 & \text{otherwise,} \end{cases}$$

where  $C_{a,r}$  is a normalising constant to ensure that  $f_{a,r}$  integrates to 1. We furthermore set

$$s_{a,r} = \exp(\beta_{a,r}) \quad (1.1)$$

and then define the entropy of  $g$  at scale  $r > 0$  with respect to the parameter  $a \geq 1$  as

$$H_a(g; r) = H(g; s_{a,r}) = H(gs_{a,r}) - H(s_{a,r}). \quad (1.2)$$

Entropy at scale  $r$  measures the amount of information  $g$  has at scale  $r$ . The parameter  $a$  is useful in order to uniformly control how far  $\beta_{a,r}$  is from a normal distribution. The reason we are working with these smoothing functions on  $G$  is to deduce in Lemma 4.6 a quantitative analogue of (1.6). It is assumed throughout this paper that the collection of random variables  $g$  and  $(s_{a,r})_{a \geq 1, r > 0}$  are independent.

Let  $\mu$  be a probability measure on  $G$  and let  $\gamma_1, \gamma_2, \dots$  be independent  $\mu$ -distributed random variables that are independent from  $(s_{a,r})_{a \geq 1, r > 0}$  and denote for  $n \geq 1$ ,

$$q_n = \gamma_1 \cdots \gamma_n.$$

Our first goal is to give a general result on the behaviour of  $H_a(q_n; r_n)$  for suitable scales  $r_n > 0$ . To do so, for a finitely supported probability measure  $\mu$  on  $G$  we define the separation rate as

$$M_n = \min \left\{ d(g, h) : g, h \in \bigcup_{i=0}^n \text{supp}(\mu^{*i}) \text{ with } g \neq h \right\}.$$

Furthermore, the random walk entropy of  $\mu$  is given as

$$h_\mu = \lim_{n \rightarrow \infty} \frac{1}{n} H(\mu^{*n}) = \inf_{n \geq 1} \frac{1}{n} H(\mu^{*n}).$$

We first make the following basic observation. If  $r_n < \frac{1}{2a} M_n$ , then all the elements in the support of  $q_n$  are separated by at least  $2ar_n$ . Therefore, the density of  $q_n s_{a,r_n}$  can be expressed as the weighted sum of the densities of  $x s_{a,r_n}$  with  $x$  ranging over the elements in the support of  $q_n$ . Thus it follows (by Lemma 2.3) that

$$H_a(q_n; r_n) = H(q_n) \geq nh_\mu. \quad (1.3)$$

Our first result is a generalisation of (1.3) for arbitrary stopped random walks  $q_{\eta_n}$  for a sequence of stopping times  $\eta_n$ . In order to deduce an analogue of (1.3), we require that our stopping times satisfy a large deviation principle.

**Definition 1.1** Let  $\eta = (\eta_n)_{n \geq 1}$  be a sequence of stopping times. Then we say that  $\eta$  satisfies the large deviation principle if  $\mathbb{E}[\eta_n] \rightarrow \infty$  as  $n \rightarrow \infty$  and for every  $\varepsilon > 0$  there exists a  $\delta > 0$  such that for all sufficiently large  $n$ ,

$$\mathbb{P}[|\eta_n - \mathbb{E}[\eta_n]| \geq \varepsilon \cdot \mathbb{E}[\eta_n]] \leq e^{-\delta \cdot \mathbb{E}[\eta_n]}.$$

To state the first theorem, we note that we use the asymptotic notation as explained at the end of the introduction.

**Theorem 1.2** *Let  $\mu$  be a finitely supported probability measure on  $G$ . Let  $\eta = (\eta_n)_{n \geq 1}$  be a sequence of stopping times satisfying the large deviation principle and write  $L_n = \mathbb{E}[\eta_n]$  for  $n \geq 1$ . Let  $a \geq 1$ ,  $\varepsilon > 0$  and let  $r_n > 0$  be a sequence satisfying for all  $n \geq 1$ ,*

$$r_n \leq a^{-1} c_G M_{\lceil (1+\varepsilon)L_n \rceil}$$

for a constant  $c_G > 0$  depending only on  $G$ . Then for all  $n \geq 1$ ,

$$H_a(q_{\eta_n}; r_n) \geq h_\mu \cdot L_n + o_{\mu, \eta, \varepsilon}(L_n).$$

Denote

$$S_n = -\frac{1}{n} \log M_n \quad \text{as well as} \quad S_\mu = \limsup_{n \rightarrow \infty} S_n.$$

In numerous concrete cases  $S_\mu$  is finite and can be bounded efficiently. Indeed, if  $G$  is a linear group and all of the entries of elements in the support of  $\mu$  lie in a number field  $K$  and have logarithmic height at most  $L$ , then  $S_\mu \ll_G L \cdot [K : \mathbb{Q}]$  as shown in [9]\*Proposition 8.10. Under the additional assumption that  $S_\mu < \infty$ , the following corollary can be deduced from Theorem 1.2.

**Corollary 1.3** *Let  $\mu$  be a finitely supported probability measure on  $G$  and assume that  $S_\mu < \infty$ . Let  $\eta = (\eta_n)_{n \geq 1}$  be a sequence of stopping times satisfying the large deviation principle and write  $L_n = \mathbb{E}[\eta_n]$  for  $n \geq 1$ . Suppose that  $a \geq 1$  and  $S > S_\mu$ . Then for any sequence  $0 < r_n < e^{-S \cdot L_n}$  as  $n \rightarrow \infty$ ,*

$$H_a(q_{\eta_n}; r_n) \geq h_\mu \cdot L_n + o_{\mu, \eta, a, S}(L_n).$$

To provide some further context, we discuss the setting from [8, 9] where Corollary 1.3 is used. A similarity is a map  $g : \mathbb{R}^d \rightarrow \mathbb{R}^d$  such that there exists a scalar  $\rho(g) > 0$ , an orthogonal matrix  $U(g) \in O(d)$  and a vector  $b(g) \in \mathbb{R}^d$  such that  $g(x) = \rho(g)U(g)x + b(g)$  for all  $x \in \mathbb{R}^d$ . Denote by  $G = \text{Sim}(\mathbb{R}^d)$  the group of similarities and let  $\mu$  be a probability measure on  $\mathbb{R}^d$  supported on finitely many contractive similarities, that is, similarities  $g$  satisfying  $\rho(g) \in (0, 1)$ . Then denote for  $\kappa > 0$  by  $\eta_\kappa$  the stopping time  $\eta_\kappa = \inf\{n \geq 1 : \rho(q_n) < \kappa\}$ .

As shown in [9]\*Lemma 3.9,  $\eta_\kappa$  satisfies a large deviation principle with  $L_n = \mathbb{E}[\eta_n] = \frac{\log \kappa^{-1}}{|\chi_\mu|} + o_\mu(\log \kappa^{-1})$ , for the logarithmic contraction rate  $\chi_\mu = \int \log \rho(g) d\mu(g)$ . Under the assumption that  $S_\mu < \infty$ , which is satisfied as discussed above when all of the entries of elements in the support of  $\mu$  are algebraic, it therefore follows by Corollary 1.3 that for  $S > S_\mu$  and  $0 < r_n < e^{-S \cdot L_n}$ ,

$$H_a(q_{\eta_\kappa}; r_n) \geq (\log \kappa^{-1}) \frac{h_\mu}{|\chi_\mu|} + o_{\mu, \eta, a, S}(\log \kappa^{-1}). \quad (1.4)$$

Results similar to (1.4) are an important component in the proof of the main theorems of [8, 9]. In fact, as discussed below, we will convert estimates such as (1.4) into variance estimates. In [8, 9], these quantitative estimates will eventually allow us to apply Berry-Esseen type results to establish absolute continuity or full dimension of stationary measures. A similar approach is used in [7] to study the Furstenberg measure of  $\mathrm{SL}_2(\mathbb{R})$ . In this paper, we provide a unified perspective on these results that can be applied to any Lie group.

We next address the second goal of this paper, which is to convert entropy estimates into variance estimates on arbitrary Lie groups. Indeed, the reader may recall that if  $Z$  is an absolutely continuous random variable on  $\mathbb{R}$  with variance  $\sigma^2$  then

$$H(Z) \leq \frac{1}{2} \log(2\pi e\sigma^2), \quad (1.5)$$

where  $H(Z)$  is the differential entropy of  $Z$  and equality holds in (1.5) if and only if  $Z$  is distributed like a centred Gaussian with variance  $\sigma^2$ . We will prove an analogue of this fact on Lie groups. To do so, for random variables  $g$  that are supported within small balls of a given point  $g_0$  we consider the covariance matrix of the Lie group logarithm applied to  $g_0^{-1}g$ . This viewpoint allows us to apply a higher dimensional analogue of (1.5) to deduce an analogous result on  $G$ .

For an  $\ell$ -dimensional random variable  $X$ , we denote by  $\mathrm{tr}(X)$  the trace of the covariance matrix of  $X$ . In particular, we use the following definition. Given  $g_0 \in G$  and a random variable  $g$  on  $G$  we define

$$\mathrm{tr}_{g_0}(g) = \mathrm{tr}(\log(g_0^{-1}g)),$$

whenever  $\log(g_0^{-1}g)$  is defined. The analogue of (1.5), which will be proved in Proposition 4.1, then amounts to

$$H(g) \leq \frac{\ell}{2} \log\left(\frac{2\pi e}{\ell} \cdot \mathrm{tr}_{g_0}(g)\right) + O_G(\varepsilon) \quad (1.6)$$

for random variables  $g$  supported on  $B_\varepsilon(g_0)$  and  $\varepsilon > 0$  sufficiently small.

To make the latter more useful, we can relate a certain notion of trace to entropy between scales. One defines the entropy between scales  $r_1, r_2 > 0$  as

$$\begin{aligned} H_a(g; r_1 | r_2) &= H(g; s_{r_1, a} | s_{r_2, a}) = H_a(g; r_1) - H_a(g; r_2) \\ &= (H(gs_{r_1, a}) - H(s_{r_1, a})) - (H(gs_{r_2, a}) - H(s_{r_2, a})). \end{aligned}$$

Roughly speaking,  $H_a(g; r_1 | r_2)$  measures how much more information  $g$  has on scale  $r_1$  than it has on scale  $r_2$ .

We furthermore define  $\mathrm{tr}(g; r)$  to be the supremum of all  $t \geq 0$  such that we can find some  $\sigma$ -algebra  $\mathcal{A}$  and some  $\mathcal{A}$ -measurable random variable  $h$  taking values in  $G$  such that

$$|\log(h^{-1}g)| \leq r \quad \text{and} \quad \mathbb{E}[\mathrm{tr}_h(g | \mathcal{A})] \geq t \cdot r^2.$$

The following general result can be deduced.

**Theorem 1.4** *Let  $g$  be a random variable taking values in  $G$ , let  $a \geq 1$  and  $r > 0$  be such that  $ar$  is sufficiently small in terms of  $G$  and assume that  $g$ ,  $s_{a,r}$  and  $s_{a,2r}$  are independent random variables. Then*

$$\text{tr}(g; 2ar) \gg_G a^{-2}(H_a(g; r|2r) - O_G(e^{-a^2/4} + a^3r)),$$

for the implied constants depending only on  $G$ .

The conditional distribution  $(g|\mathcal{A})$  becomes relevant as we can relate the entropy between scales  $H_a(g; r|2r)$  to the conditional entropy  $H(gs_{a,r_1}|gs_{a,r_1})$  as shown in Lemma 2.10 and Proposition 4.3.

In our applications [7–9] it is desirable to convert an entropy gap into lower bounds of  $\sum_{i=1}^m \text{tr}(g; s_i)$  for a sequence of scales  $s_1, \dots, s_m \in (r_1, r_2)$ . It will be also useful to assume that our scales satisfy  $s_{i+1} \geq As_i$  for some  $A > 0$ . Indeed, the latter condition is necessary to make the variance summation method [7–9] applicable, that is to combine variance bounds of stopped random walks on different scales.

Using Theorem 1.4, we deduce the following general proposition applied in [9] and [8].

**Proposition 1.5** *Let  $g$  be a  $G$ -valued random variable independent of  $(s_{a,r})_{a \geq 1, r > 0}$  and let  $0 < r_1 < r_2$ . Let  $a \geq 1$  such that  $ar_2$  is sufficiently small in terms of  $G$ . Suppose that for all  $r'_1 \in [r_1, 2r_1]$  as well as  $r'_2 \in [r_2/2, 2r_2]$  it holds for some constant  $C > 0$  that*

$$H_a(g; r'_1|r'_2) \geq C.$$

*Let  $A > 1$ . Then there exists  $s_1, \dots, s_m \in (ar_1, 4ar_2)$  where  $m = \lceil \frac{\log 4ar_2 - \log ar_1}{2 \log A} \rceil$  such that for  $N = \lceil \frac{\log r_2 - \log r_1}{\log 2} \rceil - 1$ ,*

$$\sum_{i=1}^m \text{tr}(g; s_i) \gg_G \frac{C - N \cdot O_G(e^{-\frac{a^2}{4}} + a^3r_2)}{a^2 \log A}$$

*and  $s_{i+1} \geq As_i$  for all  $1 \leq i \leq m - 1$ .*

We comment on the structure of this paper. After discussing basic properties of entropy in Sect. 2, we prove Theorem 1.2 and Corollary 1.3 in Sect. 3. In Sect. 4 we show (1.6) and Theorem 1.4. Finally, we prove Proposition 1.5 in Sect. 5.

## Notation

We use the asymptotic notation  $A \ll B$  or  $A = O(B)$  to denote that  $|A| \leq CB$  for a constant  $C > 0$ . If the constant  $C$  depends on additional parameters we add subscripts. Moreover,  $A \asymp B$  denotes  $A \ll B$  and  $B \ll A$ . For a sequence  $A_n$  and  $B_n$  we write

$A_n = o(B_n)$  to denote that  $A_n/B_n \rightarrow 0$  as  $n \rightarrow \infty$  and we add subscripts to indicate that the speed of convergence depends on certain parameters.

We write  $[n] = \{1, \dots, n\}$ .

It is assumed throughout this paper that the random variables  $(s_{a,r})_{a \geq 1, r > 0}$  as defined in (1.1) are independent. Whenever  $g$  will denote a  $G$ -valued random variable, it will be assumed that  $g$  is independent of  $(s_{a,r})_{a \geq 1, r > 0}$ . Given a probability measure  $\mu$ , we sample independent  $\mu$ -distributed random variables  $\gamma_1, \gamma_2, \dots$  that are independent from  $(s_{a,r})_{a \geq 1, r > 0}$ .

## 2 Basic Properties of Entropy on Lie Groups

In Sect. 2.1 we give definitions and discuss basic properties of entropy on  $G$ , after which we discuss the Kullback-Leibler divergence on a general measurable space  $X$  in Sect. 2.2. In Sect. 2.3 we review regular conditional distributions in order to study conditional entropy and conditional trace in Sects. 2.4 and 4.

### 2.1 Entropy and Basic Properties

For notational convenience, we denote for  $x \in [0, \infty)$  by

$$h(x) = \begin{cases} -x \log(x) & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$$

and recall that  $h$  is concave. If  $\lambda = \sum_i p_i \delta_{g_i}$  is a discrete probability measure on  $G$ , we define the Shannon entropy of  $\lambda$  as

$$H(\lambda) = \sum_i h(p_i).$$

On the other hand, given an absolutely continuous probability measure  $\lambda$  on  $G$  with density  $f_\lambda$  we define

$$H(\lambda) = \int h(f_\lambda) dm_G.$$

We extend the definition to finite positive measures  $\lambda$  that are either absolutely continuous or discrete by setting

$$H(\lambda) = ||\lambda||_1 H(\lambda/||\lambda||_1), \quad \text{where} \quad ||\lambda||_1 = \lambda(G).$$

In this subsection we collect some useful basic properties of entropy.

**Lemma 2.1** *Let  $\lambda_1, \dots, \lambda_n$  be absolutely continuous finite measures on  $G$ . Then*

$$H(\lambda_1 + \dots + \lambda_n) \geq H(\lambda_1) + \dots + H(\lambda_n).$$

**Proof** It suffices to prove the claim for  $n = 2$ . Let  $f_1$  and  $f_2$  be the densities of  $\lambda_1$  and  $\lambda_2$ . Then since  $h$  is concave

$$\begin{aligned} H(\lambda_1 + \lambda_2) &= (||\lambda_1||_1 + ||\lambda_2||_1) \int h\left(\frac{f_1 + f_2}{||\lambda_1||_1 + ||\lambda_2||_1}\right) dm_G \\ &\geq (||\lambda_1||_1 + ||\lambda_2||_1) \int \frac{||\lambda_1||_1}{||\lambda_1||_1 + ||\lambda_2||_1} h\left(\frac{f_1}{||\lambda_1||_1}\right) dm_G \\ &\quad + (||\lambda_1||_1 + ||\lambda_2||_1) \int \frac{||\lambda_2||_1}{||\lambda_1||_1 + ||\lambda_2||_1} h\left(\frac{f_2}{||\lambda_2||_1}\right) dm_G \\ &= H(\lambda_1) + H(\lambda_2). \end{aligned}$$

□

**Lemma 2.2** Let  $p = (p_1, p_2, \dots)$  be a probability vector and let  $\lambda_1, \lambda_2, \dots$  be probability measures on  $G$  either all absolutely continuous measures or all discrete measures with finite entropy such that  $||\lambda_i||_1 = p_i$ . Then

$$H\left(\sum_{i=1}^{\infty} \lambda_i\right) \leq H(p) + \sum_{i=1}^{\infty} H(\lambda_i).$$

In particular, if  $p_i = 0$  for all  $i > k$  for some  $k \geq 1$  then

$$H\left(\sum_{i=1}^k \lambda_i\right) \leq \log k + \sum_{i=1}^k H(\lambda_i).$$

**Proof** We only consider the case of absolutely continuous measures as the proof is analogous in the discrete case. Denote the densities of  $\lambda_i$  by  $f_i$ . Note that  $h(\sum_{i=1}^{\infty} a_i) \leq \sum_{i=1}^{\infty} h(a_i)$  for any  $a_1, a_2, \dots \geq 0$ . Therefore

$$\begin{aligned} H\left(\sum_{i=1}^{\infty} \lambda_i\right) &= \int h\left(\sum_{i=1}^{\infty} f_i\right) dm_G \\ &\leq \sum_{i=1}^{\infty} \int h(f_i) dm_G \\ &= \sum_{i=1}^{\infty} \int (-f_i(x) \log(p_i^{-1} f_i) - f_i(x) \log(p_i)) dm_G \\ &= \sum_{i=1}^{\infty} \int p_i h(p_i^{-1} f_i) dm_G + h(p_i) \\ &= H(p) + \sum_{i=1}^{\infty} H(\lambda_i). \end{aligned}$$

□

**Lemma 2.3** *Let  $\lambda_1$  be a discrete and  $\lambda_2$  be an absolutely continuous probability measure on  $G$ . Then*

$$H(\lambda_1 * \lambda_2) \leq H(\lambda_1) + H(\lambda_2)$$

*Suppose further that  $\lambda_1$  is supported on finitely many points with separation at least  $2r$  and that the support of  $\lambda_2$  is contained in a ball of radius  $r$ . Then*

$$H(\lambda_1 * \lambda_2) = H(\lambda_1) + H(\lambda_2).$$

**Proof** Write  $\lambda_1 = \sum_{i=1}^n p_i \delta_{g_i}$  and let  $f$  be the density of  $\lambda_2$ . Then the density of  $\lambda_1 * \lambda_2$  is given by  $\sum_{i=1}^n p_i f \circ g_i^{-1}$ . As  $h(\sum_{i=1}^n a_i) \leq \sum_{i=1}^n h(a_i)$  for any  $a_1, \dots, a_n \geq 0$ ,

$$\begin{aligned} H(\lambda_1 * \lambda_2) &= \int h \left( \sum_{i=1}^n p_i f \circ g_i^{-1} \right) dm_G \\ &\leq \sum_{i=1}^n \int h(p_i f \circ g_i^{-1}) dm_G \\ &= \sum_{i=1}^n \int (p_i f \circ g_i^{-1})(\log(p_i) + \log(f \circ g_i^{-1})) dm_G \\ &= H(\lambda_1) + H(\lambda_2). \end{aligned}$$

If  $\lambda_1$  is supported on finitely many points with separation at least  $2r$  and that the support of  $\lambda_2$  is contained in a ball of radius  $r$ , then the support of the functions  $f \circ g_i^{-1}$  is disjoint and the inequality in the second line is an equality.  $\square$

## 2.2 Kullback–Leibler Divergence

In this section we discuss Kullback–Leibler divergence on general measurable spaces  $X$ . If  $\nu \ll \mu$  are measures on  $X$ , then we define the Kullback–Leibler divergence as

$$D_{\text{KL}}(\nu \parallel \mu) = - \int \log \frac{d\nu}{d\mu} d\nu.$$

Observe that if  $\nu$  is absolutely continuous on  $G$  with respect to the Haar measure  $m_G$ , then  $H(\nu) = D_{\text{KL}}(\nu \parallel m_G)$ . We collect some basic results on the Kullback–Leibler divergence on  $X$ .

**Lemma 2.4** *Let  $\nu \ll \mu$  be measures on  $G$  and assume that  $\nu$  is a probability measure supported on a set  $A$  of positive  $\mu$  measure. Then*

$$D_{\text{KL}}(\nu \parallel \mu) \leq \log(\mu(A)).$$

**Proof** For convenience write  $d\nu = f_\nu d\mu$ . Then by Jensen's inequality,

$$\begin{aligned} D_{\text{KL}}(\nu || \mu) &= \int_A h\left(f_\nu \frac{\mu(A)}{\mu(A)}\right) d\mu = \int h(f_\nu \mu(A)) \frac{1_A}{\mu(A)} d\mu + \log(\mu(A)) \\ &\leq \log(\mu(A)). \end{aligned}$$

□

**Lemma 2.5** Assume that we can write  $X = X_1 \times \dots \times X_m$  as a product of submanifolds  $X_i \subset X$  and assume that  $m_X = m_{X_1} \times \dots \times m_{X_m}$  for a measure  $m_X$  on  $X$  and measures  $m_{X_i}$  on  $X_i$ . Let  $\mu$  be a probability measure on  $X$  with  $\mu \ll m_G$ . Denote by  $\pi_i$  the projection from  $X$  to  $X_i$  and by  $\pi_i \mu$  the pushforward of  $\mu$  under  $\pi_i$ . Then

$$D_{\text{KL}}(\mu || m_X) \leq D_{\text{KL}}(\pi_1 \mu || m_{X_1}) + \dots + D_{\text{KL}}(\pi_m \mu || m_{X_m}).$$

**Proof** It suffices to prove the claim for  $m = 2$ . Denote by  $f_\mu$  the density of  $\mu$  with respect to  $m_X$  and write

$$f_\mu^1(x_2) = \int f_\mu(x_1, x_2) dm_{X_1}(x_1) \quad \text{and} \quad f_\mu^2(x_1) = \int f_\mu(x_1, x_2) dm_{X_2}(x_2).$$

Therefore,

$$\begin{aligned} D_{\text{KL}}(\mu || m_X) &= \int \int h(f_\mu(x_1, x_2)) dm_{X_1}(x_1) dm_{X_2}(x_2) \\ &= \int \int h\left(\frac{f_\mu(x_1, x_2)}{f_\mu^2(x_1)} f_\mu^2(x_1)\right) dm_{X_1}(x_1) dm_{X_2}(x_2) \\ &= \int \int h\left(\frac{f_\mu(x_1, x_2)}{f_\mu^2(x_1)}\right) f_\mu^2(x_1) dm_{X_1}(x_1) dm_{X_2}(x_2) \\ &\quad + \int \int -\log(f_\mu^2(x_1)) f_\mu(x_1, x_2) dm_{X_1}(x_1) dm_{X_2}(x_2) \\ &\leq \int h(f_\mu^1(x_2)) dm_{X_2}(x_2) + \int h(f_\mu^2(x_1)) dm_{X_1}(x_1) \\ &= D_{\text{KL}}(\pi_1 \mu || m_{X_1}) + D_{\text{KL}}(\pi_2 \mu || m_{X_2}), \end{aligned}$$

having used that  $h$  is concave and Jensen's inequality in the penultimate line. □

**Lemma 2.6** Let  $(X, m_X)$  and  $(Y, m_Y)$  be a locally compact Hausdorff space endowed with Radon measures, and let  $\Phi : X \rightarrow Y$  be a homeomorphism with  $\Phi_* m_X = m_Y$ . Then for a measure  $\nu \ll m_X$  on  $X$  it holds that

$$D_{\text{KL}}(\Phi_* \nu || m_Y) = D_{\text{KL}}(\nu || m_X).$$

**Proof** Let  $f : Y \rightarrow \mathbb{R}$  be a continuous compactly supported function. Then

$$\int f d\Phi_* \nu = \int (f \circ \Phi) d\nu = \int (f \circ \Phi) \frac{d\nu}{dm_X} dm_X$$

as well as

$$\begin{aligned} \int f d\Phi_* \nu &= \int f \frac{d\Phi_* \nu}{dm_Y} dm_Y \\ &= \int f \frac{d\Phi_* \nu}{dm_Y} d\Phi_* m_X = \int (f \circ \Phi) \left( \frac{d\Phi_* \nu}{dm_Y} \circ \Phi \right) dm_X. \end{aligned}$$

Since  $\Phi$  is a homeomorphism, every continuous compactly supported function  $g : X \rightarrow \mathbb{R}$  can be written as  $g = f \circ \Phi$  for  $f : Y \rightarrow \mathbb{R}$  a continuous compactly supported function. Therefore, using the Riesz Representation Theorem, it holds that  $m_X$ -almost surely

$$\frac{d\Phi_* \nu}{dm_Y} \circ \Phi = \frac{d\nu}{dm_X}$$

and thus

$$\begin{aligned} D_{\text{KL}}(\Phi_* \nu || m_Y) &= - \int \log \frac{d\Phi_* \nu}{dm_Y} d\Phi_* \nu \\ &= - \int \log \left( \frac{d\Phi_* \nu}{dm_Y} \circ \Phi \right) d\nu \\ &= - \int \log \left( \frac{d\nu}{dm_X} \right) d\nu = D_{\text{KL}}(\nu || m_X). \end{aligned}$$

□

**Lemma 2.7** Let  $\lambda_1$  be a probability measure on  $X$  and let  $\lambda_2$  and  $\lambda_3$  be measures on  $X$  such that  $\lambda_1 \ll \lambda_2$  and  $\lambda_2 \ll \lambda_3$ . Let  $U \subset X$  and suppose that the support of  $\lambda_1$  is contained in  $U$ . Then

$$|D_{\text{KL}}(\lambda_1 || \lambda_2) - D_{\text{KL}}(\lambda_1 || \lambda_3)| \leq \sup_{x \in U} \left| \log \frac{d\lambda_2}{d\lambda_3} \right|.$$

**Proof** We calculate

$$\begin{aligned} |D_{\text{KL}}(\lambda_1 || \lambda_2) - D_{\text{KL}}(\lambda_1 || \lambda_3)| &= \left| \int_U \log \frac{d\lambda_1}{d\lambda_2} d\lambda_1 - \int_U \log \frac{d\lambda_1}{d\lambda_3} d\lambda_1 \right| \\ &\leq \int_U \left| \log \frac{d\lambda_1}{d\lambda_2} - \log \frac{d\lambda_1}{d\lambda_3} \right| d\lambda_1 \end{aligned}$$

$$\begin{aligned}
&= \int_U \left| \log \frac{d\lambda_2}{d\lambda_3} \right| d\lambda_1 \\
&\leq \sup_{x \in U} \left| \log \frac{d\lambda_2}{d\lambda_3} \right|.
\end{aligned}$$

□

## 2.3 Regular Conditional Distributions

In this section we review the definition of regular conditional distributions that will be used in the following subsections in order to discuss conditional entropy. On a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , we denote the conditional expectation by  $\mathbb{E}[f|\mathcal{A}]$  for  $f \in L^1(\Omega, \mathcal{F}, \mathbb{P})$  and a  $\sigma$ -algebra  $\mathcal{A} \subset \mathcal{F}$ . Given two measurable spaces  $(\Omega_1, \mathcal{A}_1)$  and  $(\Omega_2, \mathcal{A}_2)$ , recall that a Markov kernel on  $(\Omega_1, \mathcal{A}_1)$  and  $(\Omega_2, \mathcal{A}_2)$  is a map  $\kappa : \Omega_1 \times \mathcal{A}_2 \rightarrow [0, 1]$  if for any  $A_2 \in \mathcal{A}_2$ , the map  $\kappa(\cdot, A_2)$  is  $\mathcal{A}_1$ -measurable and for any  $\omega_1$  the map  $A_2 \rightarrow \kappa(\omega_1, A_2)$  is a probability measure.

**Definition 2.8** Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space and let  $\mathcal{A} \subset \mathcal{F}$  be a  $\sigma$ -algebra. Let  $(E, \xi)$  be a measurable space and let  $Y : (\Omega, \mathcal{F}) \rightarrow (E, \xi)$  be a random variable. Then we say that a Markov kernel

$$(Y|\mathcal{A}) : \Omega \times \xi \rightarrow [0, 1]$$

on  $(\Omega, \mathcal{A})$  and  $(E, \xi)$  is a **regular conditional distribution** if for all  $B \in \xi$ ,

$$(Y|\mathcal{A})(\omega, B) = \mathbb{P}[Y \in B | \mathcal{A}](\omega) = \mathbb{E}[1_{Y^{-1}(B)} | \mathcal{A}](\omega).$$

In other words,

$$\mathbb{E}[(Y|\mathcal{A})(\cdot, B)1_A] = \mathbb{P}[A \cap \{Y \in B\}]$$

for all  $A \in \mathcal{A}$ .

Regular conditional distributions exist whenever  $(\Omega, \mathcal{F}, \mathbb{P})$  is a standard probability space. To give a construction, recall (c.f. section 3 of [3]) that there exist conditional measures  $\mathbb{P}_\omega^\mathcal{A}$  uniquely characterised by

$$\mathbb{E}[f|\mathcal{A}](\omega) = \int f d\mathbb{P}_\omega^\mathcal{A}.$$

Then

$$(Y|\mathcal{A})(\omega, \cdot) = Y_* \mathbb{P}_\omega^\mathcal{A}$$

Indeed, for  $B \in \xi$ ,

$$(Y|\mathcal{A})(\omega, B) = \mathbb{E}[1_{Y^{-1}(B)} | \mathcal{A}](\omega) = \int 1_{Y^{-1}(B)} d\mathbb{P}_\omega^\mathcal{A} = \mathbb{P}_\omega^\mathcal{A}(Y^{-1}(B)) = Y_* \mathbb{P}_\omega^\mathcal{A}(B).$$

We denote by  $[Y|\mathcal{A}]$  a random variable defined on a separate probability space with law  $(Y|\mathcal{A})$ .

We recall that given two further  $\sigma$ -algebras  $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F}$ , we say that they are independent given  $\mathcal{A}$  if for all  $U \in \mathcal{G}_1$  and  $V \in \mathcal{G}_2$

$$\mathbb{P}[U \cap V|\mathcal{A}] = \mathbb{P}[U|\mathcal{A}]\mathbb{P}[V|\mathcal{A}]$$

almost surely. Similarly, two random variables  $Y_1$  and  $Y_2$  are independent given  $\mathcal{A}$  if the  $\sigma$ -algebra they generate are. Note that if  $Y_1$  is  $\mathcal{A}$ -measurable, then it is independent given  $\mathcal{A}$  to every random variable  $Y_2$ .

Given a topological group  $G$  and two measures  $\mu_1$  and  $\mu_2$  we recall that the convolution  $\mu_1 * \mu_2$  is defined as

$$(\mu_1 * \mu_2)(B) = \int \int 1_B(gh) d\mu_1(g)d\mu_2(h)$$

for any measurable set  $B \subset G$ .

**Lemma 2.9** *Let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space,  $G$  be a topological group and  $g, h$  be  $G$ -valued random variables. Let  $\mathcal{A} \subset \mathcal{F}$  be a  $\sigma$ -algebra and assume that  $g$  and  $h$  are independent given  $\mathcal{A}$ . Then the following properties hold:*

- (i)  $(gh|\mathcal{A}) = (g|\mathcal{A}) * (h|\mathcal{A})$  almost surely.
- (ii)  $[gh|\mathcal{A}] = [g|\mathcal{A}] \cdot [h|\mathcal{A}]$  almost surely.

**Proof** To show (i), we note that by assumption  $g$  and  $h$  are independent with respect to  $\mathbb{P}_\omega^\mathcal{A}$  for almost all  $\omega \in \Omega$ . This implies that for  $f : G \rightarrow \mathbb{R}$  a continuous compactly supported function,

$$\mathbb{E}_{\mathbb{P}_\omega^\mathcal{A}}[f(gh)] = \mathbb{E}_{\mathbb{P}_\omega^\mathcal{A}}[\mathbb{E}_{\mathbb{P}_\omega^\mathcal{A}}[f(gh)|h]] = \mathbb{E}_{(z_1, z_2) \sim \mathbb{P}_\omega^\mathcal{A} \times \mathbb{P}_\omega^\mathcal{A}}[f(g(z_1)h(z_2))],$$

proving (i). (ii) follows from (i) on a suitable separate probability space.  $\square$

The aim of this subsection is to prove an abstract result relating entropy between scales and the trace. To do so, we first discuss conditional entropy and conditional trace. Let  $Y$  be a random variable on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and  $\mathcal{A} \subset \mathcal{F}$  be a  $\sigma$ -algebra. Denote by  $(Y|\mathcal{A})$  the regular conditional distribution as defined in Sect. 2.3. Assuming that  $(Y|\mathcal{A})$  is almost surely absolutely continuous, we define

$$H((Y|\mathcal{A}))(\omega) = H((Y|\mathcal{A})(\omega)).$$

Recall that if  $X_1$  and  $X_2$  are two random variables, then entropy of  $X_1$  given  $X_2$  is  $H(X_1|X_2) = H(X_1, X_2) - H(X_2)$ . If  $X_1$  and  $X_2$  have finite entropy and finite joint entropy, then by [11]\*Proposition 3,

$$H(X_1|X_2) = \mathbb{E}[H((X_1|X_2))]. \quad (2.1)$$

## 2.4 Conditional Entropy

We next give an abstract definition of the entropy at a scale for a smoothing functions  $s$ . Indeed, let  $g$  and  $s$  be random variables on  $G$  and assume that  $s$  is absolutely continuous. Then the entropy at scale  $s$  is defined as

$$H(g; s) = H(gs) - H(s)$$

Moreover, if  $s_1$  and  $s_2$  are absolutely continuous smoothing functions we define the entropy between scales  $s_1$  and  $s_2$  as

$$H(g; s_1|s_2) = H(g; s_1) - H(g; s_2).$$

The following basic result on the growth of conditional entropy holds.

**Lemma 2.10** *Let  $g, s_1, s_2$  be independent random variables taking values in  $G$ . Assume that  $s_1$  and  $s_2$  are absolutely continuous with finite differential entropy and assume that  $gs_1$  and  $gs_2$  also have finite differential entropy. Then*

$$H(gs_1|gs_2) \geq H(g; s_1|s_2) + H(s_1).$$

**Proof** Note that

$$H(gs_2|gs_1) \geq H(gs_2|g, s_1) = H(gs_2|g) = H(s_2),$$

having used in the inequality that conditioning reduces entropy as in [2]\*Section 8.6, and so

$$H(gs_2, gs_1) = H(gs_2|gs_1) + H(gs_1) \geq H(gs_1) + H(s_2).$$

Therefore

$$\begin{aligned} H(gs_1|gs_2) &= H(gs_2, gs_1) - H(gs_2) \\ &\geq H(gs_1) - H(gs_2) + H(s_2) \\ &\geq H(g; s_1|s_2) + H(s_1). \end{aligned}$$

□

## 3 Entropy Growth for Stopped Random Walks

The reader may recall the definition of  $H_a(g; r)$  as given in (1.2). In this section we prove Theorem 1.2, which we restate for convenience, and Corollary 1.3.

Given a probability measure  $\mu$ , we sample independent  $\mu$ -distributed random variables  $\gamma_1, \gamma_2, \dots$  that are independent from  $(s_{a,r})_{a \geq 1, r > 0}$  and we denote

$$q_n = \gamma_1 \cdots \gamma_n.$$

**Theorem 3.1** (Theorem 1.2) Let  $\mu$  be a finitely supported probability measure on  $G$ . Let  $\eta = (\eta_n)_{n \geq 1}$  be a sequence of stopping times satisfying the large deviation principle and write  $L_n = \mathbb{E}[\eta_n]$  for  $n \geq 1$ . Let  $a \geq 1$ ,  $\varepsilon > 0$  and let  $r_n > 0$  be a sequence satisfying for all  $n \geq 1$ ,

$$r_n \leq a^{-1} c_G M_{\lceil (1+\varepsilon)L_n \rceil}$$

for a constant  $c_G > 0$  depending only on  $G$ . Then for all  $n \geq 1$ ,

$$H_a(q_{\eta_n}; r_n) \geq h_\mu \cdot L_n + o_{\mu, \eta, \varepsilon}(L_n).$$

Recall that  $H_a(q_{\eta_n}; r_n) = H(q_{\eta_n} s_{a, r_n}) - H(s_{a, r_n})$ . To give the proof idea, note that by assuming a large deviation principle, with high probability  $\eta_n \approx \mathbb{E}[\eta_n]$ . Also, by definition of  $h_\mu$ , we have that  $H(q_{L_n}) \geq h_\mu \cdot L_n$ . On the other hand,  $s_{a, r_n}$  is mostly contained in a ball around the identity with radius  $O(M_{L_n})$ , and therefore by Lemma 2.3 we have  $H(q_{L_n} \cdot s_{a, r_n}) = H(q_{L_n}) + H(s_{a, r_n})$ , which implies the claim. We proceed with a more rigorous proof.

**Proof** We note that if the assumption holds for some  $\varepsilon$ , it holds for all sufficiently small  $\varepsilon$ . Therefore it suffices to show that for all sufficiently small fixed  $\varepsilon > 0$  we have that

$$H_a(q_{\eta_n}; r_n) \geq h_\mu \cdot L_n + O_{\mu, \eta}(\varepsilon L_n) + o_{\mu, \eta, \varepsilon}(L_n).$$

So fix some  $\varepsilon > 0$  which is sufficiently small in terms of  $\mu$  and consider  $\eta'_n$  as

$$\eta'_n = \begin{cases} \lceil (1 + \varepsilon)L_n \rceil & \text{if } \eta_n > \lceil (1 + \varepsilon)L_n \rceil, \\ \lfloor (1 - \varepsilon)L_n \rfloor & \text{if } \eta_n < \lfloor (1 - \varepsilon)L_n \rfloor, \\ \eta_n & \text{otherwise.} \end{cases}$$

For a random variable  $X$  denote by  $\mathcal{L}(X)$  its law. Furthermore, given an event  $A$ , we will denote by  $\mathcal{L}(X)|_A$  the measure given by the push forward of the restriction of  $\mathbb{P}$  to  $A$  under the random variable  $X$ . Note that  $\|\mathcal{L}(X)|_A\|_1 = \mathbb{P}[A]$ .

By applying Lemma 2.1,

$$\begin{aligned} H(q_{\eta_n} s_{a, r_n}) &= H(\mathcal{L}(q_{\eta_n}) * \mathcal{L}(s_{a, r_n})) \\ &\geq H(\mathcal{L}(q_{\eta_n})|_{\eta_n = \eta'_n} * \mathcal{L}(s_{a, r_n})) + H(\mathcal{L}(q_{\eta_n})|_{\eta_n \neq \eta'_n} * \mathcal{L}(s_{a, r_n})) \\ &\geq H(\mathcal{L}(q_{\eta_n})|_{\eta_n = \eta'_n} * \mathcal{L}(s_{a, r_n})) + \mathbb{P}[\eta_n \neq \eta'_n] H(\mathcal{L}(s_{a, r_n})), \end{aligned} \quad (3.1)$$

having used that

$$H(\mathcal{L}(q_{\eta_n})|_{\eta_n \neq \eta'_n} * \mathcal{L}(s_{a, r_n})) \geq \mathbb{P}[\eta_n \neq \eta'_n] H(\mathcal{L}(s_{a, r_n})),$$

which can be shown by conditioning on  $q_n$  (as conditioning reduces entropy [2]\*Section 8.6) and using that  $q_n$  and  $s_{a, r_n}$  are independent.

We next apply that  $s_{a,r_n}$  has small support. Indeed, as  $d(s_{a,r_n}, e) \ll_G r_n a = o(M_{\lceil L_n(1+\varepsilon) \rceil})$  it follows that if  $n$  is sufficiently large,

$$d(s_{a,r_n}, \text{Id}) < \frac{1}{2} M_{\lceil L_n(1+\varepsilon) \rceil} \leq \frac{1}{2} \min_{x,y \in \text{supp}(q_{\eta'_n}), x \neq y} d(x, y).$$

In particular, by Lemma 2.3,

$$H(\mathcal{L}(q_{\eta_n})|_{\eta_n=\eta'_n} * \mathcal{L}(s_{a,r_n})) = H(\mathcal{L}(q_{\eta_n})|_{\eta_n=\eta'_n}) + \mathbb{P}[\eta_n = \eta'_n] H(\mathcal{L}(s_{a,r_n})). \quad (3.2)$$

Combining (3.2) with (3.1),

$$H(q_{\eta_n} s_{a,r_n}) \geq H(\mathcal{L}(q_{\eta_n})|_{\eta_n=\eta'_n}) + H(s_{a,r_n}).$$

It remains to estimate  $H(\mathcal{L}(q_{\eta_n})|_{\eta_n=\eta'_n})$ . Consider the random variable

$$X' = (q_{\lfloor (1-\varepsilon)L_n \rfloor}, \gamma_{\lfloor (1-\varepsilon)L_n \rfloor+1}, \gamma_{\lfloor (1-\varepsilon)L_n \rfloor+2}, \dots, \gamma_{\lceil (1+\varepsilon)L_n \rceil}).$$

As  $q_{\eta'_n}$  is completely determined by  $X'$ , we have  $H(X'|q_{\eta'_n}) = H(X') - H(q_{\eta'_n})$ .

Let  $K$  be the number of points in the support of  $\mu$ . Note that if

$$\gamma_{\lfloor (1-\varepsilon)L_n \rfloor+1}, \gamma_{\lfloor (1-\varepsilon)L_n \rfloor+2}, \dots, \gamma_{\lceil (1+\varepsilon)L_n \rceil}$$

and  $\eta'_n$  are fixed, then for any possible value of  $q_{\eta'_n}$  there is at most one choice of  $q_{\lfloor (1-\varepsilon)L_n \rfloor}$  which would lead to this value of  $q_{\eta'_n}$ . Therefore for each  $y$  in the image of  $q_{\eta'_n}$  there are at most  $(2\varepsilon L_n + 2)K^{2\varepsilon L_n + 2}$  elements  $x$  in the image of  $X'$  such that  $\mathbb{P}[X' = x \cap q_{\eta'_n} = y] > 0$ . Therefore  $(X'|q_{\eta'_n})$  is almost surely supported on less than  $(2\varepsilon L_n + 2)K^{2\varepsilon L_n + 2}$  points and hence by (2.1),

$$H(X'|q_{\eta'_n}) \leq \log \left( (2\varepsilon L_n + 2)K^{2\varepsilon L_n + 2} \right) \leq 2\varepsilon \log K \cdot L_n + o_{\mu,\varepsilon}(L_n).$$

On the other hand,

$$H(X') \geq H(q_{\lfloor L_n \rfloor}) \geq h_\mu \cdot \lfloor L_n \rfloor \quad (3.3)$$

and therefore

$$H(q_{\eta'_n}) \geq (h_\mu - 2\varepsilon \log K)L_n - o_{\mu,\varepsilon}(L_n).$$

To continue, we note that by Lemma 2.2,

$$H(q_{\eta'_n}) \leq H(\mathcal{L}(q_{\eta'_n})|_{\eta_n=\eta'_n}) + H(\mathcal{L}(q_{\eta'_n})|_{\eta_n \neq \eta'_n}) + \log 2. \quad (3.4)$$

We wish to bound  $H(\mathcal{L}(q_{\eta'_n})|_{\eta_n=\eta'_n})$  from below. By the large deviation principle,  $\mathbb{P}[\eta_n \neq \eta'_n] \leq \alpha^{L_n}$  for  $\alpha \in (0, 1)$  only depending on  $\varepsilon$  and  $\mu$ . We also know that conditional on  $\eta_n \neq \eta'_n$ , there are at most  $2K^{\lceil(1+\varepsilon)m\rceil}$  possible values for  $q_{\eta'_n}$  and therefore

$$H(\mathcal{L}(q_{\eta'_n})|_{\eta_n \neq \eta'_n}) \leq \alpha^{L_n} \log(2K^{\lceil(1+\varepsilon)L_n\rceil}) = o_{\mu, \varepsilon}(L_n).$$

This implies

$$H(\mathcal{L}(q_{\eta'_n})|_{\eta_n=\eta'_n}) \geq (h_\mu - 2\varepsilon \log K)L_n - o_{\mu, \eta, \varepsilon}(L_n),$$

concluding the proof.  $\square$

**Proof** (of Corollary 1.3) For sufficiently small  $\varepsilon > 0$  it holds that as  $L_n \rightarrow \infty$  that

$$e^{-S \cdot L_n} = o(e^{-S_{\lceil(1+\varepsilon)L_n\rceil} \cdot \lceil(1+\varepsilon)L_n\rceil}) = o(M_{\lceil(1+\varepsilon)L_n\rceil})$$

as  $S > (1 + \varepsilon)S_m$  for all sufficiently large  $m$ . The claim follows from Theorem 3.1.  $\square$

## 4 Entropy and Trace on Lie groups

In this section we prove (1.6) in Sect. 4.1 and in Sect. 4.3 we establish Theorem 1.4. In Sect. 4.2 an auxiliary result necessary for the proof of Theorem 1.4 will be shown.

### 4.1 Entropy and Trace

In this subsection we prove (1.6). Recall that given  $g_0 \in G$  and a random variable  $g$  on  $G$  we define

$$\text{tr}_{g_0}(g) = \text{tr}(\log(g_0^{-1}g)),$$

whenever  $\log(g_0^{-1}g)$  is defined.

**Proposition 4.1** *Let  $G$  be a Lie group of dimension  $\ell$ . Let  $\varepsilon > 0$  and suppose that  $g$  is a absolutely continuous random variable taking values in  $B_\varepsilon(g_0)$  for some  $g_0 \in G$ . If  $\varepsilon$  is sufficiently small depending on  $G$ ,*

$$H(g) \leq \frac{\ell}{2} \log\left(\frac{2\pi e}{\ell} \cdot \text{tr}_{g_0}(g)\right) + O_G(\varepsilon).$$

**Proof** We first note that if  $X$  is an  $\ell$ -dimensional absolutely continuous random vector, then

$$H(X) \leq \frac{\ell}{2} \log\left(\frac{2\pi e}{\ell} \cdot \text{tr}(X)\right) \tag{4.1}$$

Indeed, it follows from the one-dimensional case (1.5) that  $H(X) \leq \frac{1}{2} \log((2\pi e)^\ell \cdot |\text{Var}(X)|)$ , where  $|\text{Var}(X)|$  is the determinant of the covariance matrix. Note that by the AM-GM inequality  $|\text{Var}(X)| \leq \text{tr}(X)^\ell \ell^{-\ell}$ , which implies (4.1).

Since  $H(g_0^{-1}g) = H(g)$  and  $\text{tr}_{g_0}(g) = \text{tr}_e(g_0^{-1}g)$ , we may assume without loss of generality that  $g_0 = e$ . The density  $\frac{dm_G|_{B_\varepsilon(e)}}{d(m_g \circ \log)|_{B_\varepsilon(e)}}$  is smooth and for  $\varepsilon > 0$  sufficiently small is  $1 + O_G(\varepsilon)$  and therefore  $\sup |\log \frac{dm_G|_{B_\varepsilon(e)}}{d(m_g \circ \log)|_{B_\varepsilon(e)}}| \ll_G \varepsilon$ . Thus by Lemma 2.7,

$$|D_{\text{KL}}(g \parallel m_G) - D_{\text{KL}}(g \parallel m_g \circ \log)| \ll_G \varepsilon.$$

The claim follows since by (4.1)

$$D_{\text{KL}}(g \parallel m_g \circ \log) = D_{\text{KL}}(\log(g) \parallel m_g) = H(\log(g)) \leq \frac{\ell}{2} \log \left( \frac{2\pi e}{\ell} \text{tr}_e(g) \right).$$

□

## 4.2 Conditional Trace and Entropy Between Scales

We next define the conditional trace of a random variable on  $G$  and relate it to the entropy between scales.

**Definition 4.2** Let  $g$  be a random variable defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and taking values in  $G$ . Let  $\mathcal{A} \subset \mathcal{F}$  be a  $\sigma$ -algebra let  $g_0$  be a  $\mathcal{A}$ -measurable random variable taking values on  $G$ . Then we denote by  $\text{tr}_{g_0}(g \mid \mathcal{A})$  the  $\mathcal{A}$ -measurable function given for  $\omega \in \Omega$  by

$$\text{tr}_{g_0}(g \mid \mathcal{A})(\omega) = \text{tr}_{g_0(\omega)}((g \mid \mathcal{A})(\omega)),$$

whenever this expression is well-defined.

We note here that the variance of a measure  $\mu$  is defined as the variance of a random variable with law  $\mu$ . It follows from Proposition 4.1 that when  $(g \mid \mathcal{A})$  is almost surely absolutely continuous,

$$H((g \mid \mathcal{A})) \leq \frac{\ell}{2} \log \left( \frac{2\pi e}{\ell} \cdot \text{tr}_{g_0}(g \mid \mathcal{A}) \right) + O_G(\varepsilon). \quad (4.2)$$

**Proposition 4.3** Let  $g, s_1$  and  $s_2$  be independent absolutely continuous random variables taking values in  $G$  and suppose that  $s_1$  and  $s_2$  are supported on  $B_\varepsilon$  for some sufficiently small  $\varepsilon > 0$  and have finite differential entropy. Write  $c = \frac{\ell}{2} \log \frac{2\pi e}{\ell} \text{tr}_e(s_1) - H(s_1)$  and suppose that  $\text{tr}_e(s_1) \geq A\varepsilon^2$  for some constant  $A > 0$ . Then

$$\mathbb{E}[\text{tr}_{gs_2}(g \mid gs_2)] \geq \frac{2}{\ell} (H(g; s_1 \mid s_2) - c - O_G(A^{-1}\varepsilon)) \text{tr}_e(s_1).$$

We first prove some basic result on the trace of the product of two random variables.

**Lemma 4.4** *Let  $\varepsilon > 0$  be sufficiently small and let  $a, b$  be random variables and  $\mathcal{A}$  a  $\sigma$ -algebra. Suppose that  $b$  is independent from  $a$  and  $\mathcal{A}$  and let  $g_0$  be an  $\mathcal{A}$ -measurable random variable. Suppose that  $g_0^{-1}a$  and  $b$  are almost surely contained in  $B_\varepsilon$ . Then*

$$\text{tr}_{g_0}(ab|\mathcal{A}) = \text{tr}_{g_0}(a|\mathcal{A}) + \text{tr}_e(b) + O_G(\varepsilon^3).$$

Note that under the assumptions of Lemma 4.4 it holds by Lemma 2.9 that

$$[ab|\mathcal{A}] = [a|\mathcal{A}][b|\mathcal{A}] = [a|\mathcal{A}]b.$$

Therefore the claim follows from the following unconditional version.

**Lemma 4.5** *Let  $\varepsilon > 0$  be sufficiently small and let  $g$  and  $h$  be independent random variables taking values in  $G$ . Suppose that the image of  $g$  is contained in  $B_\varepsilon$  and the image of  $h$  is contained in  $B_\varepsilon(h_0)$  for some  $h_0 \in G$ . Then*

$$\text{tr}_{h_0}(hg) = \text{tr}_{h_0}(h) + \text{tr}_e(g) + O_G(\varepsilon^3).$$

**Proof** Let  $X = \log(h_0^{-1}h)$  and let  $Y = \log(g)$ . Then  $|X|, |Y| \leq \varepsilon$  almost surely and by Taylor's theorem there is a random variable  $E$  with  $|E| \ll \varepsilon^2$  almost surely such that

$$\log(\exp(X) \exp(Y)) = X + Y + E.$$

Therefore

$$\begin{aligned} \text{tr}_{h_0}(hg) &= \mathbb{E}[|X + Y + E|^2] - |\mathbb{E}[X + Y + E]|^2 \\ &= \mathbb{E}[|X + Y|^2] - |\mathbb{E}[X + Y]|^2 \\ &\quad + 2\mathbb{E}[(X + Y) \cdot E] + \mathbb{E}[|E|^2] - 2\mathbb{E}[X + Y]\mathbb{E}[E] - |\mathbb{E}[E]|^2 \\ &= \text{Var}[X + Y] + O_G(\varepsilon^3) = \text{Var}[X] + \text{Var}[Y] + O_G(\varepsilon^3). \end{aligned}$$

□

**Proof** (of Proposition 4.3) We note that by (2.1) and Lemma 2.10, it holds that

$$\mathbb{E}[H((gs_1|gs_2))] \geq H(g; s_1|s_2) + H(s_1)$$

and so by (4.2),

$$\mathbb{E}\left[\frac{\ell}{2} \log \frac{2\pi e}{\ell} \text{tr}_{gs_2}(gs_1|gs_2)\right] + O_G(\varepsilon) \geq H(g; s_1|s_2) + H(s_1).$$

Note that  $(gs_2)^{-1}g = s_2^{-1}$ , which is contained in  $B_\varepsilon(e)$ . Therefore by Lemma 4.4,

$$\mathrm{tr}_{gs_2}(gs_1|gs_2) \leq \mathrm{tr}_{gs_2}(g|gs_2) + \mathrm{tr}_e(s_1) + O_G(\varepsilon^3)$$

and so

$$H(g; s_1|s_2) + H(s_1) \leq \mathbb{E} \left[ \frac{\ell}{2} \log \frac{2\pi e}{\ell} \left( \mathrm{tr}_{gs_2}(g|gs_2) + \mathrm{tr}_e(s_1) + O_G(\varepsilon^3) \right) \right] + O_G(\varepsilon).$$

Thus

$$\frac{2}{\ell} (H(g; s_1|s_2) - c) \leq \mathbb{E} \left[ \log \left( 1 + \frac{\mathrm{tr}_{gs_2}(g|gs_2)}{\mathrm{tr}_e(s_1)} + O_G(A^{-1}\varepsilon) \right) \right].$$

Using that  $\log(1 + x) \leq x$  for  $x \geq 0$ , we conclude the claim.  $\square$

### 4.3 Proof of Theorem 1.4

The proof relies on the following lemma. We recall from the introduction that  $\beta_{a,r}$  is the random variable with density function  $f_{a,r} : \mathfrak{g} \rightarrow \mathbb{R}$  given by

$$f_{a,r}(x) = \begin{cases} C_{a,r} e^{-\frac{|x|^2}{2r^2}} & \text{if } |x| \leq ar, \\ 0 & \text{otherwise,} \end{cases}$$

where  $C_{a,r}$  is a normalising constant to ensure that  $f_{a,r}$  integrates to 1.

**Lemma 4.6** *The following properties hold for  $r > 0$  and  $a \geq 1$ :*

(i)  $\ell r^2 \ll \mathrm{tr}(\beta_{a,r}) \leq \ell r^2$  and for  $\eta_{a,r}$  the distribution of  $\beta_{a,r}$ ,

$$H(\eta_{a,r}) = \frac{\ell}{2} \log 2\pi e r^2 + O_\ell(e^{-a^2/4}).$$

(ii) If  $ar$  is sufficiently small,  $\ell r^2 \ll \mathrm{tr}_e(s_{a,r}) \leq \ell r^2$  and

$$H(s_{a,r}) = \frac{\ell}{2} \log 2\pi e r^2 + O_\ell(e^{-a^2/4}) + O_G(ar).$$

**Proof** We note that (ii) follows from (i). To prove (i), we deal initially with the  $r = 1$  case. Note first that

$$\int_{x \in \mathbb{R}^\ell, |x| \leq a} e^{-|x|^2/2} dx \leq \int_{x \in \mathbb{R}^\ell} e^{-|x|^2/2} dx = \prod_{i=1}^{\ell} \int_{\mathbb{R}} e^{-x_i^2/2} dx_i = (2\pi)^{\ell/2}$$

and by using spherical coordinates

$$\begin{aligned} \int_{x \in \mathbb{R}^\ell, |x| \geq a} e^{-|x|^2/2} dx &= c_\ell \int_a^\infty u^{\ell-1} e^{-u^2/2} du \\ &\ll_\ell \int_a^\infty e^{-u^2/3} du \leq \int_a^\infty e^{-au/3} du = \frac{3}{a} e^{-a^2/3} \ll_\ell e^{-a^2/4}. \end{aligned}$$

Thus we conclude

$$\begin{aligned} \int_{x \in \mathbb{R}^\ell, |x| \leq a} e^{-|x|^2/2} dx &= (2\pi)^{\ell/2} - \int_{x \in \mathbb{R}^\ell, |x| \geq a} e^{-|x|^2/2} dx \\ &\geq (2\pi)^{\ell/2} - O_\ell(e^{-a^2/4}) \end{aligned}$$

and therefore  $C_{1,a} = (2\pi)^{-\ell/2} + O_\ell(e^{-a^2/4})$ . We are now in a suitable position to calculate  $H(\eta_{1,a})$ . Indeed,

$$\begin{aligned} H(\eta_{1,a}) &= \int_{|x| \leq a} -C_{1,a} e^{-|x|^2/2} \log(C_{1,a} e^{-|x|^2/2}) dx \\ &= \int_{|x| \leq a} C_{1,a} \left( \frac{|x|^2}{2} - \log C_{1,a} \right) e^{-|x|^2/2} dx \end{aligned}$$

We calculate

$$\begin{aligned} &\int_{x \in \mathbb{R}^\ell} C_{1,a} \left( \frac{|x|^2}{2} - \log C_{1,a} \right) e^{-|x|^2/2} dx \\ &= (2\pi)^{\ell/2} C_{1,a} \left( \frac{\ell}{2} - \log C_{1,a} \right) \\ &= \left( 1 + O_\ell(e^{-a^2/4}) \right) \left( \frac{\ell}{2} \log e + \frac{\ell}{2} \log 2\pi + O_\ell(e^{-a^2/4}) \right) \\ &= \frac{\ell}{2} \log 2\pi e + O_\ell(e^{-a^2/4}). \end{aligned}$$

and again using spherical coordinates,

$$\begin{aligned} &\int_{|x| \geq a} C_{1,a} \left( \frac{|x|^2}{2} - \log C_{1,a} \right) e^{-|x|^2/2} dx \\ &= c_\ell \int_a^\infty C_{1,a} \left( \frac{u^2}{2} - \log C_{1,a} \right) u^{\ell-1} e^{-u^2/2} du \\ &\ll_\ell O_\ell(e^{-a^2/4}). \end{aligned}$$

Thus the claimed bound on  $H(\eta_{1,a})$  follows. Since  $f_{a,r}(x) = r^\ell C_{1,a} f_{1,a}(x/r)$  it follows that  $H(\eta_{a,r}) = \log(r^\ell) + H(\eta_{1,a})$  and hence the proof is complete.  $\square$

**Proof** (of Theorem 1.4) We apply Proposition 4.3 to  $s_1 = s_{a,r}$  and  $s_2 = s_{a,2r}$  and we set  $\varepsilon = \ell ar$ . By Lemma 4.6 (ii) we have that  $\text{tr}_e(s_1) \gg_G \ell r^2 \gg_G a^{-2}\varepsilon^2$  and  $c = \frac{\ell}{2} \log \frac{2\pi e}{\ell} \text{tr}_e(s_1) - H(s_1) \leq O_G(e^{-a^2/4} + ar)$ . Applying Proposition 4.3 with  $A = a^{-2}$ ,

$$\mathbb{E}[\text{tr}_{gs_2}(g|gs_2)] \gg_G r^2(H_a(g; r|2r) - O_G(e^{-a^2/4} + a^3r)).$$

On the other hand, we have that  $|\log((gs_2)^{-1}g)| = |\log s_2| \leq 2ar$  and therefore

$$\text{tr}(g; 2ar) \geq (2ar)^{-2} \mathbb{E}[\text{tr}_{gs_2}(g|gs_2)] \gg_G a^{-2}(H_a(g; r|2r) - O_G(e^{-a^2/4} + a^3r)).$$

□

## 5 From Entropy Gap to Trace Sum

In this section we prove Proposition 1.5, which we deduce from the following two propositions.

**Proposition 5.1** *Let  $g$  be a  $G$ -valued random variable independent of  $(s_{a,r})_{a \geq 1, r > 0}$  and let  $0 < r_1 < r_2$ . Let  $a \geq 1$  such that  $ar_2$  is sufficiently small in terms of  $G$ . Suppose that for all  $r'_1 \in [r_1, 2r_1]$  as well as  $r'_2 \in [r_2/2, 2r_2]$  it holds for some constant  $C > 0$  that*

$$H_a(g; r'_1|r'_2) \geq C.$$

Then

$$\int_{ar_1}^{4ar_2} \frac{1}{u} \text{tr}(g; u) du \gg_G a^{-2}(C - N \cdot (O_G(e^{-\frac{a^2}{4}} + a^3r_2)))$$

$$\text{for } N = \left\lceil \frac{\log r_2 - \log r_1}{\log 2} \right\rceil - 1.$$

**Proof** Let  $a \geq 1$  and set  $N = \left\lceil \frac{\log r_2 - \log r_1}{\log 2} \right\rceil - 1$ . Note that  $2^{N+1}r_1 \geq r_2$  as well as  $2^Nr_1 \leq r_2$ .

Given  $u \in [1, 2)$  and an integer  $1 \leq i \leq N$  denote

$$k_i(u) = H_a(g; 2^{i-1}ur_1|2^iur_1).$$

Then by Theorem 1.4, there is some constant  $c = c(G) > 0$  depending only on  $G$  such that

$$\text{tr}(g; a2^iur_1) \geq ca^{-2}(k_i(u) - O_G(e^{-a^2/4} + a^32^iur_1)). \quad (5.1)$$

Thus

$$\sum_{i=1}^N \text{tr}(g; a2^i ur_1) \geq ca^{-2} \left( \sum_{i=1}^N k_i(u) - O_G(Ne^{-\frac{a^2}{4}} + Na^3 2^N r_1) \right).$$

Note that for  $u \in [1, 2]$  we have  $a2^{N+1}ur_1 \leq 4ar_2$  and  $aur_1 \geq ar_1$ . Therefore, upon substituting  $v = a2^i ur_1$  in the third line,

$$\begin{aligned} \int_{ar_1}^{4ar_2} \frac{1}{v} \text{tr}(g; v) dv &\geq \sum_{i=1}^N \int_{a2^i ur_1}^{a2^{i+1} ur_1} \frac{1}{v} \text{tr}(g; v) dv \\ &\geq \sum_{i=1}^N \int_1^2 \frac{1}{u} \text{tr}(g; a2^i ur_1) du \\ &\geq ca^{-2} \int_1^2 \frac{1}{u} \left( \sum_{i=1}^N k_i(u) - O_G(Ne^{-\frac{a^2}{4}} + Na^3 2^N r_1) \right) du. \end{aligned} \quad (5.2)$$

Observe that by our assumption  $\sum_{i=1}^N k_i(u) = H_a(g; ur_1 | 2^N ur_1) \geq C$  and therefore the claim follows using that  $2^N r_1 \leq r_2$ .  $\square$

**Proposition 5.2** Suppose that for a  $G$ -valued random variable  $g$  and  $0 < r_1 < r_2$  we have for some constant  $C_1 > 0$  that

$$\int_{r_1}^{r_2} \frac{1}{u} \text{tr}(g; u) du \geq C_1.$$

Let  $A > 1$ . Then there exists  $s_1, \dots, s_m \in (r_1, r_2)$  where  $m = \lceil \frac{\log r_2 - \log r_1}{2 \log A} \rceil$  such that

$$\sum_{i=1}^m \text{tr}(g; s_i) \geq \frac{C_1}{4 \log A} \quad \text{and} \quad s_{i+1} \geq As_i$$

for all  $1 \leq i \leq m-1$ .

**Proof** Define  $a_1, a_2, \dots, a_{2m+1}$  by  $a_i = r_1 A^{i-1}$ . Therefore  $a_1 = r_1$  and  $a_{2m+1} \geq r_2$ . Let  $U$  and  $V$  be defined by

$$U = \bigcup_{i=1}^m [a_{2i-1}, a_{2i}) \quad \text{and} \quad V = \bigcup_{i=1}^m [a_{2i}, a_{2i+1}).$$

Without loss of generality, upon replacing  $U$  with  $V$ , by our assumption

$$\int_U \frac{1}{u} \text{tr}(g; u) du \geq C_1/2.$$

For  $i \in [m]$  let  $s_i \in (a_{2i-1}, a_{2i})$  be chosen such that

$$\text{tr}(g; s_i) \geq \frac{1}{2} \sup_{u \in (a_{2i-1}, a_{2i})} \text{tr}(g; u).$$

In particular,

$$\text{tr}(g; s_i) \geq \frac{1}{2 \log A} \int_{a_{2i-1}}^{a_{2i}} \frac{1}{u} \text{tr}(g; u) du.$$

Summing over  $i$  gives

$$\sum_{i=1}^m \text{tr}(g; s_i) \geq \frac{1}{2 \log A} \int_U \frac{1}{u} \text{tr}(g; u) du \geq \frac{C_1}{4 \log A}.$$

□

To deduce Proposition 1.5, one uses Proposition 5.2 with the range  $(ar_1, 4ar_2)$  and with

$$C_1 = a^{-2} (C - N \cdot (O_G(e^{-\frac{a^2}{4}}) + a^3 r_2)).$$

**Acknowledgements** The first-named author gratefully acknowledges support from the Heilbronn Institute for Mathematical Research. This work is part of the second-named author's PhD thesis conducted at the University of Oxford, supported by a Mathematical Institute Scholarship. We thank the anonymous referee for helpful comments and corrections.

**Author Contributions** The whole manuscript was written with equal contributions from both authors.

## References

1. Bárány, B., Hochman, M., Rapaport, A.: Hausdorff dimension of planar self-affine sets and measures. *Invent. Math.* **216**, 601–659 (2019)
2. Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. Wiley, Hoboken, NJ (2006)
3. Einsiedler, M., Ward, T.: Ergodic theory with a view towards number theory, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, MR2723325 (2011)
4. Hochman, M.: On self-similar sets with overlaps and inverse theorems for entropy. *Ann. Math.* **180**(2), 773–822 (2014)
5. Hochman, M.: On self-similar sets with overlaps and inverse theorems for entropy in  $\mathbb{R}^d$  (2017) [arxiv:1503.09043](https://arxiv.org/abs/1503.09043)
6. Johnson, O.: Information Theory and the Central Limit Theorem. Imperial College Press, London (2004)
7. Kittle, S.: Absolutely continuous Furstenberg measures (2023) [arxiv:2305.05757](https://arxiv.org/abs/2305.05757)
8. Kittle, S., Kogler, C.: Dimension of contracting on average self-similar measures (2025). [arxiv:2501.17795](https://arxiv.org/abs/2501.17795)
9. Kittle, S., Kogler, C.: On absolute continuity of inhomogeneous and contracting on average self-similar measures (2025). [arxiv:2409.18936](https://arxiv.org/abs/2409.18936)
10. Varjú, P.P.: Absolute continuity of Bernoulli convolutions for algebraic parameters. *J. Amer. Math. Soc.* **32**(2), 351–397 (2019)
11. Vigneaux, J. P.: Entropy under disintegrations, Geometric science of information, pp. 340–349 (2021)

**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.