A Sampling Method Based on LDPC Codes

Xuhong Zhang
EECS, MIT
Cambridge, MA 02139
xuhong@mit.edu

Gregory W. Wornell
EECS, MIT
Cambridge, MA 02139
gww@mit.edu

Abstract

Sampling is a powerful and popular way to tackle problems for which exact inference is computationally intractable. We propose a novel sampling scheme based on Low Density Parity Check (LDPC) codes which works for any distribution over a finite alphabet. Compared to existing methods such as Markov Chain Monte Carlo (MCMC) Methods, the proposed scheme offers a different trade-off among various features desirable of a sampling method. In particular, in addition to good scaling behaviour with dimensions, the proposed scheme has good control over independence properties and makes efficient utilization of randomness.

1 Introduction

Exact solutions of inference problems are often computationally infeasible. Monte Carlo methods, based on the idea of approximating distributions by samples, have been the workhorse for many statistical computations. Ideally, one would like independent and correct samples for any given distribution at feasible costs. Yet such a sampling scheme does not exist due to the intrinsic hardness of general inference problems. Many existing sampling techniques are widely used, each with its own strengths and suitable applications. Yet they are not without problems: for example, rejection sampling and importance sampling have poor scaling behaviour; for MCMC methods, one often has to rely solely on heuristics to determine mixing time and convergence.

This paper introduces a new sampling scheme based on LDPC codes and it provides a different trade-off among desirable features. The scheme works for target distributions with finite alphabets, and it has tractable running time, good scaling behaviour, and makes efficient use of randomness. In particular, since the proposed scheme can be viewed as an approximation on a relatively global level, it appears less susceptible to local minima and produces samples with better independence properties compared to those generated by MCMC methods.

LDPC codes, first introduced by Gallager [1] in the 60s and later revived in the mid 90s [3][4], have found extensive use in channel coding, source coding and various other problems[1]. Recently [13] proposed a compression architecture with 'model-free' encoders, where an LDPC code serves as a hash function and resulting parity values as the compressed sequence. The decoder looks for a sequence that satisfies all parity checks and is typical to the source. The success of this compression scheme suggests a way to map parity bits' values, which form an index, to typical sequences. Inspired by [13], we consider exploiting this mapping for sampling purposes by reversing the above process: randomly generate an index (i.e. parity bits), and output the corresponding typical sequence as a sample. It turns out that with adequately constructed system and appropriately chosen LDPC code, we can obtain good samples.

[1] For an introduction to LDPC codes, please refer to [8].

2 LDPC-Based Sampling Scheme

First notice that any variable over finite alphabet can be represented as a sequence of binary random variables. Thus without loss of generality, we will focus on sequences of binary variables. In particular, let $p_{\mathbf{x}}(\cdot)$ be the target distribution, where $\mathbf{x} = (x_1, x_2, ..., x_N) \in \{0,1\}^N$. The proposed scheme is defined on a *combined graph*, whose construction is described as follows:

a Given $p_{\mathbf{x}}(\cdot)$, denote its factor graph as $\mathcal{S}$. The *source graph* contains t independent, identical copies of $\mathcal{S}$. Section 3 will discuss how large t should be as a function of $p_{\mathbf{x}}(\cdot)$.

b Given the source graph, attach to it an LDPC code with Nt variable nodes and K check nodes. The factor graph of the chosen LDPC code is referred to as *code graph*.

c Source graph and code graph share the variable nodes, and together they are referred to as the *combined graph*. The figure below shows an example of a combined graph.

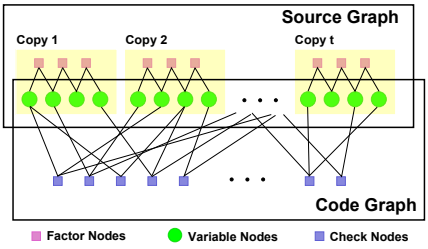

Figure 1: The schematic of an example combined graph.

Now we are ready to present the idealized LDPC-based sampling scheme:

1. Given $p_{\mathbf{x}}(\cdot)$, construct the corresponding combined graph.
2. Flip K independent fair coins and use the results as the values of parity bits.
3. Find an LDPC codeword $(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(t)})$ that maximizes $\prod_{i=1}^{t} p_{\mathbf{x}}(\mathbf{x}^{(i)})$, i.e. the most likely sequence among the ones correspond to the selected parity bits.

3 Correctness under Maximum-Likelihood Simulation

Step 3 of the proposed scheme is generally infeasible and in practice, the sum-product algorithm is implemented instead as an approximation. This section analyses the proposed scheme based on exact *Maximum-likelihood (ML) simulation*, where we assume the most likely LDPC codeword can be found. It is useful to understand this ideal case, as it provides an upper bound for the practical performance. We will start with some intuitive arguments and then present the main theorems.

Let $\mathbf{y} = (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(t)})$, where $\mathbf{x}^{(i)}$'s are drawn independently from the target distribution $p_{\mathbf{x}}(\cdot)$. We can define the set of typical sequences (i.e. typical set):

$$\mathcal{T}_{\epsilon}^{(t)} \triangleq \left\{ \mathbf{y} \in \mathcal{X}^t = \{0,1\}^{Nt} : \left| \frac{1}{t} \log_2 p_{\mathbf{y}}(\mathbf{y}) + NH_0 \right| = \left| \frac{1}{t} \sum_{i=1}^{t} \log_2 p_{\mathbf{x}}(\mathbf{x}^{(i)}) + NH_0 \right| \leq \epsilon \right\},$$

where $NH_0 = -\sum_{\mathbf{x}} p_{\mathbf{x}}(\mathbf{x}) \log_2 p_{\mathbf{x}}(\mathbf{x})$ is the entropy of $p_{\mathbf{x}}(\cdot)$. The *Asymptotic Equipartition Property* (AEP) says that for any $\delta > 0$, there exists large enough t, s.t. $\mathbb{P}(\mathcal{T}_{\epsilon}^{(t)}) \geq 1 - \delta$. Moreover, by definition, all typical sequences are roughly equi-probable with probability $\sim 2^{-tNH_0}$. As direct consequences of AEP, there are roughly 2^{tNH_0} typical sequences and sampling from $p_{\mathbf{y}}(\cdot)$ can be well approximated by sampling from the uniform distribution over the typical set. Also notice that a sample from $p_{\mathbf{y}}(\cdot)$ can be viewed as the concatenation of t samples from $p_{\mathbf{x}}(\cdot)$.

We employ an LDPC code as a hash function and partition all sequences into bins: each bin corresponds to a set of parity bits' values. It turns out that if the code rate is matched to the entropy

rate of $p_{\mathbf{x}}(\cdot)$, i.e. if we have about 2^{tNH_0} bins, a randomly chosen LDPC code is with high probability a very good hash function: *almost all bins contain exactly one typical sequence and no sequence more likely than a typical one*. Since what ML simulation does is essentially uniformly randomly pick a bin each time and output the most likely sequence in that bin, it nicely mimics the process of 'sampling from the uniform distribution over typical sequences'. To be more precise:

Theorem 1 *Given $p_{\mathbf{x}}(\cdot)$ over $\{0,1\}^N$, $\epsilon > 0, \delta > 0$, let $p_{\mathbf{y}}(\mathbf{y} = (\boldsymbol{x}^{(1)}, ..., \boldsymbol{x}^{(t)})) = \prod_{i=1}^{t} p_{\mathbf{x}}(\boldsymbol{x}^{(i)})$ and let t be large enough s.t. $\mathbb{P}(\mathcal{T}_\epsilon^{(t)}) \geq 1 - \delta/2$. Then there exists an LDPC code with parity check matrix $\mathbb{H} \in \{0,1\}^{K \times Nt}$ where $K \sim NtH_0$, for which ML simulation of the proposed sampling scheme produces a sequence in $\mathcal{T}_\epsilon^{(t)}$ with probability at least $1 - \delta$.*

In Theorem 1, if the t required for $\mathbb{P}(\mathcal{T}_\epsilon^{(t)}) \geq 1-\delta/2$ to hold is too large, e.g. if t grows exponentially with N, the proposed sampling scheme will be of little use. Theorem 2 characterizes the size of t as a function of $p_{\mathbf{x}}(\cdot)$, ϵ, and δ. In particular, the bound does not depend on N as long as the entropy per variable of the target distribution is not too small (i.e. H_0 is large compared to $1/N$).

Theorem 2 *If $H_0 < 1$, a sufficient condition for equation (1) is*

$$t > \frac{1}{H_0 + \epsilon - \log_2(1 + H_0 \ln 2 + o(\frac{1}{N}))} \log_2 \frac{4}{\delta} \sim \frac{2}{2\epsilon + H_0^2 \ln 2} \log_2 \frac{4}{\delta}$$

4 Approximate Simulation and Experiment Results

In practice, standard sum-product algorithm is used to approximately find the typical sequence that corresponds to the selected parity bits. A few remarks regarding important implementation details:

a The LDPC code is randomly generated with constant variable node degree 3 and roughly uniform check node degree. In addition, we make sure there is no size-4 cycles.

b Messages are uniformly initialized. If necessary, introduce a small fraction of degree-1 checks to avoid convergence to the trivial fixed point where all messages are $[0.5, 0.5]^T$.

c Given $p_{\mathbf{x}}(\cdot)$, H_0 is often unknown. A binary search is performed to find the 'correct' rate of LDPC code: the highest rate at which sum-product can converge works well empirically.

The proposed sampling scheme is tested on a few distributions, and the results are shown below.

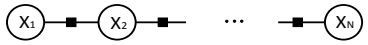

Figure 2: Schematic of Markov Chain

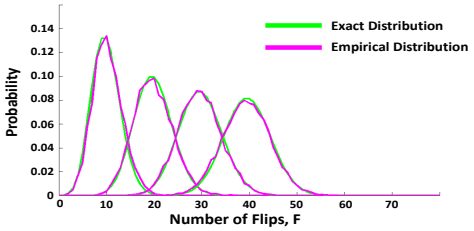

Figure 3: True and Empirical Distributions of F

Homogeneous Markov Chain Model A Markov Chain model is shown in Figure 2, and the target distribution is assumed to be homogeneous and symmetric. A pair of neighbouring variables that take different values is referred to as a 'flip' and p denotes the flipping probability. We evaluate the quality of the samples by considering the distribution of number of flips F. For Markov Chains, exact sampling is possible and the distribution of F can also be computed analytically. Figure 3 compares the true distribution (shown in green) with the empirical distribution estimated using 10000 samples from the proposed scheme (purple). The four sets of results correspond to $p =$ 0.1/0.2/0.3/0.4 respectively. As is evident from Figure 3, The empirical distribution produced by the proposed scheme follows the true distribution nicely.

Homogeneous 2D Ising Model A 2D Ising model is shown in Figure 4, and the target distribution is again assumed to be homogeneous and symmetric. Unlike Markov Chains, A 2D Ising model cannot be sampled from exactly, but some important quantities can be computed analytically. In particular, Figure 5 shows how internal energy per variable varies with temperature[2], both from analytical computation (blue line) and empirical estimation using samples generated by the proposed method (red squares). As can be seen, the estimated values agree very nicely with the true values.

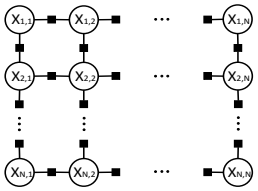

Figure 4: Schematic of 2D Ising Model

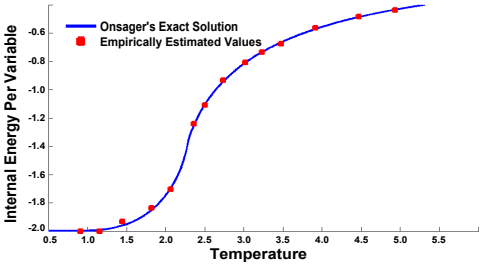

Figure 5: True and Estimated Internal Energy

5 Discussion

Choosing the 'right' sampling method is application-dependent: different methods provide different trade-offs. Table 1 compares the proposed sampling scheme with existing ones. Some particularly desirable features of the proposed scheme are highlighted below:

a Sum-product algorithm is the main tool used in the proposed scheme. Due to the distributed nature of message passing algorithms, the proposed scheme scales well with dimensions and thus can deal with high-dimensional distributions.

b Although not always important, the proposed scheme is economical in its use of 'randomness'. Randomness is only needed for generation of the parity bit values[3]. Since the LDPC code rate is roughly the entropy rate of the distribution, randomness per sample needed is close to the theoretical minimum. In contrast, MCMC methods require new randomness each iteration and the utilization of randomness is far from efficient.

c MCMC methods produce correlated samples and usually depend on heuristics to determine when convergence is reached and whether the space is well explored. The Markov Chain can be trapped in certain states when, for example, several groups of high probability state(s) are connected by unlikely state(s)[4]. In the proposed sampling scheme, however, samples are independent under ML-simulation. Moreover, since any likely state will appear in some typical sequences and all typical sequences are equally likely to be chosen, the LDPC-based sampling method seems less susceptible to the local minima problem.

	Rejection Sampling	Importance Sampling	MCMC Methods	Proposed Scheme
Applicable to Continuous Variables?	Yes	Yes	Yes	No
Require Good Proposal Distribution?	Yes	Yes	No	No
Applicable to High-Dim Data?	No	No	Yes	Yes
Guaranteed Correctness?	Yes	No	Asymptotic	Approximate
Independent Samples?	Yes	N/A	No	Approximate

Table 1: Comparing Different Sampling Methods

Acknowledgement

This work was supported, in part, by AFOSR under Grant No. FA9550-11-1-0183, and by NSF under Grant No. CCF-1319828.

[2]Temperature is uniquely determined by the choice of flipping probability p.

[3]In the experiment, randomness is also used to generate the LDPC code. But since that is only generated once, it can be viewed as an overhead. Moreover, one could in principle design good codes deterministically.

[4]In practice, various tricks are used to alleviate the problem. But such tricks are often application dependent.

References

[1] Gallager, R.G. *Low Density Parity Check Codes*, no. 21 in Research Monograph Series. Cambridge, MA: MIT Press, 1963.

[2] Berlekamp, E.R., McElieve, R.J. & van Tilborg, H.C.A. "On the Intractability of Certain Coding Problems". *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 384-386, 1978.

[3] Mackay, D.J.C. & Neal, R.M. "Near Shannon Limit Performance of Low Density Parity Check Codes". *Electron. Lett.*, vol. 32, no. 18, pp. 1645-1646, 1996.

[4] Sipser, M. & Spielman, D.A. "Expander Codes". *IEEE Trans. Inform. Theory*, vol. 42, pp. 1710-1722, 1996.

[5] Mackay, D.J.C. "Good Error-Correcting Codes Based on Very Sparse Matrices". *IEEE Trans. Inform. Theory*, vol. 45, pp. 399-431, 1999.

[6] Matsunaga, Y. & Yamamoto, H. "A Coding Theorem for Lossy Data Compression by LDPC Codes". *IEEE Trans. Inform. Theory*, vol. 49, no. 9, pp. 2225-2229, 2003.

[7] Caire, G., Shamai, S. & Verdú, S. "Noiseless Data Compression with Low-Density Parity-Check Codes". *Proc. of the IEEE Int. Symposium on Inform. Theory*, pp. 22, 2003.

[8] Ryan, W.E. "An Introduction to LDPC Codes", *CRC Handbook for Coding and Signal Processing for Recording Systems*, B. Vasic, Ed. Boca Raton, FL: CRC, 2004

[9] Miyake, S. & Muramatsu, J. "Construction of A Lossy Code Using LDPC Matrices". *Proc. of the IEEE Int. Symposium on Inform. Theory*, pp. 813-816, 2006.

[10] Costello, D.J., Jr. & Forney, G.D., Jr. "Channel Coding: The Road to Channel Capacity". *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1150-1177, 2007.

[11] Mézard, M. & Montanari, A. *Information, Physics and Computation*, Oxford University Press, 2008.

[12] Huang, Y.-z. & Wornell, G.W. "A Class of Compression Systems with Model-Free Encoding". *Proc. ITA*, 2014.

[13] Huang, Y.-z "Model-Code Separation Architectures for Compression Based on Message-Passing". PhD Thesis, Massachusetts Institute of Technology, 2015.