Supplementary Material:
A Hybrid Approach for Probabilistic Inference using Random Projections

Appendix

Proposition. *For any graphical model G and choice of parameters $M \geq 1$, $m \geq 1$, $f \in (0, \frac{1}{2})$, $p \in [0,1]$, $\frac{1}{M}\sum_{k=1}^{M}\left(\frac{2}{p+1}\right)^m Z^{(k)}$ is an unbiased estimator of the partition function Z. Furthermore, the variance of this estimator is bounded from above by the following expressions:*

$$
\begin{aligned}
Var\left[\frac{1}{M}\sum_{k=1}^{M}\left(\frac{2}{p+1}\right)^m Z^{(k)}\right] &\leq \frac{1}{M}\left(\frac{4(p+\frac{1}{2}(1-f)(1-p)^2)}{(p+1)^2}\right)^m Z^2 + \frac{1}{M}\left(\frac{2(p^2+1)}{(p+1)^2}\right)^m \sum_{x\in\mathcal{X}} w(x)^2 - \frac{Z^2}{M} \\
&\leq \frac{2}{M}\left(\frac{2(p^2+1)}{(p+1)^2}\right)^m Z^2
\end{aligned}
$$

Preliminaries

Before we prove the proposition, let us start off by proving some preliminaries. The key idea here is to relate the expected value of certain quantities involving the smoothed ($p > 0$) XOR factor potentials $\phi_{\{A_i^{(k)}, b_i^{(k)}\}}(x_{\{A_i^{(k)}\}})$ to the hard ($p = 0$) XOR factor potentials $\psi_{\{A_i^{(k)}, b_i^{(k)}\}}(x_{\{A_i^{(k)}\}})$.

Define the random variable $y_k(x) = \prod_{i=1}^m \phi_{\{A_i^{(k)}, b_i^{(k)}\}}(x_{\{A_i^{(k)}\}}) = \prod_{i=1}^m \phi_i(x)$ to be the product over all of the smoothed ($p > 0$) XOR factor potentials, where we use $\phi_i(x)$ as short-hand notation. Let $\psi_i(x)$ be the hard XOR factor potential corresponding to $\phi_i(x)$ for $p = 0$. By construction, we have that $\phi_i(x) = p + (1-p)\psi_i(x)$.

Recall that we proved $\mathbb{E}[y_k(x)] = \left(\frac{p+1}{2}\right)^m$ in Section 3.1 as follows: $\mathbb{E}[y_k(x)] = \mathbb{E}\big[\prod_{i=1}^m \phi_i(x)\big] = \prod_{i=1}^m \mathbb{E}\big[p + (1-p)\psi_i(x)\big] = \prod_{i=1}^m \big(p + (1-p)\frac{1}{2}\big) = \left(\frac{p+1}{2}\right)^m$.

Similarly, we can prove that $\mathbb{E}[y_k(x)^2] = \left(\frac{p^2+1}{2}\right)^m$ as follows: $\mathbb{E}[y_k(x)^2] = \mathbb{E}\big[\prod_{i=1}^m \phi_i(x)^2\big] = \prod_{i=1}^m \mathbb{E}[\phi_i(x)^2] = \prod_{i=1}^m \mathbb{E}\big[\big(p + (1-p)\psi_i(x)\big)^2\big] = \prod_{i=1}^m \big(p^2 + 2p(1-p)\frac{1}{2} + (1-p)^2\frac{1}{2}\big) = \left(\frac{p^2+1}{2}\right)^m$.

Given two elements $a, b \in \mathcal{X}$ with $a \neq b$, we can derive an expression for $\mathbb{E}[y_k(a)y_k(b)]$ as follows:

$$
\begin{aligned}
\mathbb{E}[y_k(a)y_k(b)] &= \mathbb{E}\left[\prod_{i=1}^m \phi_i(a)\phi_i(b)\right] \\
&= \prod_{i=1}^m \mathbb{E}[\phi_i(a)\phi_i(b)] \\
&= \prod_{i=1}^m \mathbb{E}\Big[\big(p + (1-p)\psi_i(a)\big)\big(p + (1-p)\psi_i(b)\big)\Big] \\
&= \prod_{i=1}^m \Big(p + (1-p)^2\mathbb{E}\big[\psi_i(a)\psi_i(b)\big]\Big) \\
&= \sum_{t=0}^m \binom{m}{t} p^{m-t}(1-p)^{2t}\mathbb{E}\Bigg[\prod_{\text{any } t \text{ distinct indices}} \psi_{i_1}(a)\psi_{i_1}(b)\ldots\psi_{i_t}(a)\psi_{i_t}(b)\Bigg]
\end{aligned}
$$

where we used the fact that the expected value of the product over any t distinct indices $\{i_1, ..., i_t\}$ is equal by symmetry, which allows us to simplify the last expression.

Proof

Proof. For convenience, define a scaling factor $r = \frac{2}{p+1}$.

The partition function of our randomly projected $G^{(k)}$ is $Z^{(k)} = \sum_{x\in\mathcal{X}} w(x) \prod_{i=1}^{m} \phi_{\{A_i^{(k)}, b_i^{(k)}\}}(x_{\{A_i^{(k)}\}}) = \sum_{x\in\mathcal{X}} w(x) y_k(x)$ with $\mathbb{E}[Z^{(k)}] = \sum_{x\in\mathcal{X}} w(x)\mathbb{E}[y_k(x)] = \left(\frac{p+1}{2}\right)^m Z$.

Using variance and expectation formulas, we have the following equalities:

$$
\begin{aligned}
Var\left[\frac{1}{M}\sum_{k=1}^{M} r^m Z^{(k)}\right] &= \frac{r^{2m}}{M^2}\sum_{k=1}^{M} Var[Z^{(k)}] \\
&= \frac{r^{2m}}{M^2}\sum_{k=1}^{M} Var\left[\sum_{x\in\mathcal{X}} w(x) y_k(x)\right] \\
&= \frac{r^{2m}}{M^2}\sum_{k=1}^{M}\left(\sum_{x\in\mathcal{X}} w(x)^2 Var[y_k(x)] + \sum_{x_i\in\mathcal{X}}\sum_{x_j\neq x_i} w(x_i)w(x_j) Cov\big[y_k(x_i), y_k(x_j)\big]\right) \\
&= \frac{r^{2m}}{M^2}\sum_{k=1}^{M}\Bigg(\sum_{x\in\mathcal{X}} w(x)^2\big(\mathbb{E}[y_k(x)^2] - \mathbb{E}[y_k(x)]^2\big) \\
&\quad + \sum_{x_i\in\mathcal{X}}\sum_{x_j\neq x_i} w(x_i)w(x_j)\big(\mathbb{E}\big[y_k(x_i)y_k(x_j)\big] - \mathbb{E}\big[y_k(x_i)\big]\mathbb{E}\big[y_k(x_j)\big]\big)\Bigg) \\
&= \frac{r^{2m}}{M^2}\sum_{k=1}^{M}\left(\left(\frac{p^2+1}{2}\right)^m \sum_{x\in\mathcal{X}} w(x)^2 + \sum_{x_i\in\mathcal{X}}\sum_{x_j\neq x_i} w(x_i)w(x_j)\mathbb{E}\big[(y_k(x_i)y_k(x_j)\big]\right) - \frac{Z^2}{M} \\
&= \frac{r^{2m}}{M^2}\sum_{k=1}^{M}\sum_{x_i\in\mathcal{X}}\sum_{x_j\neq x_i} w(x_i)w(x_j)\mathbb{E}\big[(y_k(x_i)y_k(x_j)\big] + \frac{r^{2m}}{M}\left(\frac{p^2+1}{2}\right)^m \sum_{x\in\mathcal{X}} w(x)^2 - \frac{Z^2}{M}
\end{aligned}
$$

Let us focus on bounding the first term in the variance expression above, V_1. Using the lemma proved earlier with $a = x_i$, $b = x_j$, and $x_j \neq x_i$, we have that

$$
\begin{aligned}
V_1 &= \frac{r^{2m}}{M^2}\sum_{k=1}^{M}\sum_{x_i\in\mathcal{X}}\sum_{x_j\neq x_i} w(x_i)w(x_j)\mathbb{E}\big[(y_k(x_i)y_k(x_j)\big] \\
&= \frac{r^{2m}}{M}\sum_{x_i\in\mathcal{X}}\sum_{x_j\neq x_i} w(x_i)w(x_j)\left(\sum_{t=0}^{m}\binom{m}{t} p^{m-t}(1-p)^{2t}\mathbb{E}\left[\prod_{\text{any } t \text{ distinct indices}} \psi_{i_1}(x_i)\psi_{i_1}(x_j)...\psi_{i_t}(x_i)\psi_{i_t}(x_j)\right]\right) \\
&= \frac{r^{2m}}{M}\sum_{t=0}^{m}\binom{m}{t} p^{m-t}(1-p)^{2t}\left(\sum_{x_i\in\mathcal{X}}\sum_{x_j\neq x_i} w(x_i)w(x_j)\mathbb{E}\left[\prod_{\text{any } t \text{ distinct indices}} \psi_{i_1}(x_i)\psi_{i_1}(x_j)...\psi_{i_t}(x_i)\psi_{i_t}(x_j)\right]\right)
\end{aligned}
$$

Hashing with Low-Density (Sparse) Parity Constraints

Recall that in the construction of our hash functions in Section 3.1, we let $A_{i,j} = 1$ if node j is included in the i^{th} XOR constraint and $A_{i,j} = 0$ otherwise, and we let b_i be the parity bit generated for the i^{th} XOR constraint. The process for randomly generating A, b ($A_{i,j}^{(k)} \overset{iid}{\sim}$ Bernoulli(f)and $b_i^{(k)} \overset{iid}{\sim}$ Bernoulli($\frac{1}{2}$)) is designed to give us sparse parity constraints. Given the set of m XOR constraints, a configuration $x \in \mathcal{X}$ satisfies the set of XOR constraints if and only if $Ax \equiv b$ (mod 2).

Adapting the proof of Theorem 3 in *Low-density Parity Constraints for Hashing-Based Discrete Integration* (Ermon et al., 2014) and letting $A_{\{i_1,...,i_t\}}$ and $b_{\{i_1,...,i_t\}}$ be the t rows of A and b, respectively, corresponding to the t parity constraints inside the expectation, the expression we have can be written as:

$$
\begin{aligned}
& \sum_{x_i \in \mathcal{X}} \sum_{x_j \neq x_i} w(x_i) w(x_j) \mathbb{E}\Big[\prod_{\text{any } t \text{ distinct indices}} \psi_{i_1}(x_i)\psi_{i_1}(x_j)...\psi_{i_t}(x_i)\psi_{i_t}(x_j) \Big] \\
= & \sum_{x_i \in \mathcal{X}} w(x_i) \sum_{x_j \neq x_i} w(x_j) \mathbb{P}[A_{\{i_1,...,i_t\}} x_i + b_{\{i_1,...,i_t\}} = \vec{0}_t, A_{\{i_1,...,i_t\}} x_j + b_{\{i_1,...,i_t\}} = \vec{0}_t] \\
= & \ 2^{-t} \sum_{x_i \in \mathcal{X}} w(x_i) \sum_{s=1}^{n} W_{x_i}(s) r^{(s,f)}(0,0) \\
= & \ 2^{-2t} \sum_{x_i \in \mathcal{X}} w(x_i) \sum_{s=1}^{n} W_{x_i}(s) \Big(1 + (1-2f)^s\Big)^t \\
\leq & \ 2^{-2t}(2-2f)^t \sum_{x_i \in \mathcal{X}} w(x_i) \sum_{s=1}^{n} W_{x_i}(s) \\
\leq & \ 2^{-t}(1-f)^t Z^2
\end{aligned}
$$

where $W_{x_i}(s)$ is the sum of $w(x_j)$ for all x_j at Hamming distance s from x_i and we used the fact that $r^{(s,f)}(0,0) \leq r^{(1,f)}(0,0)$.

Proof (continued)

The bound for our first term is

$$
\begin{aligned}
V_1 &= \frac{r^{2m}}{M} \sum_{t=0}^{m} \binom{m}{t} p^{m-t} (1-p)^{2t} \Bigg(\sum_{x_i \in \mathcal{X}} \sum_{x_j \neq x_i} w(x_i) w(x_j) \mathbb{E}\Big[\prod_{\text{any } t \text{ distinct indices}} \psi_{i_1}(x_i)\psi_{i_1}(x_j)...\psi_{i_t}(x_i)\psi_{i_t}(x_j) \Big] \Bigg) \\
&\leq \frac{r^{2m}}{M} \sum_{t=0}^{m} \binom{m}{t} p^{m-t} (1-p)^{2t} \Big(2^{-t}(1-f)^t Z^2 \Big) \\
&= Z^2 \frac{r^{2m}}{M} \sum_{t=0}^{m} \binom{m}{t} p^{m-t} \left(\frac{1}{2}(1-f)(1-p)^2 \right)^t \\
&= Z^2 \frac{r^{2m}}{M} \left(p + \frac{1}{2}(1-f)(1-p)^2 \right)^m \\
&\leq Z^2 \frac{r^{2m}}{M} \left(p + \frac{1}{2}(1-p)^2 \right)^m \\
&= Z^2 \frac{r^{2m}}{M} \left(\frac{p^2+1}{2} \right)^m
\end{aligned}
$$

The final bound for our variance becomes

$$
\begin{aligned}
Var\left[\frac{1}{M} \sum_{k=1}^{M} r^m Z^{(k)} \right] &\leq Z^2 \frac{r^{2m}}{M} \left(\frac{p^2+1}{2} \right)^m + \frac{r^{2m}}{M} \left(\frac{p^2+1}{2} \right)^m \sum_{x \in \mathcal{X}} w(x)^2 - \frac{Z^2}{M} \\
&\leq Z^2 \frac{2}{M} r^{2m} \left(\frac{p^2+1}{2} \right)^m \\
&= Z^2 \frac{2}{M} \left(\frac{2(p^2+1)}{(p+1)^2} \right)^m
\end{aligned}
$$

□

References

Ermon, Stefano, Gomes, Carla P., Sabharwal, Ashish, and Selman, Bart. Low-density parity constraints for hashing-based discrete integration. In *Proc. of the 31st International Conference on Machine Learning (ICML)*, pp. 271–279, 2014.