Statistical Inference and String Theory

Jonathan J. Heckman*

Jefferson Physical Laboratory, Harvard University, Cambridge, MA 02138, USA

Abstract

In this note we expose some surprising connections between string theory and statistical inference. We consider a large collective of agents sweeping out a family of nearby statistical models for an M-dimensional manifold of statistical fitting parameters. When the agents making nearby inferences align along a d-dimensional grid, we find that the pooled probability that the collective reaches a correct inference is the partition function of a non-linear sigma model in d dimensions. Stability under perturbations to the original inference scheme requires the agents of the collective to distribute along two dimensions. Conformal invariance of the sigma model corresponds to the condition of a stable inference scheme, directly leading to the Einstein field equations for classical gravity. By summing over all possible arrangements of the agents in the collective, we reach a string theory. We also use this perspective to quantify how much an observer can hope to learn about the internal geometry of a superstring compactification. Finally, we present some brief speculative remarks on applications to the AdS/CFT correspondence and Lorentzian signature spacetimes.

---

*e-mail: `jheckman@physics.harvard.edu`

Contents

1 Introduction

Sifting through competing interpretations of data lies at the core of quantitative approaches to statistical inference. A successful statistical model must strike a balance between the competing demands of accuracy and simplicity. A related consideration is the ability to adapt an inference scheme to new information.

In this note we show that this and related questions in statistical inference are amenable to study using well-known results from string theory and quantum field theory. Conversely, we use statistical inference to gain a different perspective on string theory. Though we couch our results in broader terms, one can also view this note as an attempt to define an approximate notion of a local observable in quantum gravity.[1] In a rather unexpected way, classical gravity will indeed emerge from our considerations.

In more formal terms, we frame the question of statistical inference as the attempt of an agent to develop a statistical model after observing independent events $E = \{e_1, ..., e_N\}$ which have been drawn from the true probability distribution. The task of the agent is to produce an accurate statistical model A depending on M continuous statistical parameters $y^{(1)}, ..., y^{(M)}$. Given two competing models A and B, we can then compare the Bayesian posterior probabilities $\Pr(A|E)$ and $\Pr(B|E)$, and select the model with the higher value.

In [1, 2], the value of $\Pr(A|E)$ was interpreted as the partition function of a statistical mechanics problem. The parameters of the statistical model specify the configuration space of a thermodynamic ensemble, with N playing the role of an inverse temperature, and the proximity of the guess from the true distribution playing the role of an energy. The competition between decreasing the energy and increasing the entropy of the ensemble translates to the competing interests of achieving a better fit to the data, but with as simple a model as possible. This thermodynamic interpretation is rather striking, and suggests a number of generalizations.

Our aim in this note is to generalize this analysis to a collective of agents, who may decide to use different statistical models to fit the data. Since the collective gets to explore many nearby models, it can infer a broader class of inference strategies, and in particular, may arrive at a different inference than any individual agent. An additional reason to consider a collective is that its inferences may be more stable against perturbations.

We define a collective as a large number of agents $K \equiv K_{\text{agent}}$ with statistical models $A_1, ..., A_K$, each of which depend on the same M parameters $y^{(1)}, ..., y^{(M)}$. Each member of the collective samples the true distribution, and receives a set of N events $E_{(i)} = \{e_1^{(i)}, ..., e_N^{(i)}\}$

---

[1]In quantum gravity, it is hard to define a local observable because selecting a point of the spacetime breaks diffeomorphism invariance. For spacetimes which asymptote to either Minkowski or Anti-de Sitter space, more precision is available because observables can be formulated in terms of boundary data.

for $i = 1, ..., K$. The pooled posterior probability of the collective is defined as:

$$Z(A_{\mathrm{coll}}|E_{\mathrm{coll}}) \equiv \Pr(A_1|E_1)...\Pr(A_K|E_K). \tag{1.1}$$

The success of the collective versus the individual agent is then obtained by comparing the geometric mean $[Z(A_{\mathrm{coll}}|E_{\mathrm{coll}})]^{1/K}$ with $\Pr(A|E)$.

Depending on the nature of the true distribution, the collective could decide on various inference strategies. For example, given two collectives A_{coll} and B_{coll}, collective A_{coll} could consist of agents trying to fit to a Gaussian, as well as other distributions which are small deformations of a Gaussian profile. Alternatively, collective B_{coll} could try fitting to a Lorentzian, and nearby distributions. The aim is to pick the collective with the higher pooled posterior probability.

In a similar vein to [1,2], our aim will be to interpret $Z(A_{\mathrm{coll}}|E_{\mathrm{coll}})$ in terms of a statistical mechanical model. When the guesses of nearby agents are arranged along a d-dimensional grid, the geometry of agents builds up an approximation to a smooth manifold Σ_{agent} with coordinates $\sigma^1, ..., \sigma^d$. Each agent makes a guess $y^{(I)}(\sigma^a)$ for $I = 1, ..., M$ and $a = 1, ..., d$. Varying over the choice of σ^a on the grid then yields a family of nearby guesses. In the limit $K \gg 1$ where the agents fill in a dense mesh, the y's define a map from the space of agents to the space of parameters:

$$y : \Sigma_{\mathrm{agent}} \to Y_{\mathrm{target}}, \tag{1.2}$$

so for each point of Σ_{agent}, we get an agent with a corresponding statistical model.

We find that in the large N limit, the probability $Z(A_{\mathrm{coll}}|E_{\mathrm{coll}})$ is a path integral:

$$Z(A_{\mathrm{coll}}|E_{\mathrm{coll}}) \propto \int [\mathcal{D}y] \ \sqrt{\det G} \ e^{-S_{\mathrm{tot}}}, \tag{1.3}$$

up to a normalization constant which we shall mostly neglect. Here, $[\mathcal{D}y] \sqrt{\det G}$ is a measure factor for the path integral, i.e., a measure for the space of all maps $y : \Sigma_{\mathrm{agent}} \to Y_{\mathrm{target}}$. The exponent S_{tot} is the action of a non-linear sigma model in d Euclidean dimensions:

$$S_{\mathrm{tot}} = \int\limits_{\Sigma_{\mathrm{agent}}} d^d\sigma \ \sqrt{\det h} \ \left(\frac{N}{2} G_{IJ} \ h^{ab} \frac{\partial y^I}{\partial \sigma^a} \frac{\partial y^J}{\partial \sigma^b} + V(y, \sigma) \right) \tag{1.4}$$

and G_{IJ} is an information metric on the space of statistical parameters Y_{target}.[2] Here and throughout, repeated subscripts and superscripts are to be summed.

The information metric is defined by picking a family of probability densities $p_A(x, \{y\})$

---

[2]The study of differential geometry on statistical parameter spaces is known as information geometry. See [3] for an excellent review.
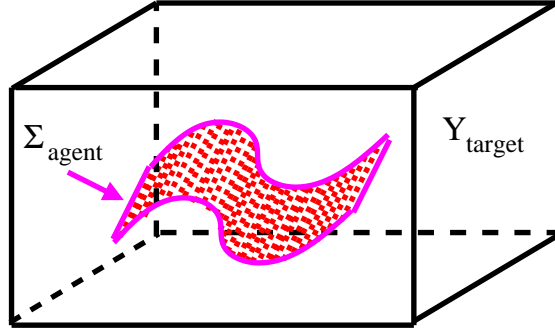
Figure 1: Depiction of a collective of agents arranged on a d-dimensional grid of Σ_{agent} making nearby guesses of statistical fitting parameters on an M-dimensional geometry Y_{target}.

and varying the parameters y:

$$G_{IJ} = \int_X d\mu(x)\ p_A \frac{\partial \log p_A}{\partial y^I} \frac{\partial \log p_A}{\partial y^J}. \tag{1.5}$$

This measures the infinitesimal proximity of p_A to the true distribution.

The kinetic term of the sigma model is simply the pullback of the information metric on Y_{target} to Σ_{agent}. Summing over all the agents yields a notion of proximity of the true distribution to the guessing strategy adopted by the entire collective.

Adding a potential energy function to the non-linear sigma model corresponds to a choice of statistical prior. This allows for local interactions between nearby neighbors on the grid. Such deformations can be either Lorentz invariant or possibly Lorentz breaking, signifying a preferred weighting of specific agents in the collective.

The d-dimensional field theory specifies a collection of agents making nearby guesses on a spatial d-dimensional grid. If we distribute the agents on a grid with a Lorentzian signature metric, we can alternatively view this as a collective of agents on a $(d-1)$-dimensional grid who update their inferences in discrete time steps. See figure 1 for a depiction.

The field theory formulation provides a tool for addressing how the collective would respond to a perturbation to the original inference, and moreover, how estimates on the parameters are correlated over the lattice of agents. If we add a linear source term $\mathcal{J}_I(\sigma)y^I(\sigma)$, we can specify initial conditions for some of the agents. Then, the one-point function will tell us about how the original inference strategy settles towards a preferred value of parameters. The higher point functions tell us about correlations between the guesses of nearby agents.

We can also ask what happens if there is a perturbation to the family of statistical models used by the collective. One might view this as the collective "changing its mind", or equivalently, reacting to a change in the true distribution. Ideally, such perturbations should

not completely destabilize the original inference, though the collective should be capable of altering its original inference. For simplicity, we mainly focus on the case of uniformly weighted agents in flat space.

Surprisingly, the answer is very sensitive to the geometry of the agents. For $d < 2$, a small change in the probability distribution generically destabilizes the original inference, that is, the collective does not reach the same inference. For $d > 2$, these perturbations wash out, and do not change the original inference. When $d = 2$, however, the collective can adjust its inference procedure.

In other words, stable statistical inference selects out two dimensions for the grid of agents. The limiting case where the overall dependence on the number of agents drops out translates to the condition of conformal invariance in the sigma model. As is well known the condition of conformal invariance leads directly to the Einstein field equations for classical gravity. Quantum fluctuations around the background metric arise from fluctuations in the inferred probability distribution. As far as we are aware, this is the first derivation of classical gravity from the condition of stable statistical inference.

The classical description breaks down at small distance in the fitting parameter space. This occurs at a statistical resolution length (in units where $2\pi\alpha' = 1/N$):

$$\ell_{\min} \sim 1/\sqrt{N} \tag{1.6}$$

with N the number of sampled events. In physics terms, this sets a cutoff for the target space theory, which we interpret as the string scale associated with the sigma model.

To reach a full string theory, we can make the stronger demand that we sum over all possible arrangements of the agents by coupling the sigma model to two-dimensional gravity. Quantum stability can be ensured by passing to the superstring. Each agent can then be viewed as the endpoint of an open string, that is, a D-instanton, and a collective of D-instantons can build up a string. From this perspective, the condition of stable inference gives an explanation for why a quantum theory of gravity must contain stringlike excitations.

Turning the discussion around, we also apply methods from statistical inference to the study of superstring theory. We ask how well an observer could hope to resolve the internal geometry of a compactified spacetime $\mathbb{R}^{3,1} \times Y_{\mathrm{internal}}$. A guess by an observer defines a probability distribution which depends on the location of the observer $y \in Y_{\mathrm{internal}}$, as well as other fitting parameters which characterize the local profile of the metric. This generates a family of probability distributions, and a corresponding information metric. Calabi-Yau compactifications can be understood in information theoretic terms as generating successively better numerics for a balanced metric (see e.g. [4–7]). Comparing the true probability distribution with the original guess, we can then quantify the amount of information an agent would gain by adjusting their initial incorrect guess of the internal geometry. We also entertain

some more speculative material on potential applications to the AdS/CFT correspondence, as well as Lorentzian signature statistical manifolds.

The rest of this note is organized to reflect the fact that some pieces may be of interest to different readers. In section 2 we briefly review some background on statistical inference and information geometry, and in section 3 we present a sigma model for statistical inference and study some of its properties using results which –though well-known to string theorists– may seem surprising. After this, we focus our attention on applications to the physical superstring. We study the information content of a superstring compactification in section 4, and briefly entertain some more speculative applications in section 5. We conclude in section 6.

2   Statistical Inference and Information Geometry

In this section we review the main concepts from statistical inference and information geometry which we shall use. Excellent resources are the reviews [3,8].

Given a probability space X with measure $\mu(x)$ we would like to define some notion of two probability distributions being nearby. A common choice is the Kullback-Leibler divergence between the true density $p_T(x)$ and a guess $p_A(x)$:

$$D_{KL}(p_T||p_A) \equiv \int_X d\mu(x)\ p_T(x)\log\frac{p_T(x)}{p_A(x)}, \tag{2.1}$$

which is also known as the relative entropy. Though it is not a true metric (as it is not symmetric), the Kullback-Leibler divergence is non-negative, and has a simple statistical interpretation as quantifying the amount of information (in the sense of Shannon [9]) an agent would gain by learning the true distribution. One can also introduce the continuum generalization of the Shannon entropy known as the differential entropy, by integrating $-p\log p$. This must be treated with more care, because this "entropy" can be negative.

Of course, the Kullback-Leibler divergence is but one way to compare the proximity of distributions. In fact, in the small distance limit, other notions of proximity produce the same notion of distance between distributions. To see this, consider a parametric family of probability densities $p_A(x,\{y\})$ depending on coordinates $\{y^I\}$ of some manifold Y_{target}:

$$p_T = p_A + \delta \equiv p_A + \delta y^I\frac{\partial p_A}{\partial y^I} \tag{2.2}$$

which are infinitesimally close to the true distribution. Expanding to second order in D_{KL}

then yields the Fisher information metric:

$$D_{KL}(p_A + \delta || p_A) \simeq G_{IJ} \, \delta y^I \delta y^J + O(\delta y^3) \tag{2.3}$$

with G_{IJ} as in equation (1.5).

A central result in information geometry is Chentsov's theorem [10]. This theorem guarantees that at least in situations where we can sample the true distribution sufficiently often, the infinitesimal form of the distance will eventually converge to the Fisher information metric, that is, equation (2.3). In this sense, the infinitesimal form of statistical inference is relatively insensitive to which finite distance measure between distributions we choose to adopt.

Now, having introduced a metric G_{IJ} for the Riemannian manifold Y_{target}, the first inclination of a physicist might be to study geodesic flows with respect to the Levi-Cevita connection ∇_{LC}. This is certainly possible to do, but it turns out to not be the only natural choice in the case of statistical inference. We can also introduce connections with torsion ∇_B which are specifically adapted to a fixed choice of affine coordinate system. We can then define another connection ∇_B^* via the implicit relation $\nabla_B + \nabla_B^* = 2\nabla_{LC}$. The use of connections with torsion is quite prevalent in the information geometry literature. This is because one is often concerned with a privileged class of probability distributions, and therefore must adapt an affine coordinate system for this specific distribution. Parallel transport with respect to the Levi-Cevita connection will generically disrupt this coordinate system, so it is common (see e.g. [3]) to label a statistical manifold by the triple $(G_{IJ}, \nabla_B, \nabla_B^*)$.

To illustrate some of the above notions, consider the case of a normal probability distribution for outcomes $x^{(I)} \in \mathbb{R}^{M-1}$. The fitting parameters are $M-1$ positions and an overall width:

$$p(x; \{y^{(1)}, ..., y^{(M-1)}, \alpha\}) = (\pi \alpha^2)^{-(M-1)/2} \exp\left[-\sum_{I=1}^{M-1} \frac{(x^{(I)} - y^{(I)})^2}{\alpha^2} \right] \tag{2.4}$$

The information metric is:

$$ds^2 = 2 \times \frac{dy_{(I)} dy^{(I)} + d\alpha^2}{\alpha^2} \tag{2.5}$$

We find it striking that this is also the metric for Euclidean Anti-de Sitter space in M dimensions.

2.1 Statistical Mechanics of Statistical Inference

We now review the statistical mechanical interpretation of statistical inference [1,2]. We start with a probability distribution T with density $p_T(x)$ which has generated N independent events $E = \{e_1, ..., e_N\}$. The aim is to find a statistical model which comes closest to approximating this distribution. Given two models A and B we can compute $\Pr(A|E)$ and

$\Pr(B|E)$. The aim is to pick the model with the higher conditional probability.

Suppose our agent has made a choice of statistical model A with probability density $p_A(x, \{y\})$ which depends on M parameters $y^{(1)}, ..., y^{(M)}$. Using Bayes' rule, the Bayesian posterior probability is:

$$\Pr(A|E) = \frac{\Pr(A)}{\Pr(E)} \frac{\int d^M y \; w(y) \; \Pr(E|A)}{\int d^M y \; w(y)}, \tag{2.6}$$

where $w(y)$ is some choice of weighting scheme for the parameters $y^{(I)}$. Here, $\Pr(A)$ is the prior probability for the model A, and $\Pr(E)$ is the probability of the event set E.

Now, the key step is that in the limit where the agent has sampled a large $N \gg 1$ number of independent events, the law of large numbers implies that $\Pr(E|A)$ approaches (in the almost sure sense) [1, 2]:

$$\Pr(E|A) \simeq \exp(-N(\mathcal{E}_A + \mathcal{E}_0)) \tag{2.7}$$

where $\mathcal{E}_A$ is the Kullback-Leibler divergence and $\mathcal{E}_0$ is the differential entropy:

$$\mathcal{E}_A = \int_X d\mu(x) \; p_T(x) \log \frac{p_T(x)}{p_A(x)} \quad \text{and} \quad \mathcal{E}_0 = -\int_X d\mu(x) \; p_T(x) \log p_T(x). \tag{2.8}$$

In other words, in the large N limit $\Pr(A|E)$ is a partition function:

$$\Pr(A|E) = \frac{\Pr(A)}{\Pr(E)} \frac{\int d^M y \; w(y) \; e^{-N(\mathcal{E}_A + \mathcal{E}_0)}}{\int d^M y \; w(y)}. \tag{2.9}$$

This provides a statistical mechanical interpretation for statistical inference [1, 2]. The number of sampled events N corresponds to the inverse temperature, and the Kullback-Leibler divergence $\mathcal{E}_A$ is an energy. Up to a y-independent shift, the quantity $\mathcal{E}_0$ can be repackaged as indicating the interaction between the guess and "statistical defects". Moreover, just as in thermodynamics, there is a competition between minimizing the energy, and maximizing the entropy. This corresponds to balancing the interests of minimizing the distance between p_A and p_T, and on the other hand, the possibility of probing more nearby configurations. In the "quenched approximation" to the partition function, we can basically drop the background value of $\mathcal{E}_0$. Indeed, we shall mostly neglect this contribution in our considerations.

If we assume that the probability density $p_A(x, \{y\})$ is nearby the true density in the sense of equation (2.2), we can further approximate the energy as:

$$\mathcal{E}_A \simeq G_{IJ} \; \delta y^I \delta y^J + O(\delta y^3). \tag{2.10}$$

Finally, in [1, 2] additional care is paid to the choice of a weighting function $w(y)$, where it is argued that the natural choice is the Jeffrey's prior, corresponding to the choice $w(y) = \sqrt{\det G}$. The integration measure is then parametrization invariant. Of course, depending on the prior information, one might make different choices.

3 Field Theory for Statistical Inference

In this section we turn to the core of our analysis, where we generalize [1, 2] to a collective of agents. Now, whereas a single agent must settle for one inference scheme, a collective can sweep out a bigger set of guesses. What we show is that the pooled posterior probability of the collective fitting the data is the partition function of a non-linear sigma model. We include some additional heuristic background on such field theories since some aspects may be unfamiliar to the reader.

We define a collective A_{coll} of K agents as a set of nearby statistical models $A_1, ..., A_K$. Each agent draws N events from the true distribution $E_i = \left\{ e_1^{(i)}, ..., e_N^{(i)} \right\}$ for $i = 1, ..., K$. The models are assumed to be nearby in the sense that they each depend on M parameters $y^{(1)}, ..., y^{(M)}$ for some guess distribution. In other words, we get K sets of fitting parameters $y_1, ..., y_K$, where each y_i itself specifies M fitting parameters. Restoring all indices, we label each such parameter as $y_j^{(I)}$ for $j = 1, ..., K$ and $I = 1, ..., M$.

For each agent, we have the Bayesian posterior probability:

$$\Pr(A_i|E_i) = \frac{\Pr(A_i)}{\Pr(E_i)} \frac{\int d^M y_i \; w_i(y_i) \; e^{-N\left(\mathcal{E}_A^{(i)} + \mathcal{E}_0^{(i)}\right)}}{\int d^M y_i \; w_i(y_i)} \tag{3.1}$$

where $\mathcal{E}_A^{(i)}$ is the energy for each agent, and $\mathcal{E}_0^{(i)}$ is the differential entropy of the true distribution. In principle, the prior probability $\Pr(A_i)$ of each agent's guess, as well as their weighting schemes $w_i(y_i)$ could vary from agent to agent.

Our interest is not in the success of any individual agent, but instead the full collective. What we want to study is the pooled posterior probability of the collective as a whole to reach some inference. Along these lines, we introduce the pooled posterior probability of the collective:

$$Z(A_{\text{coll}}|E_{\text{coll}}) \equiv \prod_{i=1}^{K} \Pr(A_i|E_i). \tag{3.2}$$

When the context is clear, we shall denote this by Z_{coll}. This definition reflects the assumption that although the agents of the collective are required to make nearby guesses, they are otherwise independent. Given two collectives A_{coll} and B_{coll} each composed of K agents, our aim will be to select the collective with the higher pooled posterior probability. We can also compare the success of a collective to that of a single agent by computing $[Z(A_{\text{coll}}|E_{\text{coll}})]^{1/K}$.

3.1   Lattice Approximation

To give the problem some more structure, we shall now assume that the agents are aligned along a d-dimensional grid, which is assumed to be a lattice approximation to some d-dimensional manifold Σ_{agent}. For each choice of point in the lattice approximation to Σ_{agent}, we can pick a value for $y^I(\sigma^a)$. In other words, we build up an approximation to the map:

$$y : \Sigma_{\text{agent}} \to Y_{\text{target}}. \tag{3.3}$$

Performing the integral $\prod\limits_{i=1}^{K} \int d^M y_i$ then sweeps out a range of possible maps.

Let us illustrate the lattice approximation for $\Sigma_{\text{agent}} = \mathrm{R}^d$. We introduce a collection of d unit normalized vectors $\overrightarrow{e}_1 = (1, 0, ..., 0),..., \overrightarrow{e}_d = (0, ..., 0, 1)$. We span a grid with the linear combinations $\ell \sum\limits_{a=1}^{d} m_a \overrightarrow{e}_a$ where ℓ is a small parameter and m_a are integers. A lattice derivative in the direction $\overrightarrow{e}_a$ corresponds to:

$$\delta_a y^J = \frac{y^J(\overrightarrow{\sigma} + \ell \overrightarrow{e}_a) - y^J(\overrightarrow{\sigma})}{2\ell} + \frac{y^J(\overrightarrow{\sigma}) - y^J(\overrightarrow{\sigma} - \ell \overrightarrow{e}_a)}{2\ell} \tag{3.4}$$

for $a = 1, ..., d$ and $J = 1, ..., M$. Discretized integration corresponds to the substitution:

$$\ell^d \sum_{i=1}^{K} \to \int d^d \sigma \ \sqrt{\det h} \tag{3.5}$$

where we have introduced a notion of distance between neighboring agents via the metric $ds^2_{\text{agent}} = h_{ab} d\sigma^a d\sigma^b$ on the agent space. To focus on the essential points, we shall typically take h_{ab} to be a constant matrix. Much of what we say generalizes to a broader class of agent space metrics.

In practice, the space Σ_{agent} will typically have finite volume. The lattice approximation therefore amounts to fixing some small unit cell of size ℓ^d, and arranging these cells to produce the total volume of Σ_{agent}:

$$\text{Vol}(\Sigma_{\text{agent}}) = K \times \ell^d. \tag{3.6}$$

The continuum limit corresponds to sending $\ell \to 0$ and $K \to \infty$ with $\text{Vol}(\Sigma_{\text{agent}})$ held fixed.

We would now like to argue that inference of the collective defines a d-dimensional quantum field theory of a very particular type. Since we are assuming the family of guesses are close to the true distribution, we can expand the energy at a point $\{\sigma^a\}$ as in equation (2.10).

10

This defines a pullback of the metric on Y_{target} to a metric on Σ_{agent}:

$$\mathcal{E}_A(\sigma) \simeq G_{IJ}\ \delta y^I \delta y^J \simeq G_{IJ}\ h^{ab} \frac{\partial y^I}{\partial \sigma^a} \frac{\partial y^J}{\partial \sigma^b} \tag{3.7}$$

where h^{ab} denotes the matrix inverse of h_{ab}. In the limit of a large number of agents, the lattice approximation tends to:

$$\prod_{i=1}^{K} \int d^M y_i\ e^{-N\mathcal{E}_A^{(i)}} \rightarrow \int [\mathcal{D}y]\, e^{-S_{\text{kin}}} \tag{3.8}$$

up to a common overall constant normalization factor. Here, the measure factor $[\mathcal{D}y]$ is a heuristic instruction to integrate over all possible maps $\Sigma_{\text{agent}} \rightarrow Y_{\text{target}}$. The contribution S_{kin} is the kinetic term for a non-linear sigma model:

$$S_{\text{kin}} = \frac{1}{\ell^d} \int d^d\sigma\ \sqrt{\det h}\ \frac{N}{2}\ G_{IJ}\ h^{ab} \frac{\partial y^I}{\partial \sigma^a} \frac{\partial y^J}{\partial \sigma^b}. \tag{3.9}$$

The kinetic energy defines a sum over the infinitesimal Kullback-Leibler divergences for all of the agents, measuring the proximity of the collective to the true distribution. This provides a first indication of how a collective inference might differ from an individual inference: Whereas individual agents might produce sharp jumps in making different inferences, the collective might instead smooth these out to only gradual changes across the entire agent space. In statistical terms, the collective gains more by exploring many nearby options than by trying to minimize the Kullback-Leibler divergence of a single agent.

The other terms of equation (3.2) also define natural objects in this quantum field theory. For example, the agent dependent weighting factor $w_i(y_i)$ lifts to a possibly position dependent potential energy:

$$\prod_{i=1}^{K} w_i(y_i) \rightarrow e^{-S_{\text{pot}}} \tag{3.10}$$

where:

$$S_{\text{pot}} = -\frac{1}{\ell^d} \int_{\Sigma_{\text{agent}}} d^d\sigma\ \sqrt{\det h}\ \ \log(w(y,\sigma)), \tag{3.11}$$

and similar considerations apply for the position dependent potential energy on agent space:

$$\prod_{i=1}^{K} \frac{\Pr(A_i)}{\Pr(E_i)} e^{-N\mathcal{E}_0^{(i)}} \rightarrow e^{-U_{\text{pos}}}. \tag{3.12}$$

Again, this makes intuitive sense; a choice of weighting scheme dictates where the statistical inference procedure may be attracted. A position dependent potential term means not all

agents are weighted equally. In most uniform weighting schemes one is essentially demanding Lorentz invariance of the continuum theory.

Assembling all of the pieces and integrating over the class of all possible maps, i.e. choices of $y(\sigma)$, we have arrived at the path integral formulation of a non-linear sigma model. Statistical inference of the collective determines a quantum field theory!

It is convenient to adhere to standard practice in field theory by rescaling the fields to eliminate the explicit factors of ℓ appearing in our continuum theory integrals. Doing so, we arrive at the final form of the partition function for a non-linear sigma model:

$$Z_{\text{coll}} \propto \int [\mathcal{D}y] \sqrt{\det G} \; e^{-S_{\text{tot}}}. \tag{3.13}$$

where:

$$S_{\text{tot}} = \int d^d\sigma \; \sqrt{\det h} \; \left(\frac{N}{2} G_{IJ} \; h^{ab} \frac{\partial y^I}{\partial \sigma^a} \frac{\partial y^J}{\partial \sigma^b} + V(y, \sigma) \right) \tag{3.14}$$

in the obvious notation. This is the action in d Euclidean dimensions for a non-linear sigma model which has been deformed by a field and position dependent potential energy. Here, we have adjusted the measure of the path integral by a factor of $\sqrt{\det G}$ so that $V = 0$ refers to the Jeffrey's prior.

3.2   Correlations and Inferences

We can now repurpose many of the methods used in quantum field theory to study inference schemes of the collective. If we perform the analytic continuation to a Lorentzian signature agent grid, we can treat the system as a $(d-1)$-dimensional grid of agents which updates along a grid of time steps.

We specify boundary conditions for the system by introducing a position dependent chemical potential for the parameters. This is a source in the field theory:

$$Z_{\text{coll}}[\mathcal{J}] = \int [\mathcal{D}y] \sqrt{\det G} \exp\left(-S_{\text{tot}} + \int d^d\sigma \; \sqrt{\det h} \; \mathcal{J}_I(\sigma) y^I(\sigma) \right). \tag{3.15}$$

Correlation functions of the quantum field theory are given by functional derivatives with respect to the source:

$$\left\langle y^{I_1}(\sigma_1) ... y^{I_m}(\sigma_m) \right\rangle_{\mathcal{J}} = \frac{\delta^m Z_{\text{coll}}[\mathcal{J}]}{\delta \mathcal{J}_{I_1}(\sigma_1) ... \delta \mathcal{J}_{I_m}(\sigma_m)}. \tag{3.16}$$

Of particular interest is the behavior of the one-point function $\left\langle y^I(\sigma) \right\rangle_{\mathcal{J}}$ in some asymptotic limit of the agent parameters σ. In Euclidean signature this tells us about the boundary behavior of the collective, and in Lorentzian signature, it tells us about the late time behavior

of the collective.

It is important to stress that our formalism implicitly assumes we are already close to the true distribution. In other words, we assume that all fluctuations in the fitting parameters are small.

In some cases the field theory will admit a semi-classical approximation, corresponding to expansion around a saddle point of Z_{coll}. Performing a functional derivative of S_{tot} with respect to $y^I(\sigma)$ yields a classical evolution equation for the y's:

$$\frac{\delta S_{\text{tot}}}{\delta y^I(\sigma)} = \mathcal{J}_I(\sigma), \tag{3.17}$$

or,

$$\nabla^2 y_I - \frac{\partial V}{\partial y^I} = \mathcal{J}_I(\sigma), \tag{3.18}$$

We view $\mathcal{J}_I(\sigma)$ and the y independent part of $\partial V/\partial y$ as specifying a prior on the guess in the statistical parameter space.

As an illustrative example, consider the case where our guess is a single centered normal distribution of fixed width. The information metric is the flat space metric on R^M, with entries proportional to the $M \times M$ identity matrix:

$$G_{IJ} \propto \text{diag}(1, ..., 1). \tag{3.19}$$

Let us further suppose that our initial weighting scheme by $w(y)$ specifies a quadratic potential $V(y) = \frac{1}{2\kappa^2} (y - y_f)^2$. This corresponds to an initial weighting scheme by a Gaussian of width κ centered on $y = y_f$. All together, this describes a free quantum field theory with a mass term. The presence of the mass term means that eventually the configuration will aim to roll towards $y = y_f$.

How it gets there is another matter, which is sensitive to the number of dimensions for the agent space. In $d \geq 3$, long range correlations between agents die off sufficiently quickly that we can use a semi-classical evolution equation. In $d \leq 2$, more care is needed in the nearly massless limit.

To further illustrate the field theory perspective, let us consider another example, given by the sum of two normal distributions for outcomes $x \in \mathrm{R}$:

$$p_{\text{sum}} = \frac{1}{2} \frac{1}{\sqrt{2\pi\alpha^2}} e^{-(x-a)^2/2\alpha^2} + \frac{1}{2} \frac{1}{\sqrt{2\pi\alpha^2}} e^{-(x-b)^2/2\alpha^2}. \tag{3.20}$$

We hold fixed the width α, but vary the parameters a and b, so our statistical manifold is parameterized by the product $\mathrm{R}_{(a)} \times \mathrm{R}_{(b)}$. In the limit where $b \simeq a + \varepsilon$, we can evaluate the

13

leading order form of the information metric:

$$G_{IJ} = \frac{1}{4\alpha^2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \frac{\varepsilon^2}{16\alpha^4} \begin{bmatrix} 1 & -3 \\ -3 & 1 \end{bmatrix}. \tag{3.21}$$

When $\varepsilon \to 0$, the information metric collapses to a rank one constant matrix. In the field theory, this means the mode $(a - b)$ does not propagate, leaving only the center of mass degree of freedom $(a + b)$. The statistical interpretation is analogous: In the limit where the two peaks coincide, we do equally well to fit the distribution to a single Gaussian; one of the degrees of freedom has become irrelevant.

3.3   Stability Under Perturbations

In this subsection we study the stability of the collective against perturbations. For simplicity, we shall assume that the potential energy $V(y)$ only depends on σ implicitly through its dependence on y. We shall also assume that the agent space is equipped with a flat metric.

Our aim will be study the growth / dissipation of perturbations to the original inference scheme as the number of agents K becomes very large. Alternatively, this can be viewed as asking what happens when the collective "changes its mind" by perturbing its initial family of guesses.

We are particularly interested in stable inference schemes. On the one hand, this means that such perturbations should not drastically change the original inference. On the other hand, this also means that the collective should be capable of changing its original inference.

The theoretical tool we use to address the response to perturbations of the collective is the renormalization group of the field theory. Roughly speaking, we can partition the original agent space into a set of ever cruder averages over nearest neighbors. The renormalization group equations track the response of perturbations as we perform this averaging procedure. Letting λ_{pert} denote a perturbation of either the background metric G_{IJ} or the potential $V(y)$, we ask whether λ_{pert} is amplified or suppressed as the number of agents K grows. In field theory terms, the scaling behavior defines a beta function:

$$\beta^{(\lambda)} = \frac{\partial \lambda_{\text{pert}}}{\partial \log K}. \tag{3.22}$$

Stability under perturbations translates to the condition that we arrive at an interacting conformal fixed point, so in particular we require $\beta^{(\lambda)} = 0$.

Rather than delve into a detailed analysis, we will instead present some heuristic intuition for the different behavior couplings can exhibit by using –basically classical– dimensional analysis arguments. The statements we make receive various quantum corrections, which we shall basically gloss over, except in subsection 3.4 when we treat the special case $d = 2$.

Since we have introduced a length scale ℓ which indicates the proximity between agents on Σ_{agent}, scaling with respect to this parameter will show up when we coarse grain the approximation on the agents. In particular, this dimensionful scale will also be related to the scaling behavior of the fitting parameters y^I. To fix the overall scaling dimension, we first note that the exponential $\exp(-S_{\text{tot}})$ is only well-defined provided S_{tot}, the action of the sigma model is dimensionless. Now, since the integral over the agent space defines a d-dimensional volume $\text{Vol}(\Sigma_{\text{agent}}) = K \times \ell^d$, it follows that the y's must have appropriate ℓ dependence to make S_{tot} dimensionless. For example, in the $d = 1$ action:

$$S_{\text{1d}} = \int d\sigma \, \frac{1}{2} \left(\frac{\partial y}{\partial \sigma} \right)^2, \tag{3.23}$$

$(\partial y / \partial \sigma)^2$ must scale as $1/\ell$ for S_{1d} to be dimensionless. Since each derivative specifies a factor of $1/\ell$ (c.f. equation 3.4), it follows that y classically scales as $\ell^{1/2}$. More generally, we can consider the case of a d-dimensional agent space. By expanding G_{IJ} around some constant value, we learn that the classical scaling of y is $y \sim \ell^{(2-d)/2}$.

Now, we are going to use this same dimensional analysis argument to compute the scaling behavior of perturbations to the original field theory. Though we do not do so here, one can give more rigorous versions of these arguments.

Consider first the possibility that the collective decides on a different inference scheme. This corresponds to a perturbation $p_A \to p_A + \delta p_A$, which will in turn show up as a change to the original information metric. In the sigma model, we track this perturbation by expanding near some point $y_{(0)}^I$. Letting $\eta^I \equiv y^I - y_{(0)}^I$ denote small variations, we can expand the kinetic term around this value:[3]

$$\partial y^I \partial y^J \times G_{IJ} = \partial y^I \partial y^J \times \left(G_{IJ}^{(0)} + \delta G_{IJK}^{(1)} \eta^K + \delta G_{IJKL}^{(2)} \eta^K \eta^L + ... \right). \tag{3.24}$$

Since η^I scales in the same way as y^I, we can work out the scaling behavior of each of these perturbations as a function of length. Each subsequent perturbation involves additional factors of η, so we have $\delta G^{(m)} \sim \ell^{m(d-2)/2}$. To compute the scaling behavior of these perturbations with respect to K, we recall that the volume of the agent space $\text{Vol}(\Sigma_{\text{agent}}) = K \times \ell^d$ is being held fixed throughout the analysis. Hence, we can solve for ℓ in terms of K, which implies each such perturbation scales with the number of agents as:

$$\delta G^{(m)} \sim K^{\gamma(m)} \quad \text{where} \quad \gamma(m) = m \times \frac{2-d}{2d}. \tag{3.25}$$

In other words, in the $K \to \infty$ limit, these perturbations die off for $d > 2$, while for $d < 2$,

---

[3]Actually, if we work with Riemann normal coordinates the variations begin at quadratic order, with $\delta G_{IJKL}^{(2)}$ proportional to the Riemann tensor.

these terms alter the long distance behavior of the field theory. For $d = 2$, such perturbations could potentially be marginal.

The interpretation in the statistical setting is quite intriguing: If we start with a given sigma model of statistical inference, we can ask whether the collective will alter its inference significantly if it makes a change to its original guessing strategy. These perturbations will show up as variations of the information metric.

When $d > 2$, the collective simply retains its original inference, the inertia of the group prevents changes.

When $d < 2$, on the other hand, we see that each small perturbation has a dramatic effect on the original inference.

When $d = 2$, we see another possibility, namely that such perturbations could be marginal, and that the collective can smoothly change its inference scheme. This is a special feature of two-dimensional non-linear sigma models.

For completeness, we perform a similar analysis in the case of a change to the weighting scheme, i.e. the potential energy $V(y)$. We consider a perturbation to $w(y)$ of the form:

$$w(y) \to w(y) \exp(\delta\lambda_m \|y\|^m), \tag{3.26}$$

which will show up as a potential term of order $\delta\lambda_m \|y\|^m$ in the field theory. Repeating the same scaling analysis, one learns that $\delta\lambda_m$ scales with the number of agents as:

$$\delta\lambda_m \sim K^{1+\gamma(m)} \tag{3.27}$$

with $\gamma(m)$ as in equation (3.25). Thus, for $d > 2$, we see that a change with $m > 2d/(d-2)$ will not alter the original inference. However, for low enough degree m's, the weighting scheme can significantly alter the outcome. Similarly, we see that for $d \leq 2$, such perturbations are always significant.

3.4   Two-Dimensional Case

Summarizing the previous subsection, we have seen that the case of a two-dimensional agent space is rather special. In this subsection we discuss in more detail the conditions for stable inference in two dimensions. From the perspective of string theory, our discussion of conformal invariance and the appearance of gravity is not new. How we managed to arrive here is another story.

We restrict our attention to the well-motivated subcase $V = 0$ corresponding to the Jeffrey's prior. To begin, we also assume a flat agent space metric h_{ab}.

Renormalization of 2d non-linear sigma models has received enormous attention in the

string theory literature. See for example [11–13] and [14] for a review. In two dimensions there is an additional coupling we can write which also has an information theoretic interpretation:

$$S_{\text{2d}} = \frac{N}{2} \int\limits_{\Sigma_{\text{agent}}} d^2\sigma \; \sqrt{\det h} \; G_{IJ} \, h^{ab} \frac{\partial y^I}{\partial \sigma^a} \frac{\partial y^J}{\partial \sigma^b} + \frac{N}{2} \int\limits_{\Sigma_{\text{agent}}} d^2\sigma \; iB_{IJ} \, \varepsilon^{ab} \frac{\partial y^I}{\partial \sigma^a} \frac{\partial y^J}{\partial \sigma^b}. \tag{3.28}$$

The coupling B_{IJ} is an additional anti-symmetric tensor on the target space, and ε^{ab} is a constant anti-symmetric tensor in the two-dimensional geometry. In the target space, a background value of B_{IJ} specifies a notion of torsion, that is, parallel transport with respect to a connection ∇_B distinct from the Levi-Cevita connection ∇_{LC} on Y_{target}. This recovers the use of connections with torsion in information geometry discussed in section 2. Readers familiar with string theory will note that we have omitted a factor of the dilaton coupling. This is because the collective works with respect to a fixed and flat agent space metric.

Assuming the response of the collective is independent of the number of agents, we arrive at the condition of conformal invariance for the 2d sigma model. Treating the values of G_{IJ}, B_{IJ} as background fields of the target space geometry, conformal invariance implies the conditions (see e.g. [14] for a review):

$$R_{IJ} - \frac{1}{2}G_{IJ}R = \frac{1}{4}\left(H_{IJ}^2 - \frac{1}{6}G_{IJ}H^2\right) \quad \text{and} \quad \nabla^I H_{IJK} = 0 \tag{3.29}$$

where R_{IJ} is the Ricci tensor for the information metric, R is the Ricci scalar, $H_{IJK} = \partial_I B_{JK} + \partial_K B_{IJ} + \partial_J B_{KI}$. Here we have not written subleading higher derivative corrections, which are suppressed by powers of $1/N$. Equations (3.29) are the Einstein field equations, coupled to a background flux. As is well-known, they are reproduced by a principle of least action on the M-dimensional target:

$$S_{\text{target}} = \int\limits_{Y_{\text{target}}} d^M y \; \sqrt{\det G} \; \left(R - \frac{1}{12}H^2\right). \tag{3.30}$$

Adding an explicit source of stress energy T_{IJ} to the Einstein equations corresponds to exposing the collective to a new source of information. In the semi-classical approximation, we can also identify the graviton as a fluctuation in the information metric, that is, a fluctuation in the family of probability distributions adopted by the collective.[4]

From these considerations alone, however, Newton's constant is simply a parameter of the theory. Indeed, our considerations here are weaker than what would follow from a full string theory. In particular, the absence of a dilaton equation of motion means that we cannot

---

[4]One might speculate on a connection to entropic gravity [15], though we do not do so here.

fix Newton's constant, and further, that the number of fitting parameters M is subject to no constraint. The price we pay for this is that our description will break down at short distances on Y_{target}, namely "the string scale":

$$\ell_{\min} \sim 1/\sqrt{N}. \tag{3.31}$$

This reflects the underlying $\sqrt{N}$ statistics from the number of sampled of events, so we typically do not demand more.

If we do demand that our description makes sense below the resolution scale of the statistics, we need the machinery of string theory. We get a string theory by summing over all arrangements of the agents, that is, by coupling our sigma model to 2d gravity on the worldsheet. Then, the dilaton Φ will automatically appear, and will couple to the worldsheet curvature via the term $\Phi \cdot \mathcal{R}_\Sigma$. The string frame metric is then the information metric. Stable inference schemes of the collective then yield the stronger condition of Weyl invariance on a curved worldsheet. We can eliminate tachyonic modes by passing to the superstring. Indeed, there is a natural extension of our analysis to the case where the agent space is a supermanifold.

4   Information and Compactification

In this section we switch gears, applying information geometry to study compactifications of the physical superstring. There are two interlinked questions we wish to study. First, we want to know whether information metrics produce accurate approximations to metrics of compactification manifolds. Second, we want to know how much information an observer would gain by correcting an incorrect guess of the internal geometry.[5]

We shall assume a compactification of the physical superstring of the form $\mathrm{R}^{3,1} \times Y$, where Y is a Calabi-Yau threefold. We expect that most of our considerations extend to other Kähler threefolds, as can occur in various F-theory compactifications.

Since Y is Calabi-Yau, and therefore a Kähler manifold, its metric is locally controlled by a Kähler potential ϕ_Y. We can split up the local coordinates y^I into holomorphic and anti-holomorphic coordinates y^i and $\overline{y}^{\overline{j}}$. The metric is then:

$$ds^2 = G_{i\overline{j}}^{\text{Kähler}}\, dy^i d\overline{y}^{\overline{j}} = \partial_i \partial_{\overline{j}} \phi_Y\, dy^i d\overline{y}^{\overline{j}}. \tag{4.1}$$

Compare this to the Fisher information metric. Assuming Y is our statistical manifold of parameters for some probability space X, we get distributions $p(x, \{y\})$, and an information

---

[5]One of the original motivations for this note was to understand in information theoretic terms the non-commutative geometry of F(uzz) theory [16], as well as the evidence presented in [17–19] for the interplay between fuzzy UV cutoffs and an emergent gravitational sector.

metric on Y:

$$G_{IJ}^{\text{Fisher}} = -\int_X d\mu(x) \, p\frac{\partial^2 \log p}{\partial y^I \partial y^J} = -\left\langle \frac{\partial^2 \log p}{\partial y^I \partial y^J} \right\rangle_p. \tag{4.2}$$

Making the further assumption that only the mixed holomorphic and anti-holomorphic terms contribute, we see that $\log p^{-1}$ defines a stochastic generalization of the Kähler potential.

To complete the circle of ideas, we need to find a collection of p's such that $G_{i\bar{j}}^{\text{Fisher}}$ is a good approximation to $G_{i\bar{j}}^{\text{Kähler}}$. Assuming that this can be done, we can check the inference abilities of an observer by comparing the true distribution to a nearby guess. This provides a measure of how well an observer can deduce the local geography of a compactification.

4.1   Construction of Metrics

Our strategy for constructing a family of information metrics will be to initially take $X = \mathbb{C}^M$, and an ambient target space $Y = \mathbb{C}^M$. Starting from this, we show how to reduce to an information metric on $Y = \mathbb{CP}^{M-1}$, and by embedding a Calabi-Yau in such a projective space, we shall produce a numerical approximation for a Calabi-Yau metric.

We first construct a family of information metrics on $Y = \mathbb{C}^M$ using the Gaussian:

$$p(x, \{s\}) = \pi^{-M} \det h \, \exp(-h_{\alpha\bar{\beta}}(x^\alpha - s^\alpha)(\overline{x}^{\bar{\beta}} - \overline{s}^{\bar{\beta}})), \tag{4.3}$$

where x^α denote holomorphic coordinates for $X = \mathbb{C}^M$, s^α denote holomorphic coordinates of the parameter manifold $Y = \mathbb{C}^M$, and $h_{\alpha\bar{\beta}}$ is an $M \times M$ invertible Hermitian matrix with positive eigenvalues. The non-vanishing entries of the information metric:

$$G_{\alpha\bar{\beta}}^{\text{Fisher}} = \left\langle h_{\alpha\overline{\alpha}'}(\overline{x}^{\overline{\alpha}'} - \overline{s}^{\overline{\alpha}'})(x^{\beta'} - s^{\beta'})h_{\beta'\bar{\beta}} \right\rangle_p = h_{\alpha\bar{\beta}}, \tag{4.4}$$

are generated by the Kähler potential:

$$\phi_{(h)} = h_{\alpha\bar{\beta}} s^\alpha \overline{s}^{\bar{\beta}}. \tag{4.5}$$

Viewing s^α as the guess of where a probe is located inside the geometry, we can also adjust the parameters $h_{\alpha\bar{\beta}}$ which specify the width of the guess. This leads to a much larger parameter space, with information metric:

$$ds^2 = h_{\alpha\bar{\beta}} ds^\alpha d\overline{s}^{\bar{\beta}} + h^{\bar{\delta}\alpha} h^{\bar{\beta}\gamma} dh_{\alpha\bar{\beta}} dh_{\gamma\bar{\delta}}. \tag{4.6}$$

To construct information metrics on other parameter spaces, we restrict the range of parameters of $Y = \mathbb{C}^M$. For example, we can define a sphere S^{2M-1} by restriction on the

parameter manifold $Y = \mathbb{C}^M$ to the fixed locus:

$$h_{\alpha\overline{\beta}} s^\alpha \overline{s}^{\overline{\beta}} = r^2. \tag{4.7}$$

The restriction to this locus (holding $X = \mathbb{C}^M$ fixed) yields a pullback of the metric on $Y = \mathbb{C}^M$ to S^{2M-1}. By a similar token, we can also pull back the probability density on $X = \mathbb{C}^M$ to S^{2M-1}. Here, we must be careful to respect the condition that p is a density, and not a scalar of X. The restriction to $\mathbb{CP}^{M-1}$ now follows by reducing along the S^1 fiber of the Hopf fibration $S^1 \to S^{2M-1} \to \mathbb{CP}^{M-1}$, so we produce an information metric on $\mathbb{CP}^{M-1}$ which agrees with the Fubini-Study metric generated by the Kähler potential:

$$\phi_{(h)}^{FS} = \log\left(h_{\alpha\overline{\beta}} s^\alpha \overline{s}^{\overline{\beta}}\right), \tag{4.8}$$

with the s^α now treated as homogeneous coordinates of $\mathbb{CP}^{M-1}$.

A similar construction also applies to the case where Y is Calabi-Yau. The idea is to fix an ample line bundle $\mathcal{L}$ over Y. Since $\mathcal{L}$ is ample, the space of sections $H^0(Y, \mathcal{L}) \simeq \mathbb{C}^M$ defines an embedding of Y into the projective space $\mathbb{CP}^{M-1}$:

$$i : Y \to \mathbb{CP}^{M-1} \quad \text{where} \quad y \mapsto s^\alpha(y). \tag{4.9}$$

This is an embedding to projective space since the s^α cannot all simultaneously vanish. One can view the Calabi-Yau as defined by a set of polynomial relations amongst the s^α. This need not be a complete intersection, since there could be relations amongst the relations (syzygies). Given the Fubini-Study Kähler form $\omega_{(h)}^{FS}$ on $\mathbb{CP}^{M-1}$, we get a $(1,1)$ form on Y via the pullback $i^*(\omega_{(h)}^{FS})$. Hence, each choice of Hermitian matrix $h_{\alpha\overline{\beta}}$ defines a candidate metric for Y. We view this choice as the inference strategy of a given agent.

To produce a best approximation, we follow the numerical procedure for constructing balanced Calabi-Yau metrics used in [6, 7]. The optimal choice is fixed under the T-map (see [4, 7]):

$$T(h)^{\alpha\overline{\beta}} = \frac{M}{\text{Vol}_Y} \int d\text{Vol}_Y \; \frac{\overline{s}^{\overline{\beta}} s^\alpha}{h_{\gamma\overline{\delta}} s^\gamma \overline{s}^{\overline{\delta}}}. \tag{4.10}$$

For a Calabi-Yau, the fixed point exists. Now, although this is not yet the $(1,1)$ form of the Calabi-Yau threefold, repeating the construction for sufficiently high degree powers of the line bundle $\mathcal{L}^k$ defines a sequence of balanced metrics which eventually converge to the Kähler form of the Calabi-Yau metric (see [4, 5]):

$$\frac{1}{k} i_k^*(\omega_k) \to \omega_{CY_3} \tag{4.11}$$

as $k \to \infty$. Observe that in this construction, the Chern class $c_1(\mathcal{L}) = [\omega_{CY_3}]$, i.e. it picks a

ray in the Kähler cone which is compatible with geometric quantization of the Calabi-Yau [4].

The statistical inference interpretation should be clear. We can make a guess both as to the location $y \in Y$ of the observer, as well as the local profile of the Calabi-Yau metric, via $h_{\alpha\bar{\beta}}$. A sequence of improved guesses, both in the choice of location, and Hermitian metric eventually converge on an adequate approximation of the internal geometry.

To illustrate how much an observer can learn by adjusting an incorrect guess of the local geometry, let us return to the case of distributions on the ambient space $\mathbb{C}^M$. Similar comments apply for all of the pullback / restriction maps. Suppose that the true distribution of the observer is centered at some point $s_{(T)}^\alpha$, and that it is designated by a choice of Hermitian matrix $h_{(T)}$. Then, if our observer makes an incorrect guess given by some other choice of $s_{(A)}^\alpha$ and $h_{(A)}$, the amount of information it would gain by adjusting its choice is the Kullback-Leibler divergence:

$$D_{KL}\left(p_T || p_A\right) = \log\left(\frac{\det h_{(T)}}{\det h_{(A)}}\right) + \text{Tr}\left(h_{(A)} \cdot h_{(T)}^{-1} - h_{(T)} \cdot h_{(T)}^{-1}\right) + \left(s_{(A)} - s_{(T)}\right) \cdot h_{(A)} \cdot \left(\bar{s}_{(A)} - \bar{s}_{(T)}\right).$$

(4.12)

The observer learns both by locating its position, and its width correctly. The supergravity limit of the compactification corresponds to a sharply localized observer with $h_{(A)} \to \infty$, and the small volume limit corresponds to sending $h_{(A)} \to 0$. In both cases, the observer learns much by adjusting its width. By the same token, once the observer get close to the true distribution, it ceases to learn very much about the internal geometry. In this sense, though there is only one true distribution, a nearby guess is "good enough".

4.2   Quantum Interpretation

The information geometry of a Calabi-Yau is also closely connected with tiling a non-commutative deformation of the geometry with nearly pointlike branes, as in F(uzz) theory [16]. In this subsection we show that the Hilbert space of fuzzy points for $\mathbb{C}^M$ induces a family of quantum information metrics for the Calabi-Yau.

Let us begin by briefly reviewing some background on quantum information metrics. Given a Hilbert space $\mathcal{H}$ and density matrix ρ which depends on continuous parameters y, we can introduce a quantum analogue of the Fisher information metric, known as the Helstrom / Bures metric (see [20, 21]):

$$G_{IJ}^{\text{HB}} \equiv \frac{1}{4} Tr_{\mathcal{H}}\left(\rho \frac{\partial L_\rho}{\partial y^{(I}} \frac{\partial L_\rho}{\partial y^{J)}}\right),$$

(4.13)

where we have chosen a convenient normalization convention. Here, $\partial L_\rho / \partial y^I$ is implicitly

defined via:

$$\frac{\partial \rho}{\partial y^I} = \frac{1}{2}\left(\rho\frac{\partial L_\rho}{\partial y^I} + \frac{\partial L_\rho}{\partial y^I}\rho\right). \tag{4.14}$$

Extending the cases treated in [16], we realize a non-commutative Calabi-Yau threefold by first constructing non-commutative $\mathbb{C}^M$. Imposing a level constraint and additional holomorphic constraints leads to non-commutative geometry on subspaces. Since the restriction maps of the previous section were specified by holomorphic conditions, we can carry over our construction to the fuzzy case.

In more detail, we first construct the fuzzy Hilbert space of points for $\mathbb{C}^M$. Introduce harmonic oscillators a^α, $\overline{a}^{\overline{\beta}}$ which are subject to the commutators:

$$\left[a^\alpha, \overline{a}^{\overline{\beta}}\right] = h^{\overline{\beta}\alpha}. \tag{4.15}$$

Next, introduce a vacuum state $|0\rangle$ such that $a^\alpha|0\rangle = 0$. We then build up a Fock space of fuzzy points for $\mathbb{C}^M$ by acting with creation operators $\overline{a}^{\overline{\beta}}$ on $|0\rangle$. Denote this Hilbert space of states by $\mathcal{H}_{\mathbb{C}^M}$. For each s^α, introduce a coherent state:

$$|s\rangle = \exp(-h_{\alpha\overline{\beta}}s^\alpha \overline{s}^{\overline{\beta}}/2)\exp(h_{\alpha\overline{\beta}}s^\alpha \overline{a}^{\overline{\beta}})|0\rangle \tag{4.16}$$

$$\langle s| = \langle 0|\exp(h_{\alpha\overline{\beta}}a^\alpha \overline{s}^{\overline{\beta}})\exp(-h_{\alpha\overline{\beta}}s^\alpha \overline{s}^{\overline{\beta}}/2). \tag{4.17}$$

These are normalized states, and form an overcomplete basis for $\mathcal{H}_{\mathbb{C}^M}$. We can introduce a further grading of $\mathcal{H}_{\mathbb{C}^M}$ by the operator:

$$\widehat{R}^2 = h_{\alpha\overline{\beta}}a^\alpha \overline{a}^{\overline{\beta}}. \tag{4.18}$$

We refer to an eigenspace of $\widehat{R}^2$ at eigenvalue r^2 as the Hilbert space $\mathcal{H}_{\mathbb{CP}^{M-1}}$ of points for fuzzy $\mathbb{CP}^{M-1}$ at radius r. In what follows we work at large radius. This is the fuzzy analogue of the Kähler quotient constraint:

$$\mathbb{CP}^{M-1} = \left\{s^\alpha \in \mathbb{C}^M : h_{\alpha\overline{\beta}}s^\alpha \overline{s}^{\overline{\beta}} = r^2 \text{ modulo } U(1) \text{ rephasings}\right\}. \tag{4.19}$$

The fuzzy Hilbert space of points for a Calabi-Yau Y is obtained by restricting to states of $\mathcal{H}_{\mathbb{CP}^{M-1}}$ annihilated by the holomorphic relations $f_i(s^\alpha) = 0$ of the commutative geometry:

$$\mathcal{H}_Y = \left\{|\psi\rangle \in \mathcal{H}_{\mathbb{CP}^{M-1}} : f_i(a^\beta)|\psi\rangle = 0\right\}, \tag{4.20}$$

where the f_i are themselves possibly subject to further relations (syzygies).

We now produce a quantum information metric for $\mathbb{CP}^{M-1}$, and pull this back to Y via the embedding $i : Y \rightarrow \mathbb{CP}^{M-1}$. We begin with a construction of the Fubini-Study metric

on $\mathbb{CP}^{M-1}$ by working with the density matrix for the family of pure states:

$$\rho(s) = |s\rangle \langle s|. \tag{4.21}$$

Observe that since $\rho^2 = \rho$, we have $2\delta\rho = \delta L_\rho$.

Thus, to compute the quantum information metric it is enough to consider the variation of $\rho(s)$ as we change $s^\alpha \to s^\alpha + \delta s^\alpha$, subject to the variation of the Kähler quotient constraint in line (4.19):

$$h_{\alpha\overline{\beta}}\delta s^\alpha \overline{s}^{\overline{\beta}} = h_{\alpha\overline{\beta}} s^\alpha \delta\overline{s}^{\overline{\beta}} = 0. \tag{4.22}$$

To leading order in δs, the normalization of $|s\rangle$ remains fixed because $\delta \exp(-h_{\alpha\overline{\beta}} s^\alpha \overline{s}^{\overline{\beta}}/2) \simeq O(\delta s^2)$. Hence,

$$\delta\rho(s) = |\delta s\rangle \langle s| + |s\rangle \langle \delta s| + O(\delta s^2) \tag{4.23}$$

where we have introduced the unnormalized states:

$$|\delta s\rangle = h_{\alpha\overline{\beta}}\delta s^\alpha \overline{a}^{\overline{\beta}} |s\rangle \quad \text{and} \quad \langle \delta s| = \langle s| h_{\alpha\overline{\beta}} a^\alpha \delta\overline{s}^{\overline{\beta}}. \tag{4.24}$$

Note that the transversality condition of line (4.22) implies $\langle s|\delta s\rangle = \langle \delta s|s\rangle = 0$. The quantum information metric therefore reduces to:

$$ds^2 = \langle \delta s|\delta s\rangle = h_{\alpha\overline{\beta}} ds^\alpha d\overline{s}^{\overline{\beta}}, \tag{4.25}$$

so we recover the Fubini-Study metric on $\mathbb{CP}^{M-1}$ due to the Kähler quotient constraint of line (4.19). Returning to our discussion in subsection 4.1, the pullback to Y then gives an approximation for the Calabi-Yau metric.

5 Speculative Remarks

In this section we entertain some speculative remarks on other possible uses of information geometry in the study of superstring theory. We take an agent to be a D-instanton, that is, the endpoint of an open string. The D-instanton comes with a local position, and possibly other parameters such as its width, when it dissolves as flux in another brane. The natural choice of Y in the supersymmetric setting is simply the moduli space of the brane.

5.1 Simplistic Holography

One of the most direct ways to probe the AdS/CFT correspondence is via D-instantons [22–25]. Indeed, the instanton density in Yang-Mills theory defines an unnormalized probability distribution, and a corresponding information metric [26] which provides another way to

interpret the AdS/CFT correspondence [27].

Rather than delve into the details of the instanton moduli for a specific gauge theory, we instead try a more simplistic approach. Given a large N_c stack of D3-branes filling R^4 in a consistent 10D geometry, we consider a D-instanton inside the worldvolume of the stack. We can view the D-instanton as independently sampling the D3-branes many times with the large number of D_{-1} / D3 strings. Assuming finite mean and variance for this sampling, the central limit theorem allows us to approximate the profile of the D-instanton by a Gaussian on R^4. In physical terms, it is actually most natural to use an unnormalized parametric family of such distributions:

$$p(x, \{y^{(1)}, ..., y^{(4)}, \alpha\}) = \frac{L^2}{2} \left(\frac{1}{\pi \alpha^2}\right)^{4/2} \exp\left[-\sum_{I=1}^{4} \frac{(x^{(I)} - y^{(I)})^2}{\alpha^2}\right]. \tag{5.1}$$

The mean of the distribution tells us the average position of the D-instanton, and the width sets a resolution length. The information metric for this distribution is 5D Euclidean AdS space of radius L:

$$ds^2 = L^2 \times \frac{dy_{(I)}dy^{(I)} + d\alpha^2}{\alpha^2}. \tag{5.2}$$

Similar considerations hold in other dimensions. Observe that our discussion only required a large number of strings to sample the stack of D3-branes, and is otherwise insensitive to the details of the gauge theory. It would be very interesting to see whether a simplistic treatment along the lines presented here could also reproduce the S^5 factor of the $\mathcal{N} = 4$ holographic dual.

5.2 Lorentzian Signature Statistics

Much of our focus in this work has been on the inference abilities of a collective of agents probing a Riemannian manifold. For applications to physics, one should eventually return to Lorentzian signature. This opens up some new possibilities as well as challenges from the perspective of information geometry.

To illustrate the main issue, consider again the normal distribution, but now in Lorentzian signature. We can of course simply write down a formal Gaussian profile on $\mathrm{R}^{M,1}$:

$$p_{\text{formal}}(x^\mu) \equiv (2\pi\alpha_t^2)^{-1/2} \left(2\pi\alpha_t^2\right)^{-M/2} \exp\left[\frac{(t - y_0)^2}{2\alpha_t^2} - \sum_{I=1}^{M} \frac{(x^{(I)} - y^{(I)})^2}{2\alpha_x^2}\right] \tag{5.3}$$

via the formal analytic continuation from $\alpha_t \to i\alpha_t$ (and taking the norm). Since analytic continuation of a D-instanton to Lorentzian signature defines a tunneling event, we take the unbounded behavior of p_{formal} to mean that while the agent eventually settles on a choice of

y_i for spatial directions, it continues to move forward in the temporal direction.

Yet another new feature of Lorentzian signature spacetimes is the possible existence of null vectors. In information theoretic terms, this means nothing is learned by moving along the null direction. It would be interesting to study the consequences for black hole event horizons and cosmological horizons.

A temporal direction in statistical inference also shows up from sequential updating of the events $e_1, ..., e_N$ received by an agent. We can then drop the distinction between N and $K^{1/d}$ by using a lattice approximation for a $(d + 1)$-dimensional field theory. It would be interesting to connect this to the field theory interpretation of Bayesian updating in [28].

6   Conclusions

In this note we have studied the interplay between statistical inference and string theory. We have shown that when a large number of agents form a collective of nearby guesses, the pooled posterior probability of an accurate inference by the collective is the partition function of a non-linear sigma model. Quite surprisingly, stability of the inference scheme against perturbations requires the collective to arrange along a two-dimensional grid. Using the well-known fact that the Einstein field equations follow from the condition of conformal invariance in 2d non-linear sigma models, we found classical gravity emerge from the condition of stable statistical inference. We have also taken some preliminary steps in applying methods from information geometry in the study of the physical superstring.

Developing more practical applications of this formalism would be quite interesting. For example, one might try to simulate a dynamical updating strategy for a collective of agents.

Much of our discussion generalizes to the quantum setting, with a quantum information metric for the non-linear sigma model. It would be interesting to further study the interpretation in the context of both quantum statistical inference and string theory.

Finally, we find it rather remarkable that stable computations by a collective requires a two-dimensional worldsheet, and moreover, that this yields a theory of gravity. It would be exciting to further develop an information theoretic formulation of such "computables".

Acknowledgements

References

[1] V. Balasubramanian, "A Geometric Formulation of Occam's Razor For Inference of Parametric Distributions," `arXiv:adap-org/9601001`.

[2] V. Balasubramanian, "Statistical Inference, Occam's Razor and Statistical Mechanics on the Space of Probability Distributions," *Neural Comp.* **9(2)** (1997) 349–368, `arXiv:cond-mat/9601030`.

[3] S.-I. Amari and H. Nagaoka, *Methods of Information Geometry*, vol. 191. Translations of Mathematical Monographs; American Mathematical Society and Oxford University Press, Providence, RI, 2000.

[4] G. Tian, "On a set of polarized Kähler metrics on algebraic manifolds," *J. Diff. Geom.* **32** (1990) 99–130.

[5] S. Zelditch, "Szego kernels and a theorem of Tian," *Int. Math. Res. Notices* **6** (1998) 317–331.

[6] S. D. Donaldson, "Some numerical results in complex differential geometry," `arXiv:math.DG/0512625`.

[7] M. R. Douglas, R. L. Karp, S. Lukic, and R. Reinbacher, "Numerical Calabi-Yau metrics," *J. Math. Phys.* **49** (2008) 032302, `arXiv:hep-th/0612075`.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, New York, NY, 1991.

[9] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.* **27** (1948) 379–423.

[10] N. N. Chentsov, *Statistical Decision Rules and Optimal Inference*. Translations of Mathematical Monographs; American Mathematical Society, Providence, RI, 1982.

[11] D. H. Friedan, "Nonlinear Models in $2 + \epsilon$ Dimensions," *Ann. Phys.* **163** (1985) 318.

[12] L. Alvarez-Gaume, D. Z. Freedman, and S. Mukhi, "The Background Field Method and the Ultraviolet Structure of the Supersymmetric Nonlinear Sigma Model," *Ann. Phys.* **134** (1981) 85.

[13] C. G. Callan Jr., E. Martinec, M. Perry, and D. Friedan, "Strings in Background Fields," *Nucl. Phys.* **B262** (1985) 593.

[14] C. G. Callan Jr. and L. Thorlacius, "Sigma Models and String Theory," *Providence 1988, Proc. of the Theor. Adv. Study Inst. in Elem. Part. Physics: Particles, Strings and Supernovae* **2** (1989) 795–878.

[15] E. P. Verlinde, "On the Origin of Gravity and the Laws of Newton," *JHEP* **1104** (2011) 029, `arXiv:1001.0785 [hep-th]`.

[16] J. J. Heckman and H. Verlinde, "Evidence for F(uzz) Theory," *JHEP* **01** (2011) 044, `arXiv:1005.3033 [hep-th]`.

[17] J. J. Heckman and H. Verlinde, "Super Yang-Mills Theory as a Twistor Matrix Model," `arXiv:1104.2605 [hep-th]`.

[18] J. J. Heckman and H. Verlinde, "Gravity Amplitudes from a Gaussian Matrix Model," `arXiv:1112.5209 [hep-th]`.

[19] J. J. Heckman and H. Verlinde, "Instantons, Twistors, and Emergent Gravity," `arXiv:1112.5210 [hep-th]`.

[20] C. W. Helstrom, "Minimum mean-squared error of estimates in quantum statistics," *Phys. Lett.* **A25** (1967) 101–102.

[21] D. Bures, "An Extension of Kakutani's Theorem on Infinite Product Measures to the Tensor Product of Semifinite w^*-Algebras," *Trans. Am. Math. Soc.* **135** (1969) 199–212.

[22] T. Banks and M. B. Green, "Non-perturbative Effects in $AdS \times S^5$ String Theory $d = 4$ SUSY Yang-Mills," *JHEP* **9805** (1998) 002, `arXiv:hep-th/9804170`.

[23] M. Bianchi, M. B. Green, S. Kovacs, and G. Rossi, "Instantons in supersymmetric Yang-Mills and D-instantons in IIB superstring theory," *JHEP* **9808** (1998) 013, `arXiv:hep-th/9807033`.

[24] N. Dorey, T. J. Hollowood, V. V. Khoze, M. P. Mattis, and S. Vandoren, "Multi-Instantons and Maldacena's Conjecture," *JHEP* **9906** (1999) 023, `arXiv:hep-th/9810243`.

[25] N. Dorey, T. J. Hollowood, V. V. Khoze, M. P. Mattis, and S. Vandoren, "Multi-Instanton Calculus and the AdS / CFT Correspondence in $\mathcal{N} = 4$ Superconformal Field Theory," *Nucl. Phys.* **B552** (1999) 88–168, `arXiv:hep-th/9901128`.

[26] N. J. Hitchin, *The Geometry and Topology of Moduli Spaces, in Global Geometry and Mathematical Physics*, vol. 1451. Springer, Heidelberg, 1988.

[27] M. Blau, K. Narain, and G. Thompson, "Instantons, the Information Metric, and the AdS / CFT Correspondence," `arXiv:hep-th/0108122`.

[28] W. Bialek, C. G. Callan Jr., and S. P. Strong, "Field Theories for Learning Probability Distributions," *Phys. Rev. Lett.* **77** (1996) 4693–4697, `arXiv:cond-mat/9607180`.