Jeroen's model

June 13, 2013

1 Notation

- Σ: partially-ordered list of symbols (types).
- T: number of symbols in grammar.
- $\mathcal{P}$: pose space.
- $\mathcal{B} = (b_1, \ldots, b_N)$: set of bricks.
- $\mathcal{P}(b) \in \mathcal{P}$: set of poses associated with brick b.
- $\mathcal{A} = (a_1, \ldots, a_M)$, $\mathcal{A} \subseteq \mathcal{B}$: set of active bricks.
- $\mathcal{A}_m = (a_1, \ldots, a_m)$, $m \leq M$ $\mathcal{A}_m \subseteq \mathcal{A}$: subset of active bricks.
- $t(b)$: type of brick b.
- $\#(C)$: the cardinality (size) of some set C. *E.g.* $\#(\mathcal{A}) = M$.
- $\mathcal{R}$: set of production rules of the form $A \rightarrow B_1, \ldots, B_k$ with $r = A, B_1, \ldots, B_k \in \Sigma$. $\mathcal{R}$ is acyclic.
- $n(b)$: number of slots/children associated with bricks of type $t(b)$.
- $c(b)$: the set of production rules with $t(b)$ in the LHS.
- $\vec{\theta}_b \in \mathbb{R}^{\#(c(b))}$: distribution over rules with $t(b) \in \mathcal{B}$ in the LHS.
- $s_b \in \{0, 1\}$: on/off state of brick b.

- $r_b \in \mathcal{R}(t(b))$: rule used by brick b.
- $\mathbf{g_b} \in (\mathcal{B} \cup \perp)^{n(b)}$: array of pointers, where each element can either point to a brick, or null (no child).
- $\vec{q}_b \in \{\mathcal{P}(b), \perp\}$: pose for brick b. $\vec{q}_b = \perp$ is a "nothing" pose that does not contribute to image evidence.
- $\mathbf{q}^{\mathcal{A}_m}$: poses for bricks $a \in \mathcal{A}_m$.
- $\mathbf{s}^{\mathcal{A}_\mathbf{m}}$: on/off state for bricks $a \in \mathcal{A}_m$.
- $\mathbf{r}^{\mathcal{A}_\mathbf{m}}$: rules for bricks $a \in \mathcal{A}_m$.
- $\mathbf{g}_b^{\mathcal{A}_m} \in (\mathcal{A}_m \cup \emptyset)^{n(b)}$: array of pointers, where each element can either point to a brick in the active set defined by $\mathcal{A}_m$, or blank (child not yet specified) denoted by $\emptyset$. Note that blank is **different** than null (no child) in that blank means "there could be a child in this slot, but we don't know if there is or which one yet" while null (no child) means "there definitely isn't a child in this slot".
- $comp(\mathbf{g}^{\mathcal{A}_m})$: the elements of $\mathbf{g}^{\mathcal{A}_m}$ for which $g_{b,k}^{\mathcal{A}_m} = \emptyset$. Note: we use "comp" as the identifier for this concept since these are the slots that may receive a child in the future; these are the slots to be"completed".
- $H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m})$: **set** of children that $(\vec{g}_b^{\mathcal{A}_m})$ points to that were orphans according to $\mathbf{X}_{m-1}$.
- $\tilde{H}^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m})$: **multi-set** of children that $(\vec{g}_b^{\mathcal{A}_m})$ points to that were orphans according to $\mathbf{X}_{m-1}$.
- $V(b)$: "before" set of bricks, which we define as $V(b) = \{b' : t(b') < t(b)\}$ where $<$ means "comes before in the partial ordering defined by Σ".
- $W(b)$: "after" set of bricks, which we define as $V(b) = \{b' : t(b') > t(b)\}$ where $>$ means "comes after in the partial ordering defined by Σ".
- $\vec{\phi}_{b,k,r} \in \mathbb{R}^{\#(W(b))+1}$: distribution over bricks to which brick b's kth slot points to if brick b is using rule r. Note that brick b may only point to bricks in its "after" set. Note that the

+1 is present since a brick/slot may point to nothing, denoted by the symbol $\perp$. *edit*: this is a bit sloppy. Each slot has a particular type, and only bricks of that type may go in this slot. The set of possibilities for each slot, in general, is a subset of the "after" set, and not the entire after set.

- $\vec{\pi} \in \mathbb{R}^T$: vector of mixing coefficients for the templates of each type of brick.
- $\mathbf{X} = \{\mathbf{s}, \mathbf{r}, \mathbf{g}, \mathbf{q}\}$: collection of hidden variables of model.
- $\mathbf{X}^{\mathcal{A}_m} = \{\mathbf{s}^{\mathcal{A}_\mathbf{m}}, \mathbf{r}^{\mathcal{A}_\mathbf{m}}, \mathbf{g}^{\mathcal{A}_\mathbf{m}}, \mathbf{q}^{\mathcal{A}_\mathbf{m}}\}$: collection of hidden variables for bricks $a \in \mathcal{A}_m$.
- $\mathbf{X}_{-r}^{\mathcal{A}_m} = \{\mathbf{s}^{\mathcal{A}_\mathbf{m}}, \mathbf{g}^{\mathcal{A}_\mathbf{m}}, \mathbf{q}^{\mathcal{A}_\mathbf{m}}\}$: collection of hidden variables for bricks $a \in \mathcal{A}_m$ except the hidden variables representing rules used by the bricks.
- $orph(\mathbf{X}_{-r}^{\mathcal{A}_m})$: the set of bricks that are orphans according to the assignments in $\mathbf{X}_{-r}^{\mathcal{A}_m}$ (see definition of "orphan" below).
- $off(\mathbf{X}_{-r}^{\mathcal{A}_m})$: the set of bricks that are off (have $s = 0$) according to the assignments in $\mathbf{X}_{-r}^{\mathcal{A}_m}$. Note that if a brick is off, then $\nexists\{a, k\}$, $a \in \mathcal{A}_m$ s.t. $g_{a,k}^{\mathcal{A}_m} = b$ where b is the off-brick in question.
- Y: the image.

Definition A brick b is an **orphan** according to the assignment of random variables for $\mathbf{X}_{-r}^{\mathcal{A}_m}$ iff $s_b = 1$ and $\nexists\{a, k\}$ s.t. $g_{a,k}^{\mathcal{A}_m} = b$ according to $\mathbf{X}_{-r}^{\mathcal{A}_m}$.

2 Model

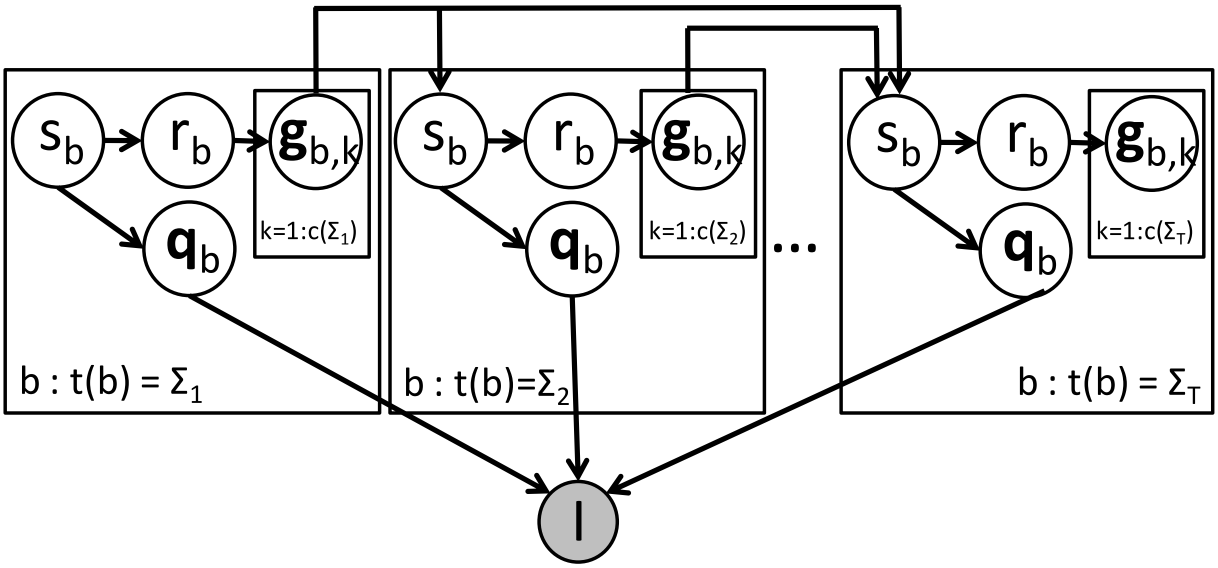

Figure 1: Graphical model

The decomposition of the joint probability is given by the graphical model in Fig. 1 and by:

$$\begin{aligned}
P(\mathbf{X}, Y) \quad &= \quad \Big(\prod_{b \in \mathcal{B}} P(s_b \mid \mathbf{g}_{V(b)}) \Big) \\
&\quad \times \Big(\prod_{b \in \mathcal{B}} P(r_b \mid s_b) \prod_{k=1}^{n(b)} P(g_{b,k} \mid r_b) \Big) \\
&\quad \times \Big(\prod_{b \in \mathcal{B}} P(\vec{q}_b \mid s_b) \Big) \\
&\quad \times \big(P(Y \mid \mathbf{q}) \big)
\end{aligned} \tag{1}$$

where we define:

$$P(s_b = 1 \mid \mathbf{g}_{V(b)}) = \begin{cases} 1, & \text{if } \exists \text{ \{a,k\} s.t. } g_{a,k} = b \text{ , } g_{a,k} \in g_{V(b)} \quad \text{(2a)} \\ \epsilon, & \text{if } \nexists \text{ \{a,k\} s.t. } g_{a,k} = b \text{ , } g_{a,k} \in g_{V(b)} \quad \text{(2b)} \end{cases}$$

$$P(r_b \mid s_b) \quad = \quad Discrete(\vec{\theta}_b) \tag{3}$$

$$P(g_{b,k} \mid r_b) \quad = \quad Discrete(\vec{\phi}_{b,k,r_b}) \tag{4}$$

$$
P(\vec{q}_b = \vec{q} \mid s_b)) \begin{cases} \dfrac{1}{\#\mathcal{P}(b)}, & \text{if } s_b = 1,\ \vec{q} \in \mathcal{P}(b) \quad (5a) \\ 0, & \text{if } s_b = 1 \text{ and } \vec{q} = \perp \quad (5b) \\ 0, & \text{if } s_b = 0,\ \vec{q} \in \mathcal{P}(b) \quad (5c) \\ 1, & \text{if } s_b = 0 \text{ and } \vec{q} = \perp \quad (5d) \end{cases}
$$

$$
\begin{aligned}
P(Y \mid \mathbf{q}) &= \prod_p P(Y_p \mid \mathbf{q}) && (6) \\
P(Y_p \mid \mathbf{q}) &\propto \sum_b \pi_{t(b)} P(Y_p \mid \vec{q}_b) I(\text{template overlaps with pixel } p) && (7) \\
P(Y_p \mid \vec{q}_b) &= \text{defn here, with Bernoulli distribution and stuff.} && (8)
\end{aligned}
$$

where $I(\cdot)$ is an indicator variable.

Localization

We define localizations in terms of the components $\big(\prod_{b\in\mathcal{B}} P(s_b \mid \mathbf{g}_{V(b)})\big)$, $\big(\prod_{b\in\mathcal{B}} P(r_b \mid s_b) \prod_{k=1}^{n(b)} P(g_{b,k} \mid r_b)\big)$, $\big(\prod_{b\in\mathcal{B}} P(\vec{q}_b \mid s_b)\big)$, and $P(Y \mid \mathbf{q})$ separately.

$$
\Big(\prod_{b\in\mathcal{B}} P(s_b \mid \mathbf{g}_{V(b)})\Big) \;\Rightarrow\; \Big(\prod_{a\in\mathcal{A}_m} P(s_a \mid \mathbf{g}^{\mathcal{A}_m}_{V(b)})\Big) \tag{9}
$$

where $\Rightarrow$ is used to mean “localiizes to”. Similarily,

$$
\begin{aligned}
\Big(\prod_{b\in\mathcal{B}} P(r_b \mid s_b) \prod_{k=1}^{n(b)} P(g_{b,k} \mid r_b)\Big) &\;\Rightarrow\; \Big(\prod_{a\in\mathcal{A}_m} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a)\Big) && (10) \\
P^{\mathcal{A}_m}(g_{a,k} \mid r_a) &= \begin{cases} 1 - \sum_{a'\in\mathcal{A}_m} P(g_{a,k} = a' \mid r_a) \text{ if } g_{a,k} = \emptyset \\ P(g_{a,k} \mid r_a) \text{ , otherwise} \end{cases} && (11)
\end{aligned}
$$

$$
\Big(\prod_{b\in\mathcal{B}} P(\vec{q}_b \mid s_b)\Big) \Rightarrow \Big(\prod_{a\in\mathcal{A}_m} P(\vec{q}_a \mid s_a)\Big) \tag{12}
$$

$$
P(Y \mid \mathbf{q}) \Rightarrow P^{\mathcal{A}_m}(Y \mid \mathbf{q}^{\mathcal{A}_\mathbf{m}}) \tag{13}
$$

where we evaluate $P^{\mathcal{A}_m}(Y \mid \mathbf{q}^{\mathcal{A}_\mathbf{m}})$ in a similar fashion to $P(Y \mid \mathbf{q})$ outlined in Eqn. 6. We note that $P^{\mathcal{A}_m}(Y \mid \mathbf{q}^{\mathcal{A}_\mathbf{m}}) \neq P(Y \mid \mathbf{q}^{\mathcal{A}_\mathbf{m}})$ since we do not perform the correct marginalizations.

We define the overall localized distribution, $P^{\mathcal{A}_m}(X^{\mathcal{A}_m}$, Y), in terms of the previously defined localization distributions:

$$
\begin{aligned}
P^{\mathcal{A}_m}(X^{\mathcal{A}_m}, Y) \;=\; & \big(\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}^{\mathcal{A}_m}_{V(b)}) \big) \\
& \times \big(\prod_{a \in \mathcal{A}_m} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \big) \\
& \times \prod_{a \in \mathcal{A}_m} P(\vec{q}_a \mid s_a) \big) \\
& \times \big(P^{\mathcal{A}_m}(Y \mid \mathbf{q}^{\mathcal{A}_\mathbf{m}}) \big)
\end{aligned} \tag{14}
$$

State representation

Representing the configuration of an image for a **non-localized** model requires specifying s_b, g_b, r_b, $\vec{q}_b$, $\forall b \in \mathcal{B}$. For a **localized** model, the same random variables could be used as the state representation (but with $\mathbf{g}^{\mathcal{A}_\mathbf{m}}$ defined over a different domain than $\mathbf{g}$). However, requiring specification of the r_b's is unnecessary and too restrictive; we may not be able to well predict the rule a brick has used by examining the set of active bricks. This is especially true if the active set only contains one brick! Instead, we represent the state representation of the localized model by the random variables $\mathbf{X}^{\mathcal{A}_m}_{-r} = \{s_a,\, g_a,\, \vec{q}_a\}$, $\forall a \in \mathcal{A}_m$, and we marginalize over r_a wherever it is used. In particular

$$
P^{\mathcal{A}_m}(\mathbf{X}^{\mathcal{A}_m}_{-r}, Y) = \sum_{\mathbf{r}} P^{\mathcal{A}_m}(\mathbf{X}^{\mathcal{A}_m}, Y) \tag{15}
$$

$$
\begin{aligned}
= \; & \big(\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}^{\mathcal{A}_m}_{V(b)}) \big) \\
& \times \big(\prod_{a \in \mathcal{A}_m} \sum_{r_a} \big(P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \big) \big) \\
& \times \prod_{a \in \mathcal{A}_m} P(\vec{q}_a \mid s_a) \\
& \times \big(P^{\mathcal{A}_m}(Y \mid \mathbf{q}^{\mathcal{A}_\mathbf{m}}) \big).
\end{aligned} \tag{16}
$$

Inference

Inference proceeds by iteratively adding one brick at a time to the active set. At any iteration, m, our goal is to maintain a representation of the localized posterior distribution $P^{\mathcal{A}_m}(\mathbf{X}_{-r}^{\mathcal{A}_m} \mid Y)$. To maintain this representation, we employ a particle filtering approach. We treat samples from the previous localized posterior distribution $P^{\mathcal{A}_{m-1}}(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid Y)$ as draws from a proposal distribution, and re-weight these particles accordingly (discuss re-weighting and evaluating marginals later). Note that $P^{\mathcal{A}_{m-1}}(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid Y) \neq P^{\mathcal{A}_m}(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid Y)$; the model defined by $P^{\mathcal{A}_m}$ is different than the one defined by $P^{\mathcal{A}_{m-1}}$.

Suppose we have sampled a previous state $\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}$, and suppose the brick we are adding to the active set is denoted b. To generate $\mathbf{X}_{-r}^{\mathcal{A}_m}$, we need to draw values for the set of random variables defined by $\mathbf{X}_{-r}^{\mathcal{A}_m} \setminus \mathbf{X}_{-r}^{\mathcal{A}_{m-1}} = \{s_b,\ \vec{q}_b,\ \mathbf{g}_{\mathbf{b}}^{\mathcal{A}_\mathbf{m}},\ comp(g^{\mathcal{A}_{m-1}})\}$. That is, we need to "expand" the representation of $\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}$ to include brick b. As such, we need to draw a sample from $P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_{\mathbf{b}}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}}) \mid \mathbf{X}_{-r}^{\mathcal{A}_{m-1}}, Y)$. We note that for $g \in comp(\mathbf{g}^{\mathcal{A}_m})$, the only possible values that can be sampled for g are $g = \emptyset$, $g = b$. That is, each not-yet-assigned brick slot can choose either to point to the newly-added active brick b, or remain empty. Note that

$$
\begin{aligned}
P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_{\mathbf{b}}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}}) \mid \mathbf{X}_{-r}^{\mathcal{A}_{m-1}}, Y) &= \frac{P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_{\mathbf{b}}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}}), \mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid Y)}{\sum_{s_b} \sum_{g_b} \sum_{\mathbf{g}_{\mathbf{b}}^{\mathcal{A}_\mathbf{m}}} \sum_{comp(g^{\mathcal{A}_{m-1}})} P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_{\mathbf{b}}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}}), \mathbf{X}_{-}^{\mathcal{A}}} \\
P^{\mathcal{A}_m}(s_b, \vec{q}_b, \mathbf{g}_{\mathbf{b}}^{\mathcal{A}_\mathbf{m}}, comp(g^{\mathcal{A}_{m-1}}), \mathbf{X}_{-r}^{\mathcal{A}_{m-1}} \mid Y) &= P^{\mathcal{A}_m}(\mathbf{X}_{-r}^{\mathcal{A}_m} \mid Y) \\
&\propto P^{\mathcal{A}_m}(\mathbf{X}_{-r}^{\mathcal{A}_m}, Y).
\end{aligned}
$$

We first note that for any setting of the random variables outlined above, we may classify it as one of three events: 1) brick b is off, 2) brick b is on, and has no parents, and 3) brick b is on, and has at least one parent. We compute the mass associated with each of these three events, with approximations where needed. We will be able to evaluate the denominator in Eqn. 17 by summing the probability mass associated with these three events. We examine these events in more detail below.

Event 1: brick b is off

For this event, there is only one setting of the random variables $\{s_b, \vec{q}_b, \vec{g}_b^{\mathcal{A}_m}, comp(g^{\mathcal{A}_{m-1}}\}$ that has non-zero probability: $s_b = 0$, $\vec{q}_b = \perp$, $g_{b,k} = \perp\ \forall k$, $g = \emptyset\ \forall g \in comp(\mathbf{g}^{\mathcal{A}_{m-1}})$. Note that $g^{\mathcal{A}_m} \neq g^{\mathcal{A}_{m-1}}$

only because $g^{\mathcal{A}_{m-1}}$ specifies the (lack of) children for the brick b, which is not specified in $g^{\mathcal{A}_{m-1}}$. The other corresponding entries of $g^{\mathcal{A}_m}$ and $g^{\mathcal{A}_{m-1}}$ are indeed equal, however. It is trivial to evaluate Eqn. 16 given a particular sampled previous state, $\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}$.

Event 2: brick b is on and has no parents

There are many settings of the random variables $\{s_b, \vec{q}_b, \mathbf{g_b^{\mathcal{A}_m}}, comp(g^{\mathcal{A}_{m-1}})\}$ that fall under this event. They can be characterized as: $s_b = 1$, $\vec{q}_b \neq \perp$ (or equivalently, $\vec{q}_b \in P(b)$), $g = \emptyset \ \forall g \in comp(\mathbf{g}^{\mathcal{A}_{m-1}})$, and $g_{b,k}^{\mathcal{A}_m} = \{\emptyset, a \text{ s.t. } a \in \mathcal{A}_m\}$, $\forall k$. So, the probability mass associated with this event is given by

$$\text{P(Event 2)} = \sum_{\vec{q}_b \in P(b)} \sum_{\vec{g}_b^{\mathcal{A}_m}} P^{\mathcal{A}_m}(s_b = 1, \vec{q}_b, \vec{g}_b^{\mathcal{A}_m}, comp(\mathbf{g}^{\mathcal{A}_{m-1}}) = \emptyset, \mathbf{X}_{-r}^{\mathcal{A}_{m-1}}, Y) \tag{20}$$

where we have used $comp(\mathbf{g}^{\mathcal{A}_{m-1}}) = \emptyset$ as a shorthand for $g = \emptyset \ \forall g \in comp(\mathbf{g}^{\mathcal{A}_{m-1}})$. Expanding, we get

$$\begin{aligned} = \quad & \sum_{\vec{g}_b^{\mathcal{A}_m}} \Big(\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) \Big) \Big(\prod_{a \in \mathcal{A}_m} \Big(\sum_{r_a} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \Big) \Big) \\ & \Big(\prod_{a \in \mathcal{A}_{m-1}} P(\vec{q}_a \mid s_a) \Big) \Big(\sum_{\vec{q}_b \in P(b)} P(\vec{q}_b \mid s_b = 1) P^{\mathcal{A}_m}(Y \mid \mathbf{q^{\mathcal{A}_{m-1}}}, \vec{q}_b) \Big). \end{aligned} \tag{21}$$

The second term in Eqn. 21, $\big(\prod_{a \in \mathcal{A}_{m-1}} P(\vec{q}_a \mid s_a)\big)\big(\sum_{\vec{q}_b \in P(b)} P(\vec{q}_b \mid s_b = 1)\big) P^{\mathcal{A}_m}(Y \mid \mathbf{q^{\mathcal{A}_{m-1}}}, \vec{q}_b)\big)$, can be evaluated by enumerating all possible values for $\vec{q}_b \in P(b)$. In practice, whether this is practical or not is dependent on $\#(P(b))$. If $\#(P(b))$ is "small enough", then $\big(\prod_{a \in \mathcal{A}_{m-1}} P(\vec{q}_a \mid s_a)\big)\big(\sum_{\vec{q}_b \neq \perp} P(\vec{q}_b \mid s_b = 1)\big) P^{\mathcal{A}_m}(Y \mid \mathbf{q^{\mathcal{A}_{m-1}}}, \vec{q}_b)\big)$ can be computed.

Turning our attention to the first term in Eqn. 21, we note that it is not possible to efficiently compute this quantity. So we come up with approximations for it.

From the definition given in Eqn. 2,

$$\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) = \epsilon^{\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_m}))} \times (1 - \epsilon)^{\#(off(\mathcal{A}_{m-1}))}. \tag{22}$$

So to evaluate $\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m})$ it is only necessary to find the number of orphan bricks in $\mathcal{A}_m$. Next, we derive a relationship between $\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}))$ and $\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_m}))$. If $s_b = 1$ and brick b does not have at least one parent, then the number of orphans increases by 1 since b itself is an orphan. If we denote the **set** of children of b given a particular $\vec{g}_b^{\mathcal{A}_m}$ as $ch(\vec{g}_b^{\mathcal{A}_m})$, then the number of orphans is decreased by $\#(ch(\vec{g}_b^{\mathcal{A}_m}) \cap (orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}))$ due to b's becoming the parent of these previously-orphaned bricks, and b being unable to be its own child. The relation between $\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_m}))$ and $\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}))$ can thus be summarized as

$$\begin{aligned} \#(orph(\mathbf{X}_{-r}^{\mathcal{A}_m})) &= \#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}})) + I(\text{b is orphan}) - \#(ch(\vec{g}_b^{\mathcal{A}_m}) \cap orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}})) && (23) \\ &= \#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}})) + I(\text{b is orphan}) - \#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m})) && (24) \end{aligned}$$

Combining Eqn. 24 with Eqn. 22 yields

$$\begin{aligned} \prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) &= \epsilon^{\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_m}))} \times (1-\epsilon)^{\#(off(\mathcal{A}_{m-1}))} && (25) \\ &= \epsilon^{\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}))+I(\text{b is orphan})-\#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m}))} \times (1-\epsilon)^{\#(s(\mathcal{A}_{m-1}))} && (26) \\ &= \epsilon^{\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}))+I(\text{b is orphan})} \times \epsilon^{-\#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m}))} \times (1-\epsilon)^{\#(off(\mathcal{A}_{m-1}))}. && (27) \end{aligned}$$

Combining Eqn. 27 and the first term of Eqn. 21, and noting for this event that $I(\text{b is orphan}) = 1$

yields

$$\begin{aligned}
& \sum_{\vec{g}_b^{\mathcal{A}_m}} \big(\prod_{a \in \mathcal{A}_m} P^{\mathcal{A}_m}(s_a \mid \mathbf{g}_{V(b)}^{\mathcal{A}_m}) \big(\prod_{a \in \mathcal{A}_m} \sum_{r_a} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \big) \big) && (28) \\
= & \sum_{\vec{g}_b^{\mathcal{A}_m}} \big(\epsilon^{\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}})) + I(\text{b is orphan})} \times (1-\epsilon)^{\#(off(\mathcal{A}_{m-1}))} \times \epsilon^{-\#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m}))} \\
& \times \big(\prod_{a \in \mathcal{A}_m} \sum_{r_a} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \big) \big) && (29) \\
= & \epsilon^{\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}))+1} \times (1-\epsilon)^{\#(off(\mathcal{A}_{m-1}))} \\
& \times \sum_{\vec{g}_b^{\mathcal{A}_m}} \big(\epsilon^{-\#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m}))} \big(\prod_{a \in \mathcal{A}_m} \sum_{r_a} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \big) \big) && (30) \\
= & \epsilon^{\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}))+1} \times (1-\epsilon)^{\#(off(\mathcal{A}_{m-1}))} \\
& \times \big(\prod_{a \in \mathcal{A}_m, a \neq b} \sum_{r_a} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \big) && (31) \\
& \times \sum_{\vec{g}_b^{\mathcal{A}_m}} \big(\epsilon^{-\#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m}))} \big(\sum_{r_b} P(r_b \mid s_b = 1) \prod_{k=1}^{n(b)} P^{\mathcal{A}_m}(g_{b,k} \mid r_b) \big) \big) \\
= & \epsilon^{\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}))+1} \times (1-\epsilon)^{\#(off(\mathcal{A}_{m-1}))} \\
& \times \big(\prod_{a \in \mathcal{A}_m, a \neq b} \sum_{r_a} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \big) && (32) \\
& \times \sum_{r_b} \big(P(r_b \mid s_b = 1) \sum_{\vec{g}_b^{\mathcal{A}_m}} \big(\epsilon^{-\#(H^{\mathbf{X}_{m-1}})}(\vec{g}_b^{\mathcal{A}_m}) \prod_{k=1}^{n(b)} P^{\mathcal{A}_m}(g_{b,k} \mid r_b) \big) \big)
\end{aligned}$$

It is intractable to compute the quantity given in Eqn. 32. Note that the evaluation of $\epsilon^{-\#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m}))}$ for a given $\vec{g}_b^{\mathcal{A}_m}$ does not factorize over the $g_{b,k}$'s. Recall that $\#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m}))$ is the number of *distinct* children pointed to $\vec{g}_b^{\mathcal{A}_m}$ that were previously orphans according to $\mathbf{X}_{m-1}$; it is the requirement that children be distinct that couples the $g_{b,k}^{\mathcal{A}_m}$'s. If we remove the distinctness requirement, and simply count the number of previously-orphan children pointed to by $\vec{g}_b^{\mathcal{A}_m}$, then the effect of each of the $g_{b,k}^{\mathcal{A}_m}$'s for a given r_b will decouple. More precisely, we use the approximation

$$\begin{aligned}
\#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m})) &\approx \#(\tilde{H}^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m})) && (33) \\
\Longrightarrow \epsilon^{-\#(H^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m}))} &\approx \epsilon^{-\#(\tilde{H}^{\mathbf{X}_{m-1}}(\vec{g}_b^{\mathcal{A}_m}))}. && (34)
\end{aligned}$$

(OK, the above isn't really true. Just because the quantities in the exponents are roughly equally doesn't necessarily mean the entire exponentiated quantity is.).

We expect the approximation given in Eqn. 34 to be a good approximation when the probability of the event $\exists\{k,k'\}, k \neq k'$ s.t. $g^{\mathcal{A}_m}_{b,k} = g^{\mathcal{A}_m}_{b,k'}$ is small. That is, it is unlikely that two slots will point to the same child.

With the approximation outlined in Eqn. 34, we can approximately compute the second term in 32. For a given r_b, we can compute the probability that each of the $n(b)$ slots will point to a previously-orphaned child separately. We then average this quantity over the r_b's. In other words, with the approximation in Eqn. 34, the computation factorizes across the slots for a given r_b:

$$\begin{aligned}
& \sum_{r_b} \big(P(r_b \mid s_b = 1) \sum_{\vec{g}^{\mathcal{A}_m}_b} \big(\epsilon^{-\#(H^{\mathbf{X}_{m-1}})(\vec{g}^{\mathcal{A}_m}_b)} \prod_{k=1}^{n(b)} P^{\mathcal{A}_m}(g_{b,k} \mid r_b)\big)\big) && (35)\\
\approx\; & \sum_{r_b} \big(P(r_b \mid s_b = 1) \sum_{\vec{g}^{\mathcal{A}_m}_b} \big(\epsilon^{-\#(\tilde{H}^{\mathbf{X}_{m-1}})(\vec{g}^{\mathcal{A}_m}_b)} \prod_{k=1}^{n(b)} P^{\mathcal{A}_m}(g_{b,k} \mid r_b)\big)\big) && (36)\\
=\; & \sum_{r_b} \big(P(r_b \mid s_b = 1) \sum_{\vec{g}^{\mathcal{A}_m}_b} \big(\prod_{k=1}^{n(b)} \epsilon^{-\#(\tilde{H}^{\mathbf{X}_{m-1}}(g^{\mathcal{A}_m}_{b,k}))} P^{\mathcal{A}_m}(g_{b,k} \mid r_b)\big)\big) && (37)\\
=\; & \sum_{r_b} \big(P(r_b \mid s_b = 1) \prod_{k=1}^{n(b)} \big(\sum_{g^{\mathcal{A}_m}_{b,k}} \epsilon^{-\#(\tilde{H}^{\mathbf{X}_{m-1}}(g^{\mathcal{A}_m}_{b,k}))} P^{\mathcal{A}_m}(g_{b,k} \mid r_b)\big)\big) && (38)\\
=\; & \sum_{r_b} \big(P(r_b \mid s_b = 1) \prod_{k=1}^{n(b)} \big(\epsilon^{-1} P^{\mathcal{A}_m}(g_{b,k} = a, a \in orph(\mathbf{X}^{\mathcal{A}_{m-1}}_{-r}) \mid r_b) && (39)\\
& \qquad + \epsilon^0 (1 - P^{\mathcal{A}_m}(g_{b,k} = a, a \in orph(\mathbf{X}^{\mathcal{A}_{m-1}}_{-r}) \mid r_b))\big)\big).
\end{aligned}$$

Notice the ˜ above the H in Eqn. 36. It is trivial to compute $P^{\mathcal{A}_m}(g_{b,k} = a, a \in orph(\mathbf{X}^{\mathcal{A}_{m-1}}_{-r}) \mid r_b)$ for each rule r_b since $P^{\mathcal{A}_m}(g_{b,k} \mid r_b)$ is a *Discrete* distribution (table of probabilities) and the set $orph(\mathbf{X}^{\mathcal{A}_{m-1}}_{-r})$ can be easily enumerated.

Combining the derived approximation, Eqn. 40, Eqn. 32, and Eqn. 21 yields

$$\begin{aligned}
P(\text{Event 2}) \quad \approx \quad & \epsilon^{\#(orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}))+1} \times (1-\epsilon)^{\#(off(\mathcal{A}_{m-1}))} \\
\times \quad & \prod_{a \in \mathcal{A}_{m-1}} \big(\sum_{r_a} P(r_a \mid s_a) \prod_{k=1}^{n(a)} P^{\mathcal{A}_m}(g_{a,k} \mid r_a) \big) \\
\times \quad & \big(\sum_{r_b} \big(P(r_b \mid s_b = 1) \prod_{k=1}^{n(b)} \big(\epsilon^{-1} P^{\mathcal{A}_m}(g_{b,k} = a, a \in orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}) \mid r_b) \\
& \qquad + \epsilon^0 (1 - P^{\mathcal{A}_m}(g_{b,k} = a, a \in orph(\mathbf{X}_{-r}^{\mathcal{A}_{m-1}}) \mid r_b)) \big) \big) \big) \\
\times \quad & \big(\prod_{a \in \mathcal{A}_{m-1}} P(\vec{q}_a \mid s_a) \big) \\
\times \quad & \sum_{\vec{q}_b \in P(b)} \big(P(\vec{q}_b \mid s_b = 1) P^{\mathcal{A}_m}(Y \mid \mathbf{q}^{\mathcal{A}_{\mathbf{m-1}}}, \vec{q}_b) \big) \big)
\end{aligned} \tag{40}$$

where we have made use of the equality $\{a : a \in \mathcal{A}_m, a \neq b\} = \{a : a \in \mathcal{A}_{m-1}\}$.

Now do alternate approximation (Jeroen's approx)...although Jeroen's approx is probably worse.

Event 3: brick b is on, and has at least 1 parent

There are many settings of the random variables $\{s_b, \vec{q}_b, \mathbf{g}_{\mathbf{b}}^{\mathcal{A}_{\mathbf{m}}}, comp(g^{\mathcal{A}_{m-1}})\}$ that fall under this event. They can be characterized as: $s_b = 1$, $\vec{q}_b \neq \perp$ (or equivalently, $\vec{q}_b \in P(b)$), $\{comp(\mathbf{g}^{\mathcal{A}_{m-1}}) : \exists \{a, k\}$ s.t. $g_{a,k}^{\mathcal{A}_m} = b\}$, and $g_{b,k}^{\mathcal{A}_m} = \{\emptyset, a$ s.t. $a \in \mathcal{A}_m\}$, $\forall k$. So, the probability mass associated with this event is given by

$$\sum_{\{comp(\mathbf{g}^{\mathcal{A}_{m-1}}) : \exists\{a,k\} \text{ s.t. } g_{a,k}^{\mathcal{A}_m} = b\}} \sum_{\vec{q}_b \in P(b)} \sum_{\vec{g}_b^{\mathcal{A}_m}} P^{\mathcal{A}_m}(s_b = 1, \vec{q}_b, \vec{g}_b^{\mathcal{A}_m}, comp(\mathbf{g}^{\mathcal{A}_{m-1}}) = \emptyset, \mathbf{X}_{-r}^{\mathcal{A}_{m-1}}, Y) \tag{41}$$

$$= \tag{42}$$