Derivation of LHL's Log-Posterior

In the MLN defining the prior component of the posterior probability, there are two rules. The first rule has infinite weight, and it states that each symbol belongs to exactly one cluster. The second rule has negative weight $-\infty < -\lambda < 0$, and it penalizes the number of cluster combinations. From that MLN, we get

$$P(\{\Gamma\}) = \frac{\exp\left(\infty \cdot n_{\{\Gamma\}} - \lambda m_{\{\Gamma\}}\right)}{Z} = \frac{\exp\left(\infty \cdot n_{\{\Gamma\}} - \lambda m_{\{\Gamma\}}\right)}{\sum_{\{\Gamma\}'} \exp\left(\infty \cdot n_{\{\Gamma\}'} - \lambda m_{\{\Gamma\}'}\right)} \tag{1}$$

where Z is the partition function; $n_{\{\Gamma\}}$ and $m_{\{\Gamma\}}$ are respectively the number of true groundings of the first and second rules for cluster assignment $\{\Gamma\}$.

We first consider the case where the first rule is violated in $\{\Gamma\}$, i.e., there is a symbol that does not belong to exactly one cluster. Note that there is a cluster assignment in which the first rule is not violated, specifically, the one where each symbol is in its own cluster. Let this cluster assignment be $\{\Gamma\}^u$. Rewriting Equation 1, we get

$$P(\{\Gamma\}) = \frac{\exp\left(-\lambda m_{\{\Gamma\}}\right)}{\exp\left(\infty \cdot (n_{\{\Gamma\}^u} - n_{\{\Gamma\}}) - \lambda m_{\{\Gamma\}^u}\right) + \sum_{\{\Gamma\}' \setminus \{\Gamma\}^u} \exp\left(\infty \cdot (n_{\{\Gamma\}'} - n_{\{\Gamma\}}) - \lambda m_{\{\Gamma\}'}\right)}. \tag{2}$$

Since $n_{\{\Gamma\}} < n_{\{\Gamma\}^u}$, $0 < \lambda < \infty$, and $0 \leq m_{\Gamma^u} < \infty$, $\exp\left(\infty \cdot (n_{\{\Gamma\}^u} - n_{\{\Gamma\}}) - \lambda m_{\{\Gamma\}^u}\right) = \infty$. Consequently, the denominator of Equation 2 is ∞, and $P(\{\Gamma\}) = 0$. Thus when the first rule is violated, the posterior $P(\{\Gamma\}|D) = 0$, and $\log P(\{\Gamma\}|D) = -\infty$.

Next we consider the case where the first rule is not violated in $\{\Gamma\}$. We divide the numerator and denominator of Equation 1 by $\exp\left(\infty \cdot n_{\{\Gamma\}}\right)$. Let $\{\Gamma\}''$ be a cluster assignment in the summation of Z. When $\{\Gamma\}''$ violates the first rule, its contribution to the summation is zero. This is because $n_{\{\Gamma\}''} < n_{\{\Gamma\}}$ and $\exp(\infty \cdot (n_{\{\Gamma\}''} - n_{\{\Gamma\}}) - \lambda m_{\{\Gamma\}''}) = 0$. When $\{\Gamma\}''$ does not violate the first rule, $n_{\{\Gamma\}''} = n_{\{\Gamma\}}$, and $\exp(\infty \cdot (n_{\{\Gamma\}''} - n_{\{\Gamma\}}) - \lambda m_{\{\Gamma\}''}) = \exp(-\lambda m_{\{\Gamma\}''})$. Consequently, we can write Equation 1 as

$$P(\{\Gamma\}) = \frac{\exp\left(-\lambda m_{\{\Gamma\}}\right)}{\sum_{\{\Gamma\}''} \exp\left(-\lambda m_{\{\Gamma\}''}\right)} = \frac{\exp\left(-\lambda m_{\{\Gamma\}}\right)}{Z'} \tag{3}$$

where the summation in the denominator is over cluster assignments that do not violate the first rule.

Taking logs, we get

$$\log P(\{\Gamma\}) = -\lambda m_{\{\Gamma\}} + K \tag{4}$$

where $K = -\log(Z')$ is a constant.

Next we derive the likelihood component of the posterior probability. Since each symbol x_i belongs to exactly one cluster γ_i, each ground atom $r(x_1, \ldots, x_n)$ is in exactly one cluster combination $(\gamma_1, \ldots, \gamma_n)$. Let $G_{r(x_1,\ldots,x_n)}$ be a set containing groundings of the atom prediction rules and the (single) grounding of the default atom prediction rule that have ground atom $r(x_1, \ldots, x_n)$ as their consequents. (An *antecedent* and *consequent* respectively appear on the left and right of the implication symbol $\Rightarrow$.) Suppose the cluster combination $(\gamma_1, \ldots, \gamma_n)$ to which $r(x_1, \ldots, x_n)$ belongs contains at least one true ground atom. Then there is exactly one grounded atom prediction rule in $G_{r(x_1,\ldots,x_n)}$ whose antecedent is true. The antecedents of all other rules in $G_{r(x_1,\ldots,x_n)}$ are false, and the rules are trivially true. Similarly, when cluster combination $(\gamma_1, \ldots, \gamma_n)$ does not contain any true ground atom, there is exactly one grounded *default* atom prediction rule in $G_{r(x_1,\ldots,x_n)}$ whose antecedent is true, and all other rules have false antecedents and are trivially true.

From the MLN defining the likelihood component, we get

$$P(D|\{\Gamma\}) = \frac{\exp\left(\sum_{i \in F} \sum_{j \in G_i} w_i g_j(D)\right)}{Z} \tag{5}$$

where Z is the partition function (different from that of Equation 1); F is a set containing all atom prediction rules and the default atom prediction rule; G_i and w_i are respectively the set of groundings and weight of the ith rule in F; and $g_j(D) = 1$ if the jth ground rule in G_i is true and $g_j(D) = 0$ otherwise.

In the numerator of Equation 5, we sum over all grounded rules. We can rewrite the equation by iterating over ground atoms $r(x_1, \ldots, x_n)$, and summing over grounded rules that have $r(x_1, \ldots, x_n)$ as their consequents.

$$P(D|\{\Gamma\}) = \frac{\exp\left(\sum_{r(x_1,\ldots,x_n) \in D} \sum_{j \in G_{r(x_1,\ldots,x_n)}} w_j g_j(D)\right)}{Z} \tag{6}$$

where $G_{r(x_1,\ldots,x_n)}$ is a set containing groundings of the atom prediction rules and the single grounding of the default atom prediction rule that have ground atom $r(x_1, \ldots, x_n)$ as their consequents; and w_j is the weight of the jth rule in $G_{r(x_1,\ldots,x_n)}$,

In $G_{r(x_1,\ldots,x_n)}$, there is exactly one grounded rule whose antecedent is true. All other grounded rules have false antecedents, and are trivially true in all worlds. Such rules cancel themselves out in the numerator and denominator of Equation 6. Hence we only need to sum over grounded rules whose antecedents are true. We can write Equation 6 as

$$P(D|\{\Gamma\}) = \frac{\exp\left(\sum_{r \in R} \sum_{c_r \in C_r} \sum_{j \in F_{c_r}} w_{c_r} g_j(r_j(x_1, \ldots, x_n))\right)}{Z'} \tag{7}$$

where R is a set of predicates; C_r is a union of cluster combinations containing at least one true grounding of predicate r, and a default cluster combination containing only false groundings of r; F_{c_r} is a set of grounded rules with cluster combination c_r in their true antecedents and a grounding of r as their consequents; w_{c_r} is the weight of the atom predication rule or default atom predication rule that has c_r in its antecedent; $r_j(x_1, \ldots, x_n)$ is the ground atom appearing as the consequent of rule j; $g_j(r_j(x_1, \ldots, x_n)) = 1$ if $r_j(x_1, \ldots, x_n)$ is true; $g_j(r_j(x_1, \ldots, x_n)) = 0$ otherwise; and Z' is the partition function.

Because a ground atom $r(x_1, \ldots, x_n)$ is in exactly one cluster combination c_r, and appears in exactly one grounded rule with c_r in its the antecedent, we can factorize Z', and write Equation 7 as

$$\begin{aligned}
P(D|\{\Gamma\}) &= \frac{\prod_{r \in R} \prod_{c_r \in C_r} \prod_{j \in F_{c_r}} \exp\left(w_{c_r} g_j(r_j(x_1, \ldots, x_n))\right)}{\prod_{r \in R} \prod_{c_r \in C_r} \prod_{j \in F_{c_r}} \sum_{r_j(x_1,\ldots,x_n) \in \{0,1\}} \exp\left(w_{c_r} g_j(r_j(x_1, \ldots, x_n))\right)} \\
&= \prod_{r \in R} \prod_{c_r \in C_r} \prod_{j \in F_{c_r}} \frac{\exp\left(w_{c_r} g_j(r_j(x_1, \ldots, x_n))\right)}{\sum_{r_j(x_1,\ldots,x_n) \in \{0,1\}} \exp\left(w_{c_r} g_j(r_j(x_1, \ldots, x_n))\right)} \\
&= \prod_{r \in R} \prod_{c_r \in C_r} \prod_{j \in F_{c_r}} \frac{\exp\left(w_{c_r} g_j(r_j(x_1, \ldots, x_n))\right)}{1 + \exp(w_{c_r})} \\
&= \prod_{r \in R} \prod_{c_r \in C_r} \left(\frac{\exp(w_{c_r})}{1 + \exp(w_{c_r})}\right)^{t_{c_r}} \left(\frac{1}{1 + \exp(w_{c_r})}\right)^{f_{c_r}}
\end{aligned} \tag{8}$$

where t_{c_r} and f_{c_r} are respectively the number of true and false ground $r(x_1, \ldots, x_n)$ atoms in cluster combination c_r.

By differentiating Equation 8 with respect to w_{c_r}, setting the derivative to 0, and solving for w_{c_r}, we find that the resulting equation is maximized when $w_{c_r} = \log(t_{c_r}/f_{c_r})$. Substituting $w_{c_r} = \log(t_{c_r}/f_{c_r})$ in Equations 8, and taking logs, we get

$$\log P(D|\{\Gamma\}) = \sum_{r\in R}\sum_{c_r\in C_r} t_{c_r}\log\left(\frac{t_{c_r}}{t_{c_r}+f_{c_r}}\right) + f_{c_r}\log\left(\frac{f_{c_r}}{t_{c_r}+f_{c_r}}\right). \tag{9}$$

Adding smoothing parameters α_r and β_r, we get

$$\log P(D|\{\Gamma\}) = \sum_{r\in R}\sum_{c_r\in C_r} t_{c_r}\log\left(\frac{t_{c_r}+\alpha_r}{t_{c_r}+f_{c_r}+\alpha_r+\beta_r}\right) + f_{c_r}\log\left(\frac{f_{c_r}+\beta_r}{t_{c_r}+f_{c_r}+\alpha_r+\beta_r}\right). \tag{10}$$

(In our experiments, we set $\alpha_r + \beta_r = 10$ and $\frac{\alpha_r}{\alpha_r+\beta_r}$ to the fraction of true groundings of r.) Separating the default cluster combination c'_r containing only false groundings of r from the set of cluster combinations C_r^+ containing at least one true grounding of r, we obtain

$$\log P(D|\{\Gamma\}) = \sum_{r\in R}\left[f_{c'_r}\log\left(\frac{f_{c'_r}+\beta_r}{f_{c'_r}+\alpha_r+\beta_r}\right) + \sum_{c_r\in C_r^+} t_{c_r}\log\left(\frac{t_{c_r}+\alpha_r}{t_{c_r}+f_{c_r}+\alpha_r+\beta_r}\right) + f_{c_r}\log\left(\frac{f_{c_r}+\beta_r}{t_{c_r}+f_{c_r}+\alpha_r+\beta_r}\right)\right]. \tag{11}$$

$\log P(\{\Gamma\}|D) = \log P(\{\Gamma\}) + \log P(D|\{\Gamma\}) + K'$ where K' is a constant. Using the values of the prior and likelihood, we get

$$\log(P(\{\Gamma\}|D) = \begin{cases} -\infty & \text{if there is a symbol that is not in exactly one cluster} \\ \sum_{r\in R}\left[f_{c'_r}\log\left(\frac{f_{c'_r}+\beta_r}{f_{c'_r}+\alpha_r+\beta_r}\right) + \sum_{c_r\in C_r^+} t_{c_r}\log\left(\frac{t_{c_r}+\alpha_r}{t_{c_r}+f_{c_r}+\alpha_r+\beta_r}\right) + f_{c_r}\log\left(\frac{f_{c_r}+\beta_r}{t_{c_r}+f_{c_r}+\alpha_r+\beta_r}\right)\right] - \lambda m_{\{\Gamma\}} + K'' & \text{otherwise} \end{cases}$$

where $K'' = K + K'$ is a constant. (When comparing candidate cluster assignments to find the one with the best log-posterior, we can ignore K'' because it is a constant.)