# A Appendix : Proofs

*Proof of Theorem 1.* Let $q \in \mathcal{Q}$. For ease of notation we define $L(q, A^{i,t}, b^{i,t})$ as $\sum_{x \in X_{A^{i,t}b^{i,t}}} q(x)\theta'\phi(x) - q(x)\log q(x)$; this is the variational lower bound using $q$ as the approximating distribution. From the non-negativity of the KL divergence we have that

$$\log Z(A^{i,t}, b^{i,t}) \geq L(q, A^{i,t}, b^{i,t})$$

and, since $q$ was arbitrary,

$$\log Z(A^{i,t}, b^{i,t}) \geq \max_{q \in \mathcal{Q}} L(q, A^{i,t}, b^{i,t})$$

Thus

$$Z(A^{i,t}, b^{i,t}) \geq \exp\left(\max_{q \in \mathcal{Q}} \left\{L(q, A^{i,t}, b^{i,t})\right\}\right) \triangleq \gamma^{i,t} \quad (12)$$

We take the expectation of both sides to yield

$$\begin{aligned}\frac{Z}{2^i} &= \mathbb{E}\left[Z(A^{i,t}, b^{i,t})\right] \\ &\geq \mathbb{E}\left[\exp\left(\max_{q \in \mathcal{Q}}\left\{L(q, A^{i,t}, b^{i,t})\right\}\right)\right] = \mathbb{E}[\gamma^{i,t}]\end{aligned}$$

which proves the first part of the theorem.

We prove the second part by using the fact that our approximating family $\mathcal{Q}$ contains the degenerate family $\mathcal{D}$, for which Theorem 2 (below) gives tight approximation bounds.

Since the conditions of Theorem 2 are satisfied, we know that equation (10) holds with probability at least $1-\delta$. From (10), since the terms in the sum are non-negative, we have that the maximum element is at least $1/(n+1)$ of the sum:

$$\begin{aligned}\max_i \exp\Bigg(&\underset{t \in [T]}{\text{Median}} - \min_{q \in \mathcal{D}} D_{KL}(q||R^i_{A^{i,t},b^{i,t}}(p)) \\ &+ \log Z(A^{i,t}, b^{i,t})\Bigg)2^{i-1} \\ &\geq \frac{1}{32} Z \frac{1}{n+1}\end{aligned}$$

Therefore there exists $m$ such that

$$\begin{aligned}&\underset{t \in [T]}{\text{Median}}\left(-\min_{q \in \mathcal{D}} D_{KL}(q||R^m_{A^{m,t},b^{m,t}}(p)) + \log Z(A^{m,t}, b^{m,t})\right) \\ &\quad + (m-1)\log 2 \\ &\quad \geq -\log 32 + \log Z - \log(n+1)\end{aligned}$$

We also have

$$\min_{q \in \mathcal{Q}} D_{KL}(q||R_{A,b}(p)) \leq \min_{q \in \mathcal{D}} D_{KL}(q||R_{A,b}(p))$$

because $\mathcal{D} \subseteq \mathcal{Q}$. Thus

$$\begin{aligned}&\underset{t \in [T]}{\text{Median}}\left(-\min_{q \in \mathcal{Q}} D_{KL}(q||R^m_{A^{m,t},b^{m,t}}(p)) + \log Z(A^{m,t}, b^{m,t})\right) \\ &\quad + (m-1)\log 2 \\ &\quad \geq -\log 32 + \log Z - \log(n+1)\end{aligned}$$

From the definition of KL divergence, we have

$$D_{KL}(q||R^m_{A,b}(p)) = -L(q, A, b) + \log Z(A, b)$$

Plugging in we get

$$\begin{aligned}&\underset{t \in [T]}{\text{Median}}\left(\log \gamma^{m,t}\right) + (m-1)\log 2 \\ &\geq -\log 32 - \log(n+1) + \log Z\end{aligned}$$

and also

$$\underset{t \in [T]}{\text{Median}}\left(\log \gamma^{m,t}\right) + m \log 2 \geq -\log 32 - \log(n+1) + \log Z$$

with probability at least $1-\delta$. Exponentiating both sides,

$$\text{Median}\left(\gamma^{m,1}, \cdots, \gamma^{m,T}\right) 2^m \geq \frac{Z}{32(n+1)}$$

and since the terms in the Median are nonzero,

$$\frac{1}{T}\sum_{t=1}^{T} \gamma^{m,t} \geq \frac{1}{2}\text{Median}\left(\gamma^{m,1}, \cdots, \gamma^{m,T}\right)$$

therefore with probability at least $1-\delta$

$$\frac{1}{T}\sum_{t=1}^{T} \gamma^{m,t} 2^m \geq \frac{Z}{64(n+1)}$$

which proves the lower bound.

From Markov's inequality we have

$$\mathbb{P}\left[Z(A^{i,t}, b^{i,t}) \geq c\mathbb{E}[Z(A^{i,t}, b^{i,t})]\right] \leq \frac{1}{c}$$

$$\mathbb{P}\left[Z(A^{i,t}, b^{i,t})2^i \geq cZ\right] \leq \frac{1}{c}$$

Since $Z(A^{i,t}, b^{i,t}) \geq \gamma^{i,t}$, setting $c = 4$ and $i = m$ yields

$$\mathbb{P}\left[\gamma^{m,t} 2^m \geq 4Z\right] \leq \frac{1}{4}$$

Applying Chernoff's inequality and selecting $T$ as in the theorem statement gives

$$\mathbb{P}\left[4Z \geq \text{Median}\left(\gamma^{m,1}, \cdots, \gamma^{m,T}\right) 2^m\right] \geq 1 - \delta$$

The claim then follows from the union bound. □

*Proof of Proposition 1.* For singleton marginals, when $k \in [m+1, n]$, $x_k$ is a free variable and thus $\mu_k = E_q[x_k] = q_k(1)$. When $k \in [1, m]$,

$$(1 - 2x_k) = (1 - 2b_k)\prod_{i=m+1}^{n}(1 - 2C_{ki}x_i)$$

Take the expectation on both side and since $x_i$ for $i \in [m+1, n]$ are free (independent) variables, we have

$$(1 - 2\mu_k) = (1 - 2b_k)\prod_{i=m+1}^{n}(1 - 2C_{ki}\mu_i)$$

That is,

$$\mu_k = \left(1 - (1 - 2b_k)\prod_{i=m+1}^{n}(1 - 2C_{ki}\mu_i)\right)/2$$

For the binary marginal $\mu_{kl}$, there are three cases: both $x_k, x_l$ are free variables; one is free and the other is constrained; both are constrained.

*For the first case*, $k, \ell \in [m+1, n]$, they are independent and thus

$$\mu_{kl} = E_q[x_k x_\ell] = \mu_k \mu_\ell$$

*For the second case*, $k \in [m+1, n]$, $\ell \in [1, m]$. Define $X^k_{C_l}$ to be the set of $x_{m+1}, \ldots, x_n$ that satisfy constraint $l$ with $x_k$ fixed to 1; that is, $X^k_{C_l} = \{x_{m+1}, \ldots, x_n | x_k = 1, -1 = (1-2x_l) = (1-2b_l)\prod_{i=m+1}^{n}(1-2C_{li}x_i)\}$. Hence

$$\mu_{kl} = Pr[x_k = 1, x_l = 1] = \sum_{x_{m+1},\ldots,x_n \in X^k_{C_l}} \prod_{i=m+1}^{n} q_i(x_i)$$

When $C_{lk} = 1$ the $1 - 2C_{lk}x_k$ term in the product constraint on $X^k_{C_l}$ is $-1$, so the product constraint is $1 = (1 - 2b_l)\prod_{i=m+1, i\neq k}^{n}(1 - 2C_{li}x_i)$. Define this constrained set as $X^{k=1}_{C_l}$ and bring $q_k(x_k = 1)$ out of the product to yield

$$\mu_{kl} = \mu_k \cdot \sum_{x_{m+1},\ldots,x_n \in X^{k=1}_{C_l}} \prod_{i=m+1, i\neq k}^{n} q_i(x_i)$$

Introducing a new binary variable $u$ that satisfies the constraint

$$(2u - 1) = (1 - 2b_l) \prod_{i\neq k, i=m+1}^{n} (1 - 2C_{li}x_i)$$

the above summation is over $x_{m+1}, \ldots x_{k-1}, x_{k+1}, \ldots, x_n, u$ such that $u = 1$. Since $P(u = 1) = E[u]$,

$$\mu_{kl} = \mu_k E[u] = \mu_k \frac{1}{2}(1 + (1 - 2b_l) \prod_{i\neq k, i=m+1}^{n} (1 - 2C_{li}\mu_i))$$

as desired.

If $C_{lk} = 0$, then $x_l$ is independent of $x_k$, so $\mu_{kl} = \mu_k \mu_l$.

*For the last case*, $k, \ell \in [1, m]$.

$$(1 - 2x_k)(1 - 2x_\ell) = (1 - 2b_k)(1 - 2b_\ell) \prod_{i=m+1}^{n} (1 - 2C_{ki}x_i) \prod_{i=m+1}^{n} (1 - 2C_{\ell i}x_i)$$

Taking the expected value of both side

$$1 - 2\mu_l - 2\mu_k + 4\mu_{kl} = (1 - 2b_l)(1 - 2b_k) \prod_{i=m+1}^{n} E[1 - x_i(2C_{ki} + 2C_{li}) + 4C_{ki}C_{li}x_i^2]$$

so

$$\mu_{kl} = \frac{1}{4}(-1 + 2\mu_k + 2\mu_l + (1 - 2b_k)(1 - 2b_l) \prod_{i=m+1}^{n} (1 - \mu_i(2C_{ki} + 2C_{li} - 4C_{ki}C_{li})))$$

Plugging in the result of $\mu_k, \mu_l$:

$$4\mu_{kl} - 1 = (1 - 2b_k)(1 - 2b_l) \prod_{i=m+1}^{n} (1 - \mu_i(2C_{ki} + 2C_{li} - 4C_{ki}C_{li})) - (1 - 2b_k) \prod_{i=m+1}^{n} (1 - 2C_{ki}\mu_i) - (1 - 2b_l) \prod_{i=m+1}^{n} (1 - 2C_{li}\mu_i))$$

By inspection, all of the marginals $\mu_k, \mu_{kl}$ are linear in any particular free marginal $\mu_{m+1}, \ldots, \mu_n$. Hence, for any free marginal $\mu_i$ the objective function 11 contains a term linear in $\mu_i$ plus the entropy of a bernoulli random variable with parameter $\mu_i$, which is concave in $\mu_i$. Thus the coordinate ascent step for $\mu_i$ can be solved in closed form, as desired.

□