GRID ENTROPY IN LAST PASSAGE PERCOLATION, A VARIATIONAL FORMULA FOR GIBBS FREE ENERGY, AND APPLICATIONS TO A "CHOOSE THE BEST OF *D* SAMPLES" MODEL

by

ALEXANDRU GATEA

A thesis submitted in conformity with
the requirements for the degree of

Doctor of Philosophy

Graduate Department of Mathematics
University of Toronto

ABSTRACT

Grid entropy in last passage percolation, a variational formula for Gibbs Free Energy, and applications to a "choose the best of D samples" model

Alexandru Gatea

Doctor of Philosophy

Graduate Department of Mathematics

University of Toronto

2022

Working in the setting of i.i.d. last-passage percolation on $\mathbb{R}^D$ with no assumptions on the underlying edge-weight distribution, we develop the notion of grid entropy: a deterministic directed norm with negative sign that measures the proportion of empirical measures of edge weights (in a fixed direction or direction-free) which converge weakly to a given target measure. We study various properties of grid entropy, including an upper bound on the sum of relative and grid entropies, upper semicontinuity in most cases, and the fact that grid entropy can be described as the negative convex conjugate of Gibbs Free Energy. We show that the direction-free case is nothing more than the direction-fixed case in the $(1, 1, \ldots, 1)$ direction. In addition, we derive a grid entropy variational formula for the point-to-point/point-to-hyperplane Gibbs Free Energies that answers a directed polymer version of a question of Hoffman. Shifting gears, we proceed to study the limiting behaviour of empirical measures in a model consisting of repeatedly taking D samples from a distribution and picking out one according to an omniscient "strategy." We show that the set of limit points of empirical measures is almost surely the same whether or not we restrict ourselves to strategies which make the choices independently of all past and future choices, and furthermore, that this set coincides with the set of measures with finite grid entropy. These sets are convex and weakly compact; we characterize their extreme points as those given by a natural "greedy" deterministic strategy and we compute the grid entropy of said extreme points to be 0. This yields a description of the set of limit points of empirical measures as the closed convex hull of measures given by a density which is $D \cdot Beta(1, D)$ distributed. We also derive a simplified version of a grid entropy-based variational formula for Gibbs Free Energy for this model, and we present the dual formula for grid entropy.

ACKNOWLEDGEMENTS

I must begin by recognizing my advisor Bálint Virág. Thank you for your patience in helping me find a research direction that interested me, for the keen insights and suggestions you have given me in this subject matter on a weekly basis, and most importantly for being a voice of reason whenever I reached an impasse. This thesis is only possible thanks to you.

Another person who has made a profound impact on mathematical journey is Professor Dan Wolczuk. I will always be indebted to you for exposing me to the beauty of mathematics at a young age and instilling in me a voracious curiosity for this field. Thank you Dan!

In addition, I wish to acknowledge every teacher, from elementary school to university, regardless of discipline, for believing in me and molding me into the person I am today. And more specifically, I am very grateful to my advisors from my undergraduate summer math research experiences, Alexandru Nica, Nico Spronk and Kenneth Davidson, and from my Master's project, John Friedlander, for providing invaluable guidance and for helping shape my mathematical research acumen.

Next, I would like to thank Duncan Dauvergne for very helpful discussions and peer review on a preliminary draft.

I must mention the entire Department staff and faculty for their hard work in making my time here memorable. In particular, I wish to thank Jemima Merisca, who from my very first day at the University of Toronto made me feel welcome and provided me with words of wisdom that proved essential to my success.

I would also like to thank the Government of Canada for generously funding my research through an NSERC Canadian Graduate Scholarship.

Out of everyone, it was my family who drove me the hardest towards success. Thank you to my parents, grandparents and aunt for your support and encouragement to pursue my dreams.

True friends are hard to come by, especially when you are busy working on your thesis and a global pandemic has everyone quarantined at home. Sophie Harrington, a billion and one thanks for your relentless positivity, kindness and companionship. Our ecclectic chats and witty storytelling brought me great joy and fulfilment.

Finally, I wish to thank my friend Frenck Figure for his infectious joie de vivre and unique taste in humour. It is a privilege having you in my life.

PUBLICATIONS

Chapter 2 and Chapter 3 have been submitted separately for publication.

CONTENTS

1

INTRODUCTION

The limiting behaviour of empirical measures is of paramount importance in percolation theory. The normalized passage time along a path, what we ultimately care about, is nothing but the identity function integrated against the empirical measure of that path. In First/Last Passage Percolation we study the minimizing/maximizing paths called geodesics.

A major open problem posed by Hoffman in [Ahl15] is whether empirical measures along geodesics in a fixed direction converge weakly. This is partially answered in [Bat20] for FPP. Bates proves that the sets $\mathcal{R}^q, \mathcal{R}$ of limiting distributions in direction q, limiting distributions in the direction-free case respectively are deterministic and derives an explicit variational formula for the limit shape of the first passage time as the minimum value of a linear functional over $\mathcal{R}^q$. When the set of minimizers is a singleton, which Bates argues happens for a dense family of edge-weight distributions, it follows that Hoffman's question is answered in the affirmative. The same argument applied to the LPP model yields analogous conclusions.

Generally, there is no known way of computing the limiting distributions along geodesics. [Mar04] showcases some cutting edge developments for the solvable Exponential LPP on $\mathbb{Z}^2$ model, including an explicit formula for weak limits of empirical measures along geodesics in this setting. For other recent work on geodesics see [AH16], [JRAS19] and [JLS20].

Our aim in Chapter 2 is to build on the work in [Bat20] in the context of LPP by studying the more general question, what *entropy* of empirical measures in a fixed direction or direction-free converge in distribution to a certain target measure. Grid entropy gives an exact, deterministic answer. Once we have a working definition, we show that grid entropy $||(q,\nu)||$ of a target measure ν in direction $q \in \mathbb{R}^D_{\geqslant 0}$ is among other things positive-homogeneous and upper semicontinuous in ν and in most cases in (q,ν), satisfies the triangle inequality as well as an amazing inequality involving relative entropy and the total entropy of paths. Furthermore, we determine that Bates' set $\mathcal{R}^q$ almost surely coincides with the set of probability measures with finite grid entropy:

$$\mathcal{R}^q = \{\nu \text{ prob meas} : ||(q,\nu)|| > -\infty\} \text{ a.s.}$$

Analogous results hold for $\mathcal{R}$ in the direction-free case; in fact, the direction-free case turns out to be just the direction-fixed case in the unit direction $(\frac{1}{D}, \ldots, \frac{1}{D})$ which maximizes the total number of paths. We then apply these results to directed polymers to derive a variational formula for the Point-to-Point/Point-to-Hyperplane Gibbs Free Energies and to give an alternate definition of grid entropy as the negative convex conjugate of Gibbs Free Energy. This concludes Chapter 2.

Now random sampling is a fundamental area of study in statistics. Countless processes can be simulated as Monte Carlo experiments, with far-reaching applications in a range of fields from biology and economics to business. A rather simple but ubiquitous and surprisingly non-trivial experiment is that of taking D samples from a distribution and picking out one of these samples according to some omniscient "strategy."

This problem is dual to K-means clustering, a well-studied process where one groups a large set of samples into a fixed K number of buckets in order to minimize an error function called the quadratic distortion (see [LP20] for state-of-the-art).

In Chapter 3, we frame this experiment in the realm of percolation theory and ask what is the limiting behaviour of empirical measures along the event-dependent path of choices. Of course, answering this question provides a gold mine of information about the model. Note also that in this setting there is only one unit direction so we apply the direction-free theory from Chapter 2.

Looking back at Chapter 2 and [Bat20], one limitation is that only empirical measures along deterministic paths are considered, rather than allowing for mixtures of paths. In Chapter 3, we seek to rectify this by studying empirical measures along *random* paths (picked according to some probabilistic "strategy"). We will show that $\mathcal{R}$ is still the set of limit points, whether or not we assume the strategy is omniscient.

Another deficiency is that grid entropy and the set $\mathcal{R}$, like limit shape and other quantities in this area, are not known to be explicitly computable in most cases. One exception is [Bat20], where it is established that replacing the lattice $\mathbb{Z}^D$ by the infinite complete D-ary tree $\mathcal{T}_D$, the set $\mathcal{R}$ of limiting distributions in the direction-free case can be precisely described as a specific sublevel of relative entropy (in terms of D). The D-ary tree is in a way the dual model to the one we explore in Chapter 3, and as one might expect, we will be able to give a description of $\mathcal{R}$ in our setting too. We will precisely characterize the extreme points of $\mathcal{R}$ as those measures whose density is $D \cdot Beta(1, D)$ distributed; $\mathcal{R}$ is then the closed convex hull of these measures. In addition, we compute the grid entropy of the extreme points of $\mathcal{R}$ to be 0, and we give a simplification of our variational formula from Chapter 2 in this model.

2

GRID ENTROPY IN LAST PASSAGE PERCOLATION AND A VARIATIONAL FORMULA FOR GIBBS FREE ENERGY

2.1 DEFINITIONS AND RESULTS

We consider north-east nearest-neighbor paths on $\mathbb{Z}^D$, and work on $\mathbb{R}^D$ by taking coordinate-wise floors. We follow a similar approach to [Bat20], in that we couple our i.i.d. edge weights τ_e to i.i.d. edge labels Unif[0,1] random variables U_e via a measurable function $\tau : [0,1] \to [0,\infty)$ satisfying $\tau_e = \tau(U_e)$. This lets us work on [0,1] at no additional cost, as we can lift everything back to $\mathbb{R}$ via the pushforward of τ. We then interest ourselves with *how* many empirical measures $\frac{1}{n}\mu_\pi = \frac{1}{n}\sum_{e\in\pi}\delta_{U_e}$ for paths $\vec{0} \to \lfloor nq \rfloor$ converge to some given target measure ν in $\mathcal{M}_+$, the set of finite non-negative Borel measures on [0,1]. We may keep the direction q fixed, or we may vary q over all points in $\mathbb{R}^D$ with the same 1-norm $||q||_1$.

To perform the counting, we consider the order statistics of the distance of the paths' empirical measures $\frac{1}{n}\mu_\pi$ to the target ν, where distance is measured via the Levy-Prokhorov metric ρ, which metrizes weak convergence of measures. That is, given a direction $q \in \mathbb{R}^D$ and a target measure ν, for every $n \in \mathbb{N}$ we let

$$\overset{1}{\min_{\pi:\vec{0}\to\lfloor nq\rfloor}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) \leqslant \overset{2}{\min_{\pi:\vec{0}\to\lfloor nq\rfloor}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) \leqslant \ldots \leqslant \overset{\#\pi:\vec{0}\to\lfloor nq\rfloor}{\min_{\pi:\vec{0}\to\lfloor nq\rfloor}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right)$$

denote the order statistics value of $\rho(\frac{1}{n}\mu_\pi, \nu)$. In the direction-free case, where we count all paths of a certain scaled length from $\vec{0}$, we let $t := ||\nu||_{TV}$ and similarly define $\overset{j}{\min_{\pi \text{ s.t. } |\pi|=\lfloor nt\rfloor}} \rho(\frac{1}{n}\mu_\pi, \nu)$ over paths of length $\lfloor nt \rfloor$ anchored at $\vec{0}$. Of course, these minimums and the paths corresponding to them are event-dependent.

We then define the grid entropy with respect to the target ν and the direction q, denoted $||(q,\nu)||$, and the direction-free grid entropy, denoted $||\nu||$, to be the critical exponent where the corresponding minimums change from converging to 0 a.s. to diverging a.s.:

$$||(q,\nu)|| := \sup\left\{\alpha \geqslant 0 : \lim_{n\to\infty} \min_{\pi:\vec{0}\to\lfloor nq\rfloor}^{\lfloor e^{\alpha n}\rfloor} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0 \text{ a.s.}\right\}$$

$$||\nu|| := \sup\left\{\alpha \geqslant 0 : \lim_{n\to\infty} \min_{\pi \text{ s.t. } |\pi|=\lfloor nt\rfloor}^{\lfloor e^{\alpha n}\rfloor} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0 \text{ a.s.}\right\}$$

It will turn out that we can replace the limits in this definition with lim inf's and get the same quantity. Note that these grid entropies lie in $\{-\infty\} \cup [0, H(q)]$, $\{-\infty\} \cup [0, \log D]$ respectively where

$$H(q) := \sum_{i=1}^{D} -q_i \log \frac{q_i}{||q||_1}$$

is the "entropy" of the total number of paths in direction q, in the sense that

$$(\#\text{paths } \vec{0} \to nq) = e^{H(q)n+o(n)}$$

Over the course of Chapter 2 we will establish two other equivalent definitions of grid entropy which avoid the annoyance of dealing with these event-dependent orderings of the paths. One of these alternate descriptions is remarkably simple: grid entropy is the negative convex conjugate of Gibbs Free Energy. We also show that the direction-free grid entropy is just the direction-fixed grid entropy in the maximizing direction $(\frac{1}{D}, \dots, \frac{1}{D})$. We summarize these in the following theorem.

Theorem A. (i) Let $q \in \mathbb{R}^D_{\geqslant 0}$ and let ν be a finite non-negative Borel measure on [0,1]. Then the direction$-q$ grid entropy $||(q,\nu)||$ as defined above is also given by

$$||(q,\nu)|| = \inf_{\epsilon>0} \lim_{n\to\infty} \frac{1}{n} \log \sum_{\pi:\vec{0}\to\lfloor nq\rfloor} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)} \text{ a.s.}$$

The expressions we take an infimum of are each directed metrics with negative sign on $\mathbb{R}^D \times \mathcal{M}_+$. That is, they take value in $[-\infty, \infty)$, evaluate to 0 when $(q,\nu) = (\vec{0}, 0)$, and satisfy the triangle inequality with the sign reversed.

Moreover, if there exists $\beta > 0$ s.t. the moment-generating function $E[e^{\beta\tau(U)}]$ is finite for $U \sim Unif[0,1]$, then direction-fixed grid entropy is the negative convex conjugate of the point-to-point β-Gibbs Free Energy in direction q (as a function of the environment-coupling function τ):

$$||(q,\nu)|| = -(G_q^\beta)^*(\nu) = -\sup_\tau [\langle \tau, \nu \rangle - G_q^\beta(\tau)]$$

where the supremum is over measurable $\tau : [0,1] \to [0,\infty)$, where $\langle \cdot, \cdot \rangle$ is the integration linear functional $\langle \tau, \nu \rangle = \int_0^1 \tau(u) d\nu$ and where the point-to-point β-Gibbs Free Energy is given by

$$G_q^\beta = \lim_{n\to\infty} \frac{1}{n} \log \sum_{\pi:\vec{0}\to\lfloor nq \rfloor} e^{\beta T(\pi)}$$

(ii) Analogous results hold for the direction-free case. Let ν be a finite non-negative Borel measure on [0,1] and let $t := ||\nu||_{TV}$. Then the direction-free grid entropy $||\nu||$ as defined above is also given by

$$\begin{aligned}
||\nu|| &= \inf_{\epsilon>0} \lim_{n\to\infty} \frac{1}{n} \log \sum_{\pi\in\mathcal{P}_{\lfloor nt \rfloor}(\vec{0})} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)} \text{ a.s.} \\
&= \sup_{q\in\mathbb{R}^D_{\geqslant 0}, ||q||_1 = t} ||(q,\nu)|| \\
&= ||(t\ell,\nu)|| \text{ where } \ell = \left(\frac{1}{D}, \dots, \frac{1}{D}\right)
\end{aligned}$$

The expressions we take an infimum of are each directed metrics with negative sign on $\mathcal{M}_t$, the set of finite non-negative Borel measures with total mass t. Moreover, if there exists $\beta > 0$ s.t.

$$E[e^{\beta\tau(U)}] < \infty \text{ for } U \sim Unif[0,1]$$

then direction-free grid entropy is the negative convex conjugate of the point-to-hyperplane β-Gibbs Free Energy:

$$||\nu|| = -(G^\beta)^*(\nu) = -\sup_\tau [\langle \tau, \nu \rangle - G^\beta(\tau)] \text{ a.s.}$$

where the supremum is over measurable $\tau : [0,1] \to [0,\infty)$ and where the point-to-hyperplane β-Gibbs Free Energy is given by

$$G^\beta = \lim_{n\to\infty} \frac{1}{n} \log \sum_{\pi \text{ s.t. } |\pi|=n} e^{\beta T(\pi)}$$

Grid entropy satisfies some rather interesting properties. The following theorem captures the highlights for direction$-q$ grid entropy; the direction-free analogues hold as well.

Theorem B. Let $q \in \mathbb{R}^D_{\geqslant 0}$. Then:
(i) Grid entropy $||(q,\nu)||$ is a directed norm with negative sign; it scales with positive-factors and it satisfies a reverse triangle inequality:

$$||(p,\xi)|| + ||(q,\nu)|| \leqslant ||(p+q,\xi+\nu)||$$

(ii) Grid entropy $||(q,\nu)||$ is upper semicontinuous in the target meausure ν always and in both (q,ν) as long as q^k either approach from the SE of q or the q^k have the same number of zero coordinates as q.
Let $\mathcal{R}^q := \{\nu \in \mathcal{M}_+ : ||(q,\nu)|| > -\infty\}$. Then
(iii) $\mathcal{R}^q$ consists only of measures ν with total variation $||\nu||_{TV} = ||q||_1$ that are absolutely continuous with respect to the Lebesgue measure Λ on [0,1]
(iv) $\mathcal{R}^q$ is weakly closed, convex and deterministic and coincides almost surely with the set of all weak accumulation points of the empirical measures in direction q
(v) Any $\nu \in \mathcal{R}^q$ satisfies the following upper bound on the sum of the grid entropy and the relative entropy with respect to Lebesgue measure on [0,1]:

$$D_{KL}(\nu||\Lambda) + ||(q,\nu)|| \leqslant \sum_{i=1}^{D} -q_i \log \frac{q_i}{||q||_1} := H(q)$$

where again, this upper bound is simply the "entropy" of the total number of paths in direction q.

Why do we care? The deterministic set $\mathcal{R}^q$ is nothing more than the LPP analogue of the set Bates takes a infimum over in [Bat20] in his variational formula for the FPP time constant. In our LPP setting, his formula becomes

$$\text{LPP time constant} := \lim_{n\to\infty} \frac{\text{last passage time for paths } \vec{0} \to \lfloor nq \rfloor}{n} = \sup_{\nu \in \mathcal{R}^q} \langle \tau, \nu \rangle \text{ a.s.}$$

We thus give a more enlightening description of these sets $\mathcal{R}^q$ in terms of these grid entropies which have all these remarkable properties. For example, the reverse triangle inequality is especially nice because other forms of entropy (such as relative entropy) do not usually exhibit one. Also, it is worth noting that the bound in (iv) is an improvement of a version proved by Bates, without the grid entropy term.

The main application of grid entropy we present is in the context of directed polymers in LPP. We derive the following variational formula for the point-to-point/point-to-hyperplane Gibbs Free Energies in direction

q/to the hyperplane $x_1 + \ldots + x_D = 1$, and we show that Bates' variational formula for the LPP time constant is simply the zero temperature analogue.

Theorem C. Fix a direction $q \in \mathbb{R}^D_{\geqslant 0}$ and an inverse temperature $\beta > 0$ s.t. the environment-coupling function τ satisfies

$$E[e^{\beta\tau(U)}] < \infty \text{ for } U \sim Unif[0,1]$$

Then the point-to-point/point-to-hyperplane Gibbs Free Energies are given by

$$G^\beta_q = \sup_{\nu \in \mathcal{R}^q} [\beta\langle \tau, \nu \rangle + ||(q, \nu)||],$$

$$G^\beta = \sup_{\nu \in \mathcal{R}^1} [\beta\langle \tau, \nu \rangle + ||\nu||] = G^\beta_\ell = \sup_{q \in \mathbb{R}^D_{\geqslant 0}, ||q||_1 = 1} G^\beta_q$$

where $\ell = (\frac{1}{D}, \ldots, \frac{1}{D})$ is the maximizing direction, and $\mathcal{R}^1 = \bigcup_{q \in \mathbb{R}^D_{\geqslant 0}, ||q||_1 = 1} \mathcal{R}^q$ is the set of Borel probability measures ν that have finite direction-free grid entropy. Moreover, these supremums are achieved by some ν in $\mathcal{R}^q, \mathcal{R}^\ell$ respectively.

As in [Bat20], it follows that the directed polymer analogue of Hoffman's question is answered in the affirmative when our variational formula has a unique maximizer, which will happen for a dense family of measurable functions τ.

Theorem D. Fix an inverse temperature $\beta > 0$ s.t.

$$E[e^{\beta\tau(U)}] < \infty \text{ for } U \sim Unif[0,1]$$

(i) Fix $q \in \mathbb{R}^D_{\geqslant 0}$ and suppose $\beta\langle \tau, \nu \rangle + ||(q, \nu)||$ has a unique maximizer $\nu \in \mathcal{R}^q$. For every n pick a path $\pi_n : \vec{0} \to \lfloor nq \rfloor$ independently and at random according to the probabilities prescribed the corresponding point-to-point β-polymer measure

$$\rho^\beta_{n,q}(d\pi) = \frac{e^{\beta T(\pi)}}{\sum_{\pi : \vec{0} \to \lfloor nq \rfloor} e^{\beta T(\pi)}} \text{ for paths } \pi : \vec{0} \to \lfloor nq \rfloor$$

Then the empirical measures $\frac{1}{n}\mu_{\pi_n}$ converge weakly to ν a.s.

(ii) Suppose $\beta\langle \tau, \nu \rangle + ||\nu||$ has a unique maximizer $\nu \in \mathcal{M}_1$. For every n pick a length n path π_n from $\vec{0}$ independently and at random according to the probabilities prescribed the corresponding point-to-hyperplane β-polymer measure

$$\rho_n^{\beta}(d\pi) = \frac{e^{\beta T(\pi)}}{\sum\limits_{\pi \text{ s.t. } |\pi|=n} e^{\beta T(\pi)}} \text{ for length } n \text{ paths } \pi \text{ from } \vec{0}$$

Then the empirical measures $\frac{1}{n}\mu_{\pi_n}$ converge weakly to ν a.s.

The plan is as follows. In Section 2.2, we describe the model and setup, and outline various facts and notions we will need over the course of this chapter. Section 2.3 will focus on developing the second definition of grid entropy as a directed norm with negative sign and showing that it is equivalent to the original definition with the $\overset{j}{\underset{\pi}{\min}}$. Then, in Section 2.4, we investigate what information we can extract from grid entropy and what properties it satisfies (Theorem B). We devote Section 2.5 to applying our results to establish a variational formula for point-to-point/point-to-hyperplane Gibbs Free Energies and study the consequences of this (namely Theorems C, D). Last but not least, we make some closing remarks about adapting our results to other models.

2.2 PRELIMINARIES

2.2.1 *Empirical Measures on the Lattice*

We begin by briefly describing the setup we will use in this chapter.

We restrict ourselves to a directed Last Passage Percolation model. Consider north-east nearest-neighbour paths on the lattice $\mathbb{Z}^D, D \geqslant 1$ with i.i.d. edge weights $\tau_e \sim \theta$ for some probability distribution θ on $\mathbb{R}$. By north-east, we of course mean that the coordinates of points on the path are nondecreasing.

For $p, q \in \mathbb{Z}^D$ we denote by $\mathcal{P}(p,q)$ the set of all NE paths $\pi : p \to q$. Similarly, for $q \in \mathbb{Z}^D$ and $t \in \mathbb{Z}_{\geqslant 0}$ we denote by $\mathcal{P}_t(q)$ the set of all NE paths from q of length t (no restriction on the endpoint).

Observe that either $q - p \notin \mathbb{Z}^D_{\geqslant 0}$ and $\mathcal{P}(p,q) = \varnothing$, or $q - p \in \mathbb{Z}_{\geqslant 0}$ and

$$|\mathcal{P}(p,q)| = \binom{||q-p||_1}{q_1 - p_1, q_2 - p_2, \cdots, q_D - p_D}$$

Here $||\cdot||_1$ is the 1-norm on $\mathbb{R}^D$ defined by $||p||_1 = \sum\limits_{i=1}^{D} |p_i|$.

On the other hand, $|\mathcal{P}_t(q)| = D^t$ trivially for any $q \in \mathbb{Z}^D$.

Note that unlike other recent work (such as [Mar04]) we do not restrict ourselves to a known solvable model. We do not impose any restrictions on the edge weight distribution θ, so that our results hold with the greatest generality possible.

We will scale the grid by n and look at the behaviour in the limit. As is standard, we will extend our initial inputs p, q to lie in $\mathbb{R}^D$ by taking coordinatewise floors of the scaled coordinates. That is, we will look at paths $\pi \in \mathcal{P}(\lfloor np \rfloor, \lfloor nq \rfloor)$ where

$$\lfloor (x_1, \ldots, x_D) \rfloor := (\lfloor x_1 \rfloor, \ldots, \lfloor x_D \rfloor)$$

Our normalized directed metrics will converge almost surely to a translation-invariant limit so it will suffice to consider the case when $p = \vec{0}$. But for now we let $\vec{p}$ be arbitrary.

Various inequalities we derive will involve the asymptotics of the number of length n NE paths from the origin in a fixed or unfixed direction. The following lemma, which is easily proved using Sterling's approximation, gives us what we want.

Lemma 2.1. *Let $k, n \in \mathbb{N}, a_i \in \mathbb{R}_{\geqslant 0}, \sum a_i = a$. Then*

$$\binom{\lfloor na \rfloor}{\lfloor na_1 \rfloor, \lfloor na_2 \rfloor, \ldots, \lfloor na_k \rfloor} = \left(\frac{a^a}{\prod\limits_{1 \leqslant i \leqslant k} a_i^{a_i}} + o(1) \right)^n$$

where we use the convention $0^0 = 1$.

Remark 2.2. In particular, this gives us that for $q \in \mathbb{R}^D_{\geqslant 0}$,

$$\lim_{n \to \infty} \frac{1}{n} \log(\#\pi : \vec{0} \to \lfloor nq \rfloor) = \lim_{n \to \infty} \frac{1}{n} \log |\mathcal{P}(\vec{0}, \lfloor nq \rfloor)| = \sum_{i=1}^{D} -q_i \log \frac{q_i}{||q||_1} := H(q)$$

We can think of $H(q)$ as the entropy of the number of paths in direction q. Note that $H(q)$ scales with positive scalars and $H(q)$ is maximized among $q \in \mathbb{R}^D_{\geqslant 0}$ with the same 1-norm by

$$q = ||q||_1 \left(\frac{1}{D}, \ldots, \frac{1}{D} \right) := ||q||_1 \ell$$

in which case $H(q) = ||q||_1 \log D$.

On the other hand, for any $t \geqslant 0$,

$$\lim_{n \to \infty} \frac{1}{n} \log(\#\pi \text{ from } \vec{0} \text{ of length } \lfloor nt \rfloor) = \lim_{n \to \infty} \frac{1}{n} \log |\mathcal{P}_{\lfloor nt \rfloor}(\vec{0})| = t \log D = H(t\ell)$$

We will look at the distribution of weights that we observe along paths $\#\pi : \lfloor np \rfloor \to \lfloor nq \rfloor$. For a path π, let the unnormalized empirical measure along π be

$$\sigma_\pi = \sum_{e \in \pi} \delta_{\tau_e}$$

Note that we will normalize by $\frac{1}{n}$ rather than $\frac{1}{|\pi|}$. This is simply for convenience in our proofs, as it will give us a certain superadditivity we do not get when we normalize by $\frac{1}{|\pi|}$.

The Glivenko-Cantelli Theorem [Dur19, Thm. 2.4.7] tells us that for any fixed infinite NE path in the grid, empirical measures along the path converge weakly to θ.

Theorem 2.3 (Glivenko-Cantelli Theorem)**.** *Let F_θ be the cumulative distribution function of θ, let $X_i \sim \theta$ be i.i.d. random variables and let*

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \leqslant x\}}$$

be the cumulative distribution functions of the empirical measures. Then

$$\sup_x |F_n(x) - F_\theta(x)| \to 0 \text{ a.s. as } n \to \infty$$

However, we are interested in the limiting behavior of the empirical measure of not one path, but of all paths from $\vec{0}$ or all paths with a given direction, as we scale the length of the paths. This will allows us to observe more than just the original measure θ.

2.2.2 *Metrics on Measures*

To gauge the distance between two measures we will be using the Levy-Prokhorov metric. We briefly introduce this metric as well as the total variation metric and we outline the relevant properties.

Consider a metric space (X, d) with Borel σ-algebra $\mathcal{B}$. We will denote by $\mathcal{M}$ the set of finite Borel measures on $(X, \mathcal{B})$, by $\mathcal{M}_+$ the set of non-negative finite Borel measures, and by $\mathcal{M}_t$ the set of Borel non-negative finite measures with total mass t for any $t \geqslant 0$. In this notation, $\mathcal{M}_1$ will be the set of Borel probability measures.

Definition 2.4. The total variation norm on $\mathcal{M}$ is defined by

$$||\mu||_{TV} = \sup_{A \in \mathcal{B}} |\mu(A)|$$

This of course gives rise to a total variation metric, given by

$$d_{TV}(\mu,\nu) = ||\mu - \nu||_{TV}$$

For example, the total variation of any measure $\mu \in \mathcal{M}_+$ will be its total mass $\mu(X)$.

Definition 2.5. For any $A \in \mathcal{B}$ and $\epsilon > 0$, the ϵ-neighborhood of A is defined to be

$$A^\epsilon := \{x \in X : d(x,a) < \epsilon \text{ for some } a \in A\}$$

Definition 2.6. The Levy-Prokhorov metric on $\mathcal{M}_+$ is defined by

$$\rho(\mu,\nu) = \inf\{\epsilon > 0 : \mu(A) \leqslant \nu(A^\epsilon) + \epsilon \text{ and } \nu(A) \leqslant \mu(A^\epsilon) + \epsilon \ \forall A \in \mathcal{B}\}$$

It is a standard result that ρ metrizes the weak convergence of measures and total variation metrizes the strong convergence of measures in $\mathcal{M}_+$. For details, see [Hub04, Sect. 2.3].

We now derive two useful inequalities involving the Levy-Prokhorov metric.

Lemma 2.7. *For $\mu, \nu \in \mathcal{M}_+$,*

$$\rho(\mu,\nu) \leqslant ||\mu - \nu||_{TV}$$

Remark 2.8. This lemma establishes that the Levy-Prokhorov metric is weaker than the total variation metric.

Remark 2.9. It is trivial to see that $\rho(\mu, 0) = ||\mu||_{TV}$.

Proof. Let $\epsilon := ||\mu - \nu||_{TV}$. If $\epsilon = 0$ then $\mu = \nu$ so $\rho(\mu,\nu) = 0$. If $\epsilon > 0$, for any $A \in \mathcal{B}$,

$$A \subseteq A^\epsilon \Rightarrow \mu(A) - \nu(A^\epsilon) \leqslant \mu(A) - \nu(A) \leqslant \sup_{A' \in \mathcal{B}} |\mu(A') - \nu(A')| = \epsilon$$

and similarly $\nu(A) - \mu(A^\epsilon) \leqslant \epsilon$ so $\rho(\mu,\nu) \leqslant \epsilon$. □

We next show that ρ satisfies a kind of subadditivity.

Lemma 2.10. *Let $\mu_1, \mu_2, \nu_1, \nu_2 \in \mathcal{M}_+$. Then*

$$\rho(\mu_1 + \mu_2, \nu_1 + \nu_2) \leqslant \rho(\mu_1, \nu_1) + \rho(\mu_2, \nu_2)$$

Remark 2.11. Note that in the case $\mu_2 = \nu_2$ the inequality becomes

$$\rho(\mu_1 + \mu_2, \nu_1 + \mu_2) \leqslant \rho(\mu_1, \nu_1)$$

Proof. For any $\epsilon_1 > \rho(\mu_1, \nu_1), \epsilon_2 > \rho(\mu_2, \nu_2)$ we have for any $A \in \mathcal{B}$,

$$\mu_1(A) \leqslant \nu_1(A^{\epsilon_1}) + \epsilon_1 \leqslant \nu_1(A^{\epsilon_1+\epsilon_2}) + \epsilon_1 \text{and}$$

$$\mu_2(A) \leqslant \nu_2(A^{\epsilon_2}) + \epsilon_2 \leqslant \nu_2(A^{\epsilon_1+\epsilon_2}) + \epsilon_2$$

$$\Rightarrow (\mu_1 + \mu_2)(A) \leqslant (\nu_1 + \nu_2)(A^{\epsilon_1+\epsilon_2}) + \epsilon_1 + \epsilon_2$$

By symmetry, the same inequality holds with μ_i, ν_i swapped. Thus

$$\rho(\mu_1 + \mu_2, \nu_1 + \nu_2) \leqslant \rho(\mu_1, \nu_1) + \rho(\mu_2, \nu_2)$$

□

2.2.3 *A Convenient Coupling of the Edge Weights*

We follow [Bat20, Sect. 2.1] in coupling the environment to uniform random variables in order to work in a compact space of measures and to connect our results with his.

The idea is to write our i.i.d. edge weights $\tau_e \sim \theta$ as

$$\tau_e = \tau(U_e)$$

for some measurable function $\tau : [0,1] \to [0,\infty)$ and i.i.d. Unif[0,1]-valued random variables $(U_e)_{e \in E(\mathbb{Z}^D)}$ on the same probability space as $(\tau_e)_{e \in E(\mathbb{Z}^D)}$. For instance, we could take the quantile function

$$\tau(x) = F_\theta^-(x) := \inf\{t \in \mathbb{R} : F_\theta(t) \geqslant x\}$$

But our results (in particular, our definition of grid entropy) will be independent of the τ chosen so we allow τ to be arbitrary (with the conditions stated above). This will come into play later in Section 2.5, when we will study the Gibbs Free Energy as a function of τ.

To distinguish between the τ_e and the U_e, we will call the former edge *weights* and the latter edge *labels*.

Let Λ denote Lebesgue measure on $[0,1]$. We tweak the definition of empirical measures in this new setup: for any NE path $\pi : \lfloor np \rfloor \to \lfloor nq \rfloor$ in $\mathbb{Z}^D$, define

$$\mu_\pi := \sum_{e \in \pi} \delta_{U_e}$$

Then we can relate Λ and the μ_π to θ and the σ_π respectively via the pushforward:

$$\theta = \tau_*(\Lambda), \sigma_\pi = \tau_*(\mu_\pi) \text{ where } \tau_*(\xi)(B) = \xi(\tau^{-1}(B))$$

$\forall B \in \mathcal{B}(\mathbb{R}^D)$ and $\forall$ measures ξ on $[0,1]$.

One advantage is of course that the set of probability measures on $[0,1]$ is weakly compact, so for any sequence of paths $\pi : \lfloor np \rfloor \to \lfloor nq \rfloor$ in the grid we will get a subsequence for which $\frac{1}{n}\mu_\pi$ converges weakly to some measure.

In the case of a continuous cumulative distribution function F_θ, we can use a lemma proved in [Bat20] to get a nice duality.

Lemma 2.12. *Given a measure θ on $\mathbb{R}$ with continuous cdf F_θ, if we let $\tau = F_\theta^- : [0,1] \to [0,\infty)$ be its quantile function then there is a probability 1 event on which $\frac{1}{n_k}\mu_{\pi_{n_k}} \Rightarrow \nu$ for some subsequence n_k and paths $\pi_{n_k} : \vec{0} \to \lfloor n_k q \rfloor$ if and only if $\tau_*(\frac{1}{n_k}\mu_{\pi_{n_k}}) \Rightarrow \tau_*(\nu)$.*

Proof. [Bat20, Lemma 6.15] establishes that there is a probability 1 event on which $\frac{1}{n_k}\mu_{\pi_{n_k}} \Rightarrow \nu$ implies $\tau_*(\frac{1}{n_k}\mu_{\pi_{n_k}}) \Rightarrow \tau_*(\nu)$. But then on the same event, given a subsequence for which $\tau_*(\frac{1}{n_k}\mu_{\pi_{n_k}}) \Rightarrow \tau_*(\nu)$, compactness gives us a convergent subsubsequence $\frac{1}{n_{k_j}}\mu_{\pi_{n_{k_j}}} \Rightarrow \xi$ hence

$$\tau_*\left(\frac{1}{n_{k_j}}\mu_{\pi_{n_{k_j}}}\right) \Rightarrow \tau_*(\xi)$$

so $\tau_*(\xi) = \tau_*(\nu)$. The fact that F_θ is continuous and the quantile function τ satisfies

$$\tau^{-1}((-\infty, x]) = [0, F_\theta(x)] \ \forall x \in \mathbb{R}$$

implies ξ and ν agree on all sets $[0, F_\theta]$ hence $\xi = \nu$. □

Thus, in the case when θ has continuous cdf, we lose no generality by doing this coupling and working with measures on [0,1].

However, even in the most general case where θ may not have continuous cdf or bounded support, our work in developing grid entropy still holds because we only use the compactness of the space of measures in later sections devoted to our variational formula for the Gibbs Free Energy.

In short, we lose nothing by restricting ourselves to the compact space of measures on $[0,1]$.

A benefit of this coupling is the following amazing result of Bates [Bat20, Lemma 6.3 and Thm 6.4]. Here we denote by $\mathcal{M}_+, \mathcal{M}_t$ the sets of finite non-negative Borel measures on [0,1] and finite non-negative Borel measures on [0,1] with total mass $t \geqslant 0$.

Theorem 2.13. *(i) Fix $q \in \mathbb{R}^D_{\geqslant 0}$. Define $\mathcal{R}^q_\infty$ to be the (event-dependent) set of measures $\nu \in \mathcal{M}_+$ for which there is a subsequence π_{n_k} of paths $\vec{0} \to \lfloor n_k q \rfloor$ with*

$\mu_{\pi_{n_k}} \Rightarrow \nu$. Then there exists a deterministic, weakly closed set $\mathcal{R}^q \subseteq \mathcal{M}_{||q||_1}$ independent of τ s.t.

$$P(\mathcal{R}^q_\infty = \mathcal{R}^q) = 1$$

(ii) Fix $t \geqslant 0$. Define $\mathcal{R}^t_\infty$ to be the (event-dependent) set of measures $\nu \in \mathcal{M}_+$ for which there is a subsequence π_{n_k} of length $\lfloor nt \rfloor$ paths from $\vec{0}$ with $\mu_{\pi_{n_k}} \Rightarrow \nu$. Then there exists a deterministic, weakly closed set $\mathcal{R}^t \subseteq \mathcal{M}_t$ independent of τ s.t.

$$P(\mathcal{R}^t_\infty = \mathcal{R}^t) = 1$$

Moreover, $\mathcal{R}^t = \bigcup_{q \in \mathbb{R}^D_{\geqslant 0}, ||q||_1 = t} \mathcal{R}^q.$

Remark 2.14. Bates proves this theorem in the setup of First Passage Percolation, but notes that it holds analogously in the Last Passage model.

Instead of looking at *all* empirical measures that have a weakly convergent subsequence we may look at only certain empirical measures with this property and the same result will apply. The proof is almost identical to Bates's original proof except for this change, so we omit it.

Corollary 2.15. *(i) Fix $q \in \mathbb{R}^D$ and $0 \leqslant \alpha \leqslant R(q)$. Define $\mathcal{R}^{q,\alpha}_\infty$ to be the (event-dependent) set of measures $\nu \in \mathcal{M}_+$ for which there is a subsequence π_{n_k} of the paths $\pi_n : \vec{0} \to \lfloor nq \rfloor$ with the $\lfloor e^{n\alpha} \rfloor$th smallest values of $\rho(\frac{1}{n}\mu_{\pi_n}, \nu)$ satisfying*

$$\frac{1}{n}\mu_{\pi_{n_k}} \Rightarrow \nu \text{ i.e. } \liminf_{n\to\infty} \min_{\pi:\vec{0}\to\lfloor nq \rfloor}^{\lfloor e^{n\alpha} \rfloor} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0$$

Then there exists a deterministic weakly closed set $\mathcal{R}^{q,\alpha} \subseteq \mathcal{M}_{||q||_1}$ s.t.

$$P(\mathcal{R}^{q,\alpha}_\infty = \mathcal{R}^{q,\alpha}) = 1$$

(ii) Fix $t \geqslant 0$ and $0 \leqslant \alpha \leqslant t \log D$. Define $\mathcal{R}^{t,\alpha}_\infty$ to be the (event-dependent) set of measures $\nu \in \mathcal{M}_+$ for which there is a subsequence π_{n_k} of the length $\lfloor tn \rfloor$ paths π_n from $\vec{0}$ with the $\lfloor e^{n\alpha} \rfloor$th smallest values of $\rho(\frac{1}{n}\mu_{\pi_n}, \nu)$ satisfying

$$\frac{1}{n}\mu_{\pi_{n_k}} \Rightarrow \nu \text{ i.e. } \liminf_{n\to\infty} \min_{\pi:|\pi|=\lfloor nt \rfloor}^{\lfloor e^{n\alpha} \rfloor} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0$$

Then there exists a deterministic weakly closed set $\mathcal{R}^{t,\alpha} \subseteq \mathcal{M}_t$ s.t.

$$P(\mathcal{R}^{t,\alpha}_\infty = \mathcal{R}^{t,\alpha}) = 1$$

Remark 2.16. Since the $\min_{\pi:\vec{0}\to\lfloor nq \rfloor}^{\lfloor e^{n\alpha} \rfloor}$ are increasing in α then the sets $\mathcal{R}^{q,\alpha}$ are decreasing in α. The same holds for $\mathcal{R}^{t,\alpha}$.

Once we develop the concept of grid entropy, we will easily relate these sets $\mathcal{R}^q, \mathcal{R}^{q,\alpha}, \mathcal{R}^t, \mathcal{R}^{t,\alpha}$ to the sets of measures with finite grid entropy in direction q, grid entropy at least α in direction q, finite direction-free length t grid entropy, direction-free length t grid entropy at least α respectively.

2.2.4 *Directed Metric Spaces*

Grid entropy will turn out to be a directed metric with negative sign. We recall what that entails.

Definition 2.17. A directed metric space with positive sign is a triple $(M, d, +)$ where M is a vector space, $d : M^2 \to (-\infty, +\infty]$ is a distance function satisfying $d(x, x) = 0$ and the usual triangle inequality $d(x, y) + d(y, z) \geqslant d(x, z)$. A directed metric space with negative sign is a triple $(M, d, -)$ such that $(M, -d, +)$ is a directed metric space with positive sign.

Standard metric spaces are clearly examples of directed metric spaces with positive sign. By weakening the conditions of a standard metric however, we allow a greater variety of distances to be defined.

Certain directed metrics give rise to directed norms in the same way certain metrics give rise to norms.

Definition 2.18. If (M, d, σ) is a directed metric with positive/negative sign such that it is translation-invariant and homogeneous with respect to positive factors, then it gives rise to a directed norm with positive/negative sign given by

$$||x|| := d(\vec{0}, x)$$

Of particular interest to us are directed norms defined in terms of the empirical measures we observe along paths between points. In the fixed direction case, our directed norms will be defined on the space of tuples consisting of a point in $\mathbb{R}^D$ (the "direction" we are observing) and a finite Borel measure on $\mathbb{R}$ (the target measure we want the empirical measures to be near). In the direction-free case, our directed norms will just be defined on the space $\mathcal{M}_+$ of finite Borel measure on $\mathbb{R}$.

2.2.5 *The Subadditive Ergodic Theorem*

The key theorem we will use to prove the existence of the scaling limit of these directed metrics is Liggett's improved version of Kingman's Subadditive Ergodic Theorem. Before stating this theorem, we recall the definitions of stationary sequences and ergodicity, as presented in [Dur19, Sect. 7].

Definition 2.19. A sequence $(Y_n)_{n\geqslant 1}$ of random variables is called stationary if the joint distributions of the shifted sequences $\{Y_{k+n} : n \geqslant 1\}$ is not dependent on $k \geqslant 0$.

As it turns out, the sequence of random variables we will be looking at will be a sequence of i.i.d. $(Y_n)_{n\geqslant 1}$, which clearly is stationary.

Definition 2.20. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $T : \Omega \to \Omega$ be a map. T is said to be measure-preserving if $P(T^{-1}(A)) = P(A)\ \forall A \in \mathcal{F}$. T is said to be ergodic if it is measure-preserving and if all T-invariant measurable sets are trivial, i.e. $P(A) \in \{0, 1\}$ whenever $A \in \mathcal{F}$ and $T^{-1}(A) = A$.

In the context of sequences, we look at the space $\Omega = \mathbb{R}^\infty$ of infinite sequences of real numbers with the σ-algebra $\mathcal{B}_\infty$ generated by

$$\{\{(y_1, y_2, \ldots) \in \Omega : y_n \in B\} : n \geqslant 1, B \in \mathcal{B}(\mathbb{R})\}$$

(where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$) and with the product probability measure μ_∞ on $\mathcal{B}_\infty$ determined by

$$\mu_\infty(B_1 \times B_2 \times \ldots \times B_n \times \mathbb{R} \times \cdots) = \prod_{i=1}^{n} \mu(B_i)$$

where $B_i \in \mathcal{B}(\mathbb{R})$ and μ is a Borel probability measure on $\mathbb{R}$. We consider the shift operator $T : \Omega \to \Omega$ given by $T(y_1, y_2, y_3, \ldots) = T(y_2, y_3, \ldots)$. T is easily seen to be measure-preserving with respect to μ_∞ since $\mu_\infty(T^{-1}(A)) = \mu_\infty(A)$ for the generating sets A of $\mathcal{B}_\infty$. When we refer to the ergodicity of a sequence of random variables, we will mean the ergodicity of this shift operator.

In our case, where we have a sequence of i.i.d. $(Y_n)_{n\geqslant 1}$, the corresponding shift operator will be ergodic. Indeed, if $T^{-1}(A) = A$ then

$$(Y_1, Y_2, \ldots) \in A \Leftrightarrow T^{n-1}(Y_1, Y_2, \ldots) = (Y_n, Y_{n+1}, \ldots) \in A\ \forall n \geqslant 1$$

so A is in the tail σ-field $\bigcap_{n\geqslant 1} \sigma(Y_n, Y_{n+1}, \ldots)$ and thus $\mu_\infty(A) \in \{0, 1\}$ by Kolmogorov's 0-1 Law.

We are now ready to state the Subadditive Ergodic Theorem in the form we will use.

Theorem 2.21 (Kingman's Subadditive Ergodic Theorem, [Lig85])**.** *Suppose $(Y_{m,n})_{0\leqslant m<n}$ are random variables satisfying*
(i) $\exists$ constant C s.t. $E|Y_{0,n}| < \infty$ and $EY_{0,n} \geqslant Cn$ for all n
(ii) $\forall k \geqslant 1$, $\{Y_{nk,(n+1)k} : n \geqslant 1\}$ is a stationary process
(iii) The joint distributions of $\{Y_{m,m+k} : k \geqslant 1\}$ are not dependent on m

(iv) $Y_{0,m+n} \leqslant Y_{0,m} + Y_{m,m+n}\ \forall m, n > 0$
Then
(a) $\lim\limits_{n\to\infty} \frac{EY_{0,n}}{n} = \inf\limits_{m\geqslant 0} \frac{EY_{0,m}}{m} := \gamma$
(b) $Y := \lim\limits_{n\to\infty} \frac{Y_{0,n}}{n}$ *exists a.s. and in* L^1*, and* $EY = \gamma$
(c) If the stationary sequences in (ii) are ergodic, then $Y = \gamma$ *a.s.*

Remark 2.22. We may replace $Y_{m,n}$ with $-Y_{m,n}$ in the statement of the theorem to obtain a version for superadditive sequences.

This theorem will prove to be the basis for the construction of our grid entropy.

2.2.6 *Relative Entropy and Sanov's Theorem*

In the next preliminary section, we wish to outline the basics of the Kullback-Leibler divergence (introduced in [KL51]) and Sanov's Theorem for large deviations. We will later use this theorem to establish a relationship between our grid entropy and this notion of relative entropy.

Definition 2.23. Let P, Q be distributions on our inherent metric space X. The Kullback-Leibler divergence or relative entropy of Q from P is defined to be

$$D_{KL}(P||Q) = \begin{cases} \int_X \log f\, dP = \int_X f \log f\, dQ, & P << Q \\ +\infty, & \text{otherwise} \end{cases}$$

where $f := \frac{dP}{dQ}$ is the Radon-Nikodym derivative and log is the natural logarithm.

Remark 2.24. [KL51] also derive several basic properties such as D_{KL} being a pre-metric. [Pos75] shows that D_{KL} is lower semicontinuous, in the sense that, given probability distributions $P_n \Rightarrow P$ and $Q_n \Rightarrow Q$, we have

$$D_{KL}(P||Q) \leqslant \liminf_{n\to\infty} D_{KL}(P_n||Q_n)$$

Our main interest in relative entropy is that it is the rate function for large deviations of empirical measures. This is captured by Sanov's Theorem.

Theorem 2.25 (Sanov's Theorem, [DS01])**.** *Consider a sequence of i.i.d. random variables* $X_i \sim \theta$ *taking values in a set* X*. Let* $\mu_n = \sum_{i=1}^n \delta_{X_i}$ *be their empirical measures. Then for any weakly closed set* $F \subset \mathcal{M}_1$ *we have*

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}\mu_n \in F\right) \leqslant -\inf_{\xi\in F} D_{KL}(\xi||\theta)$$

and for any weakly open set $G \subset \mathcal{M}_1$ we have

$$\liminf_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}\mu_n \in U\right) \geqslant -\inf_{\xi\in G} D_{KL}(\xi||\theta)$$

Remark 2.26. Since F is closed then the infimum is achieved by some $\xi \in F$. Furthermore, if $\theta \in F$ then the right-hand side of the inequality is 0, which gives us no information; however, if $\theta \notin F$, then the theorem gives an exponential bound on large deviations.

2.3 GRID ENTROPY AS A DIRECTED NORM

2.3.1 *The Plan for Deriving Direction-Fixed Grid Entropy*

For the purposes of Section 2.3, we temporarily forget our original definition of grid entropy and re-derive it as a limit of scaled directed metrics.

To summarize the setting we described in Section 2.2, we consider empirical measures $\frac{1}{n}\mu_\pi$ along NE-paths on the lattice $\mathbb{Z}^D$, where the edges have weights $\tau_e = \tau(U_e)$ for some measurable $\tau : [0,1] \to [0,\infty)$ and U_e are i.i.d. Unif[0,1] random variables. We denote by $\mathcal{M}, \mathcal{M}_+, \mathcal{M}_t$ the spaces of finite, finite non-negative, and finite non-negative with total mass $t \geqslant 0$ Borel measures respectively.

We begin with direction-fixed grid entropy. We wish to count the number of paths with empirical measure very close to the target ν. Let us try to define a distance on $\mathbb{R}^D \times \mathcal{M}_+$ by

$$d((p,\xi),(q,\nu)) = \log \#\{\text{paths } \pi : \lfloor p \rfloor \to \lfloor q \rfloor \text{ with } \mu_\pi = \nu - \xi\}$$

Note that this is $-\infty$ if there are no such paths and it is 0 if $p = q, \xi = \nu$ (since there is exactly one path $\pi : \lfloor p \rfloor \to \lfloor p \rfloor$, and it has empirical measure 0).

One glaring issue is that we are only counting paths that have exactly the target empirical measure. Since the Lebesgue measure on [0,1] is continuous, then almost surely the U_e will have different values, and thus the unnormalized empirical measures uniquely determine the paths π. It follows that almost surely $d((p,\xi),(q,\nu))$ will always be either $-\infty$ or 0. We need to change our definition of d to count paths with a empirical measures "close" to $\nu - \xi$ instead.

Another problem is that we wish to apply the Subadditive Ergodic Theorem to learn about the behavior of

$$\frac{d((np, n\xi),(nq, n\nu))}{n}$$

as $n \to \infty$. Thus we must also change our definition of d so that it is integrable (and in particular finite a.s.) when $q - p \in \mathbb{R}^D_{\geqslant 0}$.

The trick is to replace the counting of the paths exhibiting the exact target empirical measures with a "cost function" that attributes an exponential cost to each path based on how far its empirical measure is from the target. We will also have a parameter ϵ that, as it decreases to 0, takes the cost function towards the counting of the paths with the target empirical measure we tried initially.

Definition 2.27. Fix $\epsilon > 0$. Define a distance on $\mathbb{R}^D \times \mathcal{M}_+$ by

$$d^\epsilon((p,\xi),(q,\nu)) = \log \sum_{\pi \in \mathcal{P}(\lfloor p \rfloor, \lfloor q \rfloor)} e^{-\frac{1}{\epsilon}\rho(\mu_\pi, \nu - \xi)}$$

where ρ is the Levy-Prokhorov metric and where we sum over all NE paths $\pi : \lfloor p \rfloor \to \lfloor q \rfloor$.

Remark 2.28. This distance is 0 if $p = q, \xi = \nu$ and $-\infty$ if and only if $q - p \notin \mathbb{R}^D_{\geqslant 0}$.

Remark 2.29. For any path $\pi : \lfloor p \rfloor \to \lfloor q \rfloor$, the corresponding empirical measure μ_π observed must necessarily be of the form $\mu_\pi = \sum_{i=1}^{|\pi|} \delta_{a_i}$ where $|\pi| = ||\lfloor q \rfloor - \lfloor p \rfloor||_1$.

Remark 2.30. As $\epsilon \to 0$ the costs $e^{-\frac{1}{\epsilon}\rho(\mu_\pi, \nu - \xi)}$ converge to the indicators

$$\mathbf{1}_{\nu - \xi}(\mu_\pi) = \mathbf{1}_{\frac{\nu - \xi}{||\lfloor q \rfloor - \lfloor p \rfloor||_1}}\left(\frac{1}{|\pi|}\mu_\pi\right)$$

Hence the sum of the costs approaches the number of paths with empirical measures $\frac{1}{|\pi|}\mu_\pi$ precisely equal to $\frac{\nu - \xi}{||\lfloor q \rfloor - \lfloor p \rfloor||_1}$.

With this definition in mind, let us discuss the plan of attack.

First, we will prove the existence of $\lim_{n \to \infty} \frac{d^\epsilon((np, n\xi),(nq, n\nu))}{n}$ using the Subadditive Ergodic Theorem and some estimates we derive for the error terms when p, q do not have integer coordinates. Then we will take the infimum over $\epsilon > 0$ of these limits and we will define the resulting norm to be our grid entropy.

Theorem 2.31. *Fix $\epsilon > 0, \nu, \xi \in \mathcal{M}_+$ and $p, q \in \mathbb{R}^D$. Then*

$$\frac{d^\epsilon((np, n\xi),(nq, n\nu))}{n}$$

converges in probability to a constant. When $p = \vec{0}$, the convergence is pointwise a.s.

Remark 2.32. The theorem holds trivially, with the limit being $-\infty$, if $q - p \notin \mathbb{R}^D_{\geqslant 0}$ or if $q = p$ and $\nu \neq \xi$. It also holds trivially, with the limit being 0, if $q = p$ and $\nu = \xi$.

The limit given by this theorem is a directed metric with negative sign on $\mathbb{R}^D \times \mathcal{M}_+$. When we take an infimum over $\epsilon > 0$ we will still get a directed metric with negative sign.

Theorem 2.33. *For $\epsilon > 0$, $\nu, \xi \in \mathcal{M}_+$ and $p, q \in \mathbb{R}^D$ define*

$$\widetilde{d^\epsilon}((p,\xi),(q,\nu)) := \lim_{n\to\infty} \frac{d^\epsilon((np,n\xi),(nq,n\nu))}{n}$$

$$\text{and } \widetilde{d}((p,\xi),(q,\nu)) := \inf_{\epsilon>0} \widetilde{d^\epsilon}((p,\xi),(q,\nu))$$

Then each $\widetilde{d^\epsilon}$ as well as $\widetilde{d}$ are directed metrics with negative sign on $\mathbb{R}^D \times \mathcal{M}_+$.

We show this metric $\widetilde{d}$ gives rise to a norm on $\mathbb{R}^D_{\geqslant 0} \times \mathcal{M}_+$. This will finish our discussion of the direction-fixed grid entropy and we will move on to the direction-free case.

2.3.2 *The Limit Shape of d^ϵ Starting at $(\vec{0}, 0)$*

In this and the following subsection, we focus on the direction-fixed grid entropy.

To prove Theorem 2.31, we will first prove a simplified version, which we will later generalize easily.

Theorem 2.34. *Fix $\epsilon > 0, \nu \in \mathcal{M}_+$ and $q \in \mathbb{R}^D_{\geqslant 0}\backslash\{\vec{0}\}$. Then*

$$\frac{X_n^{\epsilon,q,\nu}}{n} := \frac{d^\epsilon((\vec{0},0),(nq,n\nu))}{n} \to X^{\epsilon,q,\nu} := \sup_n \frac{EX_n^{\epsilon,q,\nu}}{n} = \lim_{n\to\infty} \frac{EX_n^{\epsilon,q,\nu}}{n} \text{ a.s.}$$

Remark 2.35. As noted before, Theorem 2.31 holds trivially when $q \notin \mathbb{R}^D_{\geqslant 0}\backslash\{\vec{0}\}$, so we need not bother with this case.

We prove this theorem in stages, starting with the case when q has integer coordinates. But first, we show a useful bound on our random variables $X_n^{\epsilon,q,\nu}$.

Lemma 2.36. *Let $\epsilon > 0, \nu, \xi \in \mathcal{M}_+$ and $p, q \in \mathbb{Z}^D$ with $q - p \in \mathbb{Z}^D_{\geqslant 0}\backslash\{\vec{0}\}$. Then*

$$d^\epsilon((p,\xi),(q,\nu)) \in \left[-\frac{1}{\epsilon}(||q-p||_1 + ||\nu-\xi||_{TV}), ||q-p||_1 \log D\right]$$

Proof. Recall that any path $\pi : p \to q$ has $||q-p||_1$ edges, so $||\mu_\pi||_{TV} = ||q-p||_1$, and that the total number of paths $\pi : p \to q$ is

$$\binom{||q-p||_1}{q_1 - p_1, \cdots, q_D - p_D} \in [1, D^{||q-p||_1}]$$

Also, for any such π we have by Lemma 3.22

$$\rho(\mu_\pi, \nu - \xi) \in [0, ||\mu_\pi - (\nu - \xi)||_{TV}] \subseteq [0, ||q-p||_1 + ||\nu - \xi||_{TV}]$$

$$\Rightarrow e^{-\frac{1}{\epsilon}\rho(\mu_\pi, \nu-\xi)} \in [e^{-\frac{1}{\epsilon}(||q-p||_1 + ||\nu-\xi||_{TV})}, 1]$$

Thus

$$\begin{aligned} d^\epsilon((p,\xi),(q,\nu)) &= \log \sum_{\pi \in \mathcal{P}(p,q)} e^{-\frac{1}{\epsilon}\rho(\mu_\pi, \nu-\xi)} \\ &\in \left[-\frac{1}{\epsilon}(||q-p||_1 + ||\nu-\xi||_{TV}), ||q-p||_1 \log D \right] \end{aligned}$$

□

Lemma 2.37. *Theorem 2.34 holds with $q \in \mathbb{Z}^D_{\geqslant 0} \backslash \{\vec{0}\}$.*

Proof. We wish to use Kingman's Subadditive Ergodic Theorem (Theorem 2.21) with

$$Y_{m,n} := -d^\epsilon((mq, m\nu), (nq, n\nu)) \ \forall m \leqslant n$$

Let us now check the conditions (i)-(iv).

By Lemma 3.36,

$$Y_{0,n} = -d^\epsilon((\vec{0}, 0), (nq, n\nu)) \in \left[-n||q||_1 \log D, \frac{1}{\epsilon}(n||q||_1 + n||\nu||_{TV}) \right]$$

hence (i) holds.

Next, for every $k \geqslant 1$, the sequence

$$Y_{nk,(n+1)k} = -d^\epsilon((nkq, nk\nu), ((n+1)kq, (n+1)k\nu))$$

is i.i.d. because the distribution of the unnormalized empirical measures of paths $\pi : nkq \to (n+1)kq$ is not dependent on n (since the edge labels U_e are i.i.d.) so the distribution of the cost functions $e^{-\frac{1}{\epsilon}\rho(\mu_\pi, k\nu)}$ for $\pi : nkq \to (n+1)kq$ is not dependent on n. Thus (ii) holds. Furthermore, as discussed in section 2.2.5, $Y_{nk,(n+1)k}$ being i.i.d. implies the sequence is ergodic.

Similarly, the joint distributions of

$$\{Y_{m,m+k} : k \geqslant 1\} = \{-d^{\epsilon}((mq, m\nu), ((m+k)q, (m+k)\nu)) : k \geqslant 1\}$$

are not dependent on m since the edge labels are i.i.d. and the distribution of empirical measures of paths in a rectangle on the lattice with the difference between the top right and bottom left corners being kq is independent of the location of the rectangle. Thus we have (iii).

It remains to show (iv), namely that given $m, n > 0$,

$$\begin{aligned} &d^{\epsilon}((\vec{0}, 0), ((m+n)q, (m+n)\nu)) \\ &\geqslant d^{\epsilon}((\vec{0}, 0), (mq, m\nu)) + d^{\epsilon}((mq, m\nu), ((m+n)q, (m+n)\nu)) \end{aligned}$$

For any paths $\pi : \vec{0} \to mq$ and $\pi' : mq \to (m+n)q$, we get a unique concatenation $\pi \cdot \pi' : \vec{0} \to mq \to (m+n)q$. Its empirical measure satisfies $\mu_{\pi \cdot \pi'} = \mu_{\pi} + \mu_{\pi'}$. But the Levy-Prokhorov metric satisfies subadditivity by Lemma 3.24. Thus

$$\rho(\mu_{\pi \cdot \pi'}, (m+n)\nu) \leqslant \rho(\mu_{\pi}, m\nu) + \rho(\mu_{\pi'}, n\nu)$$

But then

$$\begin{aligned} -Y_{0,m} - Y_{m,m+n} &= \log \sum_{\pi : \vec{0} \to mq} e^{-\frac{1}{\epsilon}\rho(\mu_{\pi}, m\nu)} + \log \sum_{\pi' : mq \to (m+n)q} e^{-\frac{1}{\epsilon}\rho(\mu_{\pi'}, n\nu)} \\ &= \log \sum_{\substack{\pi : \vec{0} \to mq \\ \pi' : mq \to (m+n)q}} e^{-\frac{1}{\epsilon}\rho(\mu_{\pi}, m\nu) - \frac{1}{\epsilon}\rho(\mu_{\pi'}, n\nu)} \\ &\leqslant \log \sum_{\substack{\pi : \vec{0} \to mq \\ \pi' : mp \to (m+n)q}} e^{-\frac{1}{\epsilon}\rho(\mu_{\pi \cdot \pi'}, (m+n)\nu)} \end{aligned}$$

Not all paths $\pi''' : \vec{0} \to (m+n)q$ pass through mq so we can upper bound the expression above by removing this condition:

$$\begin{aligned} -Y_{0,m} - Y_{m,m+n} &\leqslant \log \sum_{\pi''' : \vec{0} \to (m+n)q} e^{-\frac{1}{\epsilon}\rho(\mu_{\pi'''}, (m+n)\nu)} \\ &= -Y_{0,m+n} \end{aligned}$$

Thus we can apply the Subadditive Ergodic Theorem (Theorem 2.21) to get that

$$\frac{-Y_{0,n}}{n} = \frac{X_n^{\epsilon,q,\nu}}{n}$$

converges a.s. to the constant

$$X^{\epsilon,q,\nu} := \sup_n \frac{EX_n^{\epsilon,q,\nu}}{n} = \lim_{n\to\infty} \frac{EX_n^{\epsilon,q,\nu}}{n} \in \left[-\frac{1}{\epsilon}\max(||q||_1, ||\nu||), ||q||_1 \log D \right]$$

□

The next order of business is proving the theorem for q with rational coordinates. We will find useful the following error estimate on how $X_n^{\epsilon,q,\nu}$ changes when q is perturbed in the SE direction and ν is perturbed arbitrarily.

Lemma 2.38. *Fix $q \in \mathbb{R}^D_{\geqslant 0}, \epsilon > 0$ and $\nu, \xi \in \mathcal{M}_+$. Then for any $p \in \mathbb{R}^D_{\geqslant 0}$ with $q - p \in \mathbb{R}^D_{\geqslant 0}$,*

$$X_n^{\epsilon,p,\xi} - \frac{1}{\epsilon}(n||q-p||_1 + n||\nu - \xi||_{TV} + D) \leqslant X_n^{\epsilon,p,\xi} - \frac{1}{\epsilon}(||\lfloor nq\rfloor - \lfloor np\rfloor||_1 + n\rho(\nu,\xi)) \leqslant X_n^{\epsilon,q,\nu}$$

Proof. Fix any such p. The inequality holds trivially if $p = q$ so we may assume $||q-p||_1 > 0$. Then there is at least one path $\pi' : \lfloor np\rfloor \to \lfloor nq\rfloor$, and we fix it.

For any path $\pi : \vec{0} \to \lfloor np\rfloor$, we concatenate it with π' to get a unique path $\pi \cdot \pi' : \vec{0} \to \lfloor nq\rfloor$. Note that π' consists of

$$||\lfloor nq\rfloor - \lfloor np\rfloor||_1 \leqslant n||q-p||_1 + D$$

edges so its empirical measure $\mu_{\pi'}$ has total variation at most $n||q-p||_1 + D$. Thus $\pi \cdot \pi'$ satisfies

$$\begin{aligned}\rho(\mu_{\pi\cdot\pi'}, n\nu) &\leqslant \rho(\mu_\pi + \mu_{\pi'}, \mu_\pi) + \rho(\mu_\pi, n\xi) + \rho(n\xi, n\nu)\\ &\leqslant ||\mu_{\pi'}||_{TV} + \rho(\mu_\pi, n\xi) + n\rho(\nu,\xi)\\ &\leqslant \rho(\mu_\pi, n\xi) + (||\lfloor nq\rfloor - \lfloor np\rfloor||_1 + n\rho(\nu,\xi))\end{aligned}$$

It follows that

$$X_n^{\epsilon,p,\xi} - \frac{1}{\epsilon}(n||q-p||_1 + n||\nu - \xi||_{TV} + D) \leqslant X_n^{\epsilon,p,\xi} - \frac{1}{\epsilon}(||\lfloor nq\rfloor - \lfloor np\rfloor||_1 + n\rho(\nu,\xi)) \leqslant X_n^{\epsilon,q,\nu}$$

□

Using this lemma to approximate $X_n^{\epsilon,q,\nu}$ for $q \in \mathbb{Q}^D_{\geqslant 0}$ in terms of $X_n^{\epsilon,p,\nu}$ where $p \in \mathbb{Z}^D_{\geqslant 0}$, we prove our limit theorem for q with rational coordinates.

Lemma 2.39. *Theorem 2.34 holds for $q \in \mathbb{Q}^D_{\geqslant 0} \backslash \{\vec{0}\}$.*

Proof. Let $q = \frac{(s_1,\ldots,s_D)}{t}$ for $s_i, t \in \mathbb{Z}_{\geqslant 0}$. First, we compare $X_{tr}^{\epsilon,q,\nu}, X_{tr+1}^{\epsilon,q,\nu}, \ldots, X_{tr+(t-1)}^{\epsilon,q,\nu}, X_{t(r+1)}^{\epsilon,q,\nu}$ for arbitrary r using the previous lemma. The idea is that

trq has integer coordinates so $\frac{X^{\epsilon,q,\nu}_{tr}}{tr}, \frac{X^{\epsilon,q,\nu}_{t(r+1)}}{tr}$ have the desired limits by Lemma 2.37. The rest of the $X^{\epsilon,q,\nu}_{tr+j}$ will be bounded above/below by $X^{\epsilon,q,\nu}_{t(r+1)}/X^{\epsilon,q,\nu}_{tr}$ respectively plus some small error terms which go to 0 as $r \to \infty$.

Consider any $1 \leqslant j \leqslant t$. Note that

$$X^{\epsilon,q,\nu}_{tr+j-1} = d^{\epsilon}((\vec{0},0),((tr+j-1)q,(tr+j-1)\nu)) = X^{\epsilon,\frac{tr+j-1}{tr+j}q,\frac{tr+j-1}{tr+j}\nu}_{tr+j}$$

and

$$\left|\left|q - \frac{tr+j-1}{tr+j}q\right|\right|_1 + \left|\left|\nu - \frac{tr+j-1}{tr+j}\nu\right|\right|_{TV} = \frac{||q||_1 + ||\nu||_{TV}}{tr+j}$$

By Lemma 2.38,

$$X^{\epsilon,q,\nu}_{tr+j-1} = X^{\epsilon,\frac{tr+j-1}{tr+j}q,\frac{tr+j-1}{tr+j}\nu}_{tr+j} \leqslant X^{\epsilon,q,\nu}_{tr+j} + \frac{1}{\epsilon}(||q||_1 + ||\nu||_{TV} + D)$$

Thus

$$\frac{X^{\epsilon,q,\nu}_{tr}}{r} \leqslant \frac{X^{\epsilon,q,\nu}_{tr+1}}{r} + \frac{||q||_1 + ||\nu||_{TV} + D}{\epsilon r} \leqslant \ldots \leqslant \frac{X^{\epsilon,q,\nu}_{t(r+1)}}{r} + \frac{t(||q||_1 + ||\nu||_{TV} + D)}{\epsilon r} \tag{2.1}$$

But $tq \in \mathbb{Z}^D_{\geqslant 0}$ so by Lemma 3.36, our limit theorem holds for $X^{\epsilon,q,\nu}_{tr} = X^{\epsilon,tq,t\nu}_{r}$. That is,

$$\frac{X^{\epsilon,q,\nu}_{tr}}{r} \to \sup_r \frac{EX^{\epsilon,q,\nu}_{tr}}{r} = \lim_{r\to\infty} \frac{EX^{\epsilon,q,\nu}_{tr}}{r} \text{ a.s.} \tag{2.2}$$

Taking expectations in (2.1), then taking the supremum/limit as $r \to \infty$ and using (2.2) we get

$$\lim_{r\to\infty} \frac{EX^{\epsilon,q,\nu}_{tr+j}}{r} = \sup_r \frac{EX^{\epsilon,q,\nu}_{tr+j}}{r} = \lim_{r\to\infty} \frac{EX^{\epsilon,q,\nu}_{tr}}{r} = \sup_r \frac{EX^{\epsilon,q,\nu}_{tr}}{r} \ \forall 0 \leqslant j \leqslant t-1 \tag{2.3}$$

since the error terms in (2.1) go to 0. Similarly, taking the limit as $r \to \infty$ in (2.2) and using (2.1), (2.3) we get

$$\lim_{r\to\infty} \frac{X^{\epsilon,q,\nu}_{tr+j}}{r} = \sup_r \frac{EX^{\epsilon,q,\nu}_{tr}}{r} = \lim_{r\to\infty} \frac{EX^{\epsilon,q,\nu}_{tr+j}}{r} = \sup_r \frac{EX^{\epsilon,q,\nu}_{tr+j}}{r} \text{ a.s. } \forall 0 \leqslant j \leqslant t-1$$

Multiplying everything by $\frac{1}{t}$ and using the fact that any n can be written as $tr+j$, we get

$$\frac{X^{\epsilon,q,\nu}_{n}}{n} \to \lim_{n\to\infty} \frac{EX^{\epsilon,q,\nu}_{n}}{n} = \sup_n \frac{EX^{\epsilon,q,\nu}_{n}}{n} \text{ a.s.}$$

□

Our next objective is to prove the full version of Theorem 2.34. We will need two short lemmas giving sufficient conditions for a.s. convergence of bounded random variables to the supremum of their expectations. The first of these lemmas looks at the case where we are given a particular lower bound for the liminf of all subsequences of our sequence.

Lemma 2.40. *Let Y_n be uniformly bounded random variables. If a.s.*

$$\sup_n EY_n \leqslant \liminf_{i\to\infty} Y_{n_i} \textit{ for all subsequences } (n_i)$$

then $\lim_{n\to\infty} Y_n = \sup_n EY_n = \lim_{n\to\infty} EY_n$ *a.s.*

Proof. First, taking expectations in

$$\sup_n EY_n \leqslant \liminf_{i\to\infty} Y_{n_i} \text{ a.s.}$$

and using Fatou's Lemma, we get

$$\sup_n EY_n \leqslant E\liminf_{i\to\infty} Y_{n_i} \leqslant \liminf_{i\to\infty} EY_{n_i} \leqslant \limsup_{i\to\infty} EY_{n_i} \leqslant \sup_i EY_{n_i} \leqslant \sup_n EY_n$$

so all the inequalities are equalities and

$$\sup_n EY_n = \lim_{i\to\infty} EY_{n_i} = \liminf_{i\to\infty} Y_{n_i} \text{ a.s.}$$

This holds for all subsequences (n_i), including the full sequence. Thus a.s.

$$L := \sup_n EY_n = \lim_{n\to\infty} EY_n = \liminf_{i\to\infty} Y_{n_i} \quad \text{for all subsequences } (n_i)$$

hence $Y_n \to L$ a.s.. □

The next lemma is about approximating a sequence of random variables from below by sequences that each converge to the supremum of their expectations.

Lemma 2.41. *Let Y_n be uniformly bounded random variables s.t.*
(i) For every fixed k, $Y_{n,k} \leqslant Y_n$ for large enough n
(ii) For every fixed k,

$$\lim_{n\to\infty} Y_{n,k} = \sup_n EY_{n,k} = \lim_{n\to\infty} EY_{n,k} \textit{ a.s.}$$

(iii) $\sup_n EY_n = \sup_n \sup_k EY_{n,k}$

Then $\lim_{n\to\infty} Y_n = \sup_n EY_n = \lim_{n\to\infty} EY_n$ *a.s.*

Proof. We wish to prove the hypothesis of Lemma 2.40 for Y_n. Let

$$\mathcal{F} = \left\{ \lim_{n\to\infty} Y_{n,k} \neq \sup_n EY_{n,k} \text{ for some } k \right\}$$

which has measure 0 by (ii). We claim that in the event $\mathcal{F}^C$ we have

$$\sup_n EY_n \leqslant \liminf_{i\to\infty} Y_{n_i} \text{ for all subsequences } (n_i)$$

Consider any subsequence (n_i) and any k. By (i), $Y_{n_i,k} \leqslant Y_{n_i}$ for large enough i. Taking the lim inf over i, and using the fact that we are in the event $\mathcal{F}^C$, we get

$$\sup_n EY_{n,k} = \lim_{n\to\infty} Y_{n,k} = \lim_{i\to\infty} Y_{n_i,k} \leqslant \liminf_{i\to\infty} Y_{n_i}$$

This holds for all k. Taking the supremum over k and using (iii), we get

$$\sup_n EY_n = \sup_n \sup_k EY_{n,k} = \sup_k \sup_n EY_{n,k} \leqslant \liminf_{i\to\infty} Y_{n_i}$$

as desired. Applying Lemma 2.40, we get

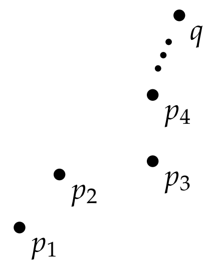
$$\lim_{n\to\infty} EY_n = \sup_n EY_n = \lim_{n\to\infty} Y_n \text{ a.s.}$$

□

We can finally prove Theorem 2.34.

Proof of Theorem 2.34.
The case $q \in \mathbb{Q}^D_{\geqslant 0} \backslash \{\vec{0}\}$ is handled by Lemma 2.39. So we may assume $q \notin \mathbb{Q}^D_{\geqslant 0}$.

We construct a sequence $p_k \in \mathbb{Q}^D_{\geqslant 0}$ as follows. For every $1 \leqslant j \leqslant D$, either $q_j \in \mathbb{Q}$ and we pick $(p_k)_j := q_j$, or $q_j \notin \mathbb{Q}$ and we pick $(p_k)_j \in \mathbb{Q}_{\geqslant 0}$ s.t. $(p_k)_j \uparrow q_j$. It follows that $p_k \in \mathbb{Q}^D_{\geqslant 0}$ with $p_{k+1} - p_k, q - p_k \in \mathbb{Q}^D_{\geqslant 0}$ $\forall k$. That is, $p_1, p_2, \ldots, q$ forms a "staircase" in $\mathbb{R}^D$.



The intuition is that $X_n^{\epsilon,p_k,\nu}$ will approximate $X_n^{\epsilon,q,\nu}$ from below. Each $\frac{X_n^{\epsilon,p_k,\nu}}{n}$ converges in n to the right limit a.s. by Lemma 2.39, and we will use Lemma 2.41 to prove that $\frac{X_n^{\epsilon,q,\nu}}{n}$ converges a.s. to the right limit.

Let us now prove the hypothesis of Lemma 2.41 with

$$Y_{n,k} := \frac{X_n^{\epsilon,p_k,\nu}}{n} - \frac{||\lfloor nq \rfloor - \lfloor np_k \rfloor||_1}{\epsilon}, Y_n := \frac{X_n^{\epsilon,q,\nu}}{n}$$

Note that these random variables are bounded by Lemma 3.36:

$$Y_{n,k}, Y_n \in \left[-\frac{1}{\epsilon}(||q||_1 + ||\nu||_{TV}) - \frac{||q||_1}{\epsilon}, ||q||_1 \log D \right]$$

First, fix n and consider any k. By Lemma 2.38,

$$X_n^{\epsilon,p_k,\nu} - \frac{||\lfloor nq \rfloor - \lfloor np_k \rfloor||_1}{\epsilon} \leqslant X_n^{\epsilon,q,\nu} = Y_n$$

so we have (i). Next, (ii) follows from Lemma 2.39:

$$\lim_{n\to\infty} Y_{n,k} = \lim_{n\to\infty} \frac{X_n^{\epsilon,p_k,\nu} - \frac{||\lfloor nq \rfloor - \lfloor np_k \rfloor||_1}{\epsilon}}{n} = \lim_{n\to\infty} \frac{X_n^{\epsilon,p_k,\nu}}{n} - \frac{||q-p_k||_1}{\epsilon}$$

$$= \sup_n \frac{EX_n^{\epsilon,p_k,\nu}}{n} - \frac{||q-p_k||_1}{\epsilon} = \lim_{n\to\infty} \frac{EX_n^{\epsilon,p_k,\nu}}{n} - \frac{||q-p_k||_1}{\epsilon} = \lim_{n\to\infty} EY_{n,k} = \sup_n EY_{n,k}$$

a.s..

It remains to show (iii). Consider any fixed n. For each $1 \leqslant j \leqslant D$ either $q_j \in \mathbb{Q}$ and $q_j = (p_k)_j \ \forall k$, or $nq_j \notin \mathbb{Z}$ and $(p_k)_j \uparrow q_j$ hence $\lfloor nq_j \rfloor = \lfloor n(p_k)_j \rfloor$ for large enough k. Thus $X_n^{\epsilon,p_k,\nu} = X_n^{\epsilon,q,\nu}$ for large enough k. It follows that

$$Y_{n,k} = X_n^{\epsilon,p_k,\nu} - \frac{||\lfloor nq \rfloor - \lfloor np_k \rfloor||_1}{\epsilon} \uparrow X_n^{\epsilon,q,\nu} = Y_n \text{ as } k \to \infty$$

By the Bounded Convergence Theorem, $EY_{n,k} \uparrow EY_n$. Thus

$$\lim_{k\to\infty} EY_{n,k} = \sup_k EY_{n,k} = EY_n \ \forall n$$

so taking the supremum over n we get (iii):

$$\sup_n \sup_k EY_{n,k} = \sup_n EY_n$$

Applying Lemma 2.41, we get

$$\lim_{n\to\infty} Y_n = \lim_{n\to\infty} EY_n = \sup_n EY_n \text{ a.s., i.e. } \lim_{n\to\infty} \frac{X_n^{\epsilon,q,\nu}}{n} = \lim_{n\to\infty} \frac{EX_n^{\epsilon,q,\nu}}{n} = \sup_n \frac{EX_n^{\epsilon,q,\nu}}{n} \text{ a.s.}$$

□

This completes the proof of the existence of the limit shape of d^ϵ when measuring the distance from $(\vec{0},0) \in \mathbb{R}^D \times \mathcal{M}_+$.

2.3.3 *The Limit Shape of d^ϵ—The General Case*

We use Lemma 2.41 to generalize Theorem 2.34 to Theorem 2.31, where d^ϵ measures distance between any two elements of $\mathbb{R}^D \times \mathcal{M}_+$.

Theorem 2.31. Fix $\epsilon > 0, \nu, \xi \in \mathcal{M}_+$ and $p, q \in \mathbb{R}^D$. Then

$$\frac{d^\epsilon((np, n\xi),(nq, n\nu))}{n}$$

converges in probability to a constant. When $p = \vec{0}$, the convergence is pointwise a.s.

Remark 2.42. During the course of the proof, we will also show that the limit shape of $\frac{d^\epsilon((np,n\xi),(nq,n\nu))}{n}$ is translation-invariant. This will help us later.

Proof. As noted before, the theorem holds trivially if $q - p \notin \mathbb{R}^D_{\geqslant 0} \backslash \{\vec{0}\}$. Thus we may assume $q - p \in \mathbb{R}^D_{\geqslant 0} \backslash \{\vec{0}\}$.

Observe that

$$d^\epsilon((np, n\xi),(nq, n\nu)) = d^\epsilon((np, 0),(nq, n\nu - n\xi)) =^d d^\epsilon((\vec{0}, 0),(\lfloor nq \rfloor - \lfloor np \rfloor, n\nu - n\xi)) \tag{2.4}$$

so it suffices to assume $\xi = 0$ and prove

$$\frac{X_n^{\epsilon, \frac{\lfloor nq \rfloor - \lfloor np \rfloor}{n}, \nu}}{n} = \frac{d^\epsilon((\vec{0}, 0),(\lfloor nq \rfloor - \lfloor np \rfloor, n\nu))}{n}$$

converge a.s. to a constant.

Our argument will mirror the one used in the proof of Theorem 2.34: we will approximate these Y_n from below by some $Y_{n,k}$ converging in k and we will apply Lemma 2.41.

First we prove some inequalities. Fix n. Observe that for $1 \leqslant j \leqslant D$,

$$n(q_j - p_j) - 1 < \lfloor n(q_j - p_j) \rfloor \leqslant n(q_j - p_j)$$

$$n(q_j - p_j) - 1 < \lfloor nq_j \rfloor - \lfloor np_j \rfloor < n(q_j - p_j) + 1$$

$$\Rightarrow (\lfloor nq_j \rfloor - \lfloor np_j \rfloor) - \lfloor n(q_j - p_j) \rfloor \in \{0, 1\} \tag{2.5}$$

and thus

$$(\lfloor nq \rfloor - \lfloor np \rfloor) - \lfloor n(q-p) \rfloor \in \mathbb{R}^D_{\geqslant 0} \text{ with } ||(\lfloor nq \rfloor - \lfloor np \rfloor) - \lfloor n(q-p) \rfloor||_1 \leqslant D \tag{2.6}$$

On the other hand, for $1 \leqslant j \leqslant D$ s.t. $p_j \neq q_j$,

$$\lfloor nq_j \rfloor - \lfloor np_j \rfloor \leqslant n(q_j - p_j) + 1 \leqslant (n+c)(q_j - p_j) \text{ for } c := \left\lceil \max_{1 \leqslant j \leqslant D \text{ s.t. } p_j \neq q_j} \left(\frac{1}{q_j - p_j} \right) \right\rceil$$

$$\Rightarrow 0 \leqslant \lfloor (n+c)(q_j - p_j) \rfloor - (\lfloor nq_j \rfloor - \lfloor np_j \rfloor) \leqslant c(q_j - p_j) \text{ by (2.5)}$$

This equation also holds trivially for j s.t. $p_j = q_j$. Thus

$$\lfloor (n+c)(q-p) \rfloor - (\lfloor nq \rfloor - \lfloor np \rfloor) \in \mathbb{R}^D_{\geqslant 0}$$

with

$$||\lfloor (n+c)(q-p) \rfloor - (\lfloor nq \rfloor - \lfloor np \rfloor)||_1 \leqslant c||q-p||_1 \tag{2.7}$$

We will prove the hypothesis of Lemma 2.41 with

$$Y_{n,k} := \frac{X_n^{\epsilon, q-p, \nu} - \frac{n}{\epsilon k}}{n}, Y_n := \frac{X_n^{\epsilon, \frac{\lfloor nq \rfloor - \lfloor np \rfloor}{n}, \nu}}{n}$$

First, by Lemma 2.38 and (2.6), for every fixed k, for large enough n we have

$$\frac{X_n^{\epsilon, q-p, \nu} - \frac{n}{\epsilon k}}{n} \leqslant \frac{X_n^{\epsilon, q-p, \nu} - \frac{D}{\epsilon}}{n} \leqslant \frac{X_n^{\epsilon, q-p, \nu} - \frac{1}{\epsilon}||(\lfloor nq \rfloor - \lfloor np \rfloor) - \lfloor n(q-p) \rfloor||_1}{n} \leqslant \frac{X_n^{\epsilon, \frac{\lfloor nq \rfloor - \lfloor np \rfloor}{n}, \nu}}{n}$$

i.e. $Y_{n,k} \leqslant Y_n$ which is precisely (i). Also, Theorem 2.34 gives (ii): for every fixed k, a.s.

$$\lim_{n \to \infty} Y_{n,k} = \lim_{n \to \infty} \frac{X_n^{\epsilon, q-p, \nu} - \frac{n}{\epsilon k}}{n} = \sup_n \frac{EX_n^{\epsilon, q-p, \nu}}{n} - \frac{1}{\epsilon k} = \lim_{n \to \infty} \frac{EX_n^{\epsilon, q-p, \nu}}{n} - \frac{1}{\epsilon k}$$

$$= \sup_n EY_{n,k} = \lim_{n \to \infty} EY_{n,k}$$

Note that this implies

$$\sup_n \frac{EX_n^{\epsilon,q-p,\nu}}{n} = \sup_n \sup_k EY_{n,k} = \sup_k \sup_n EY_{n,k} = \sup_k \lim_{n\to\infty} EY_{n,k} = \lim_{n\to\infty} \frac{EX_n^{\epsilon,q-p,\nu}}{n} \tag{2.8}$$

It remains to show (iii). Note that taking expectations in (i) immediately gives

$$\sup_n \sup_k EY_{n,k} = \sup_k \sup_n EY_{n,k} \leqslant \sup EY_n \tag{2.9}$$

We will show this is an equality. By Lemma 2.38 and (2.7), for every fixed k, for large enough n we have

$$\begin{aligned} X_n^{\epsilon,\frac{\lfloor nq\rfloor-\lfloor np\rfloor}{n},\nu} &\leqslant X_n^{\epsilon,\frac{n+c}{n}(q-p),\frac{n+c}{n}\nu} + \frac{1}{\epsilon}\left(||\lfloor (n+c)(q-p)\rfloor - (\lfloor nq\rfloor - \lfloor np\rfloor)||_1 + c||\nu||_{TV}\right) \\ &\leqslant X_{n+c}^{\epsilon,(q-p),\nu} + \frac{c||q-p||_1 + c||\nu||_{TV}}{\epsilon} \end{aligned}$$

Dividing by n, taking expectations and taking the supremum over n we get by (2.8)

$$\sup_n EY_n \leqslant \sup_n \frac{EX_{n+c}^{\epsilon,q-p,\nu}}{n} = \lim_{n\to\infty} \frac{EX_n^{\epsilon,q-p,\nu}}{n} = \sup_n \sup_k EY_{n,k} \tag{2.10}$$

which when combined with (2.9) gives (iii).

Applying Lemma 2.41, we get that a.s.,

$$\lim_{n\to\infty} Y_n = \lim_{n\to\infty} EY_n = \sup_n EY_n$$

i.e.

$$\lim_{n\to\infty} \frac{X_n^{\epsilon,\frac{\lfloor nq\rfloor-\lfloor np\rfloor}{n}}}{n} = \sup_n \frac{EX_n^{\epsilon,\frac{\lfloor nq\rfloor-\lfloor np\rfloor}{n}}}{n} = \lim_{n\to\infty} \frac{EX_n^{\epsilon,\frac{\lfloor nq\rfloor-\lfloor np\rfloor}{n}}}{n}$$

Furthermore, by (2.8), (2.9) and (3.14), we have

$$\sup_n EY_n = \sup_n \sup_k EY_{n,k} = \sup_n \frac{EX_n^{\epsilon,q-p,\nu}}{n} = \lim_{n\to\infty} \frac{EX_n^{\epsilon,q-p,\nu}}{n}$$

Combining this with (2.4), we get that $\frac{d^\epsilon((np,n\xi),(nq,n\nu))}{n}$ converges in probability to the a.s. limit of $\frac{d^\epsilon((\vec{0},0),(nq-np,n\nu-n\xi))}{n}$.

This completes the proof of Theorem 2.31 and of the fact that the limit shape is translation invariant.

□

2.3.4 *Grid Entropy as a Directed Norm*

In the previous sections we showed the existence of the limit shape of d^ϵ. We will now take the infimum as $\epsilon \downarrow 0$ and we will show the result is a directed metric with negative sign that gives rise to a norm, which we call grid entropy.

Theorem 2.34. For $\epsilon > 0$, $\nu, \xi \in \mathcal{M}_+$ and $p, q \in \mathbb{R}^D$ define

$$\widetilde{d^\epsilon}((p,\xi),(q,\nu)) := \lim_{n\to\infty} \frac{d^\epsilon((np,n\xi),(nq,n\nu))}{n},$$

$$\widetilde{d}((p,\xi),(q,\nu)) := \inf_{\epsilon>0} \widetilde{d^\epsilon}((p,\xi),(q,\nu)) \in [-\infty,\infty)$$

Then each $\widetilde{d^\epsilon}$ as well as $\widetilde{d}$ are directed metrics with negative sign on $\mathbb{R}^D \times \mathcal{M}_+$.

Remark 2.43. For any $p, q \in \mathbb{R}^D, \epsilon > 0$ and $\nu, \xi \in \mathcal{M}_+$, $d^\epsilon((np,n\xi),(nq,n\nu))$ is monotone decreasing as $\epsilon \downarrow 0$ so

$$\widetilde{d}((p,\xi),(q,\nu)) = \inf_{\epsilon>0} \widetilde{d^\epsilon}((p,\xi),(q,\nu)) = \lim_{\epsilon\downarrow 0} \widetilde{d^\epsilon}((p,\xi),(q,\nu))$$

By Lemma 3.36, for every n and $\epsilon > 0$,

$$d^\epsilon((np,n\xi),(nq,n\nu)) \in [-\infty, ||\lfloor nq \rfloor - \lfloor np \rfloor||_1 \log D]$$

so it follows that

$$\widetilde{d}((p,\xi),(q,\nu)) \in [-\infty, ||q-p||_1 \log D]$$

Once we prove that our two definitions of grid entropy are equivalent, this bound will be improved.

Remark 2.44. As was the case with Theorem 2.34, the limit $\widetilde{d}((p,\xi),(q,\nu))$ is trivially $-\infty$ if $q-p \notin \mathbb{R}^D_{\geqslant 0}$ or if $q = p$ and $\nu \neq \xi$, and it is trivially 0 if $q = p$ and $\nu = \xi$.

Proof. As noted above,

$$\widetilde{d^\epsilon}((p,\xi),(p,\xi)) = \widetilde{d}((p,\xi),(p,\xi)) = 0 \;\forall \epsilon > 0$$

It remains to prove the reverse triangle inequality. Let $p, q, r \in \mathbb{R}^D$ and $\nu, \xi, \eta \in \mathcal{M}_+$ and consider any $\epsilon > 0$ and any n. If $q - p \notin \mathbb{R}^D_{\geqslant 0}$ or

$r - q \notin \mathbb{R}^D_{\geqslant 0}$ then the following inequality holds trivially (because the right-hand side is $-\infty$)

$$d^\epsilon((np, n\xi), (nr, n\eta)) \geqslant d^\epsilon((np, n\xi), (nq, n\nu)) + d^\epsilon((nq, n\nu), (nr, n\eta)) \tag{2.11}$$

Now suppose $r - q, q - p \in \mathbb{R}^D_{\geqslant 0}$. Given paths $\pi : \lfloor np \rfloor \to \lfloor nq \rfloor$, $\pi' : \lfloor nq \rfloor \to \lfloor nr \rfloor$, we concatenate them to obtain a unique path $\pi \cdot \pi' : \lfloor np \rfloor \to \lfloor nr \rfloor$ with unnormalized empirical measure $\mu_{\pi \cdot \pi'} = \mu_\pi + \mu_{\pi'}$. By the subadditivity of the Levy-Prokhorov metric (Lemma 3.24),

$$\begin{aligned}\rho(\mu_{\pi\cdot\pi'}, n(\eta - \xi)) &= \rho(\mu_\pi + \mu_{\pi'}, n(\eta - \nu) + n(\nu - \xi)) \\ &\leqslant \rho(\mu_\pi, n(\eta - \nu)) + \rho(\mu_{\pi'}, n(\nu - \xi))\end{aligned}$$

so

$$\begin{aligned}&\left(\sum_{\pi:\lfloor np \rfloor \to \lfloor nq \rfloor} e^{-\frac{1}{\epsilon}\rho(\mu_\pi, n(\eta-\nu))}\right)\left(\sum_{\pi':\lfloor nq \rfloor \to \lfloor nr \rfloor} e^{-\frac{1}{\epsilon}\rho(\mu_{\pi'}, n(\nu-\xi))}\right) \\ &\leqslant \left(\sum_{\pi''':\lfloor np \rfloor \to \lfloor nr \rfloor} e^{-\frac{1}{\epsilon}\rho(\mu_{\pi'''}, n(\eta-\xi))}\right)\end{aligned}$$

It follows that (2.11) holds. Dividing (2.11) by n and taking the limit (in probability) as $n \to \infty$ we get

$$\widetilde{d^\epsilon}((p, \xi), (r, \eta)) \geqslant \widetilde{d^\epsilon}((p, \xi), (q, \nu)) + \widetilde{d^\epsilon}((q, \nu), (r, \eta))$$

so $\widetilde{d^\epsilon}$ is a directed metric with negative sign. Taking the limit as $\epsilon \to 0^+$, we still obtain a directed metric with negative sign. □

We proceed to show that $\widetilde{d}$ gives rise to a directed norm with negative sign.

Theorem 2.45.
(i) Each $\widetilde{d^\epsilon}$ is translation-invariant and positive-homogeneous. So is $\widetilde{d}$.
(ii) For $q \in \mathbb{R}^D, \nu \in \mathcal{M}_+$ define the grid entropy with respect to (q, ν) to be

$$||(q, \nu)|| := \widetilde{d}((\vec{0}, 0), (q, \nu))$$

Then this is a directed norm with negative sign on $\mathbb{R}^D \times \mathcal{M}_+$.

Remark 2.46. From before, $||(q, \nu)||$ is $-\infty$ if $q \notin \mathbb{R}^D_{\geqslant 0}$ or if $q = \vec{0}$ and $\nu \neq 0$, and it is 0 if $q = \vec{0}$ and $\nu = 0$.

Remark 2.47. A directed metric with negative sign is clearly concave. Thus each $\widetilde{d^\epsilon}$ as well as $||(\cdot, \cdot)||$ are concave functions on their respective domains.

Proof. (i) Fix $\epsilon > 0$. We already showed that $\widetilde{d}^\epsilon$ is translation-invariant while proving Theorem 7.

By translation-invariance, it suffices to show that $\widetilde{d}^\epsilon((\vec{0},0),(q,\nu))$ is positive-homogeneous. Consider any $c = \frac{a}{b} \in \mathbb{Q}_{>0}$. Then

$$\widetilde{d}^\epsilon((\vec{0},0),(cq,c\nu)) = \lim_{n\to\infty} \frac{d^\epsilon((\vec{0},0),(cnq,cn\nu))}{n} \text{ a.s.}$$

Looking at the subsequence consisting of multiples $n = mb$ of b, we get

$$\widetilde{d}^\epsilon((\vec{0},0),(cq,c\nu)) = \lim_{m\to\infty} \frac{d^\epsilon((\vec{0},0),(amq,am\nu))}{mb} \text{ a.s.}$$

But each $am \in \mathbb{N}$ so

$$\widetilde{d}^\epsilon((\vec{0},0),(cq,c\nu)) = \frac{a}{b}\lim_{n\to\infty} \frac{d^\epsilon((\vec{0},0),(nq,n\nu))}{n} = c\widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) \text{ a.s.}$$

Thus $\widetilde{d}^\epsilon$ is positive-homogeneous for rational factors.

Now consider any $c \in \mathbb{R}_{>0}$ and take sequences $a_k, b_k \in \mathbb{Q}_{>0}, a_k \uparrow c, b_k \downarrow c$. Consider any n, and let us look at

$$X_n^{\epsilon,cq,c\nu} = d^\epsilon((\vec{0},0),(cq,c\nu))$$

By Lemma 2.38,

$$\begin{aligned}
&X_n^{\epsilon,a_kq,a_k\nu} - \frac{1}{\epsilon}(n|c-a_k|\cdot(||q||_1 + ||\nu||_{TV}))\\
&\leqslant X_n^{\epsilon,cq,c\nu}\\
&\leqslant X_n^{\epsilon,b_kq,b_k\nu} + \frac{1}{\epsilon}(n|b_k-c|\cdot(||q||_1 + ||\nu||_{TV}))
\end{aligned}$$

Dividing by n and taking a.s. limits, and applying homogeneity for positive rational factors we get

$$\begin{aligned}
&a_k\widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) - \frac{|c-a_k|\cdot(||q||_1 + ||\nu||_{TV})}{\epsilon}\\
&\leqslant \widetilde{d}^\epsilon((\vec{0},0),(cq,c\nu))\\
&\leqslant b_k\widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) + \frac{|b_k-c|\cdot(||q||_1 + ||\nu||_{TV})}{\epsilon}
\end{aligned}$$

Taking $k \to \infty$ gives us

$$\widetilde{d}^\epsilon((\vec{0},0),(cq,c\nu)) = c\widetilde{d}^\epsilon((\vec{0},0),(q,\nu))$$

so $\widetilde{d}^\epsilon$ is homogenous with respect to any positive real factor.

Taking the infimum over $\epsilon > 0$ we get that $\widetilde{d}$ is translation-invariant and positive-homogenous.

(ii) Follows directly from (i).

□

2.3.5 *Direction-free Grid Entropy*

We now wish to develop a grid entropy for the case where we no longer restrict ourselves to paths $\pi : \vec{0} \to \lfloor nq \rfloor$ for a given direction $q \in \mathbb{R}^D$, and instead look at all length $\lfloor nt \rfloor$ paths from $\vec{0}$ for a given size parameter $t \geqslant 0$. Another way of putting this is that we look at paths from the origin to the line $x_1 + x_2 + \ldots + x_D = \lfloor nt \rfloor$. Recall that the set of all such paths is denoted $\mathcal{P}_{\lfloor nt \rfloor}(\vec{0})$.

If we try to simply repeat our previous argument, we run into a dead end because we will no longer be in a superadditive setting. The solution is to observe that the distances $\widetilde{d}^\epsilon((\vec{0},0),(q,\nu))$ are maximized over $q \in \mathbb{R}^D_{\geqslant 0}$ with $||q||_1 = t$ by $q = t(\frac{1}{D}, \ldots, \frac{1}{D}) := t\ell$. This intuitively makes sense, since this direction is the direction which has the most NE paths.

Lemma 2.48. *Fix $t \geqslant 0, \nu \in \mathcal{M}_t$. Then*

$$\sup_{q \in \mathbb{R}^D_{\geqslant 0}: ||q||_1 = t} \widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) = \widetilde{d}^\epsilon((\vec{0},0),(t\ell,\nu)) \;\forall \epsilon > 0,$$

$$\sup_{q \in \mathbb{R}^D_{\geqslant 0}: ||q||_1 = t} ||(q,\nu)|| = ||(t\ell,\nu)||$$

Remark 2.49. We will show in Section 2.4 that $||q||_1 = ||\nu||_{TV}$ is a necessary condition for $||(q,\nu)||$ to be finite, so it makes sense that we only take the supremum over $q \in \mathbb{R}^D_{\geqslant 0}$ with $||q||_1 = t$.

Proof. This is an easy consequence of the symmetries of the grid and the concavity of $\widetilde{d}^\epsilon$ and direction-fixed grid entropy. We focus on the proof for $\widetilde{d}^\epsilon$; the argument for grid entropy goes the same way.

Fix $\epsilon > 0$. By positive-homogeneity and since the $t = 0$ is trivial, we may assume $t = 1$. Suppose there exists $q \in \mathbb{R}^D_{\geqslant 0}$ s.t. $||q||_1 = 1$ and

$$\widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) > \widetilde{d}^\epsilon((\vec{0},0),(\ell,\nu)) \tag{2.12}$$

Among such q pick one that maximizes the number of coordinates which are equal $\frac{1}{D}$. Thus there are distinct $1 \leqslant i, j \leqslant D$ s.t. $q_i < \frac{1}{D} < q_j$, so we can write $\frac{1}{D}$ as a convex combination of q_i, q_j:

$$\frac{1}{D} = wq_i + (1-w)q_j \text{ for some } w \in (0,1) \tag{2.13}$$

Let $\sigma_{ij}(q)$ be q with q_i, q_j swapped. By symmetry of the grid,

$$\widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) = \widetilde{d}^\epsilon((\vec{0},0),(\sigma_{ij}(q),\nu))$$

hence by concavity of $\widetilde{d}^\epsilon$,

$$\begin{aligned}\widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) &= w\widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) + (1-w)\widetilde{d}^\epsilon((\vec{0},0),(\sigma_{ij}(q),\nu)) \\ &\leqslant \widetilde{d}^\epsilon((\vec{0},0),(wq+(1-w)\sigma_{ij}(q),\nu))\end{aligned}$$

But $wq + (1-w)\sigma_{ij}(q)$ only changes the coordinates of q in positions i, j, with q_i becoming $\frac{1}{D}$ by (2.13). Thus we have found a q satisfying (2.12) that has at least one more coordinate that is $\frac{1}{D}$ than our previous q, which we had assumed had the maximal number of such coordinates. Contradiction. Therefore $q = \ell$ as desired. □

We will now use this useful fact along with the compactness of $\{q \in \mathbb{R}^D_{\geqslant 0} : ||q||_1 = t\}$ to show that $||(t\ell,\nu)||$ is the desired direction-free grid entropy of length t.

Theorem 2.50. *Fix $t \geqslant 0, \nu \in \mathcal{M}_t$. For any $\epsilon > 0$ we have*

$$\begin{aligned}\widetilde{d}^\epsilon((\vec{0},0),(t\ell,\nu)) &= \lim_{n\to\infty} \sup_{q\in\mathbb{R}^D_{\geqslant 0}:||q||_1=t} \frac{1}{n}\log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)} \\ &= \lim_{n\to\infty} \frac{1}{n}\log \sum_{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)}\end{aligned}$$

a.s.

Proof. The statement is trivial when $t = 0$ so we may assume $t > 0$. We focus on the first equality. By Lemma 2.48 and the trivial fact that $\sup\limsup \leqslant \limsup\sup$ in general, we immediately get

$$\widetilde{d}^\epsilon((\vec{0},0),(t\ell,\nu)) \leqslant \limsup_{n\to\infty} \sup_{q\in\mathbb{R}^D_{\geqslant 0}:||q||_1=t} \frac{1}{n}\log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)} \text{ a.s.}$$

Suppose equality does not hold on some event of positive probability. Thus there exists $\delta > 0$ s.t.

$$\widetilde{d}^\epsilon((\vec{0},0),(t\ell,\nu)) + 7\delta < \limsup_{n\to\infty} \sup_{q\in\mathbb{R}^D_{\geqslant 0}:||q||_1=t} \frac{1}{n}\log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)} \tag{2.14}$$

with positive probability. Let us intersect this event with the measure 1 event that $\widetilde{d}^\epsilon((\vec{0},0),(p,\nu))$ exists and is equal to the limit in Theorem 2.31 for every $p \in \mathbb{Q}^D_{\geqslant 0}$. Denote this event by $\mathcal{E}$.

Unfortunately, the expression inside the supremum is only continuous when approaching q from the SE, and not necessarily when approaching along $\{r \in \mathbb{R}^D_{\geqslant 0} : ||r||_1 = t\}$, so we need to work with some rather unpleasant approximations.

Let (n_i) be the (event-dependent) subsequence corresponding to the lim sup. For every $i \in \mathbb{N}$ pick some (event-dependent) $q^{n_i} \in \mathbb{R}^D_{\geqslant 0}$ with $||q^{n_i}||_1 = t$ s.t.

$$\sup_{q\in\mathbb{R}^D_{\geqslant 0}:||q||_1=t} \frac{1}{n_i} \log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor n_i q\rfloor)} e^{-\frac{n_i}{\epsilon}\rho(\frac{1}{n_i}\mu_\pi,\nu)} \leqslant \delta + \frac{1}{n_i} \log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor n_i q^{n_i}\rfloor)} e^{-\frac{n_i}{\epsilon}\rho(\frac{1}{n_i}\mu_\pi,\nu)} \tag{2.15}$$

By compactness of $\{r \in \mathbb{R}^D_{\geqslant 0} : ||r||_1 = t\}$, there is a converging subsequence $q^{n_{i_j}} \to^{L^1} q'$ for some $q' \in \mathbb{R}^D_{\geqslant 0}, ||q'||_1 = t$. Pick some $q'' \in \mathbb{Q}^D_{\geqslant 0}$ s.t. $q'' - q' \in \mathbb{R}^D_{>0}$ and

$$\max\left(\frac{1}{\epsilon}||q''-q'||_1, \left(\frac{||q''||_1}{t}-1\right)\tilde{d}^\epsilon((\vec{0},0),(t\ell,\nu)), \frac{1}{\epsilon}\left|\left|\left(1-\frac{t}{||q''||_1}\right)\nu\right|\right|_{TV}\right) < \delta \tag{2.16}$$

The idea is that since $q^{n_{i_j}} \to^{L^1} q'$ and the coordinates of $q''-q'$ are *strictly* positive, then there exists J s.t. $q'' - q^{n_{i_j}} \in \mathbb{R}^D_{\geqslant 0}$ $\forall j \geqslant J$. For such j we can apply Lemma 2.38 to get

$$\frac{1}{n_{i_j}} \log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor n_{i_j} q^{n_{i_j}}\rfloor)} e^{-\frac{n_{i_j}}{\epsilon}\rho(\frac{1}{n_{i_j}}\mu_\pi,\nu)} \leqslant \frac{1}{\epsilon}\left(||q''-q^{n_{i_j}}||_1 + \frac{D}{n_{i_j}}\right) + \frac{1}{n_{i_j}} \log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor n_{i_j} q''\rfloor)} e^{-\frac{n_{i_j}}{\epsilon}\rho(\frac{1}{n_{i_j}}\mu_\pi,\nu)} \tag{2.17}$$

Recalling that we are in the event $\mathcal{E}$, pick some $J' \geqslant J$ large enough so that for any $j \geqslant J'$,

$$\frac{1}{n_{i_j}} \log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor n_{i_j} q''\rfloor)} e^{-\frac{n_{i_j}}{\epsilon}\rho(\frac{1}{n_{i_j}}\mu_\pi,\nu)} \leqslant \delta + \tilde{d}^\epsilon((\vec{0},0),(q'',\nu)) \tag{2.18}$$

and $\frac{1}{\epsilon}\left(||q'-q^{n_{i_j}}||_1 + \frac{D}{n_{i_j}}\right) < \delta$ hence

$$\frac{1}{\epsilon}\left(||q''-q^{n_{i_j}}||_1 + \frac{D}{n_{i_j}}\right) \leqslant \frac{1}{\epsilon}\left(||q'-q^{n_{i_j}}||_1 + ||q''-q'||_1 + \frac{D}{n_{i_j}}\right) < 2\delta \tag{2.19}$$

by (2.16). Putting everything together, we get

$$\begin{aligned}
&\widetilde{d}^{\epsilon}((\vec{0},0),(t\ell,\nu)) + 7\delta \\
&\leqslant \widetilde{d}^{\epsilon}((\vec{0},0),(q'',\nu)) + 4\delta \text{ by (2.14)-(2.19)} \\
&= \frac{||q''||_1}{t}\widetilde{d}^{\epsilon}\left((\vec{0},0),\left(\frac{t}{||q''||_1}q'',\frac{t}{||q''||_1}\nu\right)\right) + 4\delta \text{ by positive-homogeneity}
\end{aligned}$$

By Lemma 2.48, this is

$$\begin{aligned}
&\leqslant \frac{||q''||_1}{t}\widetilde{d}^{\epsilon}\left((\vec{0},0),\left(t\ell,\frac{t}{||q''||_1}\nu\right)\right) + 4\delta \\
&\leqslant \frac{||q''||_1}{t}\widetilde{d}^{\epsilon}((\vec{0},0),(t\ell,\nu)) + \frac{1}{\epsilon}\left|\left|\left(1-\frac{t}{||q''||_1}\right)\nu\right|\right|_{TV} + 4\delta \text{ by Lemma 2.38} \\
&\leqslant \widetilde{d}^{\epsilon}((\vec{0},0),(t\ell,\nu)) + 6\delta \text{ by (2.16)}
\end{aligned}$$

which is a contradiction. Note that even though our sequence $q^{n_{i_j}}$, the limit q'' and the J' chosen are event-dependent, the upper bound above is not by virtue of Lemma 2.48. That is what makes this argument work.

We have thus shown

$$\widetilde{d}^{\epsilon}((\vec{0},0),(t\ell,\nu)) = \lim_{n\to\infty} \sup_{q\in\mathbb{R}^D_{\geqslant 0}:||q||_1=t} \frac{1}{n}\log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)} \text{ a.s.}$$

We now proceed with the second equality. Observe that for any n, any length $\lfloor nt\rfloor$ NE path π from $\vec{0}$ ends at $\lfloor nq\rfloor$ for some $q\in\mathbb{R}^D_{\geqslant 0}$ with $||q||_1=t$. Furthermore, the number of possible different endpoints of such paths is precisely $\binom{\lfloor nt\rfloor+D-1}{D-1}=O(n^D)$, which is the number of D-tuples of non-negative integers summing to $\lfloor nt\rfloor$. Therefore

$$\begin{aligned}
&\log \sum_{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)} \\
&\in \left[\sup_{q\in\mathbb{R}^D_{\geqslant 0},||q||_1=t} \log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)}, \log\left(O(n^D)\sup_{q\in\mathbb{R}^D_{\geqslant 0},||q||_1=t}\sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)}\right)\right] \\
&= \left[\sup_{q\in\mathbb{R}^D_{\geqslant 0},||q||_1=t} d^{\epsilon}((\vec{0},n\xi),(nq,n\nu)), O(D\log n) + \sup_{q\in\mathbb{R}^D_{\geqslant 0},||q||_1=t} d^{n\epsilon}((\vec{0},n\xi),(nq,n\nu))\right]
\end{aligned}$$

When we divide by n and take the limit, the $O(D\log n)$ term goes to 0 and we get that

$$\lim_{n\to\infty}\log \sum_{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)} = \lim_{n\to\infty}\sup_{q\in\mathbb{R}^D_{\geqslant 0},||q||_1=t}\frac{d^{n\epsilon}((\vec{0},n\xi),(nq,n\nu))}{n} \text{ a.s.}$$

as desired. □

Of course, we already established that $\widetilde{d^\epsilon}((\vec{0},0),(t\ell,\nu))$ is a directed metric with negative sign. This means that if we now take the infimum over $\epsilon > 0$ we get that the length t direction-free grid entropy is precisely $||(t\ell,\nu)||$. Since this is $-\infty$ unless $t = ||\nu||_{TV}$ (see Theorem 2.57 in a later section), then we can simply let $t = ||\nu||_{TV}$ to begin with.

Definition 2.51. For $\nu \in \mathcal{M}_+$ let $t := ||\nu||_{TV}$ and define the direction-free grid entropy to be

$$||\nu|| := \inf_\epsilon \widetilde{d^\epsilon}((\vec{0},0),(t\ell,\nu)) = ||(t\ell,\nu)|| = \sup_{q\in\mathbb{R}^D} ||(q,\nu)||$$

Remark 2.52. The final equality holds by Lemma 2.48 combined with the fact that $||(q,\nu)|| = -\infty$ unless $||q||_1 = ||\nu||_{TV}$.

2.3.6 *Equivalence of Approaches*

The next order of business is to show that these definitions of direction-fixed/direction-free grid entropy are equivalent to our original, more intuitive, definitions. The key idea is that the sum

$$\sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)}$$

appearing in the definition of d^ϵ is approximately the number of paths $\vec{0} \to \lfloor nq\rfloor$ with very small $\rho(\frac{1}{n}\mu_\pi,\nu)$ and the error should disappear in the limit.

Let us recall how the $\overset{k}{\underset{\pi}{\min}}$ are defined. Fix $q \in \mathbb{R}^D, \nu \in \mathcal{M}_+$. For any $n,k \in \mathbb{N}$ with $k \leqslant (\#\pi : \vec{0} \to \lfloor nq\rfloor)$ define

$$\overset{k}{\underset{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}{\min}} \rho\left(\frac{1}{n}\mu_\pi,\nu\right)$$

to be the order statistics of $\rho(\frac{1}{n}\mu_\pi,\nu)$ as we range over all $\pi : \vec{0} \to \lfloor nq\rfloor$. In this way $\overset{1}{\underset{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}{\min}} \rho(\frac{1}{n}\mu_\pi,\nu)$, $\overset{\#\pi}{\underset{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}{\min}} \rho(\frac{1}{n}\mu_\pi,\nu)$ are just the minimum/maximum values of $\rho(\frac{1}{n}\mu_\pi,\nu)$ respectively.

Similarly, in the direction-free case, for $\nu \in \mathcal{M}_+$ we let $t := ||\nu||_{TV}$ and for any $n,k \in \mathbb{N}$ with $k \leqslant$ (# length $\lfloor nt\rfloor$ paths π from $\vec{0}$) define

$$\overset{k}{\underset{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})}{\min}} \rho\left(\frac{1}{n}\mu_\pi,\nu\right)$$

to be the k-th smallest value of $\rho(\frac{1}{n}\mu_\pi, \nu)$ when we range over *all* length $\lfloor nt \rfloor$ paths from $\vec{0}$.

We will usually omit the $\rho(\frac{1}{n}\mu_\pi, \nu)$ in $\overset{j}{\min_\pi} \rho(\frac{1}{n}\mu_\pi, \nu)$ for the sake of space.

We are now ready to prove that our definitions of grid entropy are equivalent, and moreover that the pertinent limits either converge to 0 a.s. or the corresponding $\liminf$'s are > 0 a.s.. For the sake of notation, we denote by $||(q,\nu)||, ||\nu||$ the direction-fixed/direction-free grid entropy as defined in Theorem 2.45/Theorem 2.50.

Theorem 2.53. *Fix $q \in \mathbb{R}^D, \nu \in \mathcal{M}_+$ and let $t := ||\nu||_{TV}$.*
(i) The grid entropy being finite is characterized by the existence of paths with empirical measure arbitrarily close to the target. More specifically,

$$\lim_{n\to\infty} \overset{1}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq \rfloor)}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right), \lim_{n\to\infty} \overset{1}{\min_{\pi\in\mathcal{P}_{\lfloor nt \rfloor}(\vec{0})}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) \text{ exist a.s. and}$$

$$||(q,\nu)|| \neq -\infty \Leftrightarrow \lim_{n\to\infty} \overset{1}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq \rfloor)}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0 \text{ a.s.}$$

$$||\nu|| \neq -\infty \Leftrightarrow \lim_{n\to\infty} \overset{1}{\min_{\pi\in\mathcal{P}_{\lfloor nt \rfloor}(\vec{0})}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0 \text{ a.s.}$$

Furthermore, if $||(q,\nu)|| \neq -\infty$ then

$$||(q,\nu)|| \in [0, H(q)] \text{ where } H(q) = \sum_{i=1}^{D} -q_i \log \frac{q_i}{||q||_1}$$

and if $||\nu|| \neq -\infty$ then $||\nu|| = ||(t\ell,\nu)|| \in [0, t\log D]$. In particular,

$$||\ ||q||_1\Lambda\ || \geqslant ||(q, ||q||_1\Lambda)|| \geqslant 0$$

i.e. Λ will always be among the distributions observed.
(ii) We have

$$\lim_{n\to\infty} \overset{\lfloor e^{n\alpha} \rfloor}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq \rfloor)}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0 \text{ a.s. } \forall 0 \leqslant \alpha < ||(q,\nu)||$$

$$\liminf_{n\to\infty} \overset{\lfloor e^{n\alpha} \rfloor}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq \rfloor)}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) > 0 \text{ a.s. } \forall ||(q,\nu)|| < \alpha \leqslant H(q)$$

In other words,

$$\begin{aligned}||(q,\nu)|| &= \sup\left\{\alpha \geqslant 0 : \lim_{n\to\infty} \min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}^{\lfloor e^{n\alpha}\rfloor} \rho\left(\frac{1}{n}\mu_\pi,\nu\right) = 0 \text{ a.s.}\right\} \\ &= \sup\left\{\alpha \geqslant 0 : \liminf_{n\to\infty} \min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}^{\lfloor e^{n\alpha}\rfloor} \rho\left(\frac{1}{n}\mu_\pi,\nu\right) = 0 \text{ a.s.}\right\}\end{aligned} \tag{2.20}$$

Similarly,

$$\lim_{n\to\infty} \min_{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})}^{\lfloor e^{n\alpha}\rfloor} \rho\left(\frac{1}{n}\mu_\pi,\nu\right) = 0 \text{ a.s. } \forall 0 \leqslant \alpha < ||\nu||$$

$$\liminf_{n\to\infty} \min_{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})}^{\lfloor e^{n\alpha}\rfloor} \rho\left(\frac{1}{n}\mu_\pi,\nu\right) > 0 \text{ a.s. } \forall ||\nu|| < \alpha \leqslant t\log D$$

and the direction-free grid entropy is given by

$$\begin{aligned}||\nu|| &= \sup\left\{\alpha \geqslant 0 : \lim_{n\to\infty} \min_{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})}^{\lfloor e^{n\alpha}\rfloor} \rho\left(\frac{1}{n}\mu_\pi,\nu\right) = 0 \text{ a.s.}\right\} \\ &= \sup\left\{\alpha \geqslant 0 : \liminf_{n\to\infty} \min_{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})}^{\lfloor e^{n\alpha}\rfloor} \rho\left(\frac{1}{n}\mu_\pi,\nu\right) = 0 \text{ a.s.}\right\}\end{aligned}$$

Remark 2.54. Since the $\min\limits_{\pi}^{j}$ increase in j, then the sets we are taking the supremum over in the definition in (ii) are either $\varnothing$ or of the form $[0,C)$ or $[0,C]$ where C is the corresponding grid entropy.

Remark 2.55. The limits

$$\lim_{n\to\infty} \min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}^{\lfloor e^{n\alpha}\rfloor} \rho\left(\frac{1}{n_i}\mu_\pi,\nu\right)$$

need not exist for all α, but we are saying that they do exist and are equal to 0 a.s. for $\alpha < ||(q,\nu)||$ and that the corresponding lim inf's are > 0 a.s. for $\alpha > ||(q,\nu)||$. This along with the fact the existence of the limits in (i) is completely non-trivial and will be a direct consequence of the a.s. convergence of $\frac{d^\epsilon((0,0),(nq,n\nu))}{n}$.

Proof. We will only focus on the proofs for direction-fixed grid entropy statements. The direction-free arguments are analogous.

(i) The strategy will be to provide upper and lower bounds on d^ϵ based on

$\exp[-\frac{n}{\epsilon} \min\limits_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}^{1} \rho(\frac{1}{n}\mu_\pi,\nu)]$, the leading term in the sum over π. Since $\frac{d^\epsilon((\vec{0},0),(nq,n\nu))}{n}$ converges a.s. then our analysis will give that $\widetilde{d}^\epsilon$ must be in the intersection of two intervals, one in terms of the lim sup $\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} \rho\left(\frac{1}{n}\mu_\pi,\nu\right)$ and the other in terms of the lim inf's. The only way these intervals can have non-empty intersection for small enough $\epsilon > 0$ will be if the lim sup's equal the lim inf's.

Fix $\epsilon > 0$. We consider the $\overset{j}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}$ over $\pi : \vec{0} \to \lfloor nq\rfloor$. Since $\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} \leqslant \overset{j}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}$ for all valid j, then

$$\begin{aligned}
&\frac{d^\epsilon((\vec{0},0),(nq,n\nu))}{n} \\
&= \frac{1}{n}\log \sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)} \\
&\in \left[\frac{1}{n}\log\left(\exp\left(-\frac{n}{\epsilon}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}\right)\right), \frac{1}{n}\log\left((\#\pi : \vec{0}\to\lfloor nq\rfloor)\exp\left(-\frac{n}{\epsilon}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}\right)\right)\right] \\
&= \left[-\frac{1}{\epsilon}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}\rho\left(\frac{1}{n}\mu_\pi,\nu\right), \frac{1}{n}\log(\#\pi : \vec{0}\to\lfloor nq\rfloor) - \frac{1}{\epsilon}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}\rho\left(\frac{1}{n}\mu_\pi,\nu\right)\right]
\end{aligned}$$

where each $\rho(\frac{1}{n}\mu_\pi,\nu) \geqslant 0$. Taking the a.s. lim sup or the a.s. lim inf and using Lemma 2.1 we get that a.s.,

$$\begin{aligned}
\widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) \in &\left[-\frac{1}{\epsilon}\liminf_{n\to\infty}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}, H(q) - \frac{1}{\epsilon}\liminf_{n\to\infty}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}\right] \\
&\bigcap\left[-\frac{1}{\epsilon}\limsup_{n\to\infty}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}, H(q) - \frac{1}{\epsilon}\limsup_{n\to\infty}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}\right]
\end{aligned} \tag{2.21}$$

Since this holds for arbitrarily small $\epsilon > 0$ then we must have

$$\liminf_{n\to\infty}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}\rho\left(\frac{1}{n}\mu_\pi,\nu\right) = \limsup_{n\to\infty}\overset{1}{\min}_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}\rho\left(\frac{1}{n}\mu_\pi,\nu\right) \text{ a.s.}$$

Furthermore, taking $\epsilon \to 0^+$ in (2.21), we see that

$$||(q,\nu)|| = \inf_{\epsilon>0}\widetilde{d}^\epsilon((\vec{0},0),(q,\nu)) \neq -\infty \Leftrightarrow \lim_{n\to\infty}\overset{1}{\min}_{\pi:\vec{0}\to\lfloor nq\rfloor}\rho\left(\frac{1}{n}\mu_\pi,\nu\right) = 0 \text{ a.s.}$$

Moreover, if indeed $||(q,\nu)|| \neq -\infty$ then (2.21) gives $||(q,\nu)|| \in [0, H(q)]$.

Finally, if we fix an infinite NE path π passing through all $\lfloor nq \rfloor$ then by the Glivenko-Cantelli Theorem (Theorem 3.18),

$$\rho\left(\frac{1}{n}\mu_{\pi|_{\vec{0}\to\lfloor nq\rfloor}}, ||q||_1\Lambda\right) = ||q||_1\rho\left(\frac{1}{|\pi|}\mu_{\pi|_{\vec{0}\to\lfloor nq\rfloor}}, \Lambda\right) \to 0$$

where $\mu_{\pi|_{\vec{0}\to\lfloor nq\rfloor}}$ denotes the restriction of the path to $\vec{0} \to \lfloor nq \rfloor$, so from above it follows that $||(q, ||q||_1\Lambda)|| \geqslant 0$.

(ii) If $||(q,\nu)|| = -\infty$ then by (i),

$$\lim_{n\to\infty} \overset{1}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}} \rho\left(\frac{1}{n_i}\mu_\pi, \nu\right) \text{ exists a.s. and is} > 0$$

so the statement holds trivially because the $\overset{j}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}}$ are nondecreasing in j. Thus it suffices to assume $||(q,\nu)|| \geqslant 0$ and hence

$$\overset{1}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) \to 0 \text{ a.s.} \tag{2.22}$$

Fix $\epsilon > 0$ and consider any $n \in \mathbb{N}$. By (2.22), the leading term of

$$\sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} e^{-\frac{n}{\epsilon}\rho(\frac{1}{n}\mu_\pi,\nu)}$$

is ≈ 1. So we must look at the secondary terms by writing

$$\frac{d^\epsilon((\vec{0},0),(nq,n\nu))}{n} = -\frac{1}{\epsilon}\overset{1}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}} + \frac{1}{n}\log\left[1 + \sum_{j=2}^{\#\pi} e^{-\frac{n}{\epsilon}[\overset{j}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}} - \overset{1}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}}]}\right] \tag{2.23}$$

where the summands are nonincreasing in j.

The argument will go similar to the one in (i), in that we compute some upper and lower bounds for this expression. We will fix an arbitrary $0 \leqslant C \leqslant H(q)$.

We lower bound (2.23) by truncating the sum at $j = \lfloor e^{Cn} \rfloor$ and lower bounding all the summands by the $j = \lfloor e^{Cn} \rfloor$ summand:

$$\frac{d^\epsilon((\vec{0},0),(nq,n\nu))}{n} \tag{2.24}$$

$$\geqslant -\frac{1}{\epsilon}\overset{1}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}} + \frac{1}{n}\log\left[1 + (\lfloor e^{Cn}\rfloor - 1)\exp\left[-\frac{n}{\epsilon}\left(\overset{\lfloor e^{Cn}\rfloor}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}} - \overset{1}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}}\right)\right]\right] \tag{2.25}$$

We upper bound (2.23) by upper bounding summands for $j = 2$ to $\lfloor e^{Cn} \rfloor$ to 1 and upper bounding the rest of the summands to the $j = \lfloor e^{Cn} \rfloor$ summand:

$$\frac{d^\epsilon((\vec{0},0),(nq,n\nu))}{n} \leqslant -\frac{1}{\epsilon} \min_{\pi \in \mathcal{P}(\vec{0},\lfloor nq \rfloor)}^{1} + \frac{1}{n} \log \left[1 + \lfloor e^{Cn} \rfloor \cdot 1 + (\#\pi) \exp\left[-\frac{n}{\epsilon} \left(\min_{\pi \in \mathcal{P}(\vec{0},\lfloor nq \rfloor)}^{\lfloor e^{Cn} \rfloor} - \min_{\pi \in \mathcal{P}(\vec{0},\lfloor nq \rfloor)}^{1} \right) \right] \right] \tag{2.26}$$

Consider the event-dependent sequence

$$a_n(C) := \min_{\pi \in \mathcal{P}(\vec{0},\lfloor nq \rfloor)}^{\lfloor e^{Cn} \rfloor} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) - \min_{\pi \in \mathcal{P}(\vec{0},\lfloor nq \rfloor)}^{1} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) \geqslant 0$$

Taking the lim inf / lim sup in (2.24),(2.26) and using (2.22) and Lemma 2.1 we get

$$\widetilde{d}^\epsilon((0,0),(q,\nu)) \geqslant \log \limsup_{n\to\infty} \left(1 + (\lfloor e^{Cn} \rfloor - 1) e^{-\frac{a_n(C)}{\epsilon} n}\right)^{\frac{1}{n}} = C - \frac{1}{\epsilon} \liminf_{n\to\infty} a_n(C),$$

$$\begin{aligned} \widetilde{d}^\epsilon((0,0),(q,\nu)) &\leqslant \log \liminf_{n\to\infty} \left(1 + e^{Cn} + e^{H(q)n - \frac{a_n(C)}{\epsilon} n}\right)^{\frac{1}{n}} \\ &= \max(C, H(q) - \frac{1}{\epsilon} \limsup_{n\to\infty} a_n(C)) \end{aligned}$$

a.s. for this arbitrary $0 \leqslant C \leqslant H(q)$.

In particular, for any $0 \leqslant C < ||(q,\nu)||$ we get

$$0 \leqslant ||(q,\nu)|| \leqslant \widetilde{d}^\epsilon((0,0),(q,\nu)) \leqslant \max(C, H(q) - \frac{1}{\epsilon} \limsup_{n\to\infty} a_n(C)) \text{ a.s. } \forall \epsilon > 0$$

hence

$$\lim_{n\to\infty} \min_{\pi \in \mathcal{P}(\vec{0},\lfloor nq \rfloor)}^{\lfloor e^{Cn} \rfloor} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = \lim_{n\to\infty} a_n(C) = 0 \text{ a.s.}$$

On the other hand, for any $||(q,\nu)|| < C \leqslant H(q)$,

$$||(q,\nu)|| = \inf_{\epsilon>0} \widetilde{d}^\epsilon((0,0),(q,\nu)) \geqslant C - \inf_{\epsilon>0} \frac{1}{\epsilon} \liminf_{n\to\infty} a_n(C) \text{ a.s.}$$

hence

$$\liminf_{n\to\infty} \min_{\pi \in \mathcal{P}(\vec{0},\lfloor nq \rfloor)}^{\lfloor e^{Cn} \rfloor} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = \lim_{n\to\infty} a_n(C) > 0 \text{ a.s.}$$

Equation (2.20) follows. □

Thus our approaches to grid entropy are equivalent.

Note that so far we have not made use of the coupling $\tau_e = \tau(U_e)$ or of the compactness of the space of measures on $[0,1]$. We could have developed grid entropy in the original environment, with unnormalized empirical measures σ_π and edge weight distribution θ and target measures ν on $\mathbb{R}$ in the same way. Furthermore, if θ has a continuous cdf then the value of the grid entropy would be the same either way, because of the duality between the environments, established in Lemma 2.12; that is, for any $q \in \mathbb{R}^D$ and $\nu \in \mathcal{M}_+$ we would have

$$||(q,\nu)|| = ||(q,\tau_*(\nu))||, ||\nu|| = ||\tau_*(\nu)||$$

where the grid entropies are developed on the environments $([0,1], U_e \sim \Lambda)$, $(\mathbb{R}, \tau_e \sim \theta)$ respectively. Even in the general case wher F_θ may not be continuous, [Bat20, Lemma 6.15] implies that

$$[||(q,\nu)|| \leqslant ||(q,\tau_*(\nu))||, ||\nu|| \leqslant ||\tau_*(\nu)||$$

Again though, this is just a nice fact; for practical purposes we may simply define grid entropy on the space of measures on $\mathbb{R}$ and our arguments thus far hold.

We will now reap the benefits of this coupling, by describing the sets $\mathcal{R}^q, \mathcal{R}^{q,\alpha}, \mathcal{R}^t, \mathcal{R}^{t,\alpha}$ defined in Theorem 2.13 and Corollary 2.15 in terms of direction-fixed/direction-free grid entropy.

Corollary 2.56. *Fix $q \in \mathbb{R}^D$ and $t \geqslant 0$.*
(i) The deterministic set $\mathcal{R}^q$ of limits of empirical measures in direction q whose existence is established by Theorem 2.13 (i) is precisely $\{\nu \in \mathcal{M}_+ : ||(q,\nu)|| \geqslant 0\}$. Thus

$$||(q,\nu)|| = -\infty \Leftrightarrow \lim_{n\to\infty} \overset{1}{\min_{\pi \in \mathcal{P}(\vec{0},\lfloor nq \rfloor)}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) > 0 \text{ a.s.}$$

$$\text{and } ||(q,\nu)|| \geqslant 0 \Leftrightarrow P\left(\exists \pi : \vec{0} \to \lfloor nq \rfloor \text{ with } \rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon \text{ i.o.}\right) = 1 \ \forall \epsilon > 0$$

Likewise, the deterministic set $\mathcal{R}^t$ of limits of length t empirical measures from direction-free paths, whose existence is established by Theorem 2.13(ii) is precisely

$$\{\nu \in \mathcal{M}_t : ||\nu|| \geqslant 0\} = \{\nu \in \mathcal{M}_t : ||(t\ell,\nu)|| \geqslant 0\} = \mathcal{R}^{t\ell}$$

Thus

$$||\nu|| = -\infty \Leftrightarrow \lim_{n\to\infty} \overset{1}{\min_{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) > 0 \text{ a.s.}$$

$$\text{and } ||\nu|| \geqslant 0 \Leftrightarrow P\left(\exists\, \pi \text{ s.t. } |\pi| = \lfloor nt\rfloor \text{ with } \rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon \text{ i.o.}\right) = 1\ \forall\epsilon > 0$$

(ii) Fix $0 < C \leqslant H(q)$. The set of measures with grid entropy in direction q at least/most C can be characterized in terms of the deterministic sets $\mathcal{R}^{q,\alpha}$ defined in Corollary 2.15 (i):

$$\mathcal{R}^{q,C} = \bigcap_{0\leqslant\alpha\leqslant C} \mathcal{R}^{q,\alpha} \subseteq \{\nu \in \mathcal{M}_+ : ||(q,\nu)|| \geqslant C\} \subseteq \bigcap_{0\leqslant\alpha<C} \mathcal{R}^{q,\alpha}$$

Thus

$$||(q,\nu)|| \geqslant C > 0 \Leftrightarrow P\left(\exists\lfloor e^{n\alpha}\rfloor \text{ paths } \pi : \vec{0} \to \lfloor nq\rfloor \text{ with } \rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon \text{ i.o.}\right) = 1$$

$\forall\epsilon > 0\ \forall\alpha \in (0, C)$, *and*

$$||(q,\nu)|| < C \Leftrightarrow \liminf_{n\to\infty} \overset{\lfloor e^{nC}\rfloor}{\min_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) > 0 \text{ a.s.}$$

Now fix $0 < C \leqslant H(t\ell) = t\log D$. The set of measures with length t direction-free grid entropy at least/most C can be characterized in terms of the deterministic sets $\mathcal{R}^{t,\alpha}$ defined in Corollary 2.15 (ii):

$$\mathcal{R}^{t\ell,C} = \mathcal{R}^{t,C} = \bigcap_{0\leqslant\alpha\leqslant C} \mathcal{R}^{t,\alpha} \subseteq \{\nu \in \mathcal{M}_t : ||\nu|| \geqslant C\} \subseteq \bigcap_{0\leqslant\alpha<C} \mathcal{R}^{t,\alpha}$$

Thus

$$||\nu|| \geqslant C > 0 \Leftrightarrow P\left(\exists\lfloor e^{n\alpha}\rfloor\ \pi \in \mathcal{P}_{\lfloor nt\rfloor}(\vec{0}) \text{ with } \rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon \text{ i.o.}\right) = 1\ \forall\epsilon > 0$$

$\forall\alpha \in (0, C)$, *and*

$$||\nu|| < C \Leftrightarrow \liminf_{n\to\infty} \overset{\lfloor e^{nC}\rfloor}{\min_{\pi\in\mathcal{P}_{\lfloor nt\rfloor}(\vec{0})}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) > 0 \text{ a.s.}$$

Proof. This follows immediately Theorem 2.53. □

Now that we have a firm grasp of what direction-fixed and direction-free grid entropies actually measure, we can move on examining what information they can give us.

2.4 PROPERTIES OF GRID ENTROPY

Since direction-free entropy $||\nu||$ is just $||(t\ell,\nu)||$ for $t := ||\nu||_{TV}$ then it suffices to study the properties of direction-fixed grid entropy. We are particularly interested in what the grid entropy can tell us about q and ν when $||(q,\nu)||$ is finite.

We have already established that it is necessary for q to be in $\mathbb{R}^D_{\geqslant 0}$ if we want $||(q,\nu)||$ not to be $-\infty$, and that if $q = \vec{0}$ then $||(q,\nu)|| > -\infty$ if and only if $\nu = 0$. The question now is what sorts of measures ν will be observed by the empirical measures along the direction $q \in \mathbb{R}^D_{\geqslant 0} \setminus \{\vec{0}\}$.

By positive-homogeneity of the norm, it suffices to consider $q \in \mathbb{R}^D_{\geqslant 0}$ with $||q||_1 = 1$ and study which $\nu \in \mathcal{M}_+$ give rise to finite $||(q,\nu)||$. In this section, it will be more convenient to work with the description of grid entropy given by Corollary 2.56 (ii).

Theorem 2.57. *Recall that we couple the edge weights with i.i.d. edge labels $U_e \sim \Lambda$ (the Lebesgue measure on $[0,1]$) and that*

$$H(q) := \lim_{n\to\infty} \frac{1}{n} \log(\#\pi : \vec{0} \to \lfloor nq \rfloor)$$

Consider any $q \in \mathbb{R}^D_{\geqslant 0}$ with $||q||_1 = 1$ and any $\nu \in \mathcal{M}_+$. Suppose $||(q,\nu)|| \neq -\infty$. Then:
(i) ν is a probability measure on $[0,1]$.
(ii) For any G_δ or F_σ set $A \subseteq [0,1]$, we have

$$\Lambda(A)^{\nu(A)}\Lambda(A^C)^{\nu(A^C)} \geqslant \nu(A)^{\nu(A)}\nu(A^C)^{\nu(A^C)}e^{||(q,\nu)||-H(q)}$$

(iii) From section 2.2.6 the Kullback-Leibler divergence can be defined as

$$D_{KL}(\nu||\Lambda) = \begin{cases} \int f \log f \, dx \text{ with } f := \frac{d\nu}{dx}, & \nu << \Lambda \\ +\infty, & \text{otherwise} \end{cases}$$

Then

$$||(q,\nu)|| + D_{KL}(\nu||\Lambda) \leqslant H(q)$$

so in particular, ν is absolutely continuous with respect to the Lebesgue measure on $[0,1]$.

Remark 2.58. If we take $q = \ell$ and recall that $H(\ell) = \log D$ we get the analogous statements for length 1 direction-free grid entropy.

Remark 2.59. (i) is exactly what we expect because the empirical measures $\frac{1}{n}\mu_\pi$ for paths $\pi : \vec{0} \to \lfloor nq \rfloor$ are non-negative Borel measures with total mass $\frac{|||\lfloor nq \rfloor|||}{n} \approx 1$ so in the limit we expect to only observe ν which are probability measures.

The inequalities (ii) and (iii) establish a strong relationship between Λ and ν that is necessary for (q,ν) to have finite grid entropy.

In (iii), the upper bound $H(q)$ we derive for $D_{KL}(\nu||\Lambda) + ||(q,\nu)||$ is precisely the entropy of a Categorical$(q_1,\ldots,q_D)$ distribution. It is also the upper bound we derived for grid entropy. Thus the bound from (iii) is essentially saying that as more empirical measures converge to the target distribution ν, ν must be closer to Λ. In particular, the closer q is to a permutation of $(1,0,\ldots,0)$, the fewer the number of NE paths and the smaller $||(q,\nu)|| + D_{KL}(\nu||\Lambda)$ must be.

Remark 2.60. If we do not make use of our coupling and define grid entropy directly on the space of measures on $\mathbb{R}$, (i)-(iii) hold analogously. This is trivial since we do not use weak compactness in the proof of this theorem.

Remark 2.61. (iii) was proved both by [Bat20] and [JLS20]. Our proof is completely different and quite elementary, as we make use of the heavy machinery that is grid entropy.

Remark 2.62. Recalling Theorem 2.13, (i) implies that $\mathcal{R}^q = \{\nu \in \mathcal{M}_+ : ||(q,\nu)|| \geqslant 0\}$ (and hence $\mathcal{R}^1 = \mathcal{R}^\ell$) is weakly compact.

Proof. (i) The intuition is that if $||\nu||_{TV} \neq 1$ then each $\rho(\frac{1}{n}\mu_\pi,\nu)$ is bounded below away from 0 hence it does not converge to 0.

Suppose $\nu \in \mathcal{M}_+$ is not a probability measure on $[0,1]$. By the reverse triangle inequality for ρ, for any n and any path $\pi : \vec{0} \to \lfloor nq \rfloor$,

$$\rho\left(\frac{1}{n}\mu_\pi,\nu\right) \geqslant |\rho(\mu_\pi,0) - \rho(n\nu,0)| = |\ ||\mu_\pi||_{TV} - n||\nu||_{TV}\ | = |\ ||\lfloor nq \rfloor||_1 - n||\nu||_{TV}\ |$$

Since $\frac{||\lfloor nq\rfloor||_1}{n} \to ||q||_1 = 1$ and $||\nu||_{TV} \neq 1$ then there is $\delta > 0$ s.t. for large enough n,

$$\rho\left(\frac{1}{n}\mu_\pi,\nu\right) \geqslant \frac{1}{n}|\ ||\lfloor nq \rfloor||_1 - n||\nu||_{TV}\ | > \delta$$

It follows that $\overset{1}{\min_\pi} \rho(\frac{1}{n}\mu_\pi,\nu) \nrightarrow 0$ a.s. so $||(q,\nu)|| = -\infty$. Contradiction. Therefore ν must be a probability measure on $[0,1]$.

(ii) Before proceeding with (ii) we will prove a quick claim that will also help with the proof of (iii).

Claim: Fix $\alpha \in [0,H(q)]$ and a nonempty set of measures $T \subseteq \mathcal{M}_+$. Then

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{n} \log P\left(\exists \lfloor e^{n\alpha} \rfloor \text{ paths } \pi : \vec{0} \to \lfloor nq \rfloor \text{ s.t. } \frac{1}{n}\mu_\pi \in T\right) \\ &\leqslant H(q) - \alpha + \limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}\mu_\pi \in T\right) \end{aligned}$$

In particular, if $\alpha \geqslant 0$ and

$$\alpha > H(q) + \limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}\mu_\pi \in T\right)$$

then

$$P\left(\exists \lfloor e^{n\alpha} \rfloor \text{ paths } \pi : \vec{0} \to \lfloor nq \rfloor \text{ s.t. } \frac{1}{n}\mu_\pi \in T \text{ i.o.}\right) = 0$$

Proof. Indeed, for any $n \in \mathbb{N}$ we have by Markov's Inequality

$$P\left(\sum_{\pi \in \mathcal{P}(0, \lfloor nq \rfloor)} \mathbf{1}_{\frac{1}{n}\mu_\pi \in T} \geqslant \lfloor e^{n\alpha} \rfloor\right) \leqslant \frac{1}{\lfloor e^{n\alpha} \rfloor} E\left[\sum_{\pi \in \mathcal{P}(\vec{0}, \lfloor nq \rfloor)} \mathbf{1}_{\frac{1}{n}\mu_\pi \in T}\right] = \frac{|\mathcal{P}(\vec{0}, \lfloor nq \rfloor)|}{\lfloor e^{n\alpha} \rfloor} P\left(\frac{1}{n}\mu_\pi \in T\right)$$

Taking $\limsup_{n\to\infty} \frac{1}{n} \log$ of both sides and using the definition of $H(q)$ we get the desired inequality.

Now suppose

$$\alpha > \alpha - \delta > H(q) + \limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}\mu_\pi \in T\right)$$

Then for large enough n,

$$P\left(\sum_{\pi \in \mathcal{P}(0, \lfloor nq \rfloor)} \mathbf{1}_{\frac{1}{n}\mu_\pi \in T} \geqslant \lfloor e^{n\alpha} \rfloor\right) \leqslant \exp\left[n\left(H(q) - \alpha + \delta + \limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}\mu_\pi \in T\right)\right)\right]$$

where the exponent is a strictly negative constant multiple of n. Hence, by the Borel-Cantelli Lemma,

$$P\left(\exists \lfloor e^{n\alpha} \rfloor \text{ paths } \pi : \vec{0} \to \lfloor nq \rfloor \text{ s.t. } \frac{1}{n}\mu_\pi \in T \text{ i.o.}\right) = 0$$

□ (Claim)

Now we begin the proof of (ii). The idea is to use the definition of the Levy-Prokhorov metric to unpack what $\rho(\frac{1}{n}\mu_\pi, \nu) < \epsilon$ means in terms of the values of the labels along π.

First, consider any closed $A \subseteq [0,1]$. Observe that $A = \overline{A} = \bigcap_{\epsilon > 0} A^\epsilon$ hence by continuity from above, $\nu(A^\epsilon) \downarrow \nu(A)$ and $\Lambda(A^\epsilon) \downarrow \Lambda(A)$ as $\epsilon \to 0^+$.

We will compute the probability that $\rho(\frac{1}{n}\mu_\pi, \nu) < \epsilon$ for a path $\pi : \vec{0} \to \lfloor nq \rfloor$ directly, using the definitions of ρ and μ_π, and then we will apply the Claim.

Consider $n \in \mathbb{N}$. For any path $\pi : \vec{0} \to \lfloor nq \rfloor$, by definition of the Levy-Prokhorov metric and the empirical measure μ_π,

$$\rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon \Rightarrow \nu(A) \leqslant \frac{1}{n}\mu_\pi(A^\epsilon) + \epsilon \text{ and } \frac{1}{n}\mu_\pi(A^\epsilon) \leqslant \nu((A^\epsilon)^\epsilon) + \epsilon$$

$$\Rightarrow \pi \text{ has} \geqslant \lfloor n(\nu(A) - \epsilon) \rfloor \text{ edge labels in } A^\epsilon, \leqslant \lfloor n(\nu(A^{2\epsilon}) + \epsilon) \rfloor \text{ labels in } A^\epsilon$$

$$\Rightarrow \pi \text{ has} \geqslant \lfloor n(\nu(A) - \epsilon) \rfloor \text{ edge labels in } A^\epsilon, \geqslant |||\lfloor nq \rfloor|||_1 - \lfloor n(\nu(A^{2\epsilon}) - \epsilon) \rfloor \text{ labels in } (A^\epsilon)^C$$

If $\nu(A) = 0$ or $\nu(A) = 1$ then we may completely omit the first/second half of the statement above as it is trivial. Otherwise, we take $\epsilon > 0$ small enough so that $\nu(A^{2\epsilon}) + \epsilon < 1$ and n large enough so that $n(\nu(A^{2\epsilon}) + \epsilon) < |||\lfloor nq \rfloor|||_1$. For the sake of convenience we only show the computation in the latter case; the $\nu(A) = 0$ or $\nu(A) = 1$ case is merely a simplified version of it.

Since U_e are i.i.d. Unif$[0, 1]$, then

$$P\left(\rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon\right)$$

$$\leqslant P(\pi \text{ has} \geqslant \lfloor n(\nu(A) - \epsilon) \rfloor \text{ edge labels in } A^\epsilon, \geqslant |||\lfloor nq \rfloor|||_1 - \lfloor n(\nu(A^{2\epsilon}) - \epsilon) \rfloor \text{ labels in } (A^\epsilon)^C)$$
$$\leqslant \binom{|||\lfloor nq \rfloor|||_1}{\lfloor n(\nu(A) - \epsilon) \rfloor, |||\lfloor nq \rfloor|||_1 - \lfloor n(\nu(A^{2\epsilon}) - \epsilon) \rfloor} \Lambda(A^\epsilon)^{n(\nu(A)-\epsilon)} \Lambda((A^\epsilon)^C)^{|||\lfloor nq \rfloor|||_1 - n(\nu(A^{2\epsilon})-\epsilon)} \tag{2.27}$$

by the union bound. We get the asymptotics of this binomial coefficient from Lemma 2.1:

$$\binom{|||\lfloor nq \rfloor|||_1}{\lfloor n(\nu(A) - \epsilon) \rfloor, |||\lfloor nq \rfloor|||_1 - \lfloor n(\nu(A^{2\epsilon}) - \epsilon) \rfloor}$$
$$= \left(\frac{1}{(\nu(A) - \epsilon)^{\nu(A)-\epsilon}(1 - \nu(A^{2\epsilon}) + \epsilon)^{1-\nu(A^{2\epsilon})+\epsilon}(\nu(A^{2\epsilon}) - \nu(A))^{\nu(A^{2\epsilon})-\nu(A)}} + o(1)\right)^n \tag{2.28}$$

Therefore

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(\rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon\right) \leqslant$$
$$\log \frac{\Lambda(A^\epsilon)^{\nu(A)-\epsilon}\Lambda((A^\epsilon)^C)^{1-\nu(A^{2\epsilon})+\epsilon}}{(\nu(A) - \epsilon)^{\nu(A)-\epsilon}(1 - \nu(A^{2\epsilon}) + \epsilon)^{1-\nu(A^{2\epsilon})+\epsilon}(\nu(A^{2\epsilon}) - \nu(A))^{\nu(A^{2\epsilon})-\nu(A)}}$$

Combining this with the Claim with $T = B_\epsilon(\nu)$, we get that

$$P\left(\exists \lfloor e^{n\alpha} \rfloor \text{ paths } \pi : \vec{0} \to \lfloor nq \rfloor \text{ with } \rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon \text{ i.o.}\right) = 0$$

$$\forall \alpha > H(q) + \log \frac{\Lambda(A^\epsilon)^{\nu(A)-\epsilon}\Lambda((A^\epsilon)^C)^{1-\nu(A^{2\epsilon})+\epsilon}}{(\nu(A)-\epsilon)^{\nu(A)-\epsilon}(1-\nu(A^{2\epsilon})+\epsilon)^{1-\nu(A^{2\epsilon})+\epsilon}(\nu(A^{2\epsilon})-\nu(A))^{\nu(A^{2\epsilon})-\nu(A)}}$$

By Corollary 2.56 (ii), it follows that

$$||(q,\nu)|| \leqslant$$
$$H(q) + \log \frac{\Lambda(A^\epsilon)^{\nu(A)-\epsilon}(1-\Lambda(A^\epsilon))^{1-\nu(A^{2\epsilon})+\epsilon}}{(\nu(A)-\epsilon)^{\nu(A)-\epsilon}(1-\nu(A^{2\epsilon})+\epsilon)^{1-\nu(A^{2\epsilon})+\epsilon}(\nu(A^{2\epsilon})-\nu(A))^{\nu(A^{2\epsilon})-\nu(A)}}$$

This holds for arbitrary $\epsilon > 0$. Taking $\epsilon \to 0^+$ and using $\nu(A^\epsilon) \downarrow \nu(A)$ and $\Lambda(A^\epsilon) \downarrow \Lambda(A)$,

$$1 \leqslant e^{H(q)-||(q,\nu)||}\frac{\Lambda(A)^{\nu(A)}\Lambda(A^C)^{\nu(A^C)}}{\nu(A)^{\nu(A)}\nu(A^C)^{\nu(A^C)}}$$

Therefore

$$\Lambda(A)^{\nu(A)}\Lambda(A^C)^{\nu(A^C)} \geqslant \nu(A)^{\nu(A)}\nu(A^C)^{\nu(A^C)}e^{||(q,\nu)||-H(q)} \tag{2.29}$$

This equation is symmetric in A, A^C hence it holds for all A open or closed. By continuity from below/above it holds for all G_δ and F_σ subsets of $[0,1]$.

(iii) The idea is to again use the Claim, this time to compute $\limsup_{n\to\infty} \frac{1}{n} \log P(\rho(\frac{1}{n}\mu_\pi, \nu) < \epsilon)$ using Sanov's Theorem.

Fix $\epsilon > 0$. By Sanov's Theorem (Theorem 2.25), since $\overline{B_{2\epsilon}}(\nu)$ is weakly closed then

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{|\pi|}\mu_\pi \in \overline{B_{2\epsilon}}(\nu)\right) \leqslant -\inf_{\xi \in \overline{B_{2\epsilon}}(\nu)} D_{KL}(\xi||\Lambda)$$

where the right hand side might a priori be $-\infty$.

Since $\rho(\frac{1}{|\pi|}\mu_\pi, \frac{1}{n}\mu_\pi) \leqslant |1 - \frac{|||nq|||_1}{n}| \to 0$ as $n \to \infty$, it follows that

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}\mu_\pi \in B_\epsilon(\nu)\right) &\leqslant \limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{|\pi|}\mu_\pi \in \overline{B_{2\epsilon}}(\nu)\right) \\ &\leqslant -\inf_{\xi \in \overline{B_{2\epsilon}}(\nu)} D_{KL}(\xi||\Lambda)\end{aligned}$$

By the Claim from (ii) with $T = B_\epsilon(\nu)$,

$$P\left(\exists \lfloor e^{n\alpha} \rfloor \text{ paths } \pi : \vec{0} \to \lfloor nq \rfloor \text{ with } \rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon \text{ i.o.}\right) = 0$$

$\forall \alpha > H(q) - \inf_{\xi \in \overline{B_{2\epsilon}}(\nu)} D_{KL}(\xi||\Lambda)$. Hence, by Corollary 2.56 (ii),

$$||(q,\nu)|| \leqslant H(q) - \inf_{\xi \in \overline{B_{2\epsilon}}(\nu)} D_{KL}(\xi||\Lambda)$$

Take $\epsilon = \frac{1}{k} \downarrow 0$ and let $\xi_k = arg\,min_{\xi \in \overline{B_{\frac{2}{k}}}(\nu)} D_{KL}(\xi||\Lambda)$ (the infimum over the weakly closed $\overline{B_{\frac{2}{k}}}(\nu)$ is achieved since relative entropy is lower semicontinuous). Then $\xi_k \Rightarrow \nu$ hence

$$D_{KL}(\nu||\Lambda) + ||(q,\nu)|| \leqslant \liminf_{k \to \infty} D_{KL}(\xi_k||\Lambda) + ||(q,\nu)|| \leqslant H(q)$$

as desired.

Finally, recall that $\nu \in \mathcal{R}^q$ hence $||(q,\nu)|| \geqslant 0$ so $D_{KL}(\nu||\Lambda)$ is finite and $\nu << \Lambda$. □

It is not clear that in general any probability measure ν satisfying $D_{KL}(\nu||\Lambda) \leqslant H(q)$ will result in finite grid entropies $||(q,\nu)||$ for $q \in \mathbb{R}^D_{\geqslant 0}$ with $||q||_1 = 1$.

In one specific case, however, this does hold trivially. When q is a permutation of $(1, 0, \ldots, 0)$, so that there is exactly one NE path $\vec{0} \to nq$, we get

$$||(q,\nu)|| > -\infty \Leftrightarrow D_{KL}(\nu||\Lambda) = 0 \Leftrightarrow \nu = \Lambda$$

with the value of $||(q,\Lambda)||$ being 0. Note that this completely covers the $D = 1$ case.

Also, in general we cannot explicitly compute grid entropy. But when the target measure ν is Λ, we can easily show that grid entropy is maximal. This makes sense, since the Glivenko-Cantelli Theorem says we expect that most empirical measures observed are very close to the edge label distribution.

Corollary 2.63. *Let* $q \in \mathbb{R}^D_{\geqslant 0} \backslash \{\vec{0}\}$. *Then* $||(q, q\Lambda)|| = H(q)$.

Proof. Since $H(q)$ and $||(q,\nu)||$ are positive-homogeneous (in the sense that $H(cq) = cH(q)$ and $||(cq, c\nu)|| = c||(q,\nu)|| \; \forall c > 0$), we may without loss of generality assume $||q||_1 = 1$.

Suppose $||(q,\Lambda)|| < H(q)$. Consider any $\epsilon > 0$. By Sanov's Theorem (Theorem 2.25),

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left(\frac{1}{n}\mu_\pi \in B_\epsilon(\Lambda)^C\right) \leqslant - \inf_{\xi\in B_\epsilon(\Lambda)^C} D_{KL}(\xi||\Lambda) \tag{2.30}$$

where $B_\epsilon(\Lambda)^C$ is weakly closed and relative entropy is lower semicontinuous hence the infimum is achieved by some $\xi \in B_\epsilon(\Lambda)^C$. But $\xi \neq \Lambda$ implies $-D_{KL}(\xi||\Lambda) < 0$. Fix any

$$\max(||(q,\Lambda)||, H(q) - D_{KL}(\xi||\Lambda)) < \alpha < H(q)$$

By the Claim from the proof of (ii) in the previous theorem with $T = B_\epsilon(\Lambda)^C$, we have

$$P\left(\exists\lceil e^{n\alpha}\rceil \text{ paths } \pi : \vec{0} \to \lfloor nq\rfloor \text{ s.t. } \rho\left(\frac{1}{n}\mu_\pi, \nu\right) \geqslant \epsilon \text{ i.o.}\right) = 0$$

hence

$$P\left(\exists(\#\pi : \vec{0} \to \lfloor nq\rfloor) - \lceil e^{n\alpha}\rceil \text{ paths } \pi : \vec{0} \to \lfloor nq\rfloor \text{ s.t. } \rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon \text{ eventually}\right) = 1$$

But $(\#\pi : \vec{0} \to \lfloor nq\rfloor) - \lceil e^{n\alpha}\rceil \geqslant \lfloor e^{n\alpha}\rfloor$ for large enough n since $\alpha < H(q)$, so

$$P\left(\exists\lfloor e^{n\alpha}\rfloor \text{ paths } \pi : \vec{0} \to \lfloor nq\rfloor \text{ s.t. } \rho\left(\frac{1}{n}\mu_\pi, \nu\right) < \epsilon \text{ eventually}\right) = 1$$

for this arbitrary $\epsilon > 0$, hence $\alpha \geqslant ||(q,\Lambda)||$. Contradiction. □

Continuity is another property of interest. Unfortunately, whether direction-fixed grid entropy is upper semicontinuous depends on which direction we approach our target q from. When we take a sequence q^k approaching q from the SE (in the sense that $q - q^k \in \mathbb{R}^D_{\geqslant 0}$), paths $\pi : \vec{0} \to \lfloor nq^k\rfloor$ can easily be appended with another path to get paths $\pi \cdot \pi' : \vec{0} \to \lfloor nq\rfloor$ and this will mean grid entropy is upper semicontinuous. In particular, this implies that grid entropy is upper semicontinuous in ν when we fix q and hence that direction-free grid entropy is upper semicontinuous.

On the other hand, when q^k approach q from an arbitrary direction, it matters whether the q^k have more freedom than q (in the sense that there are coordinates $(q^k)_i \neq 0$ for which the corresponding coordinate in q is $q_i = 0$). If the q^k do not have more freedom than q then upper semicontinuity follows from the first case by rescaling; otherwise, we cannot say anything.

There are some other special cases where we get lower/upper semicontinuity of grid entropy.

Note that here we no longer restrict ourselves to the case when $||q||_1 = 1$.

Theorem 2.64. *Let* $(q, \nu), (q^k, \nu^k) \in \mathbb{R}^D \times \mathcal{M}_+$ *s.t.* $q^k \to^{L^1} q, \nu^k \Rightarrow \nu$.
(i) If either $q = \vec{0}$ *or for large enough* k *we have* $q - q^k \in \mathbb{R}^D_{\geqslant 0}$ *or for large enough* k *we have* $(q^k)_i = 0$ *for all* i *s.t.* $q_i = 0$ *then*

$$\limsup_{k\to\infty} ||(q^k, \nu^k)|| \leqslant ||(q, \nu)||$$

(ii) If $||(q^k - q, \nu^k - \nu)|| \neq -\infty$*, for large enough* k *then*

$$||(q, \nu)|| \leqslant \liminf_{k\to\infty} ||(q^k, \nu^k)||$$

(iii) If there exist constants $C_k > 0$ *s.t.* $C_k \to 0$ *and s.t. for large enough* k,

$$\left|\left|\left(q - \frac{q - q^k}{C_k}, \nu - \frac{\nu - \nu^k}{C_k}\right)\right|\right| \neq -\infty$$

then

$$||(q, \nu)|| \leqslant \liminf_{k\to\infty} ||(q^k, \nu^k)||$$

Remark 2.65. In the case when $q^k = q$ for large enough k, (i) shows that direction-fixed grid entropy $||(q, \nu)||$ is always upper semicontinuous in ν. Taking $q^k = ||\nu^k||_{TV}\ell \to ||\nu||_{TV}\ell$ we get that

$$\limsup_{k\to\infty} ||\nu^k|| = \limsup_{k\to\infty} ||(||\nu^k||_{TV}\ell, \nu^k)|| \leqslant ||(||\nu||_{TV}\ell, \nu)|| = ||\nu||$$

so direction-free grid entropy is upper semicontinuous. These results improve what is shown in [Bat20], namely that $\mathcal{R}^q = \{\nu \in \mathcal{M}_+ : ||(q, \nu)|| \geqslant 0\}$, $\mathcal{R}^t = \{\nu \in \mathcal{M}_t : ||\nu|| \geqslant 0\}$ are weakly closed.

Remark 2.66. In the assumptions in (ii) and (iii) it is implicit that the parameters lie in
$\mathbb{R}^D_{\geqslant 0} \times \mathcal{M}_+$ because otherwise the grid entropy would be $-\infty$.

Remark 2.67. The conditions in (i), (ii), (iii) are not mutually exclusive so for some sequences we may combine the statements to obtain

$$||(q, \nu)|| = \lim_{k\to\infty} ||(q^k, \nu^k)||$$

Proof. For all three parts, we may assume the conditions hold for all $k \in \mathbb{N}$.
(i) **Case 1**: $q = \vec{0}$
First suppose $\nu = 0$ so that $||(q, \nu)|| = 0$. By Theorem 2.53,

$$||(q^k, \nu^k)|| \in \{-\infty\} \cup [0, H(q^k)] \ \forall k$$

hence

$$\limsup_{k\to\infty} ||(q^k, \nu^k)|| \leqslant \limsup_{k\to\infty} H(q^k) = \limsup_{k\to\infty} ||q^k||_1 \log ||q^k||_1 - \sum_{i=1}^{D} (q^k)_i \log(q^k)_i = 0$$

since $||q^k - q||_1 = ||q^k||_1 \to 0$. Thus $\limsup_{k\to\infty} ||(q^k, \nu^k)|| \leqslant ||(q, \nu)||$.

On the other hand, if $\nu \neq 0$ then for large enough k we have $||q^k||_1 \neq ||\nu^k||_{TV}$ (because otherwise $||\nu^k||_{TV} \to 0$ and hence ν, the weak limit of the ν^k, would have to be 0). But then

$$\limsup_{k\to\infty} ||(q^k, \nu^k)|| = -\infty = ||(q, \nu)||$$

by Theorem 2.57 (i).

Case 2: $q - q^k \in \mathbb{R}^D_{\geqslant 0} \ \forall k$
Recall that the infimum of a family of upper semicontinuous functions is upper semicontinuous. Thus it suffices to fix $\epsilon > 0$ and show that

$$\limsup_{k\to\infty} \widetilde{d}^\epsilon((\vec{0}, 0), (q^k, \nu^k)) \leqslant \widetilde{d}^\epsilon((\vec{0}, 0), (q, \nu)) \text{ a.s.} \tag{2.31}$$

But for every $k, n \in \mathbb{N}$ we have by Lemma 2.38 and the fact that $q - q^k \in \mathbb{R}^D_{\geqslant 0}$,

$$\frac{d^\epsilon((\vec{0}, 0), (nq^k, n\nu^k))}{n} - \frac{1}{\epsilon}\left(\frac{1}{n}||\lfloor nq \rfloor - \lfloor nq^k \rfloor||_1 + \rho(\nu, \nu^k)\right) \leqslant \frac{d^\epsilon((\vec{0}, 0), (nq^k, n\nu^k))}{n}$$

For a fixed k, taking $n \to \infty$ we get

$$\widetilde{d}^\epsilon((\vec{0}, 0), (q^k, \nu^k)) - \frac{1}{\epsilon}(||q - q^k||_1 + \rho(\nu, \nu^k)) \leqslant \widetilde{d}^\epsilon((\vec{0}, 0), (q, \nu)) \text{ a.s.}$$

(2.31) follows immediately.

Case 3: $(q^k)_i = 0$ whenever $q_i = 0$ for all $k \in \mathbb{N}$
We may assume $q \neq \vec{0}$ (because this scenario was covered by Case 1). In particular, we may assume without loss of generality that $q^k \neq \vec{0} \ \forall k$.

To show

$$\limsup_{k\to\infty} ||(q^k, \nu^k)|| \leqslant ||(q, \nu)||$$

it suffices to prove this when $||(q^k, \nu^k)|| \geqslant 0 \ \forall k$. In particular, this implies that
$(q^k)_i \geqslant 0 \ \forall k \in \mathbb{N}, 1 \leqslant i \leqslant D$ and hence $q_i \geqslant 0 \ \forall 1 \leqslant i \leqslant D$.

Now, the trick is to rescale the q^k by a factor $B_k \geqslant 1$ converging to 1 in k s.t. we land back in Case 2. More precisely, define

$$B_k := \max\left(\max_{1 \leqslant i \leqslant D \text{ s.t. } (q^k)_i \neq 0} \frac{q_i}{(q^k)_i}, \max_{1 \leqslant i \leqslant D \text{ s.t. } q_i \neq 0} \frac{(q^k)_i}{q_i} \right)$$

Note that $B_k \geqslant 0$ because $q_i, (q^k)_i \geqslant 0$. Since $q^k \neq \vec{0}$ then there is some i s.t. $(q^k)_i > 0$ so by the assumption of Case 3, $q_i > 0$. Thus $B_k \geqslant 1$.

Furthermore, $||q - q^k||_1 \to 0$ implies $B_k \to 1$. In particular, $\frac{1}{B_k} q_k \to^{L^1} q, \frac{1}{B_k}\nu^k \Rightarrow \nu$. By the choice of B_k we have for all $1 \leqslant i \leqslant D$, either $q_i = 0$ hence $B_k q_i = 0 = (q^k)_i$ or $q_i \neq 0$ so $B_k q_i \geqslant (q^k)_i$. Thus

$$B_k q - q^k \in \mathbb{R}^D_{\geqslant 0} \Rightarrow q - \frac{1}{B_k} q^k \in \mathbb{R}^D_{\geqslant 0}$$

By Case 2 and positive-homogeneity of grid entropy,

$$\limsup_{k \to \infty} ||(q^k, \nu^k)|| = \limsup_{k \to \infty} \left|\left| \left(\frac{1}{B_k} q^k, \frac{1}{B_k} \nu^k \right) \right|\right| \leqslant ||(q, \nu)||$$

as desired.

(ii) By the triangle inequality for a directed norm with negative sign,

$$||(q^k, \nu^k)|| \geqslant ||(q, \nu)|| + ||(q^k - q, \nu^k - \nu)|| \geqslant ||(q, \nu)|| \ \forall k$$

since $||(q^k - q, \nu^k - \nu)|| \neq -\infty$. Therefore

$$\liminf_{k \to \infty} ||(q^k, \nu^k)|| \geqslant ||(q, \nu)||$$

(iii) Now suppose

$$\left|\left| \left(q - \frac{q - q^k}{C_k}, \nu - \frac{\nu - \nu^k}{C_k} \right) \right|\right| \geqslant 0 \text{ and } C_k \geqslant 0 \ \forall k, C_k \to 0$$

By the reverse triangle inequality and positive-homogeneity,

$$\begin{aligned}(1 - C_k)||(q, \nu)|| &= ||(q^k - (C_k q - (q - q^k)), \nu^k - (C_k \nu - (\nu - \nu^k)))|| \\ &\leqslant ||(q^k, \nu^k)|| - C_k \left|\left| \left(q - \frac{q - q^k}{C_k}, \nu - \frac{\nu - \nu^k}{C_k} \right) \right|\right| \\ &\leqslant ||(q^k, \nu^k)||\end{aligned}$$

But $C_k \to 0$ hence

$$||(q,\nu)|| \leqslant \liminf_{k\to\infty} ||(q^k,\nu^k)||$$

□

In Lemma 2.48 we made a simple observation that is worth mentioning explicitly: $||(q,\nu)||$ is invariant under permutations of the coordinates of q because of the symmetry of the grid lattice $\mathbb{Z}^D$.

A last property we are interested in is the convexity of $\mathcal{R}^q, \mathcal{R}^t$.

Lemma 2.68. *Let $q \in \mathbb{R}^D_{\geqslant 0}, t \geqslant 0$. Then the sets*

$$\mathcal{R}^q = \{(q,\nu) \in \mathbb{R}^D_{\geqslant 0} \times \mathcal{M}_+ : ||(q,\nu)|| \geqslant 0\}, \mathcal{R}^t = \{\nu \in \mathcal{M}_t : ||\nu|| \geqslant 0\}$$

are convex.

Proof. Grid entropy is concave in by positive-homogeneity and by the reverse triangle-inequality it satisfies. □

2.5 APPLICATION: DIRECTED POLYMERS

We now apply our results to give a variational formula for the point-to-point/point-to-hyperplane Gibbs Free Energy in terms of direction-fixed/direction-free grid entropy and last passage time.

2.5.1 *Setup and Known Results*

Recall that the measurable function $\tau : [0,1] \to [0,\infty)$ determines the edge weights
$\tau_e = \tau(U_e)$. We will be talking about the last passage time constant so for the whole of Section 2.5 we impose the condition that $E|\tau(U)| < \infty$ for $U \sim Unif[0,1]$ (or equivalently that $E|\tau_e| < \infty$).

Define the last passage time between points $p, q \in \mathbb{R}^D$ to be

$$T(p,q) = \sup_{\pi:\lfloor p\rfloor\to\lfloor q\rfloor} T(\pi) = \sup_{\pi:\lfloor p\rfloor\to\lfloor q\rfloor} \sum_{e\in\pi} \tau(U_e)$$

where this is $-\infty$ if $q - p \notin \mathbb{R}^D_{\geqslant 0}$. A standard result (see [Maro4] for example) is the existence of a last passage time constant, or in other words a deterministic limiting value for the last passage time in a fixed direction $q \in \mathbb{R}^D_{\geqslant 0}$.

Theorem 2.69. *Let $q \in \mathbb{R}^D_{\geqslant 0}$. Then there is a τ-dependent time constant $\lambda_q(\tau) \in [0,\infty)$ s.t.*

$$\frac{T(\vec{0}, nq)}{n} \to \lambda_q(\tau)$$

λ_q is homogeneous and concave in q.

Remark 2.70. As was the case with grid entropy, the symmetries of the grid and concavity of passage time imply that the time constant is maximized along the direction $\ell = (\frac{1}{D}, \dots, \frac{1}{D})$:

$$\sup_{q \in \mathbb{R}^D, ||q||_1 = 1} \lambda_q = \lambda_\ell$$

Remark 2.71. Thus far the time constant is only explicitly computable when the underlying distribution θ is exponential.

It is useful to view last passage time as a linear functional. For any NE path $\pi : \lfloor np \rfloor \to \lfloor nq \rfloor$, note that

$$T(\pi) = \sum_{e \in \pi} \tau(U_e) = \int_0^1 \tau(u) d(\mu_\pi) := \langle \tau, \mu_\pi \rangle \text{ so } \frac{T(\pi)}{n} = \left\langle \tau, \frac{1}{n}\mu_\pi \right\rangle$$

where $\langle f, \nu \rangle$ denotes the linear functional that is integration of a measurable function
$f : [0,1] \to \mathbb{R}$ against a measure $\nu \in \mathcal{M}_+$. In this language, weak convergence $\nu_k \Rightarrow \nu$ is equivalent to

$$\lim_{k \to \infty} \langle f, \nu_k \rangle = \langle f, \nu \rangle \ \forall \text{ bounded, continuous functions } f$$

Of course, τ is likely not bounded and continuous. However recalling [Bat20, Lemma 6.15] , we see that there is an a.s. event on which $\frac{1}{n}\mu_{\pi_n}$ converging weakly to ν implies $(\tau)_*(\frac{1}{n}\mu_{\pi_n})$ converges weakly to $(\tau)_*(\nu)$ hence

$$\left\langle \tau, \frac{1}{n}\mu_{\pi_n} \right\rangle = \left\langle Id, (\tau)_* \left(\frac{1}{n}\mu_{\pi_n} \right) \right\rangle \to \left\langle Id, (\tau)_*(\nu) \right\rangle = \langle \tau, \nu \rangle$$

where $Id(x) = x$ is the identity function on [0,1].

By using the compactness of $\mathcal{R}^q$ and Theorem 2.13, Bates establishes a variational formula for the direction-fixed limit shape. See [Bat20, Theorem 2.1] for details. We restate this theorem here in our LPP setting (part (i)), and give an analogous variational formula for direction-free case. Bates' proof can be tweaked easily to prove this. We will give our own proofs in the next section, as this will be a direct corollary of our variational formula for the Gibbs Free Energy.

Theorem 2.72. *Let $q \in \mathbb{R}^D_{\geqslant 0}$. Recall that Theorem 2.13 gives a deterministic weakly closed set $\mathcal{R}^q$ which a.s. consists of all weak limits of subsequences of μ_π*

for paths π of increasing length. Then for any measurable $\tau : [0,1] \to [0,\infty)$, the corresponding limit shape is given by

$$\lambda_q(\tau) = \sup_{\nu \in \mathcal{R}^q} \langle \tau, \nu \rangle$$

Remark 2.73. Since $\mathcal{R}^q$ is weakly closed and nonempty (because $||q||_1 \Lambda \in \mathcal{R}^q$ from before), then the set of maximizers is nonempty:

$$\mathcal{R}^q_\tau := \{\nu \in \mathcal{R}^q : \langle \tau, \nu \rangle = \lambda_q(\tau)\} \neq \varnothing$$

As Bates remarks, when $\mathcal{R}^q_\tau$ is a singleton, empirical measures along geodesics converge weakly to this unique maximizer (or analogously unique minimizer in FPP), answering in the affirmative Hoffman's question. It is further pointed out that $\mathcal{R}^q_\tau$ is a singleton for a dense family of functions τ. See [Bat20] for more details.

Finally, let us recall the directed polymer model, still restricting ourselves to the LPP on $\mathbb{R}^D$ setting.

Definition 2.74. Fix a direction $q \in \mathbb{R}^D_{\geqslant 0}$ and an inverse temperature $\beta > 0$ such that

$$E[e^{\beta \tau(U)}] < \infty \text{ for } U \sim Unif[0,1]$$

The point-to-point β-partition function in direction q is defined to be

$$Z^\beta_{n,q} = \sum_{\pi \in \mathcal{P}(\vec{0}, \lfloor nq \rfloor)} e^{\beta T(\pi)}$$

The corresponding point-to-point β-polymer measure on the set of paths $\pi : \vec{0} \to \lfloor nq \rfloor$ is

$$\rho^\beta_{n,q}(d\pi) = \frac{1}{Z^\beta_{n,q}} e^{\beta T(\pi)}$$

and the corresponding point-to-point β- Gibbs Free Energy is defined to be

$$G^\beta_q = \lim_{n \to \infty} \frac{1}{n} \log Z^\beta_{n,q}$$

Suppressing the direction q we get point-to-hyperplane analogues (to the diagonal hyperplane $x_1 + \ldots + x_D = n$):

$$Z^\beta_n = \sum_{\pi \in \mathcal{P}_n(\vec{0})} e^{\beta T(\pi)}, \ \rho^\beta_n(d\pi) = \frac{1}{Z^\beta_n} e^{\beta T(\pi)}, \ G^\beta = \lim_{n \to \infty} \frac{1}{n} \log Z^\beta_n$$

Finally, the analogous definitions with $Z^{\beta}_{n,q}$ multiplied by $\frac{1}{\#\pi:\vec{0}\to\lfloor nq\rfloor} = \frac{1}{|\mathcal{P}(\vec{0},\lfloor nq\rfloor)|}$ or Z^{β}_{n} multiplied by $\frac{1}{D^n} = \frac{1}{|\mathcal{P}_n(\vec{0})|}$ yield quenched quantities, which we denote here by appending "Q" (as in $QZ^{\beta}_{n,q}$).

Remark 2.75. Recalling that $\#\pi : \vec{0} \to \lfloor nq\rfloor = e^{H(q)n+o(n)}$ and (#length n π from $\vec{0}$) $= D^n = e^{nH(\ell)}$ we get

$$QG^{\beta}_{q} = -H(q) + G^{\beta}_{q}, QG^{\beta} = -H(\ell) + G^{\beta}$$

Remark 2.76. All these quantities are implicitly functions of τ, the measurable $[0,1] \to [0,\infty)$ function that sets the environment. There is no other condition on τ other than measurability.

2.5.2 *Variational Formulas for PTP/PTH Gibbs Free Energies*

We derive a formula for the point-to-point Gibbs Free Energy as the supremum over measures of sum of the integration $\beta\langle\tau,\nu\rangle$ and grid entropy $||(q,\nu)||$. Since direction-free grid entropy is just grid entropy in the ℓ direction, then we will see that point-to-hyperplane Gibbs Free Energy is just the point-to-point Gibbs Free Energy in direction ℓ.

Theorem 2.77. *Fix an inverse temperature $\beta > 0$ s.t.*

$$E[e^{\beta\tau(U)}] < \infty \text{ for } U \sim Unif[0,1]$$

(i) Fix $q \in \mathbb{R}^{D}_{\geqslant 0}$. The point-to-point Gibbs Free Energy is given by

$$G^{\beta}_{q} = \sup_{\nu\in\mathcal{M}_+} [\beta\langle\tau,\nu\rangle + ||(q,\nu)||] \text{ a.s.}$$

Moreover, this supremum is achieved by some $\nu \in \mathcal{R}^q$.
(ii) The point-to-hyperplane Gibbs Free Energy is given by

$$G^{\beta} = \sup_{\nu\in\mathcal{M}_1} [\beta\langle\tau,\nu\rangle + ||\nu||] = G^{\beta}_{\ell} = \sup_{q\in\mathbb{R}^{D}_{\geqslant 0}, ||q||_1=1} G^{\beta}_{q} \text{ a.s.}$$

Remark 2.78. The fact that direction-free Gibbs Free Energy G^{β} is the supremum over $q \in \mathbb{R}^{D}_{\geqslant 0}$, $||q||_1 = 1$ of direction-fixed Gibbs Free Energies G^{β}_{q} was known previously (e.g. see [GRAS16, Theorem 2.4]). [GRAS16] also showcases some "cocycle" and "entropy" variational formulas for the Gibbs Free Energy. What is new in our paper is that G^{β}_{ℓ} achieves this supremum, and also that we have these grid entropy variational formulas.

Remark 2.79. As an immediate consequence, we get a variational formula for the quenched point-to-point/point-to-hyperplane Gibbs free energy:

$$QG_q^\beta = \sup_{\nu \in \mathcal{M}_+} [\beta\langle \tau, \nu\rangle + ||(q,\nu)||] - H(q) \text{ a.s.}$$

$$QG^\beta = \sup_{\nu \in \mathcal{M}_1} [\beta\langle \tau, \nu\rangle + ||\nu||] - H(\ell) = QG_\ell^\beta \text{ a.s.}$$

Remark 2.80. If τ is non-negative then the variational formulas are trivially positive-homogeneous.

We require a short lemma in order to prove our variational formula. The main technical difficulty will be that each direction-q grid entropy is an almost sure limit:

$$||(q,\nu)|| = \inf_{\epsilon > 0} \lim_{n\to\infty} \frac{d^\epsilon((\vec{0},0),(nq,n\nu))}{n} \text{ a.s.}$$

but we are now dealing with a whole space of measures, which is uncountable. Thus we will need to suppose these limits exist for a countable dense family of ν's and approximate the sums of $e^{\beta T(\pi)}$ over paths with empirical measure in ϵ-balls.

Lemma 2.81. *Fix $q \in \mathbb{R}^D_{\geqslant 0}, \chi \in \mathcal{M}_+, \gamma > 0$ and $m \in \mathbb{N}$. For any $n \in \mathbb{N}$,*

$$\log\Bigg(\sum_{\pi:\vec{0}\to\lfloor nq\rfloor \ s.t.\ \frac{1}{n}\mu_\pi \in B_{\frac{1}{m}}(\chi)} e^{\beta T(\pi)}\Bigg) \leqslant \beta n \Bigg(\sup_{\xi \in B_{\frac{1}{m}}(\chi)} \langle \tau, \xi\rangle\Bigg) + n\frac{1/m}{\gamma} + d^\gamma((\vec{0},0),(nq,n\chi))$$

Proof. Let us look closer at this sum. Observe that we can upper bound the indicator of the set we are summing over by a smooth function:

$$\mathbb{1}\left\{\frac{1}{n}\mu_\pi \in B_{\frac{1}{m}}(\chi)\right\} \leqslant \exp\left(n\frac{1/m}{\gamma}\right)\exp\left(-\frac{n}{\gamma}\rho\left(\frac{1}{n}\mu_\pi, \chi\right)\right)$$

Thus

$$
\begin{aligned}
&\log\Bigg(\sum_{\pi \text{ s.t. } \frac{1}{n}\mu_\pi \in B_{\frac{1}{m}}(\chi)} e^{\beta T(\pi)}\Bigg) \\
&\leqslant \beta\Bigg(\sup_{\pi \text{ s.t. } \frac{1}{n}\mu_\pi \in B_{\frac{1}{m}}(\chi)} T(\pi)\Bigg) + \log\Bigg(\sum_{\pi \text{ s.t. } \frac{1}{n}\mu_\pi \in B_{\frac{1}{m}}(\chi)} 1\Bigg) \\
&\leqslant \beta n\Bigg(\sup_{\pi \text{ s.t. } \frac{1}{n}\mu_\pi \in B_{\frac{1}{m}}(\chi)} \langle \tau, \frac{1}{n}\mu_\pi\rangle\Bigg) + \log\Bigg[\exp\Big(n\frac{1/m}{\gamma}\Big)\sum_{\pi\in\mathcal{P}(\vec{0},\lfloor nq\rfloor)} \exp\Big(-\frac{n}{\gamma}\rho\Big(\frac{1}{n}\mu_\pi,\chi\Big)\Big)\Bigg] \\
&\leqslant \beta n\Bigg(\sup_{\xi\in B_{\frac{1}{m}}(\chi)} \langle \tau, \xi\rangle\Bigg) + n\frac{1/m}{\gamma} + d^\gamma((\vec{0},0),(nq,n\chi))
\end{aligned}
$$

□

Proof of Theorem 2.77. We will focus on showing (i). The proof in the direction-free case (ii) is analogous.

(i) Our proof consists mainly of compactness arguments, where the space of measures we care about is weakly compact.

Also, note that if $q = \vec{0}$ then $Z^\beta_{n,q} = 1$ hence the variational formula holds trivially. Thus we may assume $||q||_1 > 0$.

Before beginning the proof, recall that the Levy-Prokhorov metric metrizes weak convergence of measures and that $\{\xi \in \mathcal{M}_+ : \frac{1}{2}||q||_1 \leqslant ||\xi||_{TV} \leqslant \frac{3}{2}||q||_1\} := S$ is weakly compact. It follows that $\beta\langle\tau,\cdot\rangle$ is uniformly continuous on S. More precisely,

$$\forall\epsilon > 0\ \exists\delta > 0 \text{ s.t. } \forall\xi,\xi' \in S,\ \rho(\xi,\xi') < \delta \Rightarrow |\beta\langle\tau,\xi\rangle - \beta\langle\tau,\xi'\rangle| < \epsilon \quad (2.32)$$

We split the argument into two claims.

Claim 1:

$$\lim_{n\to\infty}\frac{1}{n}\log Z^\beta_{n,q} \geqslant \sup_{\nu\in\mathcal{M}_+}[\beta\langle\tau,\nu\rangle + ||(q,\nu)||] \text{ a.s.}$$

Proof of Claim 1

Note that $||(q,\nu)|| = -\infty$ unless $\nu \in \mathcal{R}^q$. Also $\mathcal{R}^q$ is weakly closed and $||(q,\nu)||$ is upper semicontinuous in ν so there exists some $\nu \in \mathcal{R}^q$ that achieves the supremum

$$\sup_{\nu\in\mathcal{M}_+}[\beta\langle\tau,\nu\rangle + ||(q,\nu)||]$$

We will prove the claim when $||(q,\nu)|| > 0$; the case when $||(q,\nu)|| = 0$ is very similar.

For Claim 1, we restrict ourselves to the measure 1 event

$$\left\{ \lim_{n\to\infty} \min_{\pi}^{\lfloor e^{n(||(q,\nu)||-\epsilon)}\rfloor} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0 \;\forall \epsilon \in \mathbb{Q}_+ \text{ with } \epsilon < ||(q,\nu)|| \right\} \tag{2.33}$$

(this has measure 1 by Theorem 2.53). We wish to show that in this event,

$$\liminf_{n\to\infty} \frac{1}{n} \log Z^\beta_{n,q} \geqslant \beta\langle\tau,\nu\rangle + ||(q,\nu)|| - 2\epsilon \text{ for } \epsilon > 0 \text{ arbitrarily small}$$

Fix $0 < \epsilon < ||(q,\nu)||$ in $\mathbb{Q}_+$ and let $\delta > 0$ satisfy (2.32). We are in the event (2.33) so

$$\lim_{n\to\infty} \min_{\pi}^{\lfloor e^{n(||(q,\nu)||-\epsilon)}\rfloor} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0$$

Denote by π_n the (event-depedent) paths corresponding to these minimums. For large enough n we have $\rho(\frac{1}{n}\mu_{\pi_n}, \nu) < \delta$ hence there are $\lfloor e^{n(||(q,\nu)||-\epsilon)}\rfloor$ paths $\pi : \vec{0} \to \lfloor nq \rfloor$ satisfying

$$\rho\left(\frac{1}{n}\mu_\pi, \nu\right) \leqslant \rho\left(\frac{1}{n}\mu_{\pi_n}, \nu\right) < \delta \text{ so } \left|\beta\langle\tau, \frac{1}{n}\mu_\pi\rangle - \beta\langle\tau,\nu\rangle\right| < \epsilon \tag{2.34}$$

by (2.32). But then

$$Z^\beta_{n,q} \geqslant \sum_{\text{the } \lfloor e^{n(||(q,\nu)||-\epsilon)}\rfloor \text{ paths } \pi \text{ from above}} e^{\beta n\langle\tau,\frac{1}{n}\mu_\pi\rangle} \geqslant \lfloor e^{n(||(q,\nu)||-\epsilon)}\rfloor e^{n(\beta\langle\tau,\nu\rangle-\epsilon)}$$

by (2.34). It follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log Z^\beta_{n,q} \geqslant \beta\langle\tau,\nu\rangle - \epsilon + ||(q,\nu)|| - \epsilon$$

Since $\epsilon > 0$ was arbitrarily small, then taking $\epsilon \to 0^+$ we get

$$\liminf_{n\to\infty} \frac{1}{n} \log Z^\beta_{n,q} \geqslant \sup_{\nu\in\mathcal{R}^q} [\beta\langle\tau,\nu\rangle + ||(q,\nu)||] \text{ a.s.} \tag{2.35}$$

as desired. □[Claim 1]

For the second half, we will actually show a slightly stronger claim, as this will be useful in a later proof.

Claim 2: Let $W \subseteq \mathcal{M}_+$ be a weakly open, possibly empty set s.t. $\mathcal{R}^q \backslash W \neq \varnothing$. Define

$$Y^\beta_{n,q} := \sum_{\pi:\vec{0}\to\lfloor nq\rfloor \text{ s.t. } \frac{1}{n}\mu_\pi \notin W} e^{\beta T(\pi)} = \sum_{\pi:\vec{0}\to\lfloor nq\rfloor \text{ s.t. } \frac{1}{n}\mu_\pi \notin W} e^{\beta n\langle\tau,\frac{1}{n}\mu_\pi\rangle}$$

Then

$$\limsup_{n\to\infty} \frac{1}{n} \log Y^{\beta}_{n,q} \leqslant \sup_{\nu\in\mathcal{M}_+\setminus W} [\beta\langle\tau,\nu\rangle + ||(q,\nu)||] \text{ a.s.} \tag{2.36}$$

Of course, in our case $W = \varnothing$ and $Y^{\beta}_{n,q} = Z^{\beta}_{n,q}$.

Proof of Claim 2

We will be extra careful by specifying only countably many measure zero events we exclude. We do this by considering a countable dense family of target measures.

$S\setminus W$ is weakly compact. So for all $m \in \mathbb{N}$ there exists $M = M(m) \in \mathbb{N}$ and a finite $\frac{1}{m}$-net $B_{\frac{1}{m}}(\nu^m_1), \ldots, B_{\frac{1}{m}}(\nu^m_M)$ covering $S\setminus W$ with the property that $\nu^m_i \notin W$ (but the balls themselves may intersect W). From now on we restrict ourselves to the event

$$\left\{ \lim_{n\to\infty} \frac{d^{\frac{1}{\sqrt{m}}}((\vec{0},0),(nq,n\nu^m_i))}{n} = \tilde{d}^{\frac{1}{\sqrt{m}}}((\vec{0},0),(q,\nu^m_i)) \;\forall m \in \mathbb{N}, 1 \leqslant i \leqslant M(m) \right\} \tag{2.37}$$

which is a countable intersection of measure 1 events by Theorem 2.34 hence has measure 1. It is important to note that for any arbitrary measure $\xi \in \mathcal{M}_+$ we still have

$$||(q,\xi)|| = \inf_{\epsilon>0} \tilde{d}^{\epsilon}((\vec{0},0),(q,\xi))$$

because this is a non-probabilistic statement about constants; we do not however assume that these directed metrics $\tilde{d}^{\epsilon}((\vec{0},0),(q,\xi))$ are limits of the $\frac{d^{\epsilon}((\vec{0},0),(nq,n\xi))}{n}$.

Now we wish to fix an arbitrary $\epsilon > 0$ and prove that in the measure 1 event (2.37),

$$L := \limsup_{n\to\infty} \frac{1}{n} \log Y^{\beta}_{n,q} \leqslant \sup_{\nu\in\mathcal{M}_+\setminus W} [\beta\langle\tau,\nu\rangle + ||(q,\nu)||] + 5\epsilon$$

It suffices to consider a convergent subsequence $\frac{1}{n_k} \log Y^{\beta}_{n_k,q} \to L$ and show $L \leqslant$ the right hand side.

Fix any $m \in \mathbb{N}$. For large k,

$$\frac{|||n_k q]||_1}{n_k} \in \left[\frac{1}{2}||q||_1, \frac{3}{2}||q||_1\right]$$

so any path $\pi : \vec{0} \to \lfloor n_k q \rfloor$ we are summing over in $Y^\beta_{n_k,q}$ will have $\frac{1}{n_k}\mu_\pi \in S\backslash W$ hence $\frac{1}{n_k}\mu_\pi$ will be in one of the $B_{\frac{1}{m}}(\nu^m_i)$. Therefore

$$Y^\beta_{n_k,q} \leqslant \sum_{i=1}^{M(m)} \sum_{\pi \text{ s.t. } \frac{1}{n_k}\mu_\pi \in B_{\frac{1}{m}}(\nu^m_i)} e^{\beta T(\pi)} \leqslant M(m) \max_{i=1}^{M(m)} \left(\sum_{\pi \text{ s.t. } \frac{1}{n_k}\mu_\pi \in B_{\frac{1}{m}}(\nu^m_i)} e^{\beta T(\pi)} \right)$$

Note that $M(m)$ is a constant with respect to k, whereas the max above is both event-dependent and k-dependent. Thus there is some event-dependent $1 \leqslant I(m) \leqslant M(m)$ and some subsequence n_{k_j} s.t. the above maximum is achieved by the *same* $i = I(m)$ for every $n_k = n_{k_j}$. (Finite pigeonholes, infinite pigeons if you will.) Therefore

$$L = \limsup_{j\to\infty} \frac{1}{n_{k_j}} \log Y^\beta_{n_{k_j},q} \leqslant \limsup_{j\to\infty} \frac{1}{n_{k_j}} \log \left(\sum_{\pi \text{ s.t. } \frac{1}{n_{k_j}}\mu_\pi \in B_{\frac{1}{m}}(\nu^m_{I(m)})} e^{\beta T(\pi)} \right) \tag{2.38}$$

This holds for *any* m. Our strategy will be to choose a very large, event-dependent m that gives a nice event-*independent* upper bound.

Let us upper bound the expression in the limit on the right-hand side of (2.38). Lemma 2.81 with $\chi = \nu^m_{I(m)}, \gamma = \frac{1}{\sqrt{m}}$ establishes that for any $j \in \mathbb{N}$,

$$\begin{aligned} &\log \left(\sum_{\pi:\vec{0}\to\lfloor n_{k_j} q\rfloor \text{ s.t. } \frac{1}{n_{k_j}}\mu_\pi \in B_{\frac{1}{m}}(\nu^m_{I(m)})} e^{\beta T(\pi)} \right) \\ &\leqslant \beta n_{k_j} \left(\sup_{\xi \in B_{\frac{1}{m}}(\nu^m_{I(m)})} \langle \tau, \xi \rangle \right) + n_{k_j} \frac{1/m}{1/\sqrt{m}} + d^{\frac{1}{\sqrt{m}}}((\vec{0},0),(n_{k_j}q, n_{k_j}\nu^m_{I(m)}))) \end{aligned} \tag{2.39}$$

This holds for any arbitrary $m \in \mathbb{N}$. The next step is to specify a sufficiently large m.

By compactness of $S\backslash W$, the measures $\nu^m_{I(m)}$ have a weakly convergent subsequence
$\nu^{m_l}_{I(m_l)} \Rightarrow \nu' \in S\backslash W$.

Fix some $l_0 \in \mathbb{N}$ large enough so that

$$\frac{1}{m_{l_0}} + \frac{\epsilon}{\sqrt{m_{l_0}}} < \delta,\ \frac{1}{\sqrt{m_{l_0}}} < \epsilon,\ \tilde{d}^{\frac{1}{\sqrt{m_{l_0}}}}((\vec{0},0),(q,\nu')) < ||(q,\nu')|| + \epsilon \tag{2.40}$$

Consider some event-dependent $l \geqslant l_0$ large enough s.t.

$$\rho(\nu^{m_l}_{I(m_l)}, \nu') < \frac{\epsilon}{\sqrt{m_{l_0}}} \tag{2.41}$$

Since $\xi \in \overline{B_{\frac{1}{m_l}}(\nu^{m_l}_{I(m_l)})}$ implies

$$\rho(\xi,\nu') \leqslant \rho(\xi,\nu^{m_l}_{I(m_l)}) + \rho(\nu^{m_l}_{I(m_l)},\nu') \leqslant \frac{1}{m_l} + \frac{\epsilon}{\sqrt{m_{l_0}}} \leqslant \frac{1}{m_{l_0}} + \frac{\epsilon}{\sqrt{m_{l_0}}} < \delta$$

and δ satisfies (2.32) then

$$\sup_{\xi \in B_{\frac{1}{m_l}}(\nu^{m_l}_{I(m_l)})} \beta\langle\tau,\xi\rangle \leqslant \sup_{\xi \in \overline{B_{\frac{1}{m_l}}(\nu^{m_l}_{I(m_l)})}} \beta\langle\tau,\xi\rangle < \beta\langle\tau,\nu'\rangle + \epsilon \tag{2.42}$$

Next, we have

$$\frac{1}{\sqrt{m_l}} \leqslant \frac{1}{\sqrt{m_{l_0}}} < \epsilon \tag{2.43}$$

Finally, we are in the event (2.37) so for large enough j we have

$$\begin{aligned} & d^{\frac{1}{\sqrt{m_l}}}((\vec{0},0),(n_{k_j}q, n_{k_j}\nu^{m_l}_{I(m_l)})) \\ & < n_{k_j}\tilde{d}^{\frac{1}{\sqrt{m_l}}}((\vec{0},0),(q,\nu^{m_l}_{I(m_l)})) + n_{k_j}\epsilon \\ & \leqslant n_{k_j}\tilde{d}^{\frac{1}{\sqrt{m_{l_0}}}}((\vec{0},0),(q,\nu^{m_l}_{I(m_l)})) + n_{k_j}\epsilon \\ & \leqslant n_{k_j}\tilde{d}^{\frac{1}{\sqrt{m_{l_0}}}}((\vec{0},0),(q,\nu')) + \frac{n_{k_j}}{1/\sqrt{m_{l_0}}}\rho(\nu^{m_l}_{I(m_l)},\nu') + n_{k_j}\epsilon \end{aligned}$$

by Lemma 2.38

$$< n_{k_j}||(q,\nu')|| + 3n_{k_j}\epsilon \tag{2.44}$$

by (2.40),(2.41). Combining the inequalities (2.42),(2.43),(2.44) with (2.39) we get that for large enough j,

$$\log\left(\sum_{\pi:\vec{0}\to\lfloor n_{k_j}q\rfloor \text{ s.t. } \frac{1}{n_{k_j}}\mu_\pi \in B_{\frac{1}{m_l}}(\nu^{m_l}_{I(m_l)})} e^{\beta T(\pi)}\right) \leqslant n_{k_j}\left(\beta\langle\tau,\nu'\rangle + ||(q,\nu')|| + 5\epsilon\right) \tag{2.45}$$

Substituting this in (2.38), we get

$$L \leqslant \beta\langle\tau,\nu'\rangle + ||(q,\nu')|| + 5\epsilon \leqslant \sup_{\nu\in\mathcal{M}_+\setminus W} \beta\langle\tau,\nu\rangle + ||(q,\nu)|| + 5\epsilon$$

as desired. Note that even though m_l and ν' in (2.45) are event-dependent, the upper bound we get for L is *not* event-dependent, makes our argument work. □[Claim 2]

This finishes the proof of the variational formula in the fixed-direction case.

(ii) An analogous argument gives

$$G^\beta = \sup_{\nu \in \mathcal{M}_1} [\beta\langle \tau, \nu\rangle + ||\nu||]$$

But direction-free grid entropy is just grid entropy in the direction ℓ. Furthermore,
$||(\ell,\nu)|| = -\infty$ if $\nu \notin \mathcal{M}_1$. Thus

$$\begin{aligned} G^\beta &= \sup_{\nu \in \mathcal{M}_1} [\beta\langle \tau, \nu\rangle + ||\nu||] \\ &= \sup_{\nu \in \mathcal{M}_+} [\beta\langle \tau, \nu\rangle + ||(\ell,\nu)||] \\ &= G^\beta_\ell \\ &= \sup_{q \in \mathbb{R}^D_{\geqslant 0}, ||q||_1 = 1} \sup_{\nu \in \mathcal{M}_+} [\beta\langle \tau, \nu\rangle + ||(q,\nu)||] \text{ by Lemma 2.48} \\ &= \sup_{q \in \mathbb{R}^D_{\geqslant 0}, ||q||_1 = 1} G^\beta_q \end{aligned}$$

□

As an immediate corollary, we get another proof of Bates' Variational Formula (Theorem 2.72 (i)), as well as its direction-free analogue Theorem 2.72 (ii). We stress that this proof works only under the assumption that the exponential moment $E[e^{\beta\tau(U)}]$ is finite for $U \sim Unif[0,1]$.

Proof of Theorem 2.72. Fix $q \in \mathbb{R}^D_{\geqslant 0}$ and suppose for every $\alpha \geqslant 0$ we have

$$E[e^{\beta\tau(U)}] < \infty \text{ for } U \sim Unif[0,1]$$

It is standard (e.g. see [Mar04]) that for every n,

$$\lim_{\beta\to\infty} \beta^{-1} \log Z^\beta_{n,q} = Z_{n,q} = \sup_{\beta > 0} \beta^{-1} \log Z^\beta_{n,q} = T(\vec{0}, nq) := \log Z^\infty_{n,q} \text{ a.s.} \tag{2.46}$$

which we recall is the last passage time between $\vec{0}$ and $\lfloor nq \rfloor$. That is, the last passage time is the zero temperature analogue of $\log Z^\beta_{n,q}$. Thus the the zero temperature point-to-point Gibbs Free Energy is the last passage time constant λ_q:

$$G^\infty_q := \lim_{n\to\infty} n^{-1} \log Z^\infty_{n,q} = \lim_{n\to\infty} n^{-1} T(\vec{0}, nq) = \lambda_q \text{ a.s.}$$

On the other hand, another standard result (see [GRAS16]) says that the zero temperature point-to-point Gibbs Free Energy is a scaled limit of the positive temperature free energies:

$$\lim_{\beta\to\infty} \beta^{-1} G_q^\beta = \sup_{\beta>0} \beta^{-1} G_q^\beta = G_q^\infty = \lambda_q \text{ a.s} \tag{2.47}$$

Fix $N \in \mathbb{N}$. Dividing our variational formula by β and taking the supremum over $\beta \geqslant N$ we get by (2.47)

$$\begin{aligned}
\lambda_q &= \sup_{\beta\geqslant N} \beta^{-1} G_q^\beta \\
&= \sup_{\beta\geqslant N} \sup_{\nu\in\mathcal{R}^q} \left(\langle \tau, \nu\rangle + \beta^{-1}||(q,\nu)||\right) \\
&= \sup_{\nu\in\mathcal{R}^q} \sup_{\beta\geqslant N} \left(\langle \tau, \nu\rangle + \beta^{-1}||(q,\nu)||\right) \\
&= \sup_{\nu\in\mathcal{R}^q} \left(\langle \tau, \nu\rangle + N^{-1}||(q,\nu)||\right)
\end{aligned}$$

Taking $N \to \infty$ and noting that $\nu \in \mathcal{R}^q$ implies $|(q,\nu)|| \in [0, H(q)]$ is bounded, we get Bates' Variational formula:

$$\lambda_q = \sup_{\nu\in\mathcal{R}^q} \langle \tau, \nu\rangle$$

a.s.. □

In other words, Bates' variational formula is nothing more than the zero temperature analogue of our variational formula for point-to-point Gibbs Free Energy.

The other corollary of note is the answer to Hoffman's question in the direct polymer model when our variational formula has a unique maximizer. As Bates notes, this happens for a dense family of measurable functions τ (recall Remark 2.73).

Corollary 2.82. *Fix an inverse temperature $\beta > 0$ s.t.*

$$E[e^{\beta\tau(U)}] < \infty \textit{ for } U \sim Unif[0,1]$$

(i) Fix $q \in \mathbb{R}^D_{\geqslant 0}$ and suppose $\beta\langle \tau, \nu\rangle + ||(q,\nu)||$ has a unique maximizer $\nu \in \mathcal{R}^q$. For every n pick a path $\pi_n : \vec{0} \to \lfloor nq \rfloor$ independently and at random according to the probabilities prescribed the corresponding point-to-point β-polymer measure $\rho^\beta_{n,q}$. Then the empirical measures $\frac{1}{n}\mu_{\pi_n}$ converge weakly to ν a.s.

(ii) Suppose $\beta\langle\tau,\nu\rangle + ||\nu||$ has a unique maximizer $\nu \in \mathcal{M}_1$. For every n pick a length n path π_n from $\vec{0}$ independently and at random according to the probabilities prescribed the corresponding point-to-hyperplane β-polymer measure ρ_n^β. Then the empirical measures $\frac{1}{n}\mu_{\pi_n}$ converge weakly to ν a.s.

Proof. We will focus on showing (i). The proof in the direction-free case (ii) is analogous.

(i) If $q = \vec{0}$ then the paths π_n are the trivial path $\vec{0} \to \vec{0}$ and $\mathcal{R}^{\vec{0}} = \{0\}$ so $\nu = 0$, hence the result holds trivially. Thus we may assume $||q||_1 > 0$.

Since the empirical measures $\frac{1}{n}\mu_{\pi^n}$ eventually lie in the weakly compact set $S = \{\xi \in \mathcal{M}_+ : \frac{1}{2}||q||_1 \leqslant ||\xi||_{TV} \leqslant \frac{3}{2}||q||_1\}$ then by compactness the existence of an accumulation point is guaranteed so it suffices to show that the sequence $\frac{1}{n}\mu_{\pi_n}$ cannot have any accumulation point other than ν.

If $\mathcal{R}^q$ is a singleton, i.e. $\mathcal{R}^q = \{\nu\} = \{||q||_1\Lambda\}$, then by the characterization of $\mathcal{R}^q$ as the set of accumulation points of empirical measures (see Theorem 2.13), ν is the only possible accumulation point of $\frac{1}{n}\mu_{\pi_n}$ so we would be done. We can thus assume $\mathcal{R}^q$ is not a singleton. Thus there exists some $\eta_0 > 0$ s.t. $B_{\eta_0}(\nu)^C \cap \mathcal{R}^q \neq \varnothing$.

Suppose we are in the measure 1 event

$$\left\{ \limsup_{n\to\infty} \frac{1}{n} \log \sum_{\pi:\vec{0}\to\lfloor nq\rfloor \text{ s.t. } \frac{1}{n}\mu_\pi \notin B_\eta(\nu)} e^{\beta T(\pi)} \right.$$
$$\left. \leqslant \sup_{\xi\in\mathcal{M}_+\setminus B_\eta(\nu)} [\beta\langle\tau,\xi\rangle + ||(q,\xi)||] \ \forall \eta \in \mathbb{Q}_+ \cap (0,\eta_0) \right\} \tag{2.48}$$
$$\cap \left\{ \lim_{n\to\infty} \frac{1}{n} \log Z_{n,q}^\beta = \beta\langle\tau,\nu\rangle + ||(q,\nu)|| \right\}$$

This is the intersection of countably many measure 1 events by Claim 2 in the proof of Theorem 2.77. Claim 2 applies because

$$B_\eta(\nu)^C \cap \mathcal{R}^q \supseteq B_{\eta_0}(\nu)^C \cap \mathcal{R}^q \neq \varnothing \ \forall 0 < \eta < \eta_0$$

We need to prove that in the measure 1 event (2.48), the probability that the empirical measures $\frac{1}{n}\mu_{\pi^n}$ of the paths chosen independently at random according to the polymer measures do not have an accumulation point other than ν is 0.

Fix any $0 < \eta < \eta_0$ in $\mathbb{Q}_+$. We wish to show that the probability that $\frac{1}{n}\mu_{\pi_n}$ is not in $B_\eta(\nu)^C$ infinitely often is 0.

Let $\xi_\eta \in \mathcal{M}_+\setminus B_\eta(\nu)$ achieve the supremum

$$\sup_{\xi\in\mathcal{M}_+\setminus B_\eta(\nu)} [\beta\langle\tau,\xi\rangle + ||(q,\xi)||]$$

By the maximality of ν, there is some $\epsilon > 0$ s.t.

$$\beta\langle\tau, \xi_\eta\rangle + ||(q, \xi_\eta)|| < \beta\langle\tau, \nu\rangle + ||(q, \nu)|| - 3\epsilon \tag{2.49}$$

Consider any $n \in \mathbb{N}$. Recall the definition of the point-to-point β-polymer measure $\rho^\beta_{n,q}$ on the set of paths $\pi : \vec{0} \to \lfloor nq \rfloor$:

$$\rho^\beta_{n,q}(d\pi) = \frac{1}{Z^\beta_{n,q}} e^{\beta T(\pi)}$$

Thus the probability that $\frac{1}{n}\mu_{\pi_n} \notin B_\eta(\nu)$ (with respect to the β-polymer measure $\rho^\beta_{n,q}$) is

$$\frac{1}{Z^\beta_{n,q}} \sum_{\pi:\vec{0}\to\lfloor nq\rfloor \text{ s.t. } \frac{1}{n}\mu_\pi \notin B_\eta(\nu)} e^{\beta T(\pi)}$$

We are in the event (2.48), so for large enough n, this is

$$\leqslant \exp\Big(-n\beta\langle\tau, \nu\rangle - n||(q, \nu)|| + n\epsilon\Big)\exp\Big(n\beta\langle\tau, \xi_\eta\rangle + n||(q, \xi_\eta)|| + n\epsilon\Big) \leqslant e^{-n\epsilon}$$

by (2.49). $\sum_{n=1}^{\infty} e^{-n\epsilon} < \infty$ so by the Borel-Cantelli Lemma,

$$P\left(\frac{1}{n}\mu_{\pi_n} \text{ has an accumulation point in } B_{2\eta}(\nu)^C\right)$$
$$\leqslant P\left(\frac{1}{n}\mu_{\pi_n} \in B_\eta(\nu)^C \text{ for infinitely many } n\right) = 0$$

By continuity from below,

$$P\left(\frac{1}{n}\mu_{\pi_n} \text{ has an acc. point other than } \nu\right)$$
$$= \lim_{\eta\to 0^+} P\left(\frac{1}{n}\mu_{\pi_n} \text{ has an acc. point in } B_{2\eta}(\nu)^C\right) = 0$$

Therefore the empirical measures $\frac{1}{n}\mu_{\pi_n}$ converge weakly to ν a.s.. □

2.5.3 *Grid Entropy as The Negative Convex Conjugate of Gibbs Free Energy*

Another way of viewing the variational formulas in Theorem 2.77 is that the point-to-point/point-to-hyperplane Gibbs Free Energies as functions of the non-negative measurable function $\beta\tau$ are the convex conjugates of the functions $\nu \mapsto -||(q, \nu)||, \nu \mapsto -||\nu||$ on $\mathcal{M}_+, \mathcal{M}_1$ respectively.

Let us briefly recall what that means (see [Zal02] for more details).

Definition 2.83. Let X be a locally convex Hausdorff space and let X^* be its dual with respect to an inner product $\langle \cdot, \cdot \rangle$. A convex function $f : X \to [-\infty, \infty]$ is said to be proper if $f > -\infty$ and $f \not\equiv \infty$. For any proper convex function $f : X \to [-\infty, \infty]$, its convex conjugate is the function $f^* : X^* \to$ given by

$$f^*(x^*) = \sup_{x \in X} [\langle x, x^* \rangle - f(x)]$$

In our case, we have $X = \mathcal{M}_+$ (finite Borel measures on [0,1]) and X^* is the set of non-negative measurable functions $\tau : [0,1] \to [0, \infty)$ with inner product given by integration. For convenience, we absorb the β factor into τ.

Also, recall that direction-fixed/direction-free grid entropy is concave hence the maps $\nu \mapsto -||(q,\nu)||, \nu \mapsto -||\nu||$ are convex and trivially proper. Furthermore, by our continuity theorem (Thm 2.64), $-||(q,\cdot)||$ and $-||\nu||$ are lower semicontinuous in ν.

With this setup in mind, it is evident that Theorem 2.77 establishes that the point-to-point Gibbs Free Energy is precisely the convex conjugate of $-||(q,\cdot)||$ and the point-to-hyperplane Gibbs Free Energy is the convex conjugate of $-||\nu||$.

The convex conjugate has various important properties. We focus on those that are the most relevant and interesting in regards to grid entropy and Gibbs Free Energy. See [Zal02, Sect 2.3] for the full list of properties.

Corollary 2.84. *Fix $q \in \mathbb{R}^D_{\geqslant 0}$ and $\beta > 0$ s.t.*

$$E[e^{\beta \tau(U)}] < \infty \text{ for } U \sim Unif[0,1]$$

Then
(i) Gibbs Free Energies $G^\beta_q(\tau), G^\beta(\tau)$ are convex and lower semicontinuous in τ
(ii) (Order-preservation) If $q' \in \mathbb{R}^D_{\geqslant 0}$ with $||(q,\cdot)|| \leqslant ||(q',\cdot)||$ then
$G^\beta_q(\cdot) \leqslant G^\beta_{q'}(\cdot)$
(iii) (Biconjugate Duality) The convex conjugate of Gibbs Free Energy (as a function of τ) is minus grid entropy. That is,

$$(G^\beta_q)^* = -||(q,\cdot)|| \text{ and } (G^\beta)^* = -||\cdot||$$

In other words, grid entropy is equal to its biconjugate.
(iv) (Fenchel's Duality Theorem) Let $g : \mathcal{M}_+ \to [-\infty, \infty], h : \mathcal{M}_1 \to [-\infty, \infty]$

be proper convex functions s.t. $\exists \nu \in \mathcal{R}^q, \nu' \in \mathcal{R}^1$ *with* $g(\nu) < \infty, h(\nu') < \infty$ *and one of* $g, ||(q, \cdot)||$ *is continuous at* ν *and one of* $h, || \cdot ||$ *is continuous at* ν'. *Then*

$$\inf_{\nu \in \mathcal{M}_+} [g(\nu) - ||(q, \nu)||] = \sup_{\tau} [G_q^\beta(\tau) - g^*(\tau)]$$

$$\inf_{\nu \in \mathcal{M}_1} [h(\nu) - ||\nu||] = \sup_{\tau} [G^\beta(\tau) - h^*(\tau)]$$

Remark 2.85. (iii) gives yet another definition of direction-fixed/direction-free grid entropy. This one is perhaps the most remarkable of the three, as it connects it to the seemingly unrelated quantity that is Gibbs Free Energy. This more than justifies the importance and canonical nature of grid entropy.

Proof. (iii) is an immediate consequence of the Fenchel-Moreau Theorem. The rest are consequences of the properties of convex conjugates of proper convex, lower semicontinuous functions. See [Zal02, Sect 2.3] for more details on (i), (ii) and see [VT84, Sect 7.15] for more details on (iv). □

Adding an extra $-H(q)$ term everywhere we get analogous results for quenched point-to-point/point-to-hyperplane Gibbs Free Energy. The convex conjugate of QG_q^β, QG^β will be $H(q) - ||(q, \cdot)||, H(\ell) - || \cdot ||$ respectively which conveniently are always non-negative quantities.

The key takeaway is this: grid entropy and Gibbs Free Energy are intricately connected via convex duality.

2.6 EXTENSIONS TO OTHER MODELS AND OPEN QUESTIONS

Looking back at our development of grid entropy in $\mathbb{R}^D$, we never required some specific property of the lattice model other than that the number of paths $\pi : \vec{0} \to \lfloor nq \rfloor$ we consider is $O(e^{H(q)n})$ for some model-dependent constant

$$H(q) = \limsup_{n \to \infty} \frac{1}{n} \log(\#\pi : \vec{0} \to \lfloor nq \rfloor)$$

that all such π have the same length $O(n)$, and that any pair of paths $\vec{0} \to \lfloor mq \rfloor$ and $\lfloor mq \rfloor \to \lfloor (m+n)q \rfloor$ can be concatenated (thus giving us the superadditivity required to be able to apply the Subadditive Ergodic Theorem). Therefore the notion of grid entropy and all our results, including the variational formula for the Gibbs Free Energy, apply in any edge model with these properties.

In addition, if we loosen some of these conditions, then a part of our work in this chapter still holds. For example, instead of NE paths on $\mathbb{Z}^D$ consider self-avoiding walks (SAWs) on $\mathbb{Z}^D$, as done in [Bat20]. Then we cannot always concatenate a pair of SAWs $\vec{0} \to \lfloor mq \rfloor$ and $\lfloor mq \rfloor \to \lfloor (m+n)q \rfloor$ so our arguments that rely on superadditivity no longer apply. However, we may still define grid entropy by

$$||(q,\nu)|| := \sup \left\{ \alpha \geqslant 0 : \lim_{n\to\infty} \overset{\lfloor e^{\alpha n} \rfloor}{\min_{\pi:\vec{0}\to\lfloor nq \rfloor}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0 \text{ a.s.} \right\}$$

and it will be true that only target measures $\nu << \Lambda$ with total mass $||q||_1$ are observed and that we have an upper bound on the sum of relative entropy and grid entropy:

$$D_{KL}(\nu||\Lambda) + ||(q,\nu)|| \leqslant \chi_D \text{ when } ||q||_1 = 1$$

where χ_D is the connectivity constant for $\mathbb{Z}^D$ satisfying

$$\#(\text{length } n \text{ SAWs from } \vec{0}) = e^{n\chi_D + o(n)}$$

(see [Law13, Section 6.2] for details). It is not clear, however, whether any of the other properties (including this being a directed norm with negative sign) still hold.

Note however that with a small modification we can make this work. All our results hold for SAWs on $\mathbb{Z}^{D+1}$ where the first coordinate, "time," is necessarily non-decreasing because we obtain a superadditivity akin to the one for NE paths.

We end this chapter by mentioning an open question that could be the scope of future work - is grid entropy strictly concave? From our results in this chapter, answering this would immediately imply that Gibbs Free Energy is strictly convex, which is currently a major open question in this field.

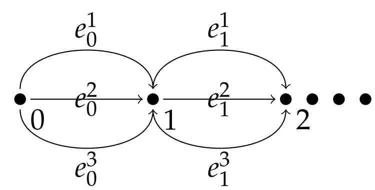
3

GRID ENTROPY IN A "CHOOSE THE BEST OF D SAMPLES" MODEL

3.1 DEFINITIONS AND RESULTS

Let us begin with our setup and goals. We consider vertices on $\mathbb{Z}_{\geqslant 0}$, D parallel edges $(e_i^1, \dots, e_i^D)$ for every $i \geqslant 0$, and an i.i.d. array of edge labels $U_i^j \sim$ Unif[0,1]. We denote by Λ the Lebesgue measure on $[0,1]$. In this setting there is exactly one unit direction, so it makes sense that we will omit mentioning the direction and use the direction-free versions of the notions and results from Chapter 2.

As before, it is convenient to work on the compact space $\mathcal{M}_1$ of probability measures on [0,1], and we may do this without losing any generality.



Next, we wish to formalize the notion of strategies. These will be probability measures on the product space coupling the environment $(U_i^j)_{i \geqslant 0, 1 \leqslant j \leqslant D}$ with the infinite target sequence of indices of the U_i corresponding to the choices $(X_i)_{i \geqslant 0}$.

Definition 3.1. A strategy is a probability measure χ on the product space $([0,1]^D)^{\mathbb{Z}_{\geqslant 0}} \times \{1, \dots, D\}^{\mathbb{Z}_{\geqslant 0}}$ s.t. the marginal distribution of the first coordinate (the environment $(U_i^j) \in ([0,1]^D)^{\mathbb{Z}_{\geqslant 0}}$) is a sequence of i.i.d. Unif[0,1].

We denote by $(J_i)_{i \geqslant 0}$ the second coordinate (a random sequence of indices) and define the random vector of choices

$$(X_0(\chi), X_1(\chi), \dots) := (U_0^{J_0}, U_1^{J_1}, \dots)$$

Denote by $\frac{1}{n}\mu_{0\to n}(\chi)$ the empirical measures of this vector:

$$\frac{1}{n}\mu_{0\to n}(\chi) = \frac{1}{n}\sum_{i=0}^{n-1}\delta_{X_i(\chi)} \in \mathcal{M}_1$$

Also, let σ_χ^i be the law of $X_i(\chi)$.

Remark 3.2. If χ conditioned on the environment is a delta mass for a.a. environments (U_i^j) with respect to the product measure $(\Lambda^{\times D})_\infty$, it means that the strategy picks exactly one sequence $(x_i)_{i\geqslant 0}$ for each set of observed labels. In other words, the strategy is deterministic.

Definition 3.3. From Chapter 2, we recall the definition of (strategy-free) empirical measures along a fixed path $\pi : 0 \to n$ consisting of edges $(e_0^{j_0}, \dots, e_{n-1}^{j_{n-1}})$ by

$$\frac{1}{n}\mu_\pi = \frac{1}{n}\sum_{i=0}^{n-1}\delta_{U_i^{j_i}}$$

An important type of strategy is one which chooses each X_k independent of $(U_i^j)_{i\neq k, 1\leqslant j\leqslant D}$, the observed values from all but the kth trial. These are χ which are infinite products of measures arising from "single-edge strategies," i.e. micro-strategies operating at the individual trial level. We call such χ "independent strategies," as they give rise to independent X_i.

Definition 3.4. A single-edge strategy is a probability measure ψ on the product space $([0,1]^D) \times \{1, \dots, D\}$ s.t. the marginal distribution of the first coordinate $(U^j)_{1\leqslant j\leqslant D}$ is a sequence of D i.i.d. Unif[0,1]. We denote by J the second coordinate (a random index) and define the random choice

$$X(\psi) := U^J$$

Also, let σ_ψ be the law of $X(\psi)$.

Remark 3.5. If ψ conditioned on $(U^j)_{1\leqslant j\leqslant D}$ is a delta mass for a.a. $(U^j)_{1\leqslant j\leqslant D}$ with respect to the product measure $\Lambda^{\times D}$, it means that the single-edge strategy picks exactly one sequence x for each set of D observed labels, so ψ is deterministic.

A single-edge strategy ψ is completely determined by the ψ-conditional expectations
$p_k : [0,1]^D \to [0,1], 1 \leqslant k \leqslant D$ defined as

$$p_k(u_1, \dots, u_D) := E_\psi[(U^j)_{1\leqslant j\leqslant D} = (u^j)_{1\leqslant j\leqslant D} | J = k]$$

We expand on this in more detail in Section 3.4.1. The key takeaway will be that we may interchangeably refer to both ψ and $\vec{p}$ as a single-edge strategy, and therefore we define $X(\vec{p}) := X(\psi), \sigma_{\vec{p}} := \sigma_\psi$.

In this chapter we are interested in the weak limit points of $\frac{1}{n}\mu_{0\to n}(\chi)$. As it turns out, this set of limit points almost surely coincides with the set of limit points of $\frac{1}{n}\mu_{0\to n}(\chi)$ over independent strategies χ only, with the distributions $\sigma_{\vec{p}}$ over single-edge strategies $\vec{p}$, as well as with the set $\mathcal{R}$ of probability measures with finite grid entropy. The following theorem will be the main objective in Section 3.3 .

Theorem E. A.s. we have

$$\begin{aligned}\mathcal{R} &= \left\{\text{limit pts of } \frac{1}{n}\mu_{0\to n}(\chi) : \text{strategies } \chi\right\} \\ &= \left\{\text{limit pts of } \frac{1}{n}\mu_{0\to n}(\chi) : \text{independent strategies } \chi\right\} \\ &= \left\{\sigma_{\vec{p}} : \text{single-edge strategies } \vec{p}\right\}\end{aligned}$$

This reduces the problem to working with single-edge strategies $\vec{p}$, which are simpler to handle than generic strategies. Furthermore, every $\sigma_{\vec{p}}$ can be achieved by a "symmetric" single-edge strategy $\vec{p}$ with the property that the p_k are uniquely determined by p_1 by applying permutations to the argument $\vec{u}$.

Now, the sets from Theorem E are convex and weakly compact. In Section 3.5 we fully characterize their extreme points.

Theorem F. Let $\vec{p}$ be a single-edge strategy. The following are equivalent:
(i) $\sigma_{\vec{p}}$ is an extreme point
(ii) Any symmetric single-edge strategy achieving $\sigma_{\vec{p}}$ must be deterministic.
(iii) $\sigma_{\vec{p}}$ has a density $f_{\vec{p}}$ which is not constant on sets of positive Λ-measure and $\sigma_{\vec{p}}$ is given by the following single-edge strategy $\vec{q}$:

$$q_k(u_1,\dots,u_D) = \mathbf{1}_{\{f_{\vec{p}}(u_k)\geqslant f_{\vec{p}}(u_i)\ \forall i\}} \text{ for } \Lambda^{\times D}\text{-a.a. } (u_1,\dots,u_D)$$

In other words, $\sigma_{\vec{p}}$ is achieved by the deterministic greedy single-edge strategy "choose whichever weight yields a higher value when evaluating the density $f_{\vec{p}}$."
(iv) $\sigma_{\vec{p}}$ has a density $f_{\vec{p}}$ s.t. $\frac{1}{D}f_{\vec{p}}(Unif[0,1]) \sim \text{Beta}(1,D)$.

As an immediate corollary, $\mathcal{R}$ is the closed convex hull of these measures.

Remark 3.6. It is important to stress that Theorem E holds even in a more general setting where the i.i.d. labels U_i^j follow some finite mean distribution on $\mathcal{R}$ that is not necessarily Unif[0,1]. The only caveat will be that the value distribution of the densities $f_{\vec{p}}$ of extreme points might not have as explicit a form as $D \cdot Beta(1,D)$.

Remark 3.7. In fact, Theorem E holds in a discrete setting as well. If the i.i.d. labels U^j are Unif$\{1, \ldots, K\}$ then the convex set of symmetric single-edge strategies forms a permutohedron of order K.

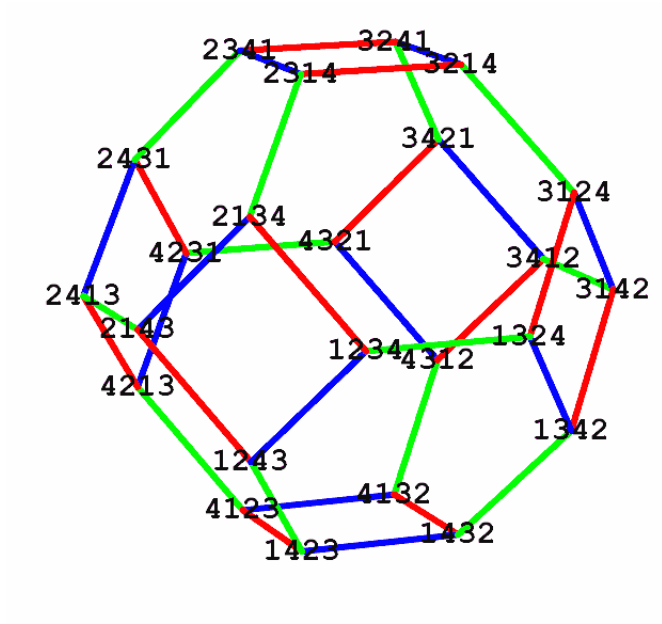


Figure 3.1: Permutohedron of order 4, [Hol]

The extreme symmetric single-edge strategies are those which choose whichever of the D labels is maximal with respect to some given ordering (e.g. the extreme point 1342 in the figure corresponds to the ordering $1 < 3 < 4 < 2$). By the discrete analogue of Theorem F, this induces a natural bijection between the extreme points of the permutohedron and the extreme points of $\mathcal{R}$, which is easily extended to a bijection between the full spaces by taking convex combinations. We will walk through an explicit example of this phenomenon in Section 3.5.5. Furthermore, going back to the continuous case, we see that the convex set of symmetric single-edge strategies there is merely the continuous analogue of the discrete permutohedron.

Generally speaking there is no known way of computing grid entropy, however, it amazingly can be computed to be 0 for these extreme points. This effectively means that a.s. the number of paths $0 \to n$ with empirical measures weakly converging to any one of these extreme points is $e^{o(n)}$.

Theorem G. Let $\sigma_{\vec{p}}$ be an extreme point. Then

$$||\sigma_{\vec{p}}|| = 0$$

Whether *every* measure in $\mathcal{R}$ with grid entropy 0 is an extreme point remains an open question.

Section 3.6 focuses on this result, as well as simplified formulas for the Gibbs Free Energy and grid entropy in this model.

Theorem H. Suppose $\tau : [0,1] \to [0,\infty)$ is a measurable function satisfying

$$E[e^{\tau(U)}] < \infty \text{ for } U \sim Unif[0,1]$$

Then Gibbs Free Energy with respect to τ is given by

$$G(\tau) = E\left[\log \sum_{j=1}^{D} e^{\tau(U^j)}\right] \text{ a.s.}$$

where U^j are i.i.d. Unif[0,1]. For all probability measures ν, grid entropy is given by

$$-||\nu|| = \sup_{\tau}\left[\langle\tau,\nu\rangle - G(\tau)\right] = \sup_{\tau}\left[\langle\tau,\nu\rangle - E\left[\log\sum_{j=1}^{D} e^{\tau(U^j)}\right]\right]$$

where the supremum is over all measurable functions $\tau : [0,1] \to [0,\infty)$ and where $\langle\tau,\nu\rangle$ denotes the integral $\int_0^1 \tau(u)d\nu$.

But first we will delve deeper into the setup and known relevant results.

3.2 PRELIMINARIES

3.2.1 *More on Strategies*

We list some miscellaneous observations about strategies.

It is trivial to see that the sets of strategies χ, of single-edge strategies ψ, of distributions σ_χ^i for strategies χ and $i \geqslant 0$, and of σ_ψ for single-edge strategies ψ are each closed under convex combinations. The extreme points of the sets of strategies/single-edge strategies are clearly the sets of deterministic strategies/single-edge strategies respectively.

Moreover, if χ is an independent strategy corresponding to a sequence (ψ_i) of single-edge strategies then $\sigma_\chi^i = \sigma_{\psi_i}$ $\forall i \geqslant 0$. In particular, this implies

$$\{\sigma_\chi^i : \text{independent strategies } \chi\} = \{\sigma_\psi : \text{single-edge strategies } \psi\}\ \forall i \geqslant 0 \tag{3.1}$$

In fact, observe that for any strategy χ and $i \geqslant 0$ if we define the strategy χ' to be the measure determined on product sets by

$$\chi'(A \times B) := \chi((0\ i)A \times (0\ i)B)\ \forall \text{ measurable } A \subseteq ([0,1]^D)^{\mathbb{Z}_{\geqslant 0}}, B \subseteq (\{1,\ldots,D\})^{\mathbb{Z}_{\geqslant 0}}$$

where $(0\ i)C$ swaps the 0th and ith coordinates of sequences in C, then this is easily checked to be a strategy with

$$\sigma^i_\chi = \sigma^0_{\chi'}$$

Therefore

$$\{\sigma^i_\chi : \text{strategies } \chi\} = \{\sigma^0_\chi : \text{strategies } \chi\}\ \forall i \geqslant 0 \tag{3.2}$$

3.2.2 *Grid Entropy*

We briefly summarize the relevant parts from Chapter 2 that we will need in this chapter.

Definition 3.8. The Levy-Prokhorov metric on the space $\mathcal{M}_+$ of finite non-negative Borel measures on $[0,1]$ is defined by

$$\rho(\mu,\nu) = \inf\{\epsilon > 0 : \mu(A) \leqslant \nu(A^\epsilon) + \epsilon \text{ and } \nu(A) \leqslant \mu(A^\epsilon) + \epsilon\ \forall A \in \mathcal{B}([0,1])\}$$

It is standard that the Levy-Prokhorov metric ρ metrizes weak convergence. Some elementary properties include that the Levy-Prokhorov metric is upper bounded by total variation and it satisfies a certain subadditivity:

$$\rho(\mu_1 + \mu_2, \nu_1 + \nu_2) \leqslant \rho(\mu_1,\nu_1) + \rho(\mu_2,\nu_2)$$

For $t \geqslant 0$ let $\mathcal{M}_t$ denote the space of non-negative Borel measures on $[0,1]$ with total mass t.

In Chapter 2, we developed the notion of grid entropy, which not just captures the convergence of empirical measures in a lattice model such as ours but the *proportion* of empirical measures converging to a certain weak limit.

Three quite different but equivalent definitions of grid entropy were given. In Chapter 3 we will work with the two most relevant to our needs. Note again that in the setting of Chapter 2, both direction-fixed and direction-free grid entropy were considered, but in the model we focus on here there is but one unit direction so we will be using the direction-free versions of the results from before.

Let us recall the major takeaways from Chapter 2.

Fix any $t \geqslant 0$ and a target measure ν. We consider the order statistics of the Levy-Prokhorov distance between ν and the empirical measures $\frac{1}{n}\mu_\pi$ varying over all $D^{\lfloor nt \rfloor}$ possible origin-anchored, length $\lfloor nt \rfloor$ paths π. That is, for every $n \in \mathbb{N}$ we let

$$\overset{1}{\min_{\pi:0\to\lfloor nt\rfloor}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) \leqslant \overset{2}{\min_{\pi:0\to\lfloor nt\rfloor}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) \leqslant \ldots \leqslant \overset{D^{\lfloor nt\rfloor}}{\min_{\pi:0\to\lfloor nt\rfloor}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right)$$

denote the order statistics value of $\rho(\frac{1}{n}\mu_\pi, \nu)$. These minimums and the paths corresponding to them are of course event-dependent. However, the following theorem from Chapter **??** shows there is a deterministic threshold where the empirical measures along the paths corresponding to these minimums change from converging a.s. to ν to a.s. diverging away from ν.

Theorem 3.9. *(i) For any target measure $\nu \in \mathcal{M}$, its grid entropy is defined to be the deterministic quantity*

$$||\nu|| := \sup\left\{\alpha \geqslant 0 : \lim_{n\to\infty} \overset{\lfloor e^{\alpha n}\rfloor}{\min_{\pi:0\to\lfloor nt\rfloor}} \rho\left(\frac{1}{n}\mu_\pi, \nu\right) = 0 \, a.s.\right\} \in \{-\infty\} \cup [0, t\log D]$$

Then grid entropy is the critical exponent where the $\overset{\lfloor e^{\alpha n}\rfloor}{\min_{\pi:0\to\lfloor nt\rfloor}} \rho(\frac{1}{n}\mu_\pi, \nu)$ *change from converging to 0 to a.s. having a* $\liminf_{n\to\infty} > 0$.
For example, the grid entropy of the original distribution, Λ, is maximal: $||\Lambda|| = t\log D$.
(ii) Grid entropy is positive-homogeneous, satisfies the reverse-triangle inequality

$$||\nu|| + ||\xi|| \leqslant ||\nu + \xi||$$

and is upper-semicontinuous in ν.
(iii) Consider the deterministic, weakly closed set

$$\mathcal{R}^t := \{\nu \in \mathcal{M} : ||\nu|| \geqslant 0\}$$

Then $\mathcal{R}^t \subseteq \mathcal{M}_t$, $\nu << \Lambda$ $\forall \nu \in \mathcal{R}^t$ and

$$\mathcal{R}^t = \left\{\text{limit pts of } \frac{1}{n}\mu_\pi \text{ for } \pi : 0 \to \lfloor nt\rfloor\right\} \text{ a.s.}$$

(iv) Grid entropy is the negative convex conjugate of Gibbs Free Energy. More concretely,

$$-||\nu|| = \sup_\tau \left[\langle \tau, \nu\rangle - G_t(\tau)\right] \; \forall \nu \in \mathcal{M}_t \tag{3.3}$$

where the supremum is taken over measurable $\tau : [0,1] \to [0,\infty)$ *satisfying*

$$E[e^{\tau(U)}] < \infty \text{ for } U \sim Unif[0,1],$$

where $\langle \tau, \nu \rangle = \int_0^1 \tau(u) d\nu$*, where* $G_t(\tau)$ *is the length t Gibbs Free Energy*

$$G_t(\tau) := \lim_{n\to\infty} \frac{1}{n} \log \sum_{\pi:0\to\lfloor nt \rfloor} e^{T(\pi)}$$

and where $T(\pi) = \sum_{e\in\pi} \tau(U_e) = \langle \tau, \mu_\pi \rangle$ *is the passage time along* π*.*

Remark 3.10. An immediate but highly non-trivial consequence of (i) is the existence of random length $\lfloor nt \rfloor$ paths whose empirical measures converge weakly to a given target $\nu \in \mathcal{R}^t$ a.s.. Indeed, the event-dependent paths corresponding to $\overset{1}{\underset{\pi:0\to\lfloor nt \rfloor}{\min}} \rho(\frac{1}{n}\mu_\pi, \nu)$ do the job.

By positive-homogeneity it suffices to work with the case $t = 1$. For the sake of notation we will drop the 1 in $\mathcal{R}^1, G_1(\tau)$ for the rest of the chapter.

Once we determine the extreme points of $\mathcal{R}$, we will compute their grid entropies to be 0. We will also give a limit-free formula for Gibbs Free Energy $G(\tau)$ in our model, which renders the convex duality formula (3.3) more practical.

3.2.3 *Measure-Preserving Bijections in* $\mathbb{R}^n$

We will encounter measure-preserving bijections in $\mathbb{R}$, so we briefly outline the required notions in this section.

Definition 3.11. Let $\mathcal{B}(\mathbb{R}^n)$ denote the Borel σ-algebra of $\mathbb{R}^n$ and let μ be a Borel probability measure on $\mathbb{R}^n$. Let $A, A' \in \mathcal{B}(\mathbb{R}^n)$ with $\mu(A') = \mu(A) > 0$. A bijection $\phi : A \to A'$ is said to be μ-measure-preserving if $\mu(B) = \mu(\phi(B))$ for all $B \in \mathcal{B}(\mathbb{R}^n), B \subseteq A$.

Nishiura proves in [Nis98, Theorem 4] the existence of a μ-measure-preserving bijection between Borel sets of equal, positive but not full μ-measure, given that μ is nonatomic. We will only need the following simplified version of this theorem.

Theorem 3.12. *Suppose* θ *is a Borel probability measure on* $[0,1]$ *that is absolutely continuous with respect to Lebesgue measure* Λ *on* $[0,1]$*. Let* $A, A' \in \mathcal{B}([0,1])$ *s.t.* $\theta(A) = \theta(A') \in (0,1)$*. Then there exists a* θ*-measure-preserving bijection* $\phi : A \to A'$*.*

Preliminaries finished, our next goal is to show that we may restrict ourselves to working with single-edge strategies without losing generality.

3.3 REDUCING THE PROBLEM

Over the course of this section, we will prove the following theorem, which allows us to reduce the problem of describing the set of limit points of the empirical measures $\frac{1}{n}\mu_{0\to n}(\chi)$ to characterizing the set of achievable distributions σ_ψ for single-edge strategies ψ.

Theorem 3.13. *A.s., the following sets are equal:*
(i) $\{$*limit pts of* $\frac{1}{n}\mu_{0\to n}(\chi)$: *strategies* $\chi\}$
(ii) $\mathcal{R} := \{\nu \in \mathcal{M}_1 : ||\nu|| \geqslant 0\}$
(iii) $\{\sigma_\chi^0$: *strategies* $\chi\}$
(iv) $\{\sigma_\psi$: *single-edge strategies* $\psi\}$
(v) $\{$*limit pts of* $\frac{1}{n}\mu_{0\to n}(\chi)$: *independent strategies* $\chi\}$
(vi) $\{$*limit pts of* $\frac{1}{n}\mu_{0\to n}(\chi)$: *i.i.d. strategies* $\chi\}$

Remark 3.14. A priori it not clear that (iii)-(iv) are weakly closed. We will show this as part of our proof.

We will prove this via a chain of inclusions. We begin with the most trivial of these.

Lemma 3.15. *A.s. we have*

$$\left\{\textit{limit pts of } \frac{1}{n}\mu_{0\to n}(\chi) : \textit{ strategies } \chi\right\} \subseteq \mathcal{R}$$

Proof. Fix a strategy χ. Suppose $\frac{1}{n_k}\mu_{0\to n_k}(\chi) \Rightarrow \nu$ for some observed edge labels $(u_i^j)_{i\geqslant 0, 1\leqslant j\leqslant D}$. Then letting $\pi_n : 0 \to n$ be the paths corresponding to the indices $(J_0(\chi), \ldots, J_{n-1}(\chi))$ conditioned on these same observed edge labels, we get $\frac{1}{n_k}\mu_{\pi_{n_k}} = \frac{1}{n_k}\mu_{0\to n_k}(\chi) \Rightarrow \nu$. Thus ν is a limit point of the $\frac{1}{n}\mu_\pi$. The desired inclusion follows from

$$\mathcal{R} = \left\{\text{limit pts of } \frac{1}{n}\mu_\pi : \text{ paths } \pi : 0 \to n\right\} \text{ a.s. by Theorem 3.9(iii)}$$

□

3.3.1 *Expected Value of Empirical Measures*

For any strategy χ,

$$E[\delta_{X_i(\chi)}] = \int_{[0,1]} \delta_x d\sigma_\chi^i(x) = \sigma_\chi^i \text{ and } E\left[\frac{1}{n}\mu_{0\to n}(\chi)\right] = \frac{1}{n}\sum_{i=0}^{n-1}\sigma_\chi^i$$

Recalling (3.2) and the fact that $\{\sigma_\chi^0 : \text{strategies } \chi\}$ is closed under convex combinations, it follows that every $E[\frac{1}{n}\mu_{0\to n}(\chi)]$ is contained in $\{\sigma_\chi^0 : \text{strategies } \chi\}$. Therefore

$$\left\{E\left[\frac{1}{n}\mu_{0\to n}(\chi)\right] : \text{ strategies } \chi\right\} = \{E[\delta_{X_0(\chi)}] : \text{ strategies } \chi\} = \{\sigma_\chi^0 : \text{ strategies } \chi\}$$

Also it is clear that

$$\{E[\delta_{X(\psi)}] : \text{single-edge strategies } \psi\} = \{\sigma_\psi : \text{ single-edge strategies } \psi\}$$

A simple Tonelli argument establishes that the closures of the two sets above coincide.

Lemma 3.16.

$$cl\{\sigma_\psi : \textit{ single-edge strategies } \psi\} = cl\{\sigma_\chi^0 : \textit{ strategies } \chi\}$$

Proof. By (3.1) $\{\sigma_\psi\} = \{\sigma_\chi^0 : \text{indep strategies } \chi\}$ so it suffices to show

$$\text{cl}\{\sigma_\psi : \text{ single-edge strategies } \psi\} \supseteq \{\sigma_\chi^0 : \text{ strategies } \chi\}$$

Given a strategy χ and observed "future" edge labels $(u_i^j)_{i\geqslant 1, 1\leqslant j\leqslant D} \in [0,1]^{\mathbb{Z}_{\geqslant 1}}$, define $\psi_{(u_i^j)_{i\geqslant 1,1\leqslant j\leqslant D}}$ to be χ conditioned on the rest of the environment $(u_i^j)_{i\geqslant 1,1\leqslant j\leqslant D}$.

The $\psi_{(u_i^j)_{i\geqslant 1,1\leqslant j\leqslant D}}$ are easily seen to be single-edge strategies. Since integrating over the entire environment $(U_i^j)_{i\geqslant 0,1\leqslant j\leqslant D}$ is equivalent to integrating over $(U_0^j)_{1\leqslant j\leqslant D}$ first and then over $(U_i^j)_{i\geqslant 1,1\leqslant j\leqslant D}$, we have

$$\sigma_\chi^0 = \int_{([0,1]^D)^{\mathbb{Z}_{\geqslant 1}}} \sigma_{\psi_{(u_i^j)_{i\geqslant 1,1\leqslant j\leqslant D}}} d\Lambda^\infty((u_i^j)_{i\geqslant 1,1\leqslant j\leqslant D})$$

By convexity and weak closure it follows that

$$\sigma_\chi^0 \in \text{cl}\{\sigma_\psi : \text{ single-edge strategies } \psi\}$$

□

Next, we show that this closure of $\{\sigma_\chi^0 : \text{strategies } \chi\}$ contains all probability measures with finite grid entropy.

Lemma 3.17. *We have*

$$\mathcal{R} := \{\nu \in \mathcal{M}_1 : ||\nu|| \geqslant 0\} \subseteq cl\{E[\delta_{X_0(\chi)}] : \textit{ strategies } \chi\}$$

as deterministic sets.

Proof. Suppose this is not the case, say $\exists \nu \in \mathcal{R} \cap (\text{cl}\{E[\delta_{X_0(\chi)}] : \text{ strategies } \chi\})^C$. Thus there exists $\epsilon > 0$ s.t. $B_\epsilon(\nu) \cap cl\{E[\delta_{X_0(\chi)}] : \chi\} = \emptyset$.

Let $\pi_n : 0 \to n$ be the environment-dependent path corresponding to $\overset{1}{\min_{\pi:0\to n}} \rho(\frac{1}{n}\mu_\pi, \nu)$. It is crucial to note that π_n depends on the observed edge labels $((U_i^j)_{0\leqslant i<n, 1\leqslant j\leqslant D})$ of the first n trials only.

By definition of grid entropy, $\frac{1}{n}\mu_{\pi_n} \Rightarrow \nu$ a.s.. In particular, we get convergence in probability:

$$\lim_{n\to\infty} P\left(\rho\left(\frac{1}{n}\mu_{\pi_n}, \nu\right) \geqslant \frac{\epsilon}{2}\right) = 0$$

Thus $\exists n \in \mathbb{N}$ s.t.

$$P\left(\rho\left(\frac{1}{n}\mu_{\pi_n}, \nu\right) \geqslant \frac{\epsilon}{2}\right) < \frac{\epsilon}{4}$$

We claim that $\rho(E[\frac{1}{n}\mu_{\pi_n}], \nu) < \epsilon$. Let $\mathcal{E}$ be the event $\{\rho(\frac{1}{n}\mu_{\pi_n}, \nu) \geqslant \frac{\epsilon}{2}\}$ so $P(\mathcal{E}) < \frac{\epsilon}{4}$. We split the expectation:

$$E\left[\frac{1}{n}\mu_{\pi_n}\right] = \int_{\mathcal{E}} \frac{1}{n}\mu_{\pi_n} dP + \int_{\mathcal{E}^C} \frac{1}{n}\mu_{\pi_n} dP$$

Since the Levy-Prokhorov metric is upper bounded by the total variation, then

$$\rho\left(\int_{\mathcal{E}} \frac{1}{n}\mu_{\pi_n} dP, \nu \cdot P(\mathcal{E})\right) \leqslant \left(\left|\left|\frac{1}{n}\mu_{\pi_n}\right|\right|_{TV} + ||\nu||_{TV}\right) P(\mathcal{E}) = 2P(\mathcal{E}) < \frac{\epsilon}{2}$$

On $\mathcal{E}^C$, we have $\rho(\frac{1}{n}\mu_{\pi_n}, \nu) < \frac{\epsilon}{2}$. By definition of ρ, for all measurable $A \in \mathcal{B}([0,1])$,

$$\left(\int_{\mathcal{E}^C} \frac{1}{n}\mu_{\pi_n} dP\right)(A) = \int_{\mathcal{E}^C} \frac{1}{n}\mu_{\pi_n}(A) dP \leqslant \int_{\mathcal{E}^C} \nu(A^{\frac{\epsilon}{2}}) + \frac{\epsilon}{2} dP \leqslant \nu(A^{\frac{\epsilon}{2}}) P(\mathcal{E}^C) + \frac{\epsilon}{2}$$

and similarly

$$\nu(A) P(\mathcal{E}^C) = \int_{\mathcal{E}^C} \nu(A) dP \leqslant \int_{G^C} \frac{1}{n}\mu_{\pi_n}(A^{\frac{\epsilon}{2}}) + \frac{\epsilon}{2} dP \leqslant \left(\int_{\mathcal{E}^C} \frac{1}{n}\mu_{\pi_n} dP\right)(A^{\frac{\epsilon}{2}}) + \frac{\epsilon}{2}$$

hence

$$\rho\left(\int_{\mathcal{E}^C} \frac{1}{n}\mu_{\pi_n} dP, \nu P(\mathcal{E}^C)\right) \leqslant \frac{\epsilon}{2}$$

By subadditivity of ρ,

$$\begin{aligned}\rho\left(E\left[\frac{1}{n}\mu_{\pi_n}\right],\nu\right) &= \rho\left(\int_{\mathcal{E}}\frac{1}{n}\mu_{\pi_n}dP + \int_{\mathcal{E}^C}\frac{1}{n}\mu_{\pi_n}dP, \nu P(\mathcal{E}) + \nu P(\mathcal{E}^C)\right)\\ &\leqslant \rho\left(\int_{\mathcal{E}}\frac{1}{n}\mu_{\pi_n}dP, \nu P(\mathcal{E})\right) + \rho\left(\int_{\mathcal{E}^C}\frac{1}{n}\mu_{\pi_n}dP, \nu P(\mathcal{E}^C)\right)\\ &< \epsilon\end{aligned}$$

It remains to construct a strategy χ for which $\frac{1}{n}\mu_{\pi_n} = \frac{1}{n}\mu_{0\to n}(\chi)$ for this fixed n. But this is trivial since conditioned on the observed edge labels $(U_i^j)_{0\leqslant i\leqslant n-1, 1\leqslant j\leqslant D}$ from the first n trials, π_n is a deterministic path $0 \to n$ (namely the one which minimizes $\rho(\frac{1}{n}\mu_\pi, \nu)$).

To be concrete, consider the product measure χ' on $([0,1]^D)^{\mathbb{Z}_{\geqslant n}} \times \{1,\ldots,D\}^{\mathbb{Z}_{\geqslant n}}$ given by

$$\chi' = \Lambda^\infty \times \delta_{(1,1,\ldots)}$$

Heuristically, χ' is a partial strategy always picking the "top" edge label U_i^1 for $i \geqslant n$.

Also let χ'' be the measure on $([0,1]^D)^{\mathbb{Z}_{0\leqslant i<n}} \times \{1,\ldots,D\}^{\mathbb{Z}_{0\leqslant i<n}}$ determined by

$$\begin{aligned}&\chi''(A \times \{(j_0,\ldots,j_{n-1})\})\\ &= \Lambda^\infty(\{(u_i^j)_{0\leqslant i<n,1\leqslant j\leqslant D} \in A : \pi_n((u_i^j)_{0\leqslant i<n,1\leqslant j\leqslant D}) = (j_0,\ldots,j_{n-1})\})\end{aligned}$$

$\forall A \in \mathcal{B}(([0,1]^D)^{\mathbb{Z}_{0\leqslant i<n}}), (j_i)_{0\leqslant i<n} \in \{1,\ldots,D\}^{\mathbb{Z}_{0\leqslant i<n}}$ where $\pi_n((u_i^j)_{0\leqslant i<n,1\leqslant j\leqslant D})$ denotes the sequence of n indices $(j_0,\ldots,j_{n-1})$ corresponding to the path π_n when the first n trials yield observed labels $(u_i^j)_{0\leqslant i<n,1\leqslant j\leqslant D}$. That is, conditioned on the observed values from the first n trials, χ'' always picks the path corresponding to π_n.

Now consider the strategy $\chi = \chi'' \times \chi'$. Then $\mu_{0\to n}(\chi) = \mu_{\pi_n}$ hence $E[\frac{1}{n}\mu_{\pi_n}] = E[\frac{1}{n}\mu_{0\to n}(\chi)]$.

This contradicts

$$B_\epsilon(\nu) \cap cl\left\{E\left[\frac{1}{n}\mu_{0\to n}(\chi)\right] : \chi\right\} = B_\epsilon(\nu) \cap cl\{E[\delta_{X_0(\chi)}] : \chi\} = \emptyset$$

□

Combining Lemmas 3.15-3.17, we get that a.s., (vi) ⊆ (v) ⊆ (i) ⊆ (ii) ⊆ cl(iii) = cl(iv) in Theorem 3.13. To complete the proof we only need to show the last inclusion cl (iv) ⊆ (vi) and the fact that (iv) is weakly closed.

3.3.2 *What Happens with Independent Strategies*

Let us now focus on the case of independent strategies. Recall that these look like $\chi = \bigtimes_{i=0}^{\infty} \psi_i$ where $(\psi_i)_{i \geqslant 0}$ are single-edge strategies.

First, consider the even simpler case of i.i.d. strategies χ, where all the ψ_i are identical, resulting in i.i.d. X_i. Then the Glivenko-Cantelli Theorem [Dur19, Thm. 2.4.7] gives that a.s. the empirical measures $\frac{1}{n}\mu_{0\to n}(\chi)$ converge weakly to the common law of the X_i's, σ_{ψ_0}.

Theorem 3.18 (Glivenko-Cantelli Theorem)**.** *Let F_γ be the cumulative distribution function of γ, let $Y_i \sim \gamma$ be i.i.d. random variables and let*

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leqslant y\}}$$

be the cumulative distribution functions of the empirical measures. Then

$$\sup_y |F_n(y) - F_\gamma(y)| \to 0 \text{ a.s. as } n \to \infty$$

As an immediate corollary we get the outstanding inclusion (vi) $\supseteq$ cl (iv) mentioned at the end of Section 3.3.1.

Corollary 3.19.

A.s. we have

$$\left\{\text{limit pts of } \frac{1}{n}\mu_{0\to n}(\chi) : \text{ i.i.d. strategies } \chi\right\} \supseteq cl\{\sigma_\psi : \text{single-edge strategies } \psi\}$$

Proof. Fix a dense subset $\mathcal{O}$ of the deterministic, weakly compact set cl (iv) s.t. $\mathcal{O} \subseteq$ (iv). For every $\sigma_\psi \in \mathcal{O}$, apply Theorem 3.18 to the i.i.d. strategy $\chi = \bigtimes_{i=0}^{\infty} \psi$ to get $\frac{1}{n}\mu_{0\to n}(\chi) \Rightarrow \sigma_\psi$ a.s.. Therefore $\mathcal{O} \subseteq$ (vi) a.s.. The inclusion follows since (vi) is weakly closed.

□

Thus we have shown that the sets (i), (ii), cl(iii), cl(iv), (v), (vi) in Theorem 3.13 are equal a.s..

Before proceeding to look further into single-edge strategies, it is worth mentioning a version of the Glivenko-Cantelli Theorem for independent but not necessarily i.d. sequences from [Wel81]. It gives further insight into the limit points of the empirical measures $\frac{1}{n}\mu_{0\to n}(\chi)$ for independent strategies χ.

Theorem 3.20. *Let Y_i be a sequence of independent random variables with distributions σ_i. Define $\overline{\sigma}_n = \frac{1}{n}(\sigma_0 + \ldots + \sigma_{n-1})$ to be the averages of these distributions and let*
$\frac{1}{n}\mu_n = \frac{1}{n}(\delta_{Y_0} + \ldots + \delta_{Y_{n-1}})$ be the empirical measures. If $\{\overline{\sigma}_n\}$ is tight then $\rho(\overline{\sigma}_n, \frac{1}{n}\mu_n) \to 0$ a.s..

In our case, $\mathcal{M}_1$ is weakly compact so any sequence of probability measures in $\mathcal{M}_1$ is tight.

Thus, for independent strategies $\chi = \bigtimes_{i=0}^{\infty} \psi_i$ with ψ_i single-edge strategies, we have
$\rho(\overline{\sigma_n}, \frac{1}{n}\mu_{0\to n}(\chi)) \to 0$ a.s. where $\overline{\sigma_n}$ denote the averages of the distributions of $X_i(\chi)$:

$$\overline{\sigma}_n = \frac{1}{n}(\sigma_\chi^0 + \ldots + \sigma_\chi^{n-1}) = \frac{1}{n}(\sigma_{\psi_0} + \ldots + \sigma_{\psi_{n-1}}) = \sigma_{\frac{1}{n}(\psi_0 + \ldots + \psi_{n-1})}$$

This not only confirms that cl(iv) and (v) are equal, but it tells us that a.s. the empirical measures $\frac{1}{n}\mu_{0\to n}(\chi)$ for independent strategies χ have the exact same limit points as the distributions $\sigma_{\frac{1}{n}(\psi_0+\ldots+\psi_{n-1})}$ corresponding to the law of X chosen according to the average of the single-edge strategies ψ_i.

3.4 SINGLE-EDGE STRATEGIES REVISITED

3.4.1 *Single-Edge Strategies in Terms of Conditional Expectations*

From what we have shown thus far, we only need to focus on single-edge strategies, as the set of limit points of empirical measures $\frac{1}{n}\mu_{0\to n}(\chi)$ coincides with

$$cl\{\sigma_\psi : \text{single-edge strategies } \psi\}$$

Rather than using the measure definition of single-edge strategies, it will be more practical to work with the vector $\vec{p}$ of conditional expectations defined as

$$p_k(u_1, \ldots, u_D) := E_\psi[(U^j)_{1\leqslant j\leqslant D} = (u^j)_{1\leqslant j\leqslant D} | J = k]$$

Then ψ evaluated on product sets is given by

$$\psi(A \times B) = \int_A \sum_{j=1}^{D} p_j(u_1, \ldots, u_D) \mathbf{1}_{\{j\in B\}}\, d\Lambda^D(u_1, \ldots, u_D) \qquad (3.4)$$

$\forall A \in \mathcal{B}([0,1]^D), B \subseteq \{1,\ldots,D\}$. We also have $\sum_{j=1}^{D} p_j \equiv 1$ everywhere.

Intuitively, each $p_k(u_1,\ldots,u_D)$ is the probability of choosing u_k when the observed samples are $(u_1,\ldots,u_D)$. This justifies the term "single-edge strategy," because $\vec{p}$ is prescribing the strategy by which we make our choice once we have the D observed samples.

Of course, the vector $\vec{p} = (p_1,\ldots,p_D)$ determines ψ by (3.4). And if all the p_j are 0-1 valued $\Lambda^{\times D}$-a.e. then the single-edge strategy ψ is deterministic.

Even though different $\vec{p}$ may give rise to the same ψ, we will conflate the two notions from now on and call both ψ and $\vec{p}$ the single-edge strategy. We will use $X(\vec{p}), X(\chi)$ and $\sigma_{\vec{p}}, \sigma_\psi$ and $F_{\vec{p}}, F_\psi$ interchangeably. At the end of the day, all that will matter is whether two vectors $\vec{p}, \vec{q}$ yield the same law $\sigma_{\vec{p}}$ of $X(\vec{p})$.

We may now write the cumulative distribution function of $\sigma_{\vec{p}}$ in terms of $\vec{p}$:

$$\begin{aligned} F_{\vec{p}}(y) &= \int_{[0,1]^D} \sum_{j=1}^{D} p_j(u_1,\ldots,u_D)\mathbf{1}_{[0,y]}(u_j)d\Lambda^D(u_1,\ldots,u_D) \\ &= \int_{[0,y]\times[0,1]^{D-1}} p_1(u_1,\ldots,u_D)d\Lambda^D(u_1,\ldots,u_D) \\ &\quad + \ldots + \int_{[0,1]^{D-1}\times[0,y]} p_D(u_1,\ldots,u_D)d\Lambda^D(u_1,\ldots,u_D) \end{aligned} \tag{3.5}$$

Therefore $\sigma_{\vec{p}} << \Lambda$ with density

$$\begin{aligned} f_{\vec{p}}(y) &= \int_{[0,1]^{D-1}} p_1(y,u_2,\ldots,u_D)d\Lambda^{D-1}(u_2,\ldots,u_D) \\ &\quad + \cdots + \int_{[0,1]^{D-1}} p_D(u_1,\ldots,u_{D-1},y)d\Lambda^{D-1}(u_1,\ldots,u_{D-1}) \end{aligned} \tag{3.6}$$

It is clear that $f_{\vec{p}} \in [0,D]$.

Let us give an example that will make everything clear. Consider the deterministic single-edge strategy $\vec{p}^{MAX}$ that always chooses the maximum of the observed edge labels. In our notation,

$$p_j^{MAX}(u_1,\ldots,u_D) = \mathbf{1}_{\{u_j \geqslant u_k \ \forall 1 \leqslant k \leqslant D\}}$$

The resulting distribution σ_{MAX} has cdf

$$F_{MAX}(y) = P(U^j \leqslant y \ \forall 1 \leqslant j \leqslant D) = y^D$$

and density

$$f_{MAX}(y) = Dy^{D-1}$$

This single-edge strategy will play a crucial role later in our description of the extreme points of $\{\sigma_{\vec{p}}\}$.

We will also be interested in single-edge strategies "scrambled" by a Λ-measure-preserving bijection.

Definition 3.21. Let $\vec{p}$ be a single-edge strategy and let $\phi : [0,1] \to [0,1]$ be a Λ-measure-preserving bijection. We define $\vec{p}^{\phi}$ to be the single-edge strategy with coordinate functions

$$p_j^{\phi}(u_1, \dots, u_D) := p_j(\phi(u_1), \dots, \phi(u_D))$$

It is clear that $\sum p_j^{\phi} \equiv 1$ still so it is a valid single-edge strategy. Furthermore, since ϕ is measure-preserving with respect to Λ, then by a change of variables in the integral formula for $f_{\vec{p}}$ we get

$$f_{\vec{p}^{\phi}} = f_{\vec{p}} \circ \phi$$

The following lemma collects some basic facts about single-edge strategies.

Lemma 3.22. *Let $\vec{p}$ be a single-edge strategy.*
(i) For any Borel set $A \in \mathcal{B}([0,1])$,

$$\Lambda(A)^D \leqslant \sigma_{\vec{p}}(A) \leqslant 1 - (1 - \Lambda(A))^D$$

In particular, $X(\vec{p}^{MAX})$ stochastically dominates $X(\vec{p})$ and $\sigma_{\vec{p}} << \Lambda, \Lambda << \sigma_{\vec{p}}$.
(ii) We have

$$\sup_{\phi} E[\sigma_{\vec{p}^{\phi}}] \leqslant E[\sigma_{MAX}]$$

where the supremum is taken over Λ-measure-preserving bijections $\phi : [0,1] \to [0,1]$.
(iii) Convex combinations of single-edge strategies $\vec{p}$ translate to convex combinations of $\sigma_{\vec{p}}, f_{\vec{p}}, F_{\vec{p}}$.

Remark 3.23. It is not clear whether the supremum in (ii) is achieved, but this is beyond the scope of this paper.

Proof. (i) If all D edge labels $u_1, \dots, u_D$ are in the set A, then so must be the edge label chosen from among them. Therefore $\Lambda(A)^D \leqslant \sigma_{\vec{p}}(A)$. The other inequality follows by replacing A with A^C.

Recalling that $F_{MAX}(y) = y^D$, we get

$$F_{MAX}(y) \leqslant F_{\vec{p}}(y) \ \forall y$$

(ii) For any Λ-measure-preserving bijection ϕ, we apply the tail integral formula for expectation and use (i) to get

$$E[\sigma_{\vec{p}^{\phi}}] = \int_0^1 1 - F_{\vec{p}^{\phi}}(t)dt \leqslant \int_0^1 1 - F_{MAX}(t)dt = E[\sigma_{MAX}]$$

(iii) For single-edge strategies $\vec{p}, \vec{q}$ and $t \in [0,1]$, $t\vec{p} + (1-t)\vec{q}$ is itself a single-edge strategy with

$$F_{t\vec{p}+(1-t)\vec{q}} = tF_{\vec{p}} + (1-t)F_{\vec{q}},\ f_{t\vec{p}+(1-t)\vec{q}} = tf_{\vec{p}} + (1-t)f_{\vec{q}},$$

$$\mu_{t\vec{p}+(1-t)\vec{q}} = t\sigma_{\vec{p}} + (1-t)\sigma_{\vec{q}}$$

□

3.4.2 "Symmetric" Single-Edge Strategies

Before proceeding further, we explain why we can restrict ourselves to single-edge strategies $\vec{p}$ with some convenient symmetries.

We will say a permutation $\iota \in S_D$ acts on a D-tuple $\vec{u}$ by applying ι to the indices:

$$\iota(\vec{u}) = (u_{\iota(1)}, \ldots, u_{\iota(D)})$$

Lemma 3.24. *Let $\vec{p}$ be any single-edge strategy. Then there exists another single-edge strategy $\vec{q}$ that gives rise to the same distribution $\sigma_{\vec{p}} = \sigma_{\vec{q}}$ s.t. $\forall x \in [0,1]$, $1 \leqslant i \leqslant D$*

$$f_{\vec{p}}(x) = f_{\vec{q}}(x) = D\int_{[0,1]^{D-1}} q_i(u_1, u_2, \ldots, x, \ldots, u_D)d\Lambda^{D-1}(u_1, \ldots, \widehat{u}_i, \ldots, u_D) \tag{3.7}$$

where the x occurs at position i, s.t.

$$q_1(\iota_i(\vec{u})) = q_2(\iota_{i-1}(\vec{u})) = \ldots = q_D(\iota_{i+1}(\vec{u}))\ \forall 1 \leqslant i \leqslant D\ \forall \vec{u} \tag{3.8}$$

where each ι_j is the cyclic shift

$$\iota_j(\vec{u}) = (u_j, u_{j+1}, \ldots, u_D, u_1, \ldots, u_{j-1})$$

and s.t.

$$q_i(\vec{u}) = q_i(\iota(\vec{u}))\ \forall 1 \leqslant i \leqslant D, \forall \vec{u}, \text{and } \forall \iota \in S_D \text{ with } \iota(i) = i \tag{3.9}$$

Remark 3.25. The new single-edge strategy $\vec{q}$ is consistent across all permutations of a tuple $\vec{u}$. That is, given an unordered tuple $(u_1, \ldots, u_D)$ we can say that $\vec{q}$ chooses each u_i with probability t_i. Then every $q_j(\iota(\vec{u}))$ for

$\iota \in S_D$ will pick out the probability $t_{\iota(j)}$ i.e. q_j outputs the probability of choosing the jth entry in its input tuple.

Furthermore, both the density $f_{\vec{q}}$ and the entire single-edge strategy $\vec{q}$ are uniquely determined by any one of the q_i. That is, given a measurable function $q_i : [0,1]^D \to [0,1]$ whose integral over $[0,1]^D$ with respect to the product measure $\Lambda^{\times D}$ is $\frac{1}{D}$, which is invariant under permutations in S_D fixing i, and which satisfies

$$\sum_j q_i(\iota_j(\vec{u})) = 1 \ \forall \vec{u}$$

we can use cyclic shifts to define a valid corresponding single-edge strategy $\vec{q}$ (that satisfies $\sum q_j \equiv 1$ and $\int_{[0,1]^D} q_j \, d\Lambda^{\times D} = \frac{1}{D}$) and we can compute $f_{\vec{q}}$ directly from q_i.

Remark 3.26. It is easy to check that $\vec{p}^{MAX}$ satisfies (3.7)-(3.9). Furthermore, if $\vec{p}$ is a symmetric single-edge strategy and $\phi : [0,1] \to [0,1]$ is a Λ-measure-preserving bijection then so is $\vec{p}^{\phi}$.

Proof. The intuition is that we take the average of the original p_i over the desired symmetries. We do this in two steps. For each $1 \leqslant i \leqslant D$ define

$$p_i'(\vec{u}) := \frac{p_1(\iota_i(\vec{u})) + p_2(\iota_{i-1}(\vec{u})) + \ldots + p_i(\iota_1(\vec{u})) + \ldots p_D(\iota_{i+1}(\vec{u}))}{D}$$

$$q_i(\vec{u}) = \frac{1}{(D-1)!} \sum_{\iota \in S_D, \iota(i)=i} p_i'(u_{\iota(1)}, u_{\iota(2)} \ldots, u_{\iota(D)})$$

A straightforward computation shows that $\vec{q}$ satisfies the required properties (3.7)-(3.9). □

From the rest of the chapter we will restrict ourselves to these symmetric single-edge strategies, which will simplify our computations.

3.4.3 *Closure of* $\{\sigma_{\vec{p}}\}$

The last remaining part of Theorem 3.13 is to show that the set of distributions $\sigma_{\vec{p}}$ is weakly closed.

Theorem 3.27. $\{\sigma_{\vec{p}} :$ *single-edge strategies* $\vec{p}\}$ *is weakly closed*

Proof. Suppose $\sigma_{\vec{p}^n} \Rightarrow \xi$ for some symmetric single-edge strategies $\vec{p}^n$. We will find a single-edge strategy yielding the distribution ξ.

Consider any n. Let $\nu_{\vec{p}^n}$ be the distribution on $[0,1]^D$ given by integration against $Dp_1^n d\Lambda^{\times D}$. Observe that (3.7) implies

$$f_{\vec{p}^n}(y) = \int_{[0,1]^{D-1}} Dp_1^n(y,u_2,\dots,u_D)d\Lambda^{D-1}(u_2,\dots,u_D)$$

so $\sigma_{\vec{p}^n}$ is just the first coordinate marginal of $\nu_{\vec{p}^n}$.

Compactness yields a weakly convergent subsequence $\nu_{\vec{p}^{n_j}} \Rightarrow \nu$. By a standard argument, since Dp_1^n are uniformly bounded by D, then $\nu << \Lambda^{\times D}$ and has a density of the form Dp_1 for some measurable function $p_1 : [0,1]^D \to [0,1]$. This along with the fact that $\int_{[0,1]^D} p_1 d\Lambda^{\times D} = \frac{1}{D}$ are enough for p_1 to give rise to a single-edge strategy $\vec{p}$ by Remark 3.25.

But then the corresponding first coordinate marginals must also converge weakly, so we get $\sigma_{\vec{p}^{n_j}} \Rightarrow \xi'$ where ξ' is the corresponding marginal of ν:

$$\xi'(A) = \int_{A\times[0,1]^{D-1}} Dp_1 d\Lambda^D = \sigma_{\vec{p}}(A)\ \forall A \in \mathcal{B}([0,1])$$

Thus $\xi' = \sigma_{\vec{p}}$. On the other hand, $\sigma_{\vec{p}^n} \Rightarrow \xi$ so by uniqueness of weak limits, $\xi = \xi' = \sigma_{\vec{p}}$. □

Therefore the set of possible distributions $\{\sigma_{\vec{p}}\}$ of $X(\vec{p})$ over single-edge strategies $\vec{p}$ almost surely coincides with the set of limit points of the empirical measures $\frac{1}{n}\mu_{0\to n}(\chi)$ over all strategies χ. In particular these have the same extreme points.

3.5 EXTREME POINTS OF $\{\sigma_{\vec{p}}\}$

Our goal in this section will be to characterize the extreme points of the possible distributions of X we can observe as we vary the underlying single-edge strategy. It turns out that the extreme points of $\{\sigma_{\vec{p}} :$ single-edge strategies $\vec{p}\}$ are precisely those $\sigma_{\vec{p}}$ which have scramblings with mean converging to the mean of σ_{MAX}, or, equivalently, those $\sigma_{\vec{p}}$ with deterministic $\vec{p}$.

Theorem 3.28. *Let $\vec{p}$ be a single-edge strategy. The following are equivalent:*
(i) $\sigma_{\vec{p}}$ is an extreme point
(ii) Any symmetric single-edge strategy achieving $\sigma_{\vec{p}}$ must be deterministic
(iii) $f_{\vec{p}}$ is not constant on sets of positive Λ-measure and $\sigma_{\vec{p}}$ is given by the following single-edge strategy $\vec{q}$:

$$q_k(u_1,\dots,u_D) = \mathbf{1}_{\{f_{\vec{p}}(u_k)\geq f_{\vec{p}}(u_i)\ \forall i\}} \text{ for } \Lambda^{\times D}\text{-a.a. } (u_1,\dots,u_D)$$

In other words, $\sigma_{\vec{p}}$ is achieved by the deterministic single-edge strategy "choose whichever weight yields a higher value when evaluating the density $f_{\vec{p}}$"

(iv) $\sup_\phi E[\sigma_{\vec{p}^\phi}] = E[\sigma_{MAX}]$
(v) For $U \sim [0,1]$ *if* $f_{\vec{p}}(U), f_{MAX}(U)$ *as* $[0,D]$*-valued random variables on the probability space* $([0,1], \mathcal{B}([0,1]), \Lambda)$*, then they have the same distribution. That is,*

$$P(f_{\vec{p}} \leqslant x) = P(f_{MAX} \leqslant x)\ \forall x$$

Remark 3.29. Since $f_{MAX}(y) = Dy^{D-1}$ on $[0,1]$ then $\frac{1}{D} f_{MAX} \sim \text{Beta}(D,1)$. Thus (v) implies that for $U \sim \text{Unif}[0,1]$, $f_{\vec{p}}(U)$ has a continuous probability distribution.

We will prove this theorem over the next couple of sections.

3.5.1 *Deterministic Single-Edge Strategies*

We begin by characterizing the extreme points of the set of distributions $\{\sigma_{\vec{p}}\}$ in terms of deterministic single-edge strategies.

Lemma 3.30. *(i)* $\Leftrightarrow$ *(ii) in Theorem 3.28*

Proof. We prove both contrapositives.

First, consider a distribution $\sigma_{\vec{p}}$ achieved by a non-deterministic symmetric single-edge strategy $\vec{p}$, i.e. a single-edge strategy satisfying (3.7)-(3.9). We will construct a set of positive $\Lambda^{\times D}$ measure on which we perturb p_1 in a way that allows us to write $\vec{p}$ as a non-trivial average of two single-edge strategies $\vec{q}, \vec{r}$.

By our assumption, there exists $0 < \epsilon < \frac{1}{2}$ s.t. the set

$$A = \{\vec{u} : p_1(\vec{u}) \in (\epsilon, 1-\epsilon), p_1(\vec{u}) \geqslant p_1(\iota_i(\vec{u}))\ \forall 1 \leqslant i \leqslant D\}$$

has positive $\Lambda^{\times D}$ measure, where ι_i are the cyclic shifts as before. Then there are $a_1 < b_1, a_2 < b_2$ s.t. $[a_1, b_1], [a_2, b_2]$ are disjoint and

$$A' = A \cap [a_1, b_1] \times [a_2, b_2]^{D-1} \text{ has positive } \Lambda^{\times D} \text{ measure}$$

The purpose of this is to ensure that the first coordinate in a tuple in A' cannot appear in any other position in another tuple in A'.

By (3.8), the fact that $p_1(\vec{u}) < 1-\epsilon$ on A', and the definition of single-edge strategies,

$$\sum_{k=2}^{D} p_1(\iota_k(\vec{u})) = \sum_{k=2}^{D} p_k(\vec{u}) = 1 - p_1(\vec{u}) > \epsilon\ \forall \vec{u} \in A'$$

Hence for any $\vec{u} \in A'$, there exists $2 \leqslant k \leqslant D$ s.t. $p_1(\iota_k(\vec{u})) > \frac{\epsilon}{D}$. It follows that there is $2 \leqslant k \leqslant D$ s.t.

$$A'' = \left\{\vec{u} \in A' : p_1(\iota_k(\vec{u})) \in \left(\frac{\epsilon}{D}, 1-\epsilon\right)\right\} \text{ has positive } \Lambda^{\times D} \text{ measure}$$

Define

$$q_1 := p_1 + \frac{\epsilon}{D!} \sum_{\iota \in S_D, \iota(1)=1} (\mathbf{1}_{\iota(A'')} - \mathbf{1}_{\iota(\iota_k^{-1}(A''))}),$$

$$r_1 := p_1 + \frac{\epsilon}{D!} \sum_{\iota \in S_D, \iota(1)=1} (-\mathbf{1}_{\iota(A'')} + \mathbf{1}_{\iota(\iota_k^{-1}(A''))})$$

It is clear that q_1, r_1 average to p_1, and are both invariant under permutations in S_D fixing 1. Since $p_1(\vec{u}) \in (\frac{\epsilon}{D}, 1-\epsilon)$ $\forall \vec{u} \in \iota(A'') \cup \iota(\iota_k^{-1}(A''))$ $\forall \iota \in S_D, \iota(1) = 1$ then both q_1, r_1 have ranges in $[0,1]$. Also, note that

$$\int_{[0,1]^D} \mathbf{1}_{\iota(B)} d\Lambda^{\times D} = \Lambda^{\times D}(B) \;\forall B \in \mathcal{B}([0,1]^D), \iota \in S_D$$

so it follows that

$$\int_{[0,1]^D} q_1 d\Lambda^{\times D} = \int_{[0,1]^D} r_1 d\Lambda^{\times D} = \int_{[0,1]^D} p_1 d\Lambda^{\times D} = \frac{1}{D}$$

Finally, note that by a simple counting argument, for any $\vec{u}$,

$$\sum_i \sum_{\iota \in S_D, \iota(1)=1} \mathbf{1}_{\iota(A'')}(\iota_i(\vec{u})) = \sum_{\iota \in S_D} \mathbf{1}_{\iota(A'')}(\vec{u}) = \sum_i \sum_{\iota \in S_D, \iota(1)=1} \mathbf{1}_{\iota(\iota_k^{-1}(A''))}(\iota_i(\vec{u}))$$

so

$$\sum_i q_1(\iota_i(\vec{u})) = \sum_i r_1(\iota_i(\vec{u})) = \sum_i p_1(\iota_i(\vec{u})) = 1 \;\forall \vec{u}$$

Thus q_1, r_1 are valid probability functions which average to p_1. By Remark 3.25, these uniquely determine single-edge strategies $\vec{q}, \vec{r}$ whose average is $\vec{p}$. Thus $f_{\vec{p}}$ is the average of $f_{\vec{q}}, f_{\vec{r}}$. It remains to show this convex combination is non-trivial.

Consider any $\vec{u} = (u_1, u_2, \ldots, u_D)$ with $u_1 \in [a_1, b_1]$. For any $\iota \in S_D, \iota(1) = 1$,

$$\iota^{-1}(k) \neq 1 \Rightarrow u_{\iota^{-1}(k)} \in [a_2, b_2] \Rightarrow u_{\iota^{-1}(k)} \notin [a_1, b_1] \Rightarrow \vec{u} \notin \iota(\iota_k^{-1}(A''))$$

It follows that

$$q_1(\vec{u}) = p_1(\vec{u}) + \frac{\epsilon}{D!} \sum_{\iota \in S_D, \iota(1)=1} \mathbf{1}_{\iota(A'')}(\vec{u}) \geqslant p_1(\vec{u}) + \frac{\epsilon}{D!} \mathbf{1}_{A''}(\vec{u})$$

so by (3.7),

$$f_{\vec{q}}(u_1) \geqslant f_{\vec{p}}(u_1) + \frac{\epsilon}{(D-1)!} \int_{[0,1]^{D-1}} \mathbf{1}_{A''}(u_1, u_2, \ldots, u_D) \, d\Lambda^{\times(D-1)}(u_2, \ldots, u_D)$$

Since $\Lambda^{\times D}(A'') > 0$ then there is $\delta > 0$ and a Borel set $B \subset [a_1, b_1]$ with $\Lambda(B) > 0$ s.t. the integral above is $\geqslant \delta$ for $u_1 \in B$. It follows that $f_{\vec{q}} \geqslant f_{\vec{p}} + \frac{\epsilon}{(D-1)!}\delta$ on B so the convex combination is nontrivial. Therefore $\sigma_{\vec{p}}$ is not an extreme point.

Now suppose $\sigma_{\vec{p}}$ is not an extreme point, say $\sigma_{\vec{p}} = t\sigma_{\vec{q}} + (1-t)\sigma_{\vec{r}}$ for some symmetric single-edge strategies $\vec{q}, \vec{r}$ with $\sigma_{\vec{q}} \neq \sigma_{\vec{r}}$. Then $\sigma_{\vec{p}}$ is achieved by the single-edge strategy $t\vec{q} + (1-t)\vec{r}$, which trivially satisfies (3.7)-(3.9) so is symmetric. Since $\vec{q}, \vec{r}$ have range in $[0,1]$ and differ on some set of positive $\Lambda^{\times D}$-measure then on this same set $t\vec{q} + (1-t)\vec{r}$ is not $\{0,1\}$-valued. Thus $\sigma_{\vec{p}}$ is achieved by a non-deterministic single-edge strategy satisfying (3.7)-(3.9). □

3.5.2 *"Weight Tuples" of Single-Edge Strategies*

$\sigma_{\vec{p}}$ being an extreme point will give us further properties relating to $(D+1)$-tuples $(u_1, \ldots, u_{D+1})$, but these properties hold a.s. and we must be extra careful about which tuples living in measure 0 sets we are excluding. We need some setup to address this issue.

Let $\vec{p}$ be a symmetric single-edge strategy s.t. $\sigma_{\vec{p}}$ is an extreme point. By Lemma 3.30, p_1 is 0-1 valued $\Lambda^{\times D}$-a.e.. Also, the set of tuples in $[0,1]^D$ with two or more duplicates has $\Lambda^{\times D}$-measure 0. We will let

$$S := \{\text{ordered } (u_1, \ldots, u_D) \in \mathbb{R}^D \text{ duplicate-free s.t. } p_1(u_1, \ldots, u_D) \in \{0,1\}\}$$

henceforth be the set of "good" ordered D-tuples we consider. Note that $\Lambda^D(S) = 1$. Once we prove more a.s. properties of the elements of S, we will amend this definition of S accordingly.

Combining the fact that p_1 is 0-1 valued on S and Remark 3.25, we see that p_1 is equivalent to a choice function taking in *unordered* tuples $(u_1, \ldots, u_D)$ and outputting one of the coordinates u_j; then for any $\iota \in S_D$, $p_1(\iota(\vec{u})) = \mathbf{1}_{\{j=\iota(1)\}}$ i.e. p_1 returns whether or not the first coordinate of the input ordered tuple is the choice associated with the corresponding unordered tuple.

Let

$$T = \left\{ \begin{array}{l} \text{ordered } (u_1, \ldots, u_{D+1}) \in [0,1]^{D+1} \text{ s.t.} \\ \text{all } (D+1)! \text{ ordered subtuples of size } D \text{ are in } S \end{array} \right\}$$

It is clear that $\Lambda^{\times(D+1)}(T) = 1$ and the tuples in T contain no duplicate entries. T is the set of "good" ordered $(D+1)$-tuples we will restrict ourselves to.

For ordered $(D+1)$-tuples in T, we wish to study the possible sets of choices we can make for the $(D+1)!$ ordered subtuples of size D. Since the choice for some $\vec{u} \in [0,1]^D$ is also the choice for $\iota(\vec{u})$ for all $\iota \in S_D$, then rather than keeping a factor of $D!$ everywhere, we will instead consider the possible sets of choices we can make for the $D+1$ *unordered* subtuples of size D. We can encode these choices as an ordered "weight tuple" $(w_1, \ldots, w_{D+1})$ where w_j is the number of times u_j is the choice made. In terms of the symmetric single-edge strategy $\vec{p}$,

$$w_j = w_j(u_1, \ldots, u_{D+1}) := \sum_{1 \leqslant k \leqslant D+1, k \neq j} p_1(u_j, u_1, \ldots, \widehat{u_j}, \ldots, \widehat{u_k}, \ldots, u_{D+1}) \tag{3.10}$$

For example, the weight tuple $(D, 1, 0, \ldots, 0)$ corresponds to the choices

$$(u_1, \ldots, \widehat{u_k}, \ldots, u_{D+1}) \mapsto u_1 \ \forall 2 \leqslant k \leqslant D+1 \text{ and } (u_2, \ldots, u_{D+1}) \mapsto u_2$$

Note that each u_j appears in exactly D of the $D+1$ D-subtuples and there is exactly one choice for each subtuple so

$$w_i \in \{0, 1, \ldots, D\} \ \forall 1 \leqslant i \leqslant D+1, \sum w_i \equiv D+1 \tag{3.11}$$

Even though a weight tuple may correspond to more than one set of choices for the $D+1$ D-subtuples, it is easily checked that every valid weight tuple (satisfying (3.11)) corresponds to at least one set of choices.

The following flow chart summarizes this process:

Start with ordered $(D+1)$-tuple $(u_1, \ldots, u_{D+1}) \in T$

↓

Consider the choice for each unordered D-subtuple $(u_1, \ldots, \widehat{u_k}, \ldots, u_{D+1})$.
Count occurrences of each u_j among these choices.

↓

Get an ordered $(D+1)$-weight tuple $(w_1, \ldots, w_{D+1})$ satisfying (3.11)

One last observation we make about the weight tuples is that they can be used to compute the density $f_{\vec{p}}$ directly. Fix any $1 \leqslant j \leqslant D+1$. Observe that (3.7) with a change of variables gives

$$\begin{aligned} f_{\vec{p}}(t) &= D \int_{[0,1]^{D-1}} p_1(t, u_2, \ldots, u_D) d\Lambda^{D-1}(u_2, \ldots, u_D) \\ &= \sum_{1 \leqslant k \leqslant D+1, k \neq j} \int_{[0,1]^{D-1}} p_1(t, u_1, \ldots, \widehat{u}_j, \ldots, \widehat{u}_k, \ldots, u_{D+1}) \\ &\qquad d\Lambda^{D-1}(u_1, \ldots, \widehat{u}_j, \ldots, \widehat{u}_k, \ldots, u_{D+1}) \end{aligned}$$

But Λ is a probability measure so we can integrate everything against du_k without changing the value:

$$\begin{aligned} & f_{\vec{p}}(t) \\ &= \sum_{1 \leqslant k \leqslant D+1, k \neq j} \int_{[0,1]^D} p_1(t, u_1, \ldots, \widehat{u}_j, \ldots, \widehat{u}_k, \ldots, u_{D+1}) d\Lambda^D(u_1, \ldots, \widehat{u}_j, \ldots, u_{D+1}) \\ &= \int_{[0,1]^D} w_j(u_1, \ldots, t, \ldots, u_{D+1}) d\Lambda^D(u_1, \ldots, \widehat{u}_j, \ldots, u_{D+1}) \text{ by (3.10)} \end{aligned} \tag{3.12}$$

where t occurs in position j. Note that the above formula holds for any t for which

$$\{(u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_{D+1}) | (u_1, \ldots, u_{j-1}, t, u_{j+1}, \ldots, u_{D+1}) \in T\}$$

has $\Lambda^{\times D}$ – measure 1, i.e. it holds for t in a set of Λ-measure 1.

3.5.3 *"Weight Tuples" of the Extreme Points*

We claim that $\sigma_{\vec{p}}$ being an extreme point implies that $\Lambda^{\times(D+1)}$-almost all weight tuples must be a permutation of $(D, 1, 0, \ldots, 0)$. That means that almost all $(D+1)$-tuples in T will have one coordinate that "dominates" the others in terms of the choice function, and this behaviour will mean precisely that $f_{\vec{p}}$ can be approximated by "scrambles" of f_{MAX}.

Lemma 3.31. *Suppose $\vec{p}$ is a symmetric, single-edge strategy s.t. $\sigma_{\vec{p}}$ is an extreme point. For $\Lambda^{\times(D+1)}$-a.a. ordered tuples in T, the corresponding weight tuple is a permutation of $(D, 1, 0, \ldots, 0)$.*

Proof. We will sketch the details of the proof in the case $D \geqslant 3$. The case $D = 2$ is similar.

Suppose the claim is false, say the set U of ordered $(D+1)$-tuples $(u_1, \ldots, u_{D+1}) \in T$ whose weight tuple $(w_1, \ldots, w_{D+1})$ is not a permutation of $(D, 1, 0, \ldots, 0)$ has positive $\Lambda^{\times(D+1)}$-measure. Note that U is clearly invariant under permutations in S_{D+1}.

We will follow a similar approach as in the proof of Lemma 3.30, in that we will seek write $\vec{p}$ as a convex combination of two new single-edge strategies $\vec{q}, \vec{r}$ obtained by perturbing $\vec{p}$.

First, we will further restrict U. There exist $a_i < b_i$ s.t. $[a_i, b_i]$ are pairwise disjoint and s.t.

$$U' := U \cap [a_1, b_1] \times \cdots \times [a_{D+1}, b_{D+1}] \text{ has positive } \Lambda^{\times(D+1)} \text{ measure}$$

In this way, each coordinate in a tuple in U' cannot appear in any other position in another tuple in U'.

Now consider any $(u_1, \ldots, u_{D+1}) \in U'$ so it has a maximal weight $1 \leqslant w_K \leqslant D-1$ in its weight tuple. If $w_K \geqslant 2$ then either there exists $i \neq K$ s.t. $w_i \geqslant 2$ or there are at least two $i \neq K$ s.t. $w_i = 1$ (this is because $\sum w_m = D+1$ and $w_K \leqslant D-1$). In either case, we can pick $I \neq K$ so that the choice for $(u_1, \ldots, \widehat{u_K}, \ldots, u_{D+1})$ is not u_I, hence there is another D-subtuple containing u_K for which the choice is u_I. Furthermore, since $w_K \geqslant 2$, then there also is a D-subtuple containing u_I for which the choice is u_K. On the other hand, if $w_K = 1$ then $w_m = 1\ \forall m$. If $(u_{m_1}, \ldots, u_{m_D})$ is the D-subtuple for which the choice is u_K then of the remaining $D-1 \geqslant 2$ coordinates u_{m_n} with $m_n \neq K$ at least one, call it u_I, is the choice for a D-subtuple other than $(u_1, \ldots, \widehat{u_K}, \ldots, u_{D+1})$, i.e. for a D-subtuple containing u_K.

Since this holds for all $\vec{u} \in U'$ then there exist distinct $1 \leqslant I, J, K, L \leqslant D+1$ s.t.

$$U'' := \left\{ \vec{u} \in U' \;\middle|\; \begin{array}{l} 1 \leqslant w_I, \text{ and } w_m \leqslant w_K\ \forall m, \text{ and} \\ \text{the choice for the } D\text{-subtuple } (u_1, \ldots, \widehat{u_J}, \ldots, u_{D+1}) \text{ is } u_I, \text{ and} \\ \text{the choice for the } D\text{-subtuple } (u_1, \ldots, \widehat{u_L}, \ldots, u_{D+1}) \text{ is } u_K \end{array} \right\}$$

has positive $\Lambda^{\times(D+1)}$ measure.

The idea is that we write the weight tuples for $(D+1)$-tuples in U'' as averages of two different weight tuples, and use these new weight tuples to obtain two new nontrivial, valid single-edge strategies whose densities average to $f_{\vec{p}}$.

Consider any $\vec{u} \in U''$. We can write the corresponding weight tuple $(w_1, \ldots, w_{D+1})$ as an average of two other weight tuples:

$$(\ldots, w_I, \ldots, w_K, \ldots) = \frac{1}{2}(\ldots, w_I + 1, \ldots, w_K - 1, \ldots) + \frac{1}{2}(\ldots, w_I - 1, \ldots, w_K + 1, \ldots)$$

The weight tuple $(\ldots, w_I + 1, \ldots, w_K - 1, \ldots)$ can be achieved simply by changing the choice of $(u_1, \ldots, \widehat{u_L}, \ldots, u_{D+1})$ from u_K to u_I, whereas the weight tuple $(\ldots, w_I - 1, \ldots, w_K + 1, \ldots)$ can be achieved by changing the choice of $(u_1, \ldots, \widehat{u_J}, \ldots, u_{D+1})$ from u_I to u_K. It is trivial to see that these two new tuples are also valid weight tuples that satisfy (3.11) so they correspond to new single-edge strategies $\vec{q}, \vec{r}$ respectively. By (3.12), we see that $f_{\vec{p}}(y)$ is the average of $f_{\vec{q}}(y), f_{\vec{r}}(y)$.

The only difference in the $D = 2$ case is that the new weight tuples will be achieved by making two changes to the choice function rather than one.

It remains to check that this average is non-trivial. Let

$$V := \{u_K : \exists u_1, \ldots, \widehat{u_K}, \ldots, u_{D+1} \text{ s.t. } (u_1, \ldots, u_{D+1}) \in U''\}$$

and for $x \in [0,1]$ let

$$W^x := \{(u_1, \ldots, \widehat{u_K}, \ldots, u_{D+1}) : \ (u_1, \ldots, , u_{K-1}, x, u_{K+1}, \ldots, u_{D+1}) \in U''\}$$

Then $\Lambda^{(D+1)\times}(U'') > 0$ implies $\Lambda(V) > 0$ and for Λ-a.a. $u_K \in V$, $\Lambda^{\times D}(W^{u_K}) > 0$.

Let us consider how the function $w_K(\cdot, \ldots, \cdot, u_K, \cdot, \ldots, \cdot) : [0,1]^D \to \{0, 1, \ldots, D\}$ changes from $\vec{p}$ to $\vec{r}$ for any given fixed $u_K \in V$. On W^{u_K}, $w_K(\cdot, \ldots, \cdot, u_K, \cdot, \ldots, \cdot)$ increases by 1 by construction of $\vec{r}$. Note that u_K can only appear in the Kth coordinate of a tuple in U'' (by construction of U'') hence $w_K(\cdot, \ldots, \cdot, u_K, \cdot, \ldots, \cdot)$ either remains unchanged or increases by 1 off of W^{u_K} (because the only way it decreases by 1 is if u_K had appeared at index I in a tuple in U''). Thus, for any $u_K \in V$,

$$f_{\vec{r}}(u_K) \geqslant f_{\vec{p}}(u_K) + \Lambda^{\times D}(W^{u_K})$$

by (3.12). It follows that $\sigma_{\vec{r}} \neq \sigma_{\vec{p}}$ so this is indeed a nontrivial convex combination. This contradicts the assumption that $\sigma_{\vec{p}}$ was an extreme point. Therefore $\Lambda^{\times(D+1)}$-a.a. weight tuples must be a permutation of $(D, 1, 0, \ldots, 0)$. □

We amend our definition of "good" $(D+1)$-tuples:

$$T' := \{(u_1, \ldots, u_{D+1}) \in T : (w_1, \ldots, w_{D+1}) \text{ is a perm. of } \ (D, 1, 0, \ldots, 0)\}$$

We still have $\Lambda^{\times(D+1)}(T') = 1$ (provided of course that $\sigma_{\vec{p}}$ is an extreme point).

We now prove the (i) $\Rightarrow$ (ii) direction of Theorem 3.28. We split the proof into several claims.

Lemma 3.32. *Let $\sigma_{\vec{p}}$ be an extreme point.*
(i) Suppose $(u_1, \ldots, u_{D+1}) \in T'$ and u_i has weight D in $(u_1, \ldots, u_{D+1})$. Then for

any $1 \leqslant j \leqslant D+1$, $j \neq i$ *and any* u'_j *s.t.* $(u_1, \dots, u_{j-1}, u'_j, u_{j+1}, \dots, u_{D+1}) \in T'$, *either* u_i *or* u'_j *has weight* D *in* $(u_1, \dots, u_{j-1}, u'_j, u_{j+1}, \dots, u_{D+1})$.

(ii) Suppose $(u_1, u'_1, u_2, \dots, u_D) \in T'$. *Then in the corresponding weight tuple,* u'_1 *has weight* D *and* u_1 *has weight 0 if and only if*

$$p_1(u_1, u_2, u_3, \dots, u_D) < p_1(u'_1, u_2, u_3, \dots, u_D)$$

(iii) For $\Lambda^{\times 2}$*-a.a.* (u_1, u'_1), *if*

$$p_1(u_1, u_2, u_3, \dots, u_D) < p_1(u'_1, u_2, u_3, \dots, u_D)$$

on a set of positive $\Lambda^{\times(D-1)}$ *measure then*

$$p_1(u_1, u_2, u_3, \dots, u_D) \leqslant p_1(u'_1, u_2, u_3, \dots, u_D)$$

for $\Lambda^{\times(D-1)}$*-a.a.* $(u_2, \dots, u_D)$.

(iv) For $\Lambda^{\times 2}$*-a.a.* (u_1, u'_1), *if* $f_{\vec{p}}(u_1) \leqslant f_{\vec{p}}(u'_1)$ *then*

$$p_1(u_1, u_2, u_3, \dots, u_D) \leqslant p_1(u'_1, u_2, u_3, \dots, u_D)$$

for $\Lambda^{\times(D-1)}$*-a.a.* $(u_2, \dots, u_D)$.

(v) $f_{\vec{p}}$ *is not constant on any set of positive* Λ *measure.*

(vi) For $\Lambda^{\times D}$*-a.a.* $(u_1, \dots, u_D)$,

$$p_1(u_1, \dots, u_D) \geqslant \mathbf{1}_{\{f_{\vec{p}}(u_1) \geqslant f_{\vec{p}}(u_i)\ \forall i\}}$$

(vii) For $\Lambda^{\times D}$*-a.a.* $(u_1, \dots, u_D)$,

$$p_1(u_1, \dots, u_D) = \mathbf{1}_{\{f_{\vec{p}}(u_1) \geqslant f_{\vec{p}}(u_i)\ \forall i\}}$$

Remark 3.33. We give heuristics to aid in understanding these claims:
(i) If u_i is the dominant choice in $(u_1, \dots, u_{D+1})$ then u_i can never be dominated by one of the u_j for $i \neq j$ in tuples containing u_i and u_j.
(ii) Dominance can be determined by evaluating p_1.]] (iii) If u'_1 dominates u_1 once then u'_1 will always dominate u_1.
(iv) The ordering imposed by domination coincides with the ordering imposed by the values of $f_{\vec{p}}$.
(vii) u_1 is the choice in $(u_1, \dots, u_D)$ if and only of u_1 maximizes the value of $f_{\vec{p}}$.

Proof. (i) Let $1 \leqslant i \leqslant D$ be the index for which u_i has weight D in $(u_1, \dots, u_{D+1})$. Recall that this means we have the choices

$$(u_1, \dots, \widehat{u_\ell}, \dots, u_{D+1}) \mapsto u_i\ \forall \ell \neq i$$

In particular, taking $\ell = j$, this means u_i has weight 1 or D in $(u_1, \dots, u_{j-1}, u'_j, u_{j+1}, \dots, u_{D+1})$ (because this $(D+1)$-tuple is in T'). If it is D, we are done. If it is 1, the fact that we have the choice

$$(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_{D+1}) \mapsto u_i$$

implies no $u_k, k \neq j$ can have weight D so u'_j must have weight D in $(u_1, \dots, u_{j-1}, u'_j, u_{j+1}, \dots, u_{D+1})$.

(ii) Recall that p_1 is 0-1 valued on D-subtuples of tuples in T'. We have the following sequence of if-and-only-if statements:

$$\begin{aligned} & p_1(u_1, u_2, u_3, \dots, u_D) < p_1(u'_1, u_2, u_3, \dots, u_D) \\ \Leftrightarrow\ & p_1(u_1, u_2, u_3, \dots, u_D) = 0, p_1(u'_1, u_2, u_3, \dots, u_D) = 1 \\ \Leftrightarrow\ & (u_1, u_2, u_3, \dots, u_D) \mapsto u_i \text{ for some } 2 \leqslant i \leqslant D \text{ and } (u'_1, u_2, u_3, \dots, u_D) \mapsto u'_1 \end{aligned}$$

Now if u'_1 has weight D and u_1 has weight 0 in $(u_1, u'_1, u_2, \dots, u_D)$, then we must have

$$(u_1, u_2, u_3, \dots, u_D) \mapsto u_i \text{ for some } 2 \leqslant i \leqslant D \text{ and } (u'_1, u_2, u_3, \dots, u_D) \mapsto u'_1 \tag{3.13}$$

On the other hand, if we know (3.13) then the weight of u_i cannot be D since the choice for $(u'_1, u_2, \dots, u_i, \dots, u_D)$ is u'_1. Thus u_i has weight 1, so u'_1 has weight D and u_1 has weight 0.

(iii) Consider u_1, u'_1 s.t. $(u_1, u'_1, u_2, \dots, u_D) \in T'$ for $\Lambda^{\times(D-1)}$-a.a. $(u_2, \dots, u_D)$. This clearly holds for $\Lambda^{\times 2}$-a.a. (u_1, u'_1) with $u_1 \neq u'_1$. Let

$$U = \left\{ (u_2, \dots, u_D) : \begin{array}{l} (u_1, u'_1, u_2, \dots, u_D) \in T', \\ p_1(u_1, u_2, u_3, \dots, u_D) < p_1(u'_1, u_2, u_3, \dots, u_D) \end{array} \right\}$$

so $\Lambda^{\times(D-1)}(U) > 0$ by assumption.

Consider any $(u_2, \dots, u_D) \in U$ and u'_2 s.t. $(u_1, u'_1, u'_2, u_3, \dots, u_D) \in T'$. By (ii), u'_1 has weight D and u_1 has weight 0 in the weight tuple for $(u_1, u'_1, u_2, u_3, \dots, u_D)$. By (i), it follows that either u'_1 or u'_2 has weight D in $(u_1, u'_1, u'_2, u_3, \dots, u_D)$. By the contrapositive of (ii),

$$p_1(u_1, u'_2, u_3, \dots, u_D) \leqslant p_1(u'_1, u'_2, u_3, \dots, u_D)$$

Note that for $\Lambda^{\times 2}$-a.a. (u_1, u'_1) it is true that for $\Lambda^{\times(D-1)}(U) > 0$ of $(u_2, \dots, u_D) \in U$ and Λ-a.a. $u'_2 \in \mathbb{R}$ we have $(u_1, u'_1, u'_2, u_3, \dots, u_D) \in T'$ and the inequality above holds.

We repeat this argument, "replacing" $u_2, u_3, \dots, u_D$ with Λ-almost arbitrary $u'_2, u'_3, \dots, u'_D$ and noting the invariant that one of $u'_1, \dots, u'_k$ has weight of D in the weight tuple of $(u_1, u_1', u_2', \ldots, u_k', u_{k+1}, \ldots, u_D) \in T'$ (which means we may apply (ii) to replace u_{k+1} with Λ-almost arbitrary u_{k+1}'). Thus, for $\Lambda^{\times 2}$-a.a. (u_1, u_1'),

$$p_1(u_1, u_2', u_3', \ldots, u_D') \leqslant p_1(u_1', u_2', u_3', \ldots, u_D')$$

for $\Lambda^{(D-1)}$-a.a. $(u_2', \ldots, u_D')$.

(iv) This follows immediately from (iii) and (3.7).

(v) Suppose $f_{\vec{p}}$ is constant on a set V with $\Lambda(V) > 0$. By (iv), can restrict V to a set V' with $\Lambda(V') = \Lambda(V) > 0$ s.t. for all distinct $u_1, u_2, \ldots, u_{D+1} \in V'$, $(u_1, \ldots, u_{D+1}) \in T'$ and

$$p_1(u_i, u_2, \ldots, \widehat{u}_i, \ldots, \widehat{u}_j, \ldots, u_{D+1}) = p_1(u_j, u_2, \ldots, \widehat{u}_i, \ldots, \widehat{u}_j, \ldots, u_{D+1}) \ \forall i, j$$

We claim this is impossible.

Consider distinct $u_1, \ldots, u_{D+1} \in V'$. Suppose u_i has weight D and u_k has weight 1 in the corresponding weight tuple. Pick $j \notin \{i, k\}$ (which exists since $D + 1 \geqslant 3$). Then u_j has weight 0 in the corresponding weight tuple. But this implies the contradiction

$$1 = p_1(u_i, u_2, \ldots, \widehat{u}_i, \ldots, \widehat{u}_j, \ldots, u_{D+1}) = p_1(u_j, u_2, \ldots, \widehat{u}_i, \ldots, \widehat{u}_j, \ldots, u_{D+1}) = 0$$

Therefore $f_{\vec{p}}$ is not constant on any set of positive Λ-measure.

(vi) It suffices to show that for Λ-a.a. u_1, we have

$$p_1(u_1, u_2, \ldots, u_D) = 1$$

for all but a $\Lambda^{\times(D-1)}$-null subset of tuples $(u_2, \ldots, u_D) \in \left(f_{\vec{p}}^{-1}([0, f_{\vec{p}}(u_1)])\right)^{D-1}$.

Suppose not, say there is a set G of u_1 of Λ-positive measure on which

$$p_1(u_1, u_2, \ldots, u_D) = 0$$

for a tuples $(u_2, \ldots, u_D)$ in a set $H_{u_1} \subseteq \left(f_{\vec{p}}^{-1}([0, f_{\vec{p}}(u_1)])\right)^{D-1}$ of $\Lambda^{\times(D-1)}$-positive measure. By (iv), for Λ-a.a. $u_1 \in G$,

$$0 = p_1(u_1, u_2, \ldots, u_D) \geqslant p_1(u_{D+1}, u_2, \ldots, u_D)$$

for Λ-a.a. $u_{D+1} \in f_{\vec{p}}^{-1}([0, f_{\vec{p}}(u_1)])$ and $\Lambda^{\times(D-1)}$-a.a. $(u_2, \ldots, u_D) \in H_{u_1}$. In particular, there exists $2 \leqslant i \leqslant D$ and a subset $H_{u_1}' \subset H_{u_1}$ of positive

Λ^D-measure s.t. the choice for $(u_1, u_2, \ldots, u_D)$ is u_i and s.t. for Λ-a.a. $u_{D+1} \in f_{\vec{p}}^{-1}([0, f_{\vec{p}}(u_1)])$ we have

$$p_1(u_1, u_2, \ldots, u_D) = p_1(u_{D+1}, u_2, \ldots, u_D) = 0 \tag{3.14}$$

For each such tuple consider the weight of u_i in $(u_1, \ldots, u_{D+1})$; it must be 1 or D and if it were 1 then the choice $(u_1, u_2, \ldots, u_D) \mapsto u_i$ implies none of $u_2, \ldots, u_D$ can have weight D so u_{D+1} must have weight D, contradicting $p_1(u_{D+1}, u_2, \ldots, u_D) = 0$. Thus u_i has weight D in $(u_1, \ldots, u_{D+1})$ so we must have the choice $(u_{D+1}, u_2, \ldots, u_D) \mapsto u_i$. Therefore

$$p_1(u_i, u_2, \ldots, u_{i-1}, u_{D+1}, u_{i+1}, \ldots, u_D) = 1 \tag{3.15}$$

Since (3.14) holds for Λ-a.a. $u_{D+1} \in f_{\vec{p}}^{-1}([0, f_{\vec{p}}(u_1)])$ and $\Lambda^{\times(D-1)}$-a.a. $(u_2, \ldots, u_D) \in H'_{u_1}$, then there is a set of tuples $(u_2, \ldots, u_{D+1})$ of positive $\Lambda^{\times D}$-measure for which (3.14) (hence (3.15)) holds for both $(u_2, \ldots, u_{D+1})$ and $(u_2, \ldots, u_{i-1}, u_{D+1}, u_{i+1}, \ldots, u_D, u_i)$ (i.e. it holds with the values u_i, u_{D+1} swapped). Combining the two sets of (3.14), (3.15) gives the contradiction

$$0 = p_1(u_i, u_2, \ldots, u_{i-1}, u_{D+1}, u_{i+1}, \ldots, u_D) = 1$$

(vii) By (v) and (vi), for $\Lambda^{\times D}$-a.a. $(u_1, \ldots, u_D)$, $f_{\vec{p}}(u_i)$ are distinct and

$$p_1(u_k, u_1, \ldots, \widehat{u}_k, \ldots, u_D) \geqslant \mathbf{1}_{\{f_{\vec{p}}(u_k) \geqslant f_{\vec{p}}(u_i)\ \forall i\}}\ \forall k$$

Consider any such $(u_1, \ldots, u_D)$ and let u_k maximize $f_{\vec{p}}$. Then

$$p_1(u_k, u_1, \ldots, \widehat{u}_k, \ldots, u_D) = \mathbf{1}_{\{f_{\vec{p}}(u_k) \geqslant f_{\vec{p}}(u_i)\ \forall i\}} = 1$$

so the choice for $(u_1, \ldots, u_D)$ is u_k. It follows that for all other $j \neq k$,

$$p_1(u_j, u_1, \ldots, \widehat{u}_j, \ldots, u_D) = 0 = \mathbf{1}_{\{f_{\vec{p}}(u_j) \geqslant f_{\vec{p}}(u_i)\ \forall i\}}$$

We have shown that the a.s. inequality in (vi) is an a.s. equality. □

The next lemma shows that (ii) implies (iii) in Theorem 3.28.

Lemma 3.34. *Suppose $\vec{p}$ is a single-edge strategy s.t.*
(i) $f_{\vec{p}}$ is not constant on any set of positive Λ measure, and
(ii) $p_k(u_1, \ldots, u_D) = \mathbf{1}_{\{f_{\vec{p}}(u_k) \geqslant f_{\vec{p}}(u_i)\ \forall i\}}$ for $\Lambda^{\times D}$-a.a. $(u_1, \ldots, u_D)$ $\forall 1 \leqslant k \leqslant D$
Then $\sup_\phi E[\sigma_{\vec{p}^\phi}] = E[\sigma_{MAX}]$ where the supremum is taken over Λ-measure-preserving bijections $\phi : [0,1] \to [0,1]$.

Proof. We begin by constructing a sequence (ϕ_n) of Λ-measure-preserving bijections for which

$$\mathbf{1}_{\{f_{\vec{p}}(\phi_n(u_1)) \geqslant f_{\vec{p}}(\phi_n(u_i))\ \forall i\}} \to \mathbf{1}_{\{u_1 \geqslant u_i\ \forall i\}}\ \Lambda^{\times D}\text{-a.s.} \tag{3.16}$$

Consider any $n \in \mathbb{N}$. By (i), inverse images of singletons $f_{\vec{p}}(\{c\})$ have Λ-measure 0. There exist $0 = a_0^n \leqslant a_1^n \leqslant \cdots \leqslant a_{D2^n}^n = 1$ s.t.

$$\Lambda([a_i^n, a_{i+1}^n)) = \Lambda\left(f_{\vec{p}}^{-1}\left(\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)\right)\right)\ \forall 0 \leqslant i \leqslant D2^n - 1 \tag{3.17}$$

The indices stop at $D2^n - 1$ since the range of $f_{\vec{p}}$ is a subset of $[0, D]$. Clearly, we can make it so that the partition for $n+1$ refines the one for n, for all n.

We construct ϕ_n by pasting together Λ-measure-preserving bijections we get from Theorem 3.12 between sets

$$[a_i^n, a_{i+1}^n) \to f_{\vec{p}}^{-1}\left(\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)\right)$$

for $0 \leqslant i \leqslant D2^n - 1$ for which $f_{\vec{p}}^{-1}([\frac{i}{2^n}, \frac{i+1}{2^n}))$ has positive Λ-measure. Then for each such i, $f_{\vec{p}\phi_n} = f_{\vec{p}} \circ \phi_n$ is a map

$$[a_i^n, a_{i+1}^n) \to \left[\frac{i}{2^n}, \frac{i+1}{2^n}\right)$$

Let us show (3.16). It suffices to show that for $\Lambda^{\times 2}$-a.a. (u_1, u_2) we have

$$\mathbf{1}_{\{f_{\vec{p}}(\phi_n(u_1)) \geqslant f_{\vec{p}}(\phi_n(u_2))\}} \to \mathbf{1}_{\{u_1 \geqslant u_2\}}$$

since multiplying $D-1$ of these sequences of indicators gives (3.16).

By (i), the sets $f_{\vec{p}}^{-1}(\{c\})$ have Λ-measure 0 hence it follows that for $\Lambda^{\times 2}$-a.a. (u_1, u_2) we have $f_{\vec{p}}(u_1) \neq f_{\vec{p}}(u_2)$. Consider any such u_1, u_2. Since $\Lambda([a_i^n, a_{i+1}^n)) \downarrow 0$ as $n \to \infty$ (because $\Lambda <<$ Lebesgue measure), then for large enough n, u_1 and u_2 are in different intervals of the form $[a_i^n, a_{i+1}^n)$ and they will stay in these intervals (because the partitions get progressively more refined). Thus $f_{\vec{p}} \circ \phi_n(u_1), f_{\vec{p}} \circ \phi_n(u_2)$ are in different intervals of the form $[\frac{i}{2^n}, \frac{i+1}{2^n})$ and the order is preserved:

$$u_1 < u_2 \Leftrightarrow u_1 \in [a_{i_1}^n, a_{i_1+1}^n), u_2 \in [a_{i_2}^n, a_{i_2+1}^n) \text{ with } i_1 < i_2 \Leftrightarrow f_{\vec{p}} \circ \phi_n(u_1) < f_{\vec{p}} \circ \phi_n(u_2)$$

Therefore we get (3.16).

Recall that $\vec{p}^{\phi_n}$ are themselves symmetric single-edge strategies. Using (ii) and (3.5), it follows that for any $t \in [0,1]$,

$$\begin{aligned} 1 - F_{\vec{p}^{\phi_n}}(t) &= D \int_{[t,1]\times[0,1]^{D-1}} p_1(\phi_n(\vec{u}))d\Lambda^D(u_1,\dots,u_D) \\ &= D \int_{[t,1]\times[0,1]^{D-1}} \mathbf{1}_{\{f_{\vec{p}}(\phi_n(u_1)) \geqslant f_{\vec{p}}(\phi_n(u_i))\ \forall i\}} d\Lambda^D(u_1,\dots,u_D) \end{aligned} \tag{3.18}$$

By (3.16) and the Bounded Convergence Theorem, this last expression converges to the integral of the density f_{MAX}:

$$\begin{aligned} &D \int_{[t,1]\times[0,1]^{D-1}} \mathbf{1}_{\{f_{\vec{p}}(\phi_n(u_1)) \geqslant f_{\vec{p}}(\phi_n(u_i))\ \forall i\}} d\Lambda^D(u_1,\dots,u_D) \\ &\to D \int_{[t,1]\times[0,1]^{D-1}} \mathbf{1}_{\{u_k \geqslant u_i\ \forall i\}} d\Lambda^D(u_1,\dots,u_D) \\ &= \int_t^1 f_{MAX}(u_1)du_1 \\ &= 1 - F_{MAX}(t) \end{aligned}$$

But we already established that $F_{\vec{q}} \geqslant F_{MAX}$ for any single-edge strategy $\vec{q}$. Therefore $F_{\vec{p}^{\phi_n}} \to F_{MAX}$ pointwise from below hence $E[\sigma_{\vec{p}^{\phi_n}}] \to E[\sigma_{MAX}]$ by the Bounded Convergence Theorem combined with the tail integral formula for expectation. □

3.5.4 *Theorem 3.28*

We will prove the rest of Theorem 3.28, but first we make some useful observations about σ_{MAX}.

Lemma 3.35. *Let f^{-}_{MAX} be the quantile function*

$$f^{-}_{MAX}(t) = \inf\{x : t \leqslant f_{MAX}(x)\}$$

Then treating f_{MAX} itself as a $[0,D]$-valued random variable on the probability space
$([0,1], \mathcal{B}([0,1]), \Lambda)$,

$$P(f_{MAX} \leqslant t) = f^{-}_{MAX}(t)\ \forall t$$

In particular,

$$P(f_{MAX} \leqslant f_{MAX}(y)) = y\ \forall y$$

Proof. This is trivial since the density $f_{MAX}(y) = Dy^{D-1}$ is invertible so the quantile function f^{-}_{MAX} is just its inverse.

□

To prove Theorem 3.28, we need one more short lemma exploring what $\sup_{\phi} E[\sigma_{\vec{p}^{\phi}}] = E[\sigma_{MAX}]$ tells us.

Lemma 3.36. *Let $\vec{p}$ be a single-edge strategy and $\phi_n : [0,1] \to [0,1]$ be Λ-measure-preserving bijections s.t. $E[\sigma_{\vec{p}^{\phi}}] \uparrow E[\sigma_{MAX}]$. Then there is a subsequence ϕ_{n_j} s.t. $F_{\vec{p}^{\phi_{n_j}}}(y) \to F_{MAX}(y)$, $f_{\vec{p}^{\phi_{n_j}}}(y) = f_{\vec{p}} \circ \phi_{n_j}(y) \to f_{MAX}(y)$ Λ-a.e. and almost uniformly on compact subsets of $[0,1]$.*

Proof. By the tail integral formula for expectation,

$$\int_0^1 F_{\vec{p}^{\phi_n}}(t) - F_{MAX}(t)dt = E[\sigma_{MAX}] - E[\sigma_{\vec{p}^{\phi}}] \downarrow 0$$

where the integrands $F_{\vec{p}^{\phi_n}}(t) - F_{MAX}(t)$ are non-negative from Lemma 3.22 (i), hence this is L^1 convergence. It follows that there exists a subsequence n_j s.t. $F_{\vec{p}^{\phi_{n_j}}}(t) \to F_{MAX}(t)$ a.e. and a.u. on compact subsets of $[0,1]$. The latter convergence gives us that $f_{\vec{p}^{\phi_{n_j}}}(t) \to f_{MAX}(t)$ a.e. and a.u. on compact subsets of $[0,1]$. □

We are now ready to prove Theorem 3.28 in its entirety.

Theorem 3.28. Let $\vec{p}$ be a single-edge strategy. The following are equivalent:
(i) $\sigma_{\vec{p}}$ is an extreme point
(ii) Any symmetric single-edge strategy achieving $\sigma_{\vec{p}}$ must be deterministic
(iii) $f_{\vec{p}}$ is not constant on sets of positive Λ-measure and $\sigma_{\vec{p}}$ is given by the following single-edge strategy $\vec{q}$:

$$q_k(u_1, \dots, u_D) = \mathbf{1}_{\{f_{\vec{p}}(u_k) \geq f_{\vec{p}}(u_i)\ \forall i\}} \text{ for } \Lambda^{\times D}\text{-a.a. } (u_1, \dots, u_D)$$

In other words, $\sigma_{\vec{p}}$ is achieved by the deterministic single-edge strategy "choose whichever weight yields a higher value when evaluating the density $f_{\vec{p}}$"
(iv) $\sup_{\phi} E[\sigma_{\vec{p}^{\phi}}] = E[\sigma_{MAX}]$
(v) For $U \sim [0,1]$ if we treat $f_{\vec{p}}(U), f_{MAX}(U)$ as $[0,D]$-valued random variables on the probability space $([0,1], \mathcal{B}([0,1]), \Lambda)$, then they have the same distribution. That is,

$$P(f_{\vec{p}} \leq t) = P(f_{MAX} \leq t) = \left(\frac{t}{D}\right)^{\frac{1}{D-1}} \ \forall t \in [0,1]$$

Proof. (i) ⇔ (ii), (i) ⇒ (iii), (iii) ⇒ (iv) These are given by Lemmas 3.30, 3.32, 3.34.

(iv) ⇒ (i) Suppose (iv) holds but $\sigma_{\vec{p}}$ is not an extreme point. Since Radon-Nikodym derivatives are additive, then

$$f_{\vec{p}} = \alpha f_{\vec{q}} + (1-\alpha) f_{\vec{r}}$$

for $\alpha \in (0,1)$ and single-edge strategies $\vec{q}, \vec{r}$ s.t. $f_{\vec{p}} \neq f_{\vec{q}}$ on a set of positive Λ measure.

Take ϕ_n Λ-measure-preserving bijections s.t. $E[\sigma_{\vec{p}^{\phi_n}}] \uparrow E[\sigma_{MAX}]$. Then

$$\alpha E[\sigma_{\vec{q}^{\phi_n}}] + (1-\alpha) E[\sigma_{\vec{r}^{\phi_n}}] \uparrow \alpha E[\sigma_{MAX}] + (1-\alpha) E[\sigma_{MAX}]$$

where $E[\sigma_{\vec{q}^{\phi_n}}], E[\sigma_{\vec{r}^{\phi_n}}] \leqslant E[\sigma_{MAX}]$. It follows that

$$E[\sigma_{\vec{q}^{\phi_n}}], E[\sigma_{\vec{r}^{\phi_n}}] \uparrow E[\sigma_{MAX}]$$

By Lemma 3.36, there is a subsequence ϕ_{n_j} s.t.

$$f_{\vec{q}} \circ \phi_{n_j}, f_{\vec{r}} \circ \phi_{n_j} \to f_{MAX}$$

Λ-a.e. and Λ-a.u. on compact subsets of $[0,1]$.

Since $f_{\vec{q}} \neq f_{\vec{r}}$ on a set of positive Λ-measure then $\exists \epsilon > 0$ s.t.

$$Q := \{x : |f_{\vec{q}}(x) - f_{\vec{r}}(x)| \geqslant \epsilon\}$$

has measure $\delta := \Lambda(Q) > 0$. Now, by the choice of ϕ_{n_j}, there is a subset $R \subset [0,1]$ of Λ measure $\leqslant \frac{\delta}{2}$ s.t. $f_{\vec{q}} \circ \phi_{n_j}, f_{\vec{r}} \circ \phi_{n_j} \to f_{MAX}$ uniformly on $[0,1] \backslash R$. In particular, $\exists J \in \mathbb{N}$ s.t. for all $j \geqslant J$ and $y \in [0,1] \cap R$ we have

$$|f_{\vec{q}} \circ \phi_{n_j}(y) - f_{MAX}(y)|, |f_{\vec{r}} \circ \phi_{n_j}(y) - f_{MAX}(y)| < \frac{\epsilon}{2} \Rightarrow |f_{\vec{q}} \circ \phi_{n_j}(y) - f_{\vec{r}} \circ \phi_{n_j}(y)| < \epsilon$$

Thus

$$\phi_{n_J}([0,1] \backslash R) \subseteq [0,1] \backslash Q \tag{3.19}$$

But ϕ_{n_J} is Λ-measure-preserving and

$$\begin{aligned} \Lambda(\phi_{n_J}([0,1] \backslash R)) &= \Lambda([0,1]) - \Lambda(R) \\ &\geqslant \Lambda([0,1] - \frac{\delta}{2} \\ &> \Lambda([0,1]) - \Lambda(Q) \\ &= \Lambda([0,1] \backslash Q) \end{aligned}$$

Contradiction of (3.19). Therefore $\sigma_{\vec{p}}$ is an extreme point.

(iv) $\Rightarrow$ (v) Take ϕ_n Λ-measure-preserving bijections s.t. $E[\sigma_{\vec{p}^{\phi_n}}] \uparrow E[\sigma_{MAX}]$. By Lemma 3.36 there is a subsequence ϕ_{n_j} s.t. $f_{\vec{p}^{\phi_{n_j}}} = f_{\vec{p}} \circ \phi_{n_j} \to f_{MAX}$ Λ-a.e and a.u. on compact subsets of $[0,1]$.

Now for any $z \in [0,1]$,

$$P(f_{\vec{p}} \leqslant z) = \Lambda(f_{\vec{p}}^{-1}([0,z])) = \Lambda((f_{\vec{p}^{\phi_{n_j}}})^{-1}([0,z]))$$

We claim this last expression equals $\Lambda(f_{MAX}^{-1}([0,z])) = P(f_{MAX} \leqslant z)$.

Suppose $P(f_{\vec{p}} \leqslant z) \neq P(f_{MAX} \leqslant z)$ for some z, say
$P(f_{\vec{p}} \leqslant z) < \Lambda(f_{MAX}^{-1}([0,z]))$; the $>$ case is very similar. By continuity from below and since $\Lambda(f_{MAX}^{-1}(\{z\})) = 0$, there exists $\delta > 0$ s.t.

$$\Lambda(f_{\vec{p}}^{-1}([0,z])) < \Lambda(f_{MAX}^{-1}([0,z-\delta]))$$

By a.u. convergence on $[0,1]$, there is a subset $R \subset [0,1]$ off which $f_{\vec{p}} \circ \phi_{n_j} \to f_{MAX}$ uniformly s.t. $\Lambda(R)$ is sufficiently small so that for all j,

$$\Lambda((f_{\vec{p}} \circ \phi_{n_j})^{-1}([0,z])) = \Lambda(f_{\vec{p}}^{-1}([0,z])) < \Lambda(f_{MAX}^{-1}([0,z-\delta]) \cap [0,1] \backslash R) \tag{3.20}$$

Uniform convergence gives a J s.t. for all $j \geqslant J$ and all
$t \in f_{MAX}^{-1}([0,z-\delta]) \cap [0,1] \backslash R$,

$$f_{\vec{p}} \circ \phi_{n_j}(t) \in (f_{MAX}(t) - \delta, f_{MAX}(t) + \delta) \cap [0,\infty) \subseteq [0,z]$$

(here we use the fact that $f_{\vec{p}}$ is a non-negative function). Thus

$$f_{MAX}^{-1}([0,z-\delta]) \cap [0,1] \backslash R \subseteq (f_{\vec{p}} \circ \phi_{n_j})^{-1}([0,z])$$

which is a contradiction of (3.20). Therefore

$$P(f_{\vec{p}} \leqslant z) = P(f_{MAX} \leqslant z) = f_{MAX}^{-}(z)$$

where the last inequality holds by Lemma 3.35. The expression on the right is precisely $(\frac{z}{D})^{\frac{1}{D-1}}$.

(v) $\Rightarrow$ (iv) We will construct Λ-measure-preserving bijections $\phi_n : [0,1] \to [0,1]$ s.t.
$E[\sigma_{\vec{p}^{\phi_n}}] \to E[\sigma_{MAX}]$.

Consider any n. By (v) combined with Lemma 3.35, $\forall z \in \{0, 1, \ldots, 2^n - 1\}$ we have

$$\begin{aligned}
&\Lambda\left(f_{\vec{p}}^{-1}\left(\left[f_{MAX}\left(\frac{z}{2^n}\right), f_{MAX}\left(\frac{z+1}{2^n}\right)\right)\right)\right) \\
&= P\left(f_{MAX}\left(\frac{z}{2^n}\right) \leqslant f_{\vec{p}} < f_{MAX}\left(\frac{z+1}{2^n}\right)\right) \\
&= P\left(f_{MAX}\left(\frac{z}{2^n}\right) \leqslant f_{MAX} < f_{MAX}\left(\frac{z+1}{2^n}\right)\right) \\
&= \Lambda\left(\left[\frac{z}{2^n}, \frac{z+1}{2^n}\right)\right)
\end{aligned}$$

We construct ϕ_n by pasting together Λ-measure-preserving bijections we get from Theorem 3.12 between sets

$$\left[\frac{z}{2^n}, \frac{z+1}{2^n}\right) \to f_{\vec{p}}^{-1}\left(\left[f_{MAX}\left(\frac{z}{2^n}\right), f_{MAX}\left(\frac{z+1}{2^n}\right)\right)\right)$$

Then for each z, $f_{\vec{p}^{\phi_n}} = f_{\vec{p}} \circ \phi_n$ is a map

$$\left[\frac{z}{2^n}, \frac{z+1}{2^n}\right) \to \left[f_{MAX}\left(\frac{z}{2^n}\right), f_{MAX}\left(\frac{z+1}{2^n}\right)\right)$$

We now compute the expectation of each $X(\vec{p}^{\phi_n})$:

$$\begin{aligned}
E[X(p^{MAX})] &\geqslant E[X_0(\vec{p}^{\phi_n})] \\
&= \sum_{z=0}^{2^n-1} \int_{\frac{z}{2^n}}^{\frac{z+1}{2^n}} f_{\vec{p}}(\phi_n(t))t\, dt \\
&\geqslant \sum_{z=0}^{2^n-1} \int_{\frac{z}{2^n}}^{\frac{z+1}{2^n}} f_{MAX}\left(\frac{z}{2^n}\right)t\, dt
\end{aligned}$$

But $f_{MAX}(t)t$ is nondecreasing hence is of bounded variation on $[0,1]$ hence this lower bound converges to $E[X(p^{MAX})]$ as $n \to \infty$. Thus

$$E[\sigma_{MAX}] = E[X(p^{MAX})] = \sup_n E[X(\vec{p}^{\phi_n})] = \sup_\phi E[\sigma_{\vec{p}^\phi}]$$

as desired. □

By the Krein-Milman Theorem, $\mathcal{R}$ is exactly the convex hull of its extreme points. Thus we obtain a description of the set of limit points of empirical measures.

As a final remark, observe that we only used the compactness of the space $\mathcal{M}_1$ of probability measures on [0,1] to reduce from generic strategies to single-edge strategies. Furthermore, the fact that Λ has nice formulas for its cdf and pdf was convenient but unnecessary. The proof of this theorem could be tweaked to hold with Λ replaced by an arbitrary distribution θ on $\mathbb{R}$ with finite mean. We could thus get a similar characterization for the extreme points of $\{\sigma_\psi$: single-edge strategies $\psi\}$. The only caveats would be that this set might not coincide with {limit points of $\frac{1}{n}\mu_{0\to n}(\chi)$: strategies $\chi\}$ and that the value distribution of the densities $f_{\vec{p}}$ of extreme points might not have as nice a form as $D \cdot Beta(1, D)$.

3.5.5 *The Discrete Case with Example*

The same argument with the weight tuples can be used to show a discrete version of Theorem 3.28, where the i.i.d. labels U^j are $Unif\{1, \ldots, K\}$. In this modified setting, the convex set of symmetric single-edge strategies is the permutohedron of order K and there is a natural extreme point-preserving bijection between the permutohedron and $\mathcal{R}$, the set of achievable distributions $\{\sigma_{\vec{p}}\}$.

As an example, let us work through the $D = 2, K = 4$ discrete case.

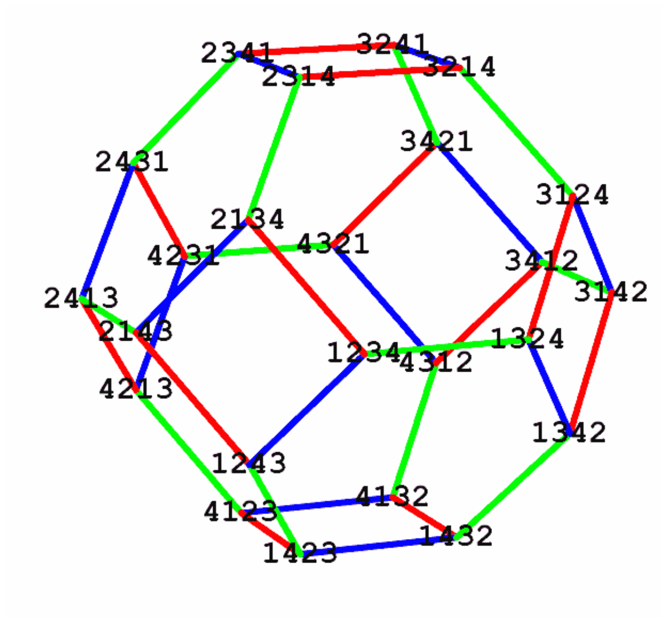


Figure 3.2: Permutohedron of order 4, [Hol]

The extreme points of the permutohedron of order 4 coincide with the extreme points of the set of symmetric single-edge strategies as follows: 1342 corresponds to the single-edge strategy choosing whichever of the samples $u^1, u^2 \in \{1,2,3,4\}$ is maximal with respect to the ordering $1 < 3 < 4 < 2$. It is plain to see that all the deterministic symmetric single-edge strategies correspond to such an ordering.

The $\vec{p}^{MAX}$ strategy here corresponds to the ordering 1234 and yields the distribution

$$\sigma_{MAX} = \frac{1}{16}\delta_1 + \frac{3}{16}\delta_2 + \frac{5}{16}\delta_3 + \frac{7}{16}\delta_4$$

The discrete version of Theorem 3.28 gives that the set of extreme points of $\mathcal{R}$ are those given by deterministic symmetric-edge strategies, and that their density have the same value distribution as f_{MAX}. That is, we get a natural bijection between the extreme points of the permutohedron and of $\mathcal{R}$:

$$1342 \mapsto \sigma_{1342} = \frac{1}{16}\delta_1 + \frac{3}{16}\delta_3 + \frac{5}{16}\delta_4 + \frac{7}{16}\delta_2$$

Since convex combinations of single-edge strategies translate to convex combinations of the distributions $\sigma_{\vec{p}}$ then this bijection extends to a bijection between the permutohedron and $\mathcal{R}$. Of course, the permutohedron is the same for different D but the bijection depends on D.

3.6 GRID ENTROPY IN THIS MODEL

In this short section, we compute the grid entropy of the extreme points described in the previous section and we describe grid entropy in general using a simplified formula for the Gibbs Free Energy.

3.6.1 *Grid Entropy of Extreme Points*

Recall that the extreme points of $\{\sigma_{\vec{p}}\}$ are given by the single-edge strategy of choosing whichever edge label maximizes the density $f_{\vec{p}}$. We show that such extreme points $\sigma_{\vec{p}}$ have grid entropy 0.

However, we first need a short lemma about the partial averages of the expectations of order statistics being bounded away from 0.

Lemma 3.37. *Let $Z_i \sim \theta, i \geqslant 0$ be i.i.d. random variables s.t. θ is a distribution on $[0,\infty)$ with cdf F_θ satisfying $F_\theta(0) = 0$. Consider the order statistics $Z_{0:n} \leqslant \cdots \leqslant Z_{n:n}$. Then there exists a constant $C > 0$ s.t. for large enough n,*

$$\frac{1}{n}\sum_{k=0}^{\lfloor n\epsilon\rfloor} E[Z_{k:n}] \geqslant C$$

Remark 3.38. We allow for the case where some of these expected values may be ∞.

Proof. Observe that half the terms in this sum are bounded below by $E[Z_{\lfloor \frac{n\epsilon}{2} \rfloor:n}]$ hence

$$\frac{1}{n}\sum_{k=0}^{\lfloor n\epsilon \rfloor} E[Z_{k:n}] \geqslant \frac{1}{n}\left\lfloor \frac{n\epsilon}{2} \right\rfloor E[Z_{\lfloor \frac{n\epsilon}{2} \rfloor:n}]$$

Thus it suffices to find $C > 0$ s.t. for large n, $E[Z_{\lfloor \frac{n\epsilon}{2} \rfloor:n}] \geqslant \frac{1}{2}C$.

By right-continuity of the cdf F_θ, we can take $C > 0$ s.t. $F_\theta(C) \leqslant \frac{1}{8}\epsilon$.

Consider the i.i.d. Bernoulli random variables $\mathbf{1}_{\{Z_i \leqslant C\}}$ with success probability $F_\theta(C)$. Then Markov's Inequality yields

$$P(Z_{\lfloor \frac{n\epsilon}{2} \rfloor:n} \leqslant C) = P\left(\#\text{successes in Bin}(n, F_\theta(C)) \geqslant \left\lfloor \frac{n\epsilon}{2} \right\rfloor\right) \leqslant \frac{nF_\theta(C)}{\lfloor \frac{n\epsilon}{2} \rfloor} \leqslant \frac{1}{2}$$

for large n by the choice of C. It follows that

$$E[Z_{\lfloor \frac{n\epsilon}{2} \rfloor:n}] \geqslant \frac{1}{2}C$$

for large n, which completes the proof. □

Theorem 3.39. *Fix $\tau : [0,1] \to [0,\infty)$ measurable and bounded s.t. τ is not constant on sets of positive Λ measure, and consider the single-edge strategy of picking the maximal $\tau(U^j)$ over $1 \leqslant j \leqslant D$, given by*

$$\vec{p}(u_1, \dots, u_D) = \mathbf{1}_{\{\tau(u_k) \geqslant \tau(u_i)\ \forall i\}}$$

Then $||\sigma_{\vec{p}}|| = 0$.

Remark 3.40. The types of strategies considered in this theorem are deterministic hence the resulting $\sigma_{\vec{p}}$ are all extreme points by Theorem 3.28. On the other hand, Theorem 3.28 establishes that all extreme points can be realized as $\sigma_{\vec{p}}$ for a single-edge strategy $\vec{p}$ choosing whichever observed label maximizes the density $f_{\vec{p}} \in [0, D]$ where $f_{\vec{p}}(\text{Unif}[0,1])$ is the $D \cdot \text{Beta}(D, 1)$ distribution. Thus Theorem 3.39 captures what happens for all extreme points.

Proof. Suppose $||\sigma_{\vec{p}}|| > \delta > 0$. Let $\alpha = \frac{\delta}{\ln 2}$. For $n \in \mathbb{N}$ and $1 \leqslant m_n \leqslant \lfloor e^{n\delta} \rfloor = \lfloor 2^{n\alpha} \rfloor$ consider the event-dependent paths π_{n,m_n} corresponding to

$$\overset{m_n}{\underset{\pi:0\to n}{\min}}\, \rho\left(\frac{1}{n}\mu_\pi, \sigma_{\vec{p}}\right)$$

Since $\delta < ||\sigma_{\vec{p}}||$ then by definition of grid entropy,

$$\overset{\lfloor e^{n\delta} \rfloor}{\underset{\pi:0\to n}{\min}}\, \rho\left(\frac{1}{n}\mu_\pi, \sigma_{\vec{p}}\right) \to 0 \text{ a.s.}$$

hence $\frac{1}{n}\mu_{\pi_{n,m_n}} \Rightarrow \sigma_{\vec{p}}$ a.s. regardless of the sequence (m_n).

Also for $i \geqslant 0$ define the random variables

$$Y_i := \max_{1\leqslant j\leqslant D}(\tau(U_i^j)), Y_i' = \text{2nd} \max_{1\leqslant j\leqslant D}(\tau(U_i^j)), Z_i := Y_i - Y_i'$$

Now it is a classic result that for $\epsilon > 0$,

$$\sum_{i=0}^{\lfloor n\epsilon\rfloor}\binom{n}{i} \leqslant 2^{nL} \text{ with } L = L(\epsilon) = \epsilon\log\epsilon + (1-\epsilon)\log(1-\epsilon)$$

Take $\epsilon > 0$ small enough so that $L + \epsilon\log(D-1) < \frac{1}{2}\alpha$ and take $N \in \mathbb{N}$ so that $\forall n \geqslant N$, $2^{nL}(D-1)^{n\epsilon} < \lfloor 2^{n\alpha}\rfloor$.

Consider any $n \geqslant N$. Then the number of paths $\pi : 0 \to n$ with $< \lceil n\epsilon\rceil$ of its edges not having the maximal edge label in their trial is at most

$$\sum_{i=0}^{\lfloor n\epsilon\rfloor}\binom{n}{i}(D-1)^i \leqslant 2^{nL}(D-1)^{n\epsilon} < \lfloor 2^{n\alpha}\rfloor$$

By the Pigeonhole Principle, there is an event-dependent path $\pi_{n,m_n} : 0 \to n$ s.t. at least $\lceil n\epsilon\rceil$ of its edges do not have the maximal edge label in their trial. Let π_{n,m_n} have edges $e_0^{j_0}, \ldots, e_{n-1}^{j_{n-1}}$ and let $I_n \subseteq \{0, \ldots, n-1\}, |I_n| = \lceil n\epsilon\rceil$ be a set of indices i for which $\tau(U_i^{j_i}) \neq Y_i$.

We compute an upper bound for the passage time along π_{n,m_n} (with respect to τ) by splitting the sum over the edges with index in I_n and those not in I_n:

$$\begin{aligned}\langle\tau, \frac{1}{n}\mu_{\pi_{n,m_n}}\rangle &= \frac{1}{n}\sum_{i=0}^{n-1}\tau(U_i^{j_i})\\ &\leqslant \frac{1}{n}\sum_{i\in I_n^C}\max_{1\leqslant j\leqslant D}(\tau(U_i^j)) + \frac{1}{n}\sum_{i\in I_n}\text{2nd}\max_{1\leqslant j\leqslant D}(\tau(U_i^j))\\ &= \frac{1}{n}\sum_{i\in I_n^C}Y_i + \frac{1}{n}\sum_{i\in I_n}Y_i'\end{aligned}$$

On the other hand, the passage time along the τ-optimal path $\pi_{0\to n}(\vec{p})$ is

$$\langle\tau, \frac{1}{n}\mu_{\pi_n(\vec{p})}\rangle = \frac{1}{n}\sum_{i=0}^{n-1}\max_{1\leqslant j\leqslant D}(\tau(U_i^j)) = \frac{1}{n}\sum_{i=0}^{n-1}Y_i$$

hence

$$\langle \tau, \frac{1}{n}\mu_{\pi_n(\vec{p})}\rangle - \langle \tau, \frac{1}{n}\mu_{\pi_{n,m_n}}\rangle \geqslant \frac{1}{n}\sum_{i\in I_n}(Y_i - Y_i') = \frac{1}{n}\sum_{i\in I_n} Z_i \geqslant \frac{1}{n}\sum_{k=0}^{\lfloor n\epsilon \rfloor} Z_{k:n} \quad (3.21)$$

Now Z_i are i.i.d., non-negative and satisfy

$$\begin{aligned} P(Z_i = 0) &= P\left(\max_{1\leqslant j\leqslant D}(\tau(U_i^j)) = \text{2nd} \max_{1\leqslant j\leqslant D}(\tau(U_i^j))\right) \\ &\leqslant P(\exists 1 \leqslant j_1 < j_2 \leqslant D \text{ s.t. } \tau(U_i^{j_1}) = \tau(U_i^{j_2})) \\ &= 0 \end{aligned}$$

since τ is not constant on sets of positive Λ measure. Thus we can take the expectation in (3.21) and apply Lemma 3.37 to get that $\exists C > 0$ s.t.

$$E\left[\langle \tau, \frac{1}{n}\mu_{\pi_n(\vec{p})}\rangle - \langle \tau, \frac{1}{n}\mu_{\pi_{n,m_n}}\rangle\right] \geqslant \frac{1}{n}\sum_{k=0}^{\lfloor n\epsilon \rfloor} E[Z_{k:n}] \geqslant C \quad (3.22)$$

for large n.

Now, by assumption, $\frac{1}{n}\mu_{\pi_{n,m_n}} \Rightarrow \sigma_{\vec{p}}$ a.s.. On the other hand, by nature of the model, the τ-optimal length m path $\pi_m(\vec{p})$ contains the τ-optimal length n path $\pi_n(\vec{p})$ for any $m \geqslant n$; thus we can apply the Glivenko-Cantelli Theorem (Theorem 3.18) to get that the empirical measures $\frac{1}{n}\mu_{\pi_n(\vec{p})}$ converge weakly to $\sigma_{\vec{p}}$ a.s.. Recall from Section **??** that a.s. the pushforward τ_* preserves weak limits of empirical measures so

$$\begin{aligned} \langle \tau, \frac{1}{n}\mu_{\pi_n(\vec{p})}\rangle - \langle \tau, \frac{1}{n}\mu_{\pi_{n,m_n}}\rangle &= \langle 1, \tau_*(\frac{1}{n}\mu_{\pi_n(\vec{p})})\rangle - \langle 1, \tau_*(\frac{1}{n}\mu_{\pi_{n,m_n}})\rangle \\ &\to \langle 1, \tau_*(\sigma_{\vec{p}})\rangle - \langle 1, \tau_*(\sigma_{\vec{p}})\rangle \\ &= 0 \end{aligned}$$

a.s.. But τ is bounded so by the Bounded Convergence Theorem we get

$$E\left[\langle \tau, \frac{1}{n}\mu_{\pi_n(\vec{p})}\rangle - \langle \tau, \frac{1}{n}\mu_{\pi_{n,m_n}}\rangle\right] = 0$$

which contradicts (3.22). Thus $||\sigma_{\vec{p}}|| = 0$. □

3.6.2 *Grid Entropy via Gibbs Free Energy*

Suppose $\tau : [0,1] \to [0,\infty)$ is a measurable function satisfying

$$E[e^{\tau(U)}] < \infty \text{ for } U \sim Unif[0,1]$$

From the definition of Gibbs Free Energy,

$$G(\tau) := \lim_{n\to\infty} \frac{1}{n} \log \sum_{\pi:0\to n} e^{T(\pi)} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \sum_{j=1}^{D} e^{\tau(U_i^j)} \text{ a.s.}$$

But $\log \sum_{j=1}^{D} e^{\tau(U_i^j)}$ are i.i.d. in i hence by the SLLN,

$$G(\tau) = E\left[\log \sum_{j=1}^{D} e^{\tau(U^j)}\right] \text{ a.s.}$$

where U^j are i.i.d. Unif[0,1].

Of course, grid entropy is simply the negative convex conjugate of Gibbs Free Energy by Theorem 3.9:

$$-||\nu|| = \sup_{\tau} \left[\langle \tau, \nu \rangle - G(\tau)\right] = \sup_{\tau} \left[\langle \tau, \nu \rangle - E\left[\log \sum_{j=1}^{D} e^{\tau(U^j)}\right]\right] \; \forall \nu \in \mathcal{M} \tag{3.23}$$

where the supremum is over all measurable functions $\tau : [0,1] \to [0,\infty)$ and where $\langle \tau, \nu \rangle$ denotes the integral $\int_0^1 \tau(u) d\nu$.

3.7 NEXT STEPS

We have characterized the extreme points of the set of limit points of empirical measures and have shown that these extreme points have grid entropy 0. Recalling that grid entropy is upper semicontinuous, a natural next question to ask is whether it is strictly convex in the model used in this chapter. A simpler version of this question is whether only the extreme points have grid entropy 0. These are questions well worth exploring in the future.

BIBLIOGRAPHY

[AH16] Daniel Ahlberg and Christopher Hoffman, *Random coalescing geodesics in first-passage percolation*, arXiv preprint:1609.02447 (2016), 1–74.

[Ahl15] D Ahlberg (ed.), *Problem lists: first passage percolation*, American Institute of Mathematics, Aug 2015.

[Bat20] Erik Bates, *Empirical distributions, geodesic lengths, and a variational formula in first-passage percolation*, arXiv preprint:2006.12580 (2020), 1–93.

[DS01] Jean-Dominique Deuschel and Daniel W Stroock, *Large deviations*, vol. 342, American Mathematical Soc., 2001.

[Dur19] Rick Durrett, *Probability: theory and examples*, vol. 49, Cambridge university press, 2019.

[GRAS16] Nicos Georgiou, Firas Rassoul-Agha, and Timo Seppäläinen, *Variational formulas and cocycle solutions for directed polymer and percolation models*, Communications in Mathematical Physics **346** (2016), no. 2, 741–779.

[Hol] Alexander E. Holroyd, A Sorting Networks Picture Gallery, https://personal.math.ubc.ca/~holroyd/sort/ [Accessed: July 2022].

[Hub04] Peter J Huber, *Robust statistics*, vol. 523, John Wiley & Sons, 2004.

[JLS20] Christopher Janjigian, Wai-Kit Lam, and Xiao Shen, *Tail bounds for the averaged empirical distribution on a geodesic in first-passage percolation*, arXiv preprint:2010.08072 (2020), 1–36.

[JRAS19] Christopher Janjigian, Firas Rassoul-Agha, and Timo Seppäläinen, *Geometry of geodesics through busemann measures in directed last-passage percolation*, arXiv preprint:1908.09040 (2019), 1–59.

[KL51] Solomon Kullback and Richard A Leibler, *On information and sufficiency*, The annals of mathematical statistics **22** (1951), no. 1, 79–86.

[Law13] Gregory F Lawler, *Intersections of random walks*, Springer Science & Business Media, 2013.

[Lig85] Thomas M Liggett, *An improved subadditive ergodic theorem*, The Annals of Probability **13** (1985), no. 4, 1279–1285.

[LP20] Yating Liu and Gilles Pagès, *Convergence rate of optimal quantization and application to the clustering performance of the empirical measure.*, J. Mach. Learn. Res. **21** (2020), no. 86, 1–36.

[Mar04] James B Martin, *Limiting shape for directed percolation models*, The Annals of Probability **32** (2004), no. 4, 2908–2937.

[Nis98] Togo Nishiura, *Measure-preserving maps of* $\mathbb{R}^n$, Real Analysis Exchange (1998), 837–842.

[Pos75] Edward Posner, *Random coding strategies for minimum entropy*, IEEE Transactions on Information Theory **21** (1975), no. 4, 388–391.

[VT84] Jan Van Tiel, *Convex analysis: an introductory text*, Wiley, 1984.

[Wel81] Jon A Wellner, *A glivenko-cantelli theorem for empirical measures of independent but non-identically distributed random variables*, Stochastic Processes and their Applications **11** (1981), no. 3, 309–312.

[Zal02] Constantin Zalinescu, *Convex analysis in general vector spaces*, World scientific, 2002.

COLOPHON

This thesis was typeset using the typographical look-and-feel `classicthesis` developed by André Miede and Ivo Pletikosić.

The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*".