

On Sampling from Multivariate Distributions*

Zhiyi Huang and Sampath Kannan

University of Pennsylvania, Philadelphia PA 19104, USA,
{hzhiyi, kannan}@cis.upenn.edu

Abstract. Let X_1, X_2, \dots, X_n be a set of random variables. Suppose that in addition to the prior distributions of these random variables we are also given *linear constraints relating them*. We ask for necessary and sufficient conditions under which we can efficiently sample the constrained distributions, find constrained marginal distributions for each of the random variables, etc. We give a tight characterization of the conditions under which this is possible. The problem is motivated by a number of scenarios where we have separate probabilistic inferences in some domain, but domain knowledge allows us to relate these inferences. When the joint prior distribution is a product distribution, the linear constraints have to be carefully chosen and are crucial in creating the lower bound instances. No such constraints are necessary if arbitrary priors are allowed.

Keywords: sampling, algorithm, complexity

1 Introduction

Suppose we are solving a crossword puzzle, sudoku or other grid puzzle. Looking at the column in which a grid cell lies, we could come up with a random variable X (with known prior distribution) that is the value in the cell. Similarly, looking at the row in which the cell lies, we could come up with another random variable Y for the value in the cell. Of course, these are two random variables for the value in one cell, and hence we have the constraint $X = Y$. If we could come up with a set of such random variables and constraints on them and compute the constrained distributions on these random variables, we would go a long way towards solving these puzzles.

More usefully, consider the following scenarios where we make related inferences:

In bioinformatics we have programs that attempt to locate the starting position of a gene along a genome. These programs typically produce a distribution on positions. We also have programs that infer the location of other genomic elements, such as transcription factor binding sites, and again produce distributions on the possible positions of these elements. Biological domain knowledge

* This material is based upon work supported by the National Science Foundation under Grant No. 0716172 and ONR MURI Grant N000140710907.

tells us something about how far apart the gene and the binding site can be, and this can be modeled as a constraint between these random variables.

In image segmentation and processing, say of human images, the positions of limbs, torsos, heads, etc., are identified probabilistically. Anatomical constraints are then the linear constraints between these random variables, and determining their constrained distributions helps localize the image more precisely.

We model these problems as follows: X_1, X_2, \dots, X_n are a given set of random variables. We assume we know their prior joint probability distribution F . (Later in this paper we consider the special case where the random variables are originally independent and this joint distribution is a product distribution.) Without loss of generality, we assume that each random variable has a support that is a subset of $[0, 1]$. The linear constraints on the random variables define a feasible region, which is a convex body K .

Thus we will study the problem of sampling vectors in the unit hypercube $[0, 1]^n$ according to some multivariate distribution F subject to the condition that we want to sample only inside a convex body K . Here, F is a multivariate distribution of an n -dimensional random vector $\mathbf{X} = (X_1, \dots, X_n) \in [0, 1]^n$ (We will always assume that F is defined on the entire unit hypercube.) We seek to answer the following question: *What is the necessary and sufficient condition for F such that the sampling can be done in polynomial time?*

Model and definition. We will follow the notation in [1] in general. For every vector $\mathbf{x} = (x_1, \dots, x_n) \in [0, 1]^n$, we let $F(\mathbf{x})$ denote the probability density of \mathbf{x} in distribution F . For every $i \in [n]$, we let $F_i(\cdot)$ denote the marginal density function of the i^{th} coordinate. We will let $f(\cdot)$ and $f_i(\cdot)$ denote the functions $\log F(\cdot)$ and $\log F_i(\cdot)$.

We let $K(\mathbf{x})$ denote the membership indicator function of the convex body K , that is, $K(\mathbf{x}) = 1$ if $\mathbf{x} \in K$; and $K(\mathbf{x}) = 0$ otherwise.

While the algorithms for sampling (e.g. [1]) only require oracle access to F and K , our lower bound results in this paper hold even if $F(\mathbf{x})$ and $K(\mathbf{x})$ are explicitly given in closed form and can be computed in polynomial time.

Lipschitz condition. We will assume that $f(\mathbf{x})$ satisfies the Lipschitz condition for a polynomially large Lipschitz constant α , that is, $|f(\mathbf{x}) - f(\mathbf{x}')| \leq \alpha \|\mathbf{x} - \mathbf{x}'\|_\infty$. This parameter specifies how smooth function $f(\cdot)$ is.

Log-concavity. Previous research showed that the log-concavity of the distribution plays an important role in efficient sampling algorithms. The distribution F is β -close to log-concave if for any $\mathbf{x}, \mathbf{y} \in [0, 1]^n$, and $\lambda \in [0, 1]$, $f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) \geq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{y}) - \beta$.

Problem statement. Concretely, we consider the following problem $\text{SAMPLE}(\epsilon, \alpha, \beta)$: Sample $\mathbf{X} \in \mathbb{R}^n$ from the constrained distribution $F|_K$ with error at most ϵ , where K is a convex body and F satisfies the Lipschitz condition for Lipschitz constant α and is β -close to log-concave. That is, sample \mathbf{X} from some distribution \tilde{F} , such that $\forall \mathbf{x} : \left| \tilde{F}(\mathbf{X} = \mathbf{x}) - F(\mathbf{X} = \mathbf{x} | \mathbf{X} \in K) \right| \leq \epsilon$.

Related work. Inspired by the work by Dyer, Frieze, and Kannan [4] that gave a polynomial time algorithm for estimating the volume of a convex body, Applegate and Kannan [1] proposed an efficient sampling algorithm based on fast convergence rate of a carefully chosen random walk. More concretely, they showed that there is an algorithm for $\text{SAMPLE}(\epsilon, \alpha, \beta)$ that runs in $\tilde{O}(n^3 \alpha^2 e^{2\beta} \log(\frac{1}{\epsilon}))$ time if $K = [0, 1]^n$. This is a polynomial-time algorithm on the unit cube if the distribution F is $O(\log n)$ -close to log-concave and the sampling error ϵ is at most inverse exponentially small. Their result implicitly implies that there is a polynomial time algorithm for general convex bodies K via a simple reduction as follows [6]: Let $d_K(\mathbf{x})$ denote the distance between \mathbf{x} and the closest point in the convex body K . Consider the smoothed membership function K_γ for some large enough γ (i.e. $\gamma \gg \alpha$): $K_\gamma(\mathbf{x}) = e^{-\gamma d_K(\mathbf{x})}$. The function K_γ is log-concave. So the product of functions F and K_γ has the same distance from log-concavity as F does. Moreover, sampling with density proportional to $F(\mathbf{x})K_\gamma(\mathbf{x})$ approximates the constrained distribution $F|_K$ very well. Hence, we can use the sampling algorithm by Applegate and Kannan to handle the case of general convex bodies K via the above reduction.

There have been a series of improvements in the running time [5, 8–10]. Very recently, Chandrasekaran, Deshpande, and Vempala [3] showed that there exists a polynomial time algorithm for sampling according to harmonic concave density functions, a slightly broader family of functions than log-concave functions.

On the other hand, Koutis [7] proved that there is no polynomial time sampling algorithm if the density function is allowed to be $\tilde{\Omega}(\log^3 n)$ -far from log-concave, unless there is a $2^{o(n)}$ algorithm for the HAMILTONIAN PATH problem.

Our results. We close the gap between the upper and lower bounds above by showing that there is no polynomial time algorithm for sampling from density functions that are $\omega(\log n)$ -far from log-concave, unless there exists a $2^{o(n)}$ algorithm for 3SAT. (Section 3.1)

Next, we turn to the more restricted case where the distribution F is a product distribution. The study of this special case is triggered by the fact that in many applications each random variable is the result of an independent experiment or algorithm. Domain knowledge relating these random variables is applied later. Note that if $K = [0, 1]^n$, then we can tolerate the product distribution F to be $O(\log n)$ -far from log-concave *on each coordinate* since we can sample each coordinate independently. The independence of random variables seems to provide more power to sampling algorithms. Surprisingly, we show that in fact independence does not help too much in the sense that for general convex bodies K , there is no polynomial time algorithm for sampling from product distributions that are $\omega(\log n \log \log n)$ -far from log-concave, unless there is a $2^{o(n)}$ algorithm for 3SAT. (Section 3.2)

2 Warm-up: Discrete Case

It is fairly easy to show hardness if the random variables are discrete. The intuitive idea is to consider a uniform distribution on the vertices of a hypercube.

These vertices represent assignments to a 3SAT instance and we will use constraints that enforce that each clause is satisfied (i.e. consider the linear constraints in the standard LP-relaxation). We also add a special variable Z that serves as a “switch” on the constraints. More precisely, for each of the linear constraints $X_i + X_j + X_k \geq 1$, we will convert it into $X_i + X_j + X_k + Z \geq 1$. Hence, when $Z = 1$ the constraints are “off” and any vertex of the hypercube is feasible; but when $Z = 0$, the only feasible vertices correspond to satisfying assignments. Then, computing the marginal distribution of Z is equivalent to solving 3SAT. The details of the construction and the proof of its correctness are deferred to the full version.

3 Continuous Case

Now we turn to the continuous case. Instead of attacking the original sampling problem itself, we will consider the following easier problem of

INTEGRATION (δ, α, β) : *Compute the integral $\int_{x \in K} F(x) dx$ up to multiplicative error δ , where F satisfies the Lipschitz condition for Lipschitz constant α and is β -close to log-concave. That is, compute an approximate integral \tilde{I} , such that $(1 + \delta)^{-1} \int_{x \in K} F(x) dx \leq \tilde{I} \leq (1 + \delta) \int_{x \in K} F(x) dx$.*

Proposition 1. *If there is a poly-time algorithm for **SAMPLE** $((\frac{\epsilon}{24\alpha n})^n, \alpha, \beta)$, then there is a polynomial time algorithm for **INTEGRATION** $(\epsilon, \alpha, \beta)$.*

Proposition 1 indicates that it suffices to show our lower bound for the problem of computing the approximate integral with an $1 + \epsilon$ error for some constant ϵ . The readers are referred to [1] or the full version of this paper for the proof of this proposition.

3.1 Correlated Distributions

In this section, we will consider the general case where the distribution F could be correlated. We develop the following Theorem 1 that closes the gap in the previous results.

Theorem 1. *For any $\beta = \omega(\log n)$, there is no polynomial time algorithm for the problem **INTEGRATION** $(\frac{1}{30}, 2\beta n, \beta)$ even when $K = [0, 1]^n$, unless there exists a $2^{o(n)}$ algorithm for 3SAT.*

The following problem is not directly related to our problems. But we will use it as an intermediate problem in our proofs.

GAP-#3SAT $(g(n))$: *Decide whether a given 3SAT instance with n variables has no feasible assignment or has at least $2^n/n^{g(n)}$ feasible assignments.*

High-level sketch of our approach. Let us first consider the following overly simplified approach. Divide the hypercube $[0, 1]^n$ into 2^n smaller hypercubes via n hyperplanes $x_i = \frac{1}{2}$, $i = 1, 2, \dots, n$. Each small hypercube contains exactly one integral point. We will hardwire a 3SAT instance with the function F as follows: For every $\mathbf{x} \in [0, 1]^n$, let $[\mathbf{x}]$ denote the integral point that is in the same small hypercube with \mathbf{x} (if \mathbf{x} is on the boundary of two or more small hypercubes, define $[\mathbf{x}]$ to be an arbitrary one of the corresponding integral points); let $F(\mathbf{x}) = 1$ if $[\mathbf{x}]$ is a satisfying assignment for the 3SAT instance and let $F(\mathbf{x}) = 0$ otherwise. It is easy to see that the integral value equals the number of satisfying assignments times $\frac{1}{2^n}$. However, the function F in this approach is highly discontinuous and infinitely far from log-concave. Therefore, we will proceed by considering a smoothed version of this function.

Proof (Proof of Theorem 1). In the following discussion, we assume that $g(n) = \omega(1)$ is a monotone and slowly-growing function of n . We will prove Theorem 1 via the following two lemmas.

Lemma 1. *If there is a poly-time algorithm for INTEGRATION $(\frac{1}{30}, 2\beta n, \beta)$ where $\beta = g(n) \log n$, then there is a poly-time algorithm for GAP-#3SAT $(g(n))$.*

Proof. Suppose there is a polynomial time algorithm for INTEGRATION $(\frac{1}{30}, 2\beta n, \beta)$ where $\beta = g(n) \log n$. Let I be a GAP-#3SAT $(g(n))$ instance with n variables X_1, \dots, X_n . We will demonstrate how to encode I with an instance of INTEGRATION $(\frac{1}{30}, 2\beta n, \beta)$.

We consider an n -dimensional distribution of variables X_1, \dots, X_n . For any $\mathbf{x} = (x_1, \dots, x_n) \in [0, 1]^n$, we let $[\mathbf{x}] = ([x]_1, \dots, [x]_n) \in \{0, 1\}^n$ denote the rounded version of \mathbf{x} , such that for each $1 \leq i \leq n$, we let $[x]_i = 1$ if $x_i \geq \frac{1}{2}$; let $[x]_i = 0$ if $x_i < \frac{1}{2}$.

We will let $\hat{f}(\mathbf{x})$ denote the following function:

$$\hat{f}(\mathbf{x}) = \begin{cases} 0 & , \text{ if } \|\mathbf{x} - [\mathbf{x}]\|_\infty > \frac{1}{2} - \frac{1}{2n}, \\ g(n) \log n & , \text{ if } \|\mathbf{x} - [\mathbf{x}]\|_\infty < \frac{1}{2} - \frac{1}{n}, \\ (n - 1 - 2n \|\mathbf{x} - [\mathbf{x}]\|_\infty) g(n) \log n & , \text{ otherwise.} \end{cases}$$

$\hat{f}(\mathbf{x})$ is a piecewise-linear function that satisfies the Lipschitz condition with Lipschitz constant $2n g(n) \log n$. Further, we have that $\max_{\mathbf{x}} \hat{f}(\mathbf{x}) - \min_{\mathbf{x}} \hat{f}(\mathbf{x}) = g(n) \log n$. Therefore, for every \mathbf{x}, \mathbf{y} and λ , we have $\hat{f}(\lambda \mathbf{x} + (1 - \lambda) \mathbf{y}) \geq \lambda \hat{f}(\mathbf{x}) + (1 - \lambda) \hat{f}(\mathbf{y}) - g(n) \log n$.

So if we let $\hat{F}(\mathbf{x}) = 2^{\hat{f}(\mathbf{x})}$, then $\hat{F}(\mathbf{x})$ is $g(n) \log n$ -close to log-concave. Furthermore, the value of $\hat{F}(\mathbf{x})$ is large if \mathbf{x} is “close” to being integral, that is, $\|\mathbf{x} - [\mathbf{x}]\|_\infty < \frac{1}{2} - \frac{1}{n}$; and it is small if \mathbf{x} is far from being integral.

Next, we will hardwire the GAP-#3SAT $(g(n))$ instance I by constructing the following distribution F based on \hat{F} . Essentially, we will “iron” the function

value to 1 for the areas that do not correspond to a satisfying assignment of I .

$$F(\mathbf{x}) = \begin{cases} \hat{F}(\mathbf{x}) & , \text{ if } [\mathbf{x}] \text{ is a satisfying assignment for } I; \\ 1 & , \text{ otherwise.} \end{cases}$$

It is easy to verify that $F(x)$ is also $g(n) \log n$ -close to log-concave and $\log F(x)$ satisfies Lipschitz condition with Lipschitz constant $2n g(n) \log n = 2\beta n$. On the one hand, if I is not satisfiable, then clearly $\int_{x \in [0,1]^n} F(x) dx = 1$. On the other hand, if I has at least $2^n n^{-g(n)}$ satisfying assignments, then $\int_{x \in [0,1]^n} F(x) dx = 1 + \int_{\mathbf{x}: I([\mathbf{x}])=1} (\hat{F}(\mathbf{x}) - 1) dx \geq 1 + (n^{g(n)} - 1) \frac{2^n}{n^{g(n)}} (\frac{1}{2} - \frac{1}{n})^n = 1 + (1 - n^{-g(n)}) (1 - \frac{2}{n})^n > 1 + \frac{1}{2e^2}$.

The last inequality holds for sufficiently large n because $1 - n^{-g(n)} \rightarrow 1$ and $(1 - \frac{2}{n})^n \rightarrow e^{-2}$ as n approaches ∞ . By our assumption, there is an polynomial time algorithm for the problem INTEGRATION $(\frac{1}{30}, 2\beta n, \beta)$ for $\beta = g(n) \log n$. We can use this algorithm to distinguish these two cases because $(1 + \frac{1}{30})^2 < 1 + \frac{1}{2e^2}$. So we can solve GAP-#3SAT($g(n)$) in polynomial time.

Lemma 2. *If there is a polynomial time algorithm for GAP-#3SAT($g(n)$), then there exists a $2^{O(n/g(n))}$ algorithm for 3SAT.*

Proof. The proof of this lemma relies on padding redundant variables to a 3SAT instance. Suppose there is a polynomial time algorithm A for GAP-#3SAT($g(n)$). Let I be a 3SAT instance with n variables X_1, \dots, X_n and m clauses C_1, \dots, C_m . We shall let I^* denote the 3SAT instance with the same set of clauses and $N - n$ redundant variables X_{n+1}, \dots, X_N , where $N = 2^{n/g(n)}$. We have that $\#3SAT(I^*) = 2^{N-n} \cdot \#3SAT(I)$.

Hence, if I is not satisfiable, then $\#3SAT(I^*) = 0$, otherwise, the number of feasible assignments for I^* is at least $\#3SAT(I^*) \geq 2^{N-n} = 2^N / N^{g(n)} \geq 2^N / N^{g(N)}$. By our assumption, we can use algorithm A to distinguish these two cases. The running time is $\text{POLY}(N) = 2^{O(n/g(n))}$.

Theorem 1 follows easily from these two lemmas.

3.2 Product Distributions

The previous reduction does not apply to product distributions, unless we allow *each of the component distributions* to deviate from log-concave by $\omega(\log n)$. There are two places where the previous approach heavily relies on the correlation of the distributions: First, hardwiring a 3SAT instance with the function F requires the function F to be correlated. Second, the construction of \hat{F} , the function that simulates a discrete hypercube by taking on a large value at points that are “close” to integral in *all* coordinates and a small value otherwise, is also highly correlated. If we construct a product distribution whose total deviation from log-concave is at most $O(\log n)$, then on average each dimension is only $O(\frac{\log n}{n})$ apart from being log-concave. So the areas with only a few fractional

entries will have density that is close to that of the almost integral points, and hence contribute too large a fraction in the integration.

On the other hand, the algorithm for efficient sampling only works if the *total* deviation from log-concave is $O(\log n)$. In this section we close the gap, showing a lower bound that matches the upper bound. More precisely, we will show the following theorem.

Theorem 2. *For any $\beta = \omega(\log N \log \log N)$, there is no polynomial time algorithm that solves the problem INTEGRATION $(1, \beta n^2, \beta)$ for N -variate product distributions subject to a convex body K , unless there exists a $2^{o(n)}$ randomized algorithm for 3SAT.*

Main idea. To get around the first obstacle mentioned above, we will simply encode the 3SAT instance with linear inequalities as we did in the discrete case. The challenge is to overcome the second obstacle. The main idea is to embed an n -dimensional hypercube into an N -dimensional hypergrid where $N > n$ such that any fractional point in the n -dimensional hypercube has a lot of fractional entries in the N -dimensional hypergrid. The purpose of such an embedding is to create a larger gap between the density of the near-integral parts and the parts where at least one of the coordinates is far from integral for a product distribution. The embedding will be in the flavor of a linear error-correcting code: In error-correcting codes the goal is to ensure that changing one bit in the original string will change many bits in the encoded string; in our embedding scheme, the goal is to ensure that one fractional coordinate in the n -dimensional hypercube will lead to many fractional coordinates of the embedded N dimensional hypergrid. Indeed, the embedding scheme used in our construction is inspired by random linear error-correcting codes. For instance, Figure 1 shows how to embed a 2-dimensional cube into a 3-dimensional grid.

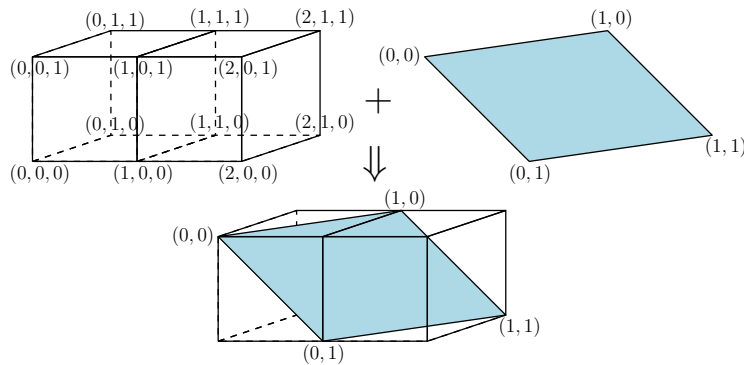


Fig. 1. Embedding 2-dimensional cube into a 3-dimensional grid

In the following discussion, we will again assume that $g(\cdot) = \omega(1)$ is a slowly growing monotone function. Suppose there is a polynomial time algorithm that

tolerates $O(g(N) \log N \log \log N)$ deviation from log-concavity. Then, we will prove Theorem 2 by showing that there is a $2^{n/g(n)}$ algorithm for 3SAT. The key technical step in our construction will be embedding an n -dimensional hypercube into a $2^{n/g(n)}$ -dimensional hypergrid. For the sake of exposition, we will omit the Lipschitz condition in the hard instance in this paper. The Lipschitz condition can be easily fixed by considering a smoothed version of our instance (e.g. a piecewise linear density function).

Some notation. We let $N = 2^{n/g(n)}$. Hence, we have that $\log N = n/g(n)$ and $\log n = \Theta(\log \log N)$. We let $\tau = 2^{\frac{6n \log n}{N}}$. So $\tau^N = n^{6n}$. For any real number $z \in \mathbb{R}$, we let $\lfloor z \rfloor$ denote the nearest integer value to z , that is, $\lfloor z \rfloor = \lfloor z \rfloor$ if $z - \lfloor z \rfloor \leq \frac{1}{2}$; and $\lceil z \rceil = \lceil z \rceil$ otherwise. We let $\{z\} \stackrel{def}{=} |z - \lfloor z \rfloor|$ denote the distance between z and $\lfloor z \rfloor$.

The basic instance. We define the *basic instance* Π as follows. Let $Z_i \in [0, n]$, $i \in \{0, 1\}^n$, be 2^n i.i.d. random variables. Here we relax the support of the distributions from $[0, 1]$ to $[0, n]$ in order to simplify notation. It is clear that such a relaxation is without loss of generality for constructing hard instances. For each $i \in \{0, 1\}^n$, Z_i follows the distribution with density: $F_Z(Z_i = z_i) = c\tau$ if $0 \leq \{z_i\} \leq \frac{1}{10}$ and $F_Z(Z_i = z_i) = c$ otherwise. Here $c = \frac{5}{(\tau+4)^n}$ is the normalization factor. Hence, points that are close to integer have larger density.

By the above definition, the distribution of each Z_i is $\log \tau = \frac{6n \log n}{N}$ -close to log-concave. So the joint distribution of any N of these Z_i 's is $6n \log n$ -close to log-concave. By definition of N , we get $6n \log n = O(g(N) \log N \log \log N)$.

Fact 1 *The joint distribution of any N of the Z_i 's is $O(g(N) \log N \log \log N)$ -close to log-concave.*

We will consider another n i.i.d. random variables $X_1, \dots, X_n \in [0, 1]$. Each X_i follows a uniform distribution on interval $[0, 1]$, i.e. $F_{\mathbf{X}}(X_i = x_i) = 1$ for any $x_i \in [0, 1]$. Let us consider the joint distribution of (\mathbf{Z}, \mathbf{X}) subject to:

$$\forall i = (i_1, \dots, i_n) \in \{0, 1\}^n : \quad Z_i = \sum_{j=1}^n i_j X_j^1 . \quad (1)$$

Note that under these constraints, the value of \mathbf{z} is uniquely determined by the value of \mathbf{x} . The joint density function subject to these constraints is proportional to $F(\mathbf{x}, \mathbf{z}) \stackrel{def}{=} F_{\mathbf{X}}(\mathbf{x}) F_Z(\mathbf{z}) = \prod_{i \in \{0, 1\}^n} F_Z(z_i)$.

Properties of the basic instance. We will prove two properties of the basic instance. Lemma 3 and Corollary 1 state that in the n -dimensional hypercube defined by variables \mathbf{X} , the density of any *almost integral point* is relatively

¹ For the sake of exposition, we use equality constraints in our construction of the hard instance. Strictly speaking, this integral is always 0 in N -dimensional space since it is over a body of lower dimension. But we can easily fix this by replacing each equality constraint such as $Z_i = \sum_{j=1}^n i_j X_j$ by two inequality constraints $Z_i - \sum_{j=1}^n i_j X_j \geq \delta$ and $Z_i - \sum_{j=1}^n i_j X_j \leq \delta$ for some sufficiently small δ .

large. Here, almost integral points are points \mathbf{x} 's that satisfy $\{x_j\} \leq \frac{1}{n^2}$ for all $1 \leq j \leq n$. In contrast, Lemma 4 and Corollary 2 states that the density of the any *fractional point* is small. Fractional points are points \mathbf{x} 's such that there exists $1 \leq j \leq n$ such that $\{x_j\} \geq \frac{1}{4}$.

Lemma 3. *Suppose $\mathbf{x} \in [0, 1]^n$ satisfies that for any $1 \leq j \leq n$, $\{x_j\} \leq \frac{1}{n^2}$. Then, for each $i \in \{0, 1\}^n$, the corresponding $\{z_i\} \leq \frac{1}{10}$.*

Proof. Since $\{x_j\} \leq \frac{1}{n^2}$ for all $j \in [n]$, for any $i = (i_1, \dots, i_n) \in \{0, 1\}^n$, we have that $\{z_i\} = \left\{ \sum_{j=1}^n i_j x_j \right\} \leq \sum_{j=1}^n \{i_j x_j\} \leq \sum_{j=1}^n \frac{1}{n^2} = \frac{1}{n} < \frac{1}{10}$.

By Lemma 3 and the definition of F_Z , we have the the following corollary.

Corollary 1. *Suppose $\mathbf{x} \in [0, 1]^n$ satisfies that for any $1 \leq j \leq n$, $\{x_j\} \leq \frac{1}{n^2}$. Then $\forall i \in \{0, 1\}^n : F_Z(Z_i = z_i) = c\tau$. So $F(\mathbf{x}, \mathbf{z}) = (c\tau)^{2^n}$.*

Now we move to the second property.

Lemma 4. *Suppose $\mathbf{x} \in [0, 1]^n$ satisfies that there exists $1 \leq j \leq n$, $\{x_j\} \geq \frac{1}{4}$. Then, for at least half of $i \in \{0, 1\}^n$, the corresponding z_i satisfies $\{z_i\} \geq \frac{1}{8}$.*

Proof. For any $i_{-j} \in \{0, 1\}^{n-1}$, let $i = (i_{-j}, i_j = 0)$ and $i' = (i_{-j}, i_j = 1)$. By constraint (1), we have $z_{i'} - z_i = x_j$. So we get that $\{z_{i'}\} + \{z_i\} \geq \{x_j\} \geq \frac{1}{4}$. Hence, either $\{z_i\} \geq \frac{1}{8}$ or $\{z_{i'}\} \geq \frac{1}{8}$. Therefore, we can pair up the variables z_i 's such that in each pair at least one variable lies in the interval $[\frac{1}{8}, \frac{7}{8}]$. This finishes our proof.

By Lemma 4 and definition of the F_Z , we get the following.

Corollary 2. *Suppose $\mathbf{x} \in [0, 1]^n$ satisfies that there exists $1 \leq j \leq n$, $\{x_j\} \geq \frac{1}{4}$. Then, for at least half of $i \in \{0, 1\}^n$, we have $F_Z(Z_i = z_i) = c$ and hence $F(\mathbf{x}, \mathbf{z}) \leq c^{2^{n-1}} (c\tau)^{2^{n-1}}$.*

The random basic sub-instance. We want to embed the n -dimensional hypercube into an $N = 2^{n/g(n)}$ -dimensional hypergrid. So we cannot afford to use all of the Z_i 's in our hard instance construction. Instead, we will use N randomly and independently chosen Z_i 's and show that properties similar to Corollary 1 and Corollary 2 hold with high probability.

A random basic sub-instance is constructed as follows. We will consider N i.i.d. random variables $\hat{Z}_1, \dots, \hat{Z}_N$ each of which follows the density function F_Z . Moreover, for each $1 \leq k \leq N$, we randomly choose $i \in \{0, 1\}^n$ and impose a constraint between \hat{Z}_k and \mathbf{X} that is the same as the constraint between Z_i and \mathbf{X} (See (1)), i.e. $\sum_{j=1}^n i_j X_j = \hat{Z}_k$. The joint density function is proportional to $\hat{F}(\mathbf{x}, \hat{\mathbf{z}}) \stackrel{def}{=} F_{\mathbf{X}}(\mathbf{x}) F_Z(\hat{\mathbf{z}}) = \prod_{k=1}^N F_Z(\hat{z}_k)$.

Properties of the random basic sub-instance. We now show that with high probability a random basic sub-instance has some good properties: The almost integral points have large density; the fractional points have small density. The former can be formalized as the Lemma 5, which follows directly from Lemma 3.

Lemma 5. Suppose $\mathbf{x} \in [0, 1]^n$ satisfies that for any $1 \leq j \leq n$, $\{x_j\} \leq \frac{1}{n^2}$. Then, for any $1 \leq i \leq N$, the corresponding \hat{z}_i satisfies that $\{\hat{z}_i\} \leq 1/10$. Hence, for every $1 \leq i \leq N$, $F_{\mathbf{Z}}(\hat{Z}_i = \hat{z}_i) = c\tau$, and $\hat{F}(\mathbf{x}, \hat{\mathbf{z}}) = (c\tau)^N$.

The other property is less trivial. We first state it formally as follows.

Lemma 6. With high probability, the following holds: If $\mathbf{x} \in [0, 1]^n$ satisfies that there exists $1 \leq j \leq n$ such that $\{x_j\} \geq \frac{1}{4}$, then at least a third of the \hat{z}_k , $1 \leq k \leq N$, satisfy $\{\hat{z}_k\} \geq \frac{1}{10}$. Hence, $F_{\mathbf{Z}}(\hat{Z}_k = \hat{z}_k) = c$, $\hat{F}(\mathbf{x}, \hat{\mathbf{z}}) \leq c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}}$.

Proof. We divide the n -dimensional unit hypercube defined by \mathbf{x} into $M = (4n^2)^n$ small hypercubes with edge length $\frac{1}{4n^2}$. Let $\mathbf{x}^1, \dots, \mathbf{x}^M$ denote the centers of these small hypercubes. Further, we let $\hat{\mathbf{z}}^1, \dots, \hat{\mathbf{z}}^M$ be the corresponding $\hat{\mathbf{Z}}$ variables. We say a cube is fractional if its center is fractional.

We will first show the following claim.

Claim. With high probability, for every $1 \leq \ell \leq M$ such that \mathbf{x}^ℓ is fractional, at least a $\frac{1}{3}$ fraction of the \hat{z}_k^ℓ 's satisfies that $\{\hat{z}_k^\ell\} \geq \frac{1}{8}$.

Proof (Proof of Claim 3.2). Let Y_k^ℓ be the indicator of whether $\{\hat{z}_k^\ell\} \geq \frac{1}{8}$. Then, by Lemma 4, each \hat{z}_k^ℓ , $1 \leq k \leq N$, satisfies that $\{\hat{z}_k^\ell\} \geq \frac{1}{8}$ with probability at least half. So we have $\mathbf{E}[Y_k^\ell] \geq \frac{1}{2}$ and hence $\mathbf{E}[\sum_{k=1}^N Y_k^\ell] = \frac{N}{2}$. Furthermore, since Y_k^ℓ 's are binary variables, we get that $\sigma[Y_k^\ell] \leq 1$ and $\sigma[\sum_{k=1}^N Y_k^\ell] \leq \sqrt{N}$. By Chernoff-Hoeffding bound, for all ℓ such that \mathbf{x}^ℓ is fractional, the probability of $\sum_{k=1}^N Y_k^\ell < \frac{N}{3} \leq \mathbf{E}[\sum_{k=1}^N Y_k^\ell] - \frac{\sqrt{N}}{6} \sigma[\sum_{k=1}^N Y_k^\ell]$ is at most $2^{-(\sqrt{N}/6)^2/2} \ll \frac{1}{M}$. So by union bound, we get that with high probability, $\sum_{k=1}^N Y_k^\ell \geq \frac{N}{3}$ for $1 \leq \ell \leq M$.

Now we are ready to prove Lemma 6. We will show that if the property in Claim 3.2 holds, then the property in Lemma 6 holds.

Suppose we break ties on the boundary of small hypercubes in a manner that the boundary points are allocated to the most fractional hypercube adjacent to it. Then, we can easily verify the following fact holds.

Fact 2 *If a point is fractional, then it lies in a fractional small hypercube.*

For any fractional point \mathbf{x} , let $\hat{\mathbf{z}}$ be the corresponding $\hat{\mathbf{Z}}$ variables. Suppose it lies in a small cube centered at \mathbf{x}^ℓ . Then, by Fact 2 we know \mathbf{x}^ℓ is fractional. By our assumption that the property in Claim 3.2 holds, we get that at least a $\frac{1}{3}$ fraction of the corresponding \hat{z}_k^ℓ 's satisfies $\{\hat{z}_k^\ell\} \geq \frac{1}{8}$. Note that $\{\hat{z}_k^\ell - \hat{z}_k\} \leq \sum_{j=1}^n \{x_j^\ell - x_j\} \leq \sum_{j=1}^n \frac{1}{n^2} = \frac{1}{n} < \frac{1}{40}$. So, we get that $\{\hat{z}_k\} \geq \{\hat{z}_k^\ell\} - \{\hat{z}_k^\ell - \hat{z}_k\} \geq \frac{1}{8} - \frac{1}{40} = \frac{1}{10}$.

Therefore, with high probability, for every fractional point \mathbf{x} , at least a third of the corresponding \hat{z}_k 's satisfies $\{\hat{z}_k\} \geq \frac{1}{10}$. Hence, $F_{\mathbf{Z}}(\hat{Z}_k = \hat{z}_k) = c$, and $\hat{F}(\mathbf{x}, \hat{\mathbf{z}}) \leq c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}}$.

Construction of the hard instance. We will consider a random basic sub-instance such that Lemma 5 and Lemma 6 holds. By the relation between sampling and integration, we only need to show that estimating the integral of f subject to arbitrary convex constraint K is computationally hard, subject to our complexity assumption. We will prove this by encoding an arbitrary 3SAT instance by adding carefully chosen linear inequality constraints in the random basic sub-instance.

Consider any 3SAT instance with n variables. We will abuse notation by letting X_1, \dots, X_n denote not only the random variables in the random basic sub-instance but the n variables in the 3SAT instance as well. We will see that they are indeed closely related to each other in our construction. For each constraint $L_1 \vee L_2 \vee L_3$, where L_k 's are literals of the form X_j or $\neg X_j$, we add a linear inequality constraint $L_1 + L_2 + L_3 \geq \frac{3}{4}$. Here, for any literal L_k of the form X_j , we will replace L_k by X_j in this inequality constraint; and for any literal L_k of the form $\neg X_j$, we replace L_k by $(1 - X_j)$.

Proof outline. For this instance, we can prove that if we can estimate the integral, then we solve the 3SAT instance. It suffices to show that the values of the integral in these two cases are well separated. We will proceed in three steps: First, we show that the contribution of the fractional part in the integral is relatively small; next, we will show that any of the 2^n integral parts has contribution comparable to that of the fractional part; finally, we will prove that an integer part is not excluded by the above constraints if and only if it corresponds to a feasible assignment. Therefore, the 3SAT instance is satisfiable if and only if the integral is much larger than the contribution solely by the fractional part. This will finish the proof of Theorem 2.

Proof (Proof of Theorem 2). Now we are ready to proof Theorem 2. Let us proceed to the first step. We will let \mathcal{F} and \mathcal{I}_i , $1 \leq i \leq 2^n$, denote the set of fractional and integer parts, that is, $\mathcal{F} = \{\mathbf{x} \in [0, 1]^n : \exists j, \{x_j\} \geq \frac{1}{4}\}$, and $\mathcal{I}_i = \{\mathbf{x} \in [0, 1]^n : \forall j, \{x_j\} < \frac{1}{4}, [x_j] = i_j\}$ for every $i(i_1, \dots, i_n) \in \{0, 1\}^n$.

Lemma 7. $\int_{\mathbf{x} \in \mathcal{F}} F(\mathbf{x}) d\mathbf{x} \leq c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}}$.

Proof. By Lemma 6, we get that for any $\mathbf{x} \in \mathcal{F}$, $F(\mathbf{x}) \leq c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}}$. Hence, we have that $\int_{\mathbf{x} \in \mathcal{F}} F(\mathbf{x}) d\mathbf{x} \leq \int_{\mathbf{x} \in \mathcal{F}} c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}} d\mathbf{x} \leq \int_{\mathbf{x} \in [0, 1]^n} c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}} d\mathbf{x} = c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}}$.

Now we turn to the contribution of any integer part \mathcal{I}_i .

Lemma 8. For any $i \in \{0, 1\}^n$, we have $\int_{\mathbf{x} \in \mathcal{I}_i} F(\mathbf{x}) d\mathbf{x} \geq c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}}$.

Proof. We will let $\mathcal{I}_i^* \subseteq \mathcal{I}_i$ denote the almost integral part in \mathcal{I}_i , that is, $\mathcal{I}_i^* = \{\mathbf{x} \in [0, 1]^n : \forall j, \{x_j\} \leq \frac{1}{n^2}, [x_j] = i_j\}$. By Lemma 5, we have that for any $\mathbf{x} \in \mathcal{I}_i^*$, $F(\mathbf{x}) \geq (c\tau)^N = \tau^{\frac{N}{3}} c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}} = n^{2n} c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}}$. Thus, we have $\int_{\mathbf{x} \in \mathcal{I}_i} F(\mathbf{x}) d\mathbf{x} \geq \int_{\mathbf{x} \in \mathcal{I}_i^*} F(\mathbf{x}) d\mathbf{x} \geq \int_{\mathbf{x} \in \mathcal{I}_i^*} n^{2n} c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}} d\mathbf{x} = c^{\frac{N}{3}} (c\tau)^{\frac{2N}{3}}$.

Finally we will prove Lemma 9 which states that any integer point subject to the above constraints corresponds to a feasible assignment for the 3SAT instance.

Lemma 9. *For any $\mathbf{x} \in K$ such that $\{x_j\} < 1/4$ for all $1 \leq j \leq n$, we have that $[x_1], \dots, [x_n]$ is a feasible assignment for the 3SAT instance.*

Proof. For any constraint $L_1 \vee L_2 \vee L_3$, we have that $L_1 + L_2 + L_3 \geq 3/4$. So at least one of the L_k satisfies $L_k \geq 1/4$. But $\{x_j\} < 1/4$, so the value of $[x_j]$ will satisfy this constraint. Since this is true for any constraint, we get that $[x_1], \dots, [x_n]$ is a feasible assignment.

As we discuss earlier, Lemma 7, Lemma 8, and Lemma 9 prove Theorem 2.

4 Conclusions and Open Problems

While we have shown that discrete distributions are hard in general, for the discrete distributions that arise in the motivating applications in computational biology and image processing, it looks possible to fit a log-concave continuous distribution and use this to find the constrained distributions efficiently. On the other hand, the discrete distributions that arise in the puzzle-solving applications seem difficult to fit to nice continuous distributions.

We leave open the question of a deterministic construction showing the hardness of sampling from product distributions whose total deviation from log-concave is $\omega(\log n)$.

References

1. Applegate, D., Kannan, R.: Sampling and integration of near log-concave functions. In: Proceedings of the 23rd Annual ACM Symposium on Theory of Computing. pp. 156–163. ACM (1991)
2. Chandrasekaran, K., Dadush, D., Vempala, S.: Thin partitions: Isoperimetric inequalities and a sampling algorithm for star shaped bodies. In: Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms. pp. 1630–1645. Society for Industrial and Applied Mathematics (2010)
3. Chandrasekaran, K., Deshpande, A., Vempala, S.: Sampling s-concave functions: The limit of convexity based isoperimetry. Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques pp. 420–433 (2009)
4. Dyer, M., Frieze, A., Kannan, R.: A random polynomial-time algorithm for approximating the volume of convex bodies. *Journal of the ACM* 38(1), 1–17 (1991)
5. Frieze, A., Kannan, R., Polson, N.: Sampling from log-concave distributions. *The Annals of Applied Probability* 4(3), 812–837 (1994)
6. Kannan, R.: Personal communication (2011)
7. Koutis, I.: On the complexity of approximate multivariate integration. Tech. rep., Carnegie Mellon University (2005)
8. Lovasz, L., Vempala, S.: Fast algorithms for log-concave functions: Sampling, rounding, integration and optimization. In: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science. pp. 57–68. IEEE (2006)
9. Lovasz, L., Vempala, S.: Hit-and-run from a corner. *SIAM Journal on Computing* 35(4), 985–1005 (2006)
10. Lovasz, L., Vempala, S.: The geometry of log-concave functions and sampling algorithms. *Random Structures and Algorithms* 30(3), 307–358 (2007)