Supplementary Material for
Early Gains Matter: A Case for Preferring Generative over Discriminative Crowdsourcing Models

Paul Felt, Eric Ringger, Kevin Seppi, Kevin Black, Robbie Haertel
Brigham Young University
Provo, UT 84602, USA

Appendix A: Derivation of mean-field variational updates

Introduction. Here we derive mean field variational updates for MOMRESP. Although this derivation is largely a mechanical exercise, it is our belief that there is a contingent of crowdsourcing practitioners whose background is more practical than theoretical and who may appreciate seeing the mechanics of mean-field variational inference presented in a high level of detail for a model they are familiar with. The updates for LOGRESP involve so much overlap with those for MOMRESP that we leave them as an exercise for the interested reader.

Problem Setup. Given some posterior distribution p^* over variables $\boldsymbol{z}$, our goal is to search among some family of simpler approximate tractable models Q and identify the $q(\boldsymbol{z}) \in Q$ that most closely resembles $p^*(\boldsymbol{z})$. If we choose Q to be the set of fully factored models such that $q(\boldsymbol{z}) = \prod q_i(z_i)$ (the mean-field assumption) then the q that minimizes KL divergence $KL(q||p^*)$ can be shown to have the following form:

$$\log q_i(z_i) = E_{q_{-i}}[\log \tilde{p}] + \text{const} \tag{8}$$

where $\tilde{p}$ is the unnormalized posterior and $q_{-i} = \{q_{i'} : i' \neq i\}$. For a derivation of this property, see Chapter 21.3 of Kevin Murphy's excellent reference, *Machine Learning: A Probabilistic Perpective*.

Notation. We adopt slightly different notation here than is used in the paper. Variables that represent non-scalars (e.g., vectors or matrices) after resolving subscripts are bolded. That is, we might use $\boldsymbol{z}$ to denote a matrix, $\boldsymbol{z}_i$ a vector, and z_{ij} a scalar. Furthermore, in order to clearly distinguish among variational distributions each is given a unique distributional name rather than simply being distinguished by its arguments as in the main paper. Sums are simplified by providing only a single subscript. $\sum_i$ is short for $\sum_{i \in N}$, $\sum_j$ is short for $\sum_{j \in J}$, and so on.

Mean-Field Variational Update Equations

The unnormalized posterior $\tilde{p}$ required by Equation 8 is proportional to the full joint. Therefore we begin by writing out the full (unnormalized) joint according to MOM-RESP by starting with Equation 3 from the main paper, plugging in distributional forms, and then simplifying by omitting constants and combining terms from conjugate distributions:

$$\tilde{p}(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y}|\boldsymbol{x},\boldsymbol{a}) \propto p(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y},\boldsymbol{x},\boldsymbol{a}) \tag{9}$$

$$\propto \left(\prod_k \theta_k^{b_k^{(\theta)}+n_k^{(\theta)}-1} \right) \left(\prod_j \prod_k \prod_{k'} \gamma_{jkk'}^{b_{jkk'}^{(\gamma)}+n_{jkk'}^{(\gamma)}-1} \right) \left(\prod_k \prod_f \phi_{kf}^{b_{kf}^{(\phi)}+n_{kf}^{(\phi)}-1} \right) \tag{10}$$

$$\cdot \left(\prod_i \prod_j \frac{|\boldsymbol{a}_{ij}|_1!}{\prod_k a_{ijk}!} \right) \left(\prod_i \frac{|\boldsymbol{x}_i|_1!}{\prod_f x_{if}!} \right)$$

where $n_k^{(\theta)} = \sum_i \mathbb{1}(y_i = k)$, $n_{jkk'}^{(\gamma)} = \sum_i a_{ijk'}\mathbb{1}(y_i = k)$, and $n_{kf}^{(\phi)} = \sum_i x_{if}\mathbb{1}(y_i = k)$. That is, $n_k^{(\theta)}$ is the number of instances labeled k, $n_{jkk'}^{(\gamma)}$ is the number of times that annotator j chose annotation k' on instances with true label k, and $n_{kf}^{(\phi)}$ is the number of times feature f occurs with instances having label k.

We next take the log of Equation 10 to get the unnormalized posterior $\tilde{p}$. Note that in this step the multinomial coefficients consituting the last two terms of Equation 10 are absorbed into the constant of proportionality because they are constant in the context of posterior inference of $p^*(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y}|\boldsymbol{x},\boldsymbol{a})$ where $\boldsymbol{a}$ and $\boldsymbol{x}$ are fixed and known.

$$\log \tilde{p} = \sum_k (b_k^{(\theta)}+n_k^{(\theta)}-1)\log\theta_k + \sum_j \sum_k \sum_{k'} (b_{jkk'}^{(\gamma)}\gamma_j + n_{jkk'}^{(\gamma)}-1)\log\gamma_{jkk'} \tag{11}$$

$$+ \sum_k \sum_f (b_{kf}^{(\phi)}+n_{kf}^{(\phi)}-1)\log\phi_{kf}$$

Let the approximate distribution q be fully factored:

$$q(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y}) = \pi(\theta)\prod_j\prod_k \nu(\gamma_{jk})\prod_k \lambda(\phi_k)\prod_i g(y_i) \tag{12}$$

We derive the mean-field update equation for each factor in turn. The update equation will lead us to a concrete functional form for each variational distribution with appropriate variational paramters. The derivations make heavy use of the following five properties of expectations:

$$E_{p(x,y)}[af(x)+bg(y)] = aE_{p(x,y)}[f(x)]+bE_{p(x,y)}[g(y)] \qquad \text{linearity} \tag{13}$$

$$E_{p(x,y)}[f(x)] = E_{p(x)}[f(x)] \qquad \text{marginal} \tag{14}$$

$$E_{p(x)}[\mathbb{1}(x=k)] = p(k) \qquad \text{delta} \tag{15}$$

$$E_{p(x,y)}[f(x)g(y)] = E_{p(x)}[f(x)]E_{p(y)}[g(y)]$$
$$\text{iff } p(x,y) = p(x)p(y) \qquad \text{independence} \tag{16}$$

$$E_{p(x)}[K] = K \qquad\qquad constant \qquad (17)$$

Mean-field update and functional form for $\pi(\theta)$

$$\log \pi(\theta) = E_{\mathbf{v},\mathbf{g},\boldsymbol{\lambda}}[\log \tilde{p}] \qquad (18)$$

$$= E_{\mathbf{v},\mathbf{g},\boldsymbol{\lambda}}\Big[\sum_k (b_k^{(\theta)} + n_k^{(\theta)} - 1)\log \theta_k\Big] + const \qquad (19)$$

$$= \sum_k E_{\mathbf{g}}[(b_k^{(\theta)} + n_k^{(\theta)} - 1)]\log \theta_k + const \qquad (20)$$

$$= \sum_k (b_k^{(\theta)} + E_{\mathbf{g}}[n_k^{(\theta)}] - 1)\log \theta_k + const \qquad (21)$$

$$= \sum_k \Big(b_k^{(\theta)} + \sum_i E_{g_i}[\mathbb{1}(y_i = k)] - 1\Big)\log \theta_k + const \qquad (22)$$

$$= \sum_k \Big(b_k^{(\theta)} + \sum_i g_i(k) - 1\Big)\log \theta_k + const \qquad (23)$$

$$\pi(\theta) \propto \prod_k \theta_k^{b_k^{(\theta)} + \sum_i g_i(k) - 1} \qquad (24)$$

$$= Dirichlet(\boldsymbol{\alpha}^{(\theta)}) \qquad (25)$$

where $\alpha^{(\theta_k)} = b_k^{(\theta)} + \sum_i g_i(k)$.

Explanation. Equation 18 instantiates Equation 8 for π. Equation 19 plugs in the functional form of the posterior from Equation 11 and applies the linearity of expectations (Equation 13) to distribute the expectation over the sum. Terms not involving θ are absorbed into a constant. Equation 20 again applies the linearity of expectation over addition and multiplication and then applies the marginal property of expectations (Equation 14) so that the expectation is with respect to only the variational distribution $\mathbf{g}$ needed to compute the expectation of $n_k^{(\theta)}$. Equation 21 applies the linearity of expectations and simplifies the expectations of constants (Equation 17). Equation 22 substitutes the definition of $n_k^{(\theta)}$ and applies the linearity and marginal properties of expectations. Equation 23 simplifies expectations applied to delta functions (Equation 15). We then recognize the kernel of a log Dirichlet distribution.

The remaining derivations follow very similar lines of reasoning, so we only offer explanations where they differ from the patterns seen above. In particular, $v_{jk}(\gamma_{jk})$ and $\lambda_k(\phi_k)$ are nearly identical so no explanation will be given.

Mean-field update and functional form for $v_{jk}(\gamma_{jk})$

$$\log v_{jk}(\gamma_{jk}) = E_{\mathbf{v}_{-jk},\mathbf{g},\boldsymbol{\lambda},\pi}[\log \tilde{p}]$$

$$= E_{\mathbf{v}_{-jk},\mathbf{g},\boldsymbol{\lambda},\pi}\Big[\sum_{j'}\sum_{k''}\sum_{k'}(b_{j'k''k'}^{(\gamma)} + n_{j'k''k'}^{(\gamma)} - 1)\log \gamma_{j'k''k'}\Big] + const$$

$$= \sum_{j'}\sum_{k''}\sum_{k'} E_{\mathbf{g}}[(b_{j'k''k'}^{(\gamma)} + n_{j'k''k'}^{(\gamma)} - 1)\log \gamma_{j'k''k'}] + const$$

$$= \sum_{k'}(b_{jkk'}^{(\gamma)} + E_{\mathbf{g}}[n_{jkk'}^{(\gamma)}] - 1)]\log \gamma_{jkk'} + const$$

$$= \sum_{k'} (b_{jkk'}^{(\gamma)} + \sum_i a_{ijk'} E_{g_i}[\mathbb{1}(y_i = k)] - 1) \log \gamma_{jkk'} + const$$

$$= \sum_{k'} (b_{jkk'}^{(\gamma)} + \sum_i a_{ijk'} g_i(k) - 1) \log \gamma_{jkk'} + const$$

$$\nu_{jk}(\gamma_{jk}) \propto \prod_{k'} \gamma_{jkk'}^{b_{jkk'}^{(\gamma)} + \sum_i a_{ijk'} g_i(k) - 1}$$

$$= Dirichlet(\boldsymbol{\alpha}_{jk}^{(\gamma)})$$

where $\alpha_{jkk'}^{(\gamma)} = b_{jkk'}^{(\gamma)} + \sum_i a_{ijk'} g_i(k)$.

Mean-field update and functional form for $\lambda_k(\phi_k)$

$$\log \lambda_k(\phi_k) = E_{\boldsymbol{\nu},\boldsymbol{g},\boldsymbol{\lambda}_{-k},\pi}[\log \tilde{p}] + const$$

$$= E_{\boldsymbol{\nu},\boldsymbol{g},\boldsymbol{\lambda}_{-k},\pi} \left[\sum_{k'} \sum_f (b_{k'f}^{(\phi)} + n_{k'f}^{(\phi)} - 1) \log \phi_{k'f} \right] + const$$

$$= E_{\boldsymbol{g},\boldsymbol{\lambda}_{-k}} \left[\sum_f (b_{kf}^{(\phi)} + n_{kf}^{(\phi)} - 1) \right] \log \phi_{kf} + const$$

$$= \sum_f (b_{kf}^{(\phi)} + E_{\boldsymbol{g}}[n_{kf}^{(\phi)}] - 1) \log \phi_{kf} + const$$

$$= \sum_f (b_{kf}^{(\phi)} + E_{\boldsymbol{g}}[\sum_i x_{if} \mathbb{1}(y_i = k)] - 1) \log \phi_{kf} + const$$

$$= \sum_f (b_{kf}^{(\phi)} + \sum_i x_{if} g_i(k) - 1) \log \phi_{kf} + const$$

$$\lambda_k(\phi_k) \propto \prod_f \phi_{kf}^{b_{kf}^{(\phi)} + \sum_i x_{if} g_i(k) - 1}$$

$$= Dirichlet(\boldsymbol{\alpha}_k^{(\phi)})$$

where $\alpha_{kf} = b_{kf}^{(\phi)} + \sum_i x_{if} g_i(k)$

Mean-field update and functional form for $g_i(y_i)$

$$\log g_i(y_i) = E_{\boldsymbol{\nu},\boldsymbol{g}_{-i},\boldsymbol{\lambda},\pi}[\log \tilde{p}] + const$$

$$= E_{\boldsymbol{\nu},\boldsymbol{g}_{-i},\boldsymbol{\lambda},\pi}[\sum_k (b_k^{(\theta)} + n_k^{(\theta)} - 1) \log \theta_k + \sum_j \sum_k \sum_{k'} (b_{jkk'}^{(\gamma)} + n_{jkk'}^{(\gamma)} - 1) \log \gamma_{jkk'}$$

$$+ \sum_k \sum_f (b_{kf}^{(\phi)} + n_{kf}^{(\phi)} - 1) \log \phi_{kf}] + const$$

$$= \sum_k E_{\boldsymbol{g}_{-i},\pi_k}[(b_k^{(\theta)} + n_k^{(\theta)} - 1) \log \theta_k] + \sum_j \sum_k \sum_{k'} E_{\boldsymbol{\nu}_{jk},\boldsymbol{g}_{-i}}[(b_{jkk'}^{(\gamma)} + n_{jkk'}^{(\gamma)} - 1) \log \gamma_{jkk'}]$$

$$+ \sum_{k'} \sum_f E_{\boldsymbol{g}_{-i},\lambda_k}[(b_{kf}^{(\phi)} + n_{kf}^{(\phi)} - 1) \log \phi_{kf}] + const$$

4

$$= \sum_k (b_k^{(\theta)} + E_{\boldsymbol{g}_{-i}}[n_k^{(\theta)}] - 1)E_{\pi_k}[\log\theta_k] + \sum_j \sum_k \sum_{k'} (b_{jkk'}^{(\gamma)} + E_{\boldsymbol{g}_{-i}}[n_{jkk'}^{(\gamma)}] - 1)E_{\nu_{jk}}[\log\gamma_{jkk'}]$$

$$(26)$$

$$+ \sum_k \sum_f (b_{kf}^{(\phi)} + E_{\boldsymbol{g}_{-i}}[n_{kf}^{(\phi)}] - 1)E_{\lambda_k}[\log\phi_{kf}] + const$$

$$= \sum_k (b_k^{(\theta)} + E_{\boldsymbol{g}_{-i}}[\sum_{i'}\mathbb{1}(y_{i'} = k)] - 1)E_{\pi_k}[\log\theta_k]$$

$$+ \sum_j \sum_k \sum_{k'} (b_{jkk'}^{(\gamma)} + E_{\boldsymbol{g}_{-i}}[\sum_{i'} a_{i'jk'}\mathbb{1}(y_{i'} = k)] - 1)E_{\nu_{jk}}[\log\gamma_{jkk'}]$$

$$+ \sum_k \sum_f E_{\boldsymbol{g}_{-i}}[\sum_{i'} x_{i'f}\mathbb{1}(y_{i'} = k)]E_{\lambda_k}[\log\phi_{kf}] + const$$

$$= \sum_k E_{\boldsymbol{g}_{-i}}[\sum_{i'}\mathbb{1}(y_{i'} = k)]E_{\pi_k}[\log\theta_k]$$

$$+ \sum_j \sum_k \sum_{k'} E_{\boldsymbol{g}_{-i}}[\sum_{i'} a_{i'jk'}\mathbb{1}(y_{i'} = k)]E_{\nu_{jk}}[\log\gamma_{jkk'}]$$

$$+ \sum_k \sum_f x_{if}\mathbb{1}(y_i = k)E_{\lambda_k}[\log\phi_{kf}] + const \qquad (27)$$

$$= \sum_k \mathbb{1}(y_i = k)E_{\pi_k}[\log\theta_k] + \sum_j \sum_k \sum_{k'} a_{ijk'}\mathbb{1}(y_i = k)E_{\nu_{jk}}[\log\gamma_{jkk'}]$$

$$+ \sum_k \sum_f x_{if}\mathbb{1}(y_i = k)E_{\lambda_k}[\log\phi_{kf}] + const$$

$$= \sum_k \mathbb{1}(y_i = k)\Big[E_{\pi_k}[\log\theta_k] + \sum_j \sum_{k'} a_{ijk'}E_{\nu_{jk}}[\log\gamma_{jkk'}] + \sum_f x_{if}E_{\lambda_k}[\log\phi_{kf}]\Big] + const$$

$$(28)$$

$$g_i(y_i) \propto \prod_k \exp\Big[E_{\pi_k}[\log\theta_k] + \sum_j \sum_{k'} a_{ijk'}E_{\nu_{jk}}[\log\gamma_{jkk'}] + \sum_f x_{if}E_{\lambda_k}[\log\phi_{kf}]\Big]^{\mathbb{1}(y_i=k)}$$

$$\propto \prod_k \alpha_{ik}^{(y)\mathbb{1}(y_i=k)}$$

$$= Categorical(\boldsymbol{\alpha}_i^{(y)})$$

where $\alpha_{ik}^{(y)} = E_{\pi_k}[\log\theta_k] + \sum_j \sum_{k'} a_{ijk'}E_{\nu_{jk}}[\log\gamma_{jkk'}] + \sum_f x_{if}E_{\lambda_k}[\log\phi_{kf}]$. As noted in the main paper, the expected value of a log term with respect to a Dirichlet distribution, such as $E_{\pi_k}[\log\theta_k]$, can be computed analytically as $\psi(\alpha_k^{(\theta)}) - \psi(\sum_{k'}\alpha_{k'}^{(\theta)})$, where ψ is the digamma function (and similarly for $E_{\nu_{jk}}[\log\gamma_{jkk'}]$ and $E_{\lambda_k}[\log\phi_{kf}]$).

Explanation. This derivation differs slightly in a couple of locations from what we have seen before. In Equation 26 we use the independence property of expectations (Equation 16) to separate expectations with recept to $\boldsymbol{g}$ from expectations with respect to approximate distributions $\boldsymbol{\pi}$, $\boldsymbol{\nu}$, and $\boldsymbol{\lambda}$. We can do this because all distributions in our approximate model are independent of one another. Equation 27 distributes Dirichlet expectation terms such as $E_{\nu_{jk}}[\log\gamma_{jkk'}]$ and then absorbs terms that do not involve y_i into the constant. Equation 28 combines sums over k and then factors out the common multiplier $\mathbb{1}(y_i = k)$.

Lower Bound

The mean-field variational updates we derived above are designed to minimize the objective $KL(q||p^*)$. Computing values of this objective is highly useful during optimization. Tracking the objective's rate of change after each iteration of the optimization algorithm allows us to assess convergence. Second, the objective function provides a powerful debugging tool: if the objective does not improve after an update (modulo floating point noise after convergence) then there is a bug in that particular update. Unfortunately, directly computing $KL(q||p^*)$ is intractable because it involves evaluating the intractable denominator of $p^*(\theta,\gamma,\phi,y|x,a) = \frac{p(\theta,\gamma,\phi,y,x,a)}{p(a,x)} = \frac{p(\theta,\gamma,\phi,y,x,a)}{\int p(\theta,\gamma,\phi,y,x,a)d\theta,\gamma,\phi,y}$. However, because $KL(q||p^*) = KL(q||p) + p(x,a)$ and $p(x,a)$ is constant, minimizing $KL(q||p^*)$ is equivalent to minimizing $KL(q||p)$, which is tractable to evaluate. Finally, it is typical to treat mean-field optimization as a maximization rather than a minimization problem. We therefore use $-KL(q||p)$ as the objective function for the purposes of tracking convergence and debugging. $-KL(q||p)$ is commonly referred to as a lower bound because it is a lower bound on the log marginal likelihood of the model $p(x,a)$, giving it additional theoretical interest. However, its immediate practical value is as the objective function being optimized by mean-field variational updates.

We first break $-KL(q||p)$ into two manageable parts: entropy $H(q)$, and cross entropy $H(q|p)$:

$$
\begin{aligned}
-KL(q||p) &= \int q(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y})\log\frac{p(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y},\boldsymbol{x},\boldsymbol{a})}{q(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y})}d\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y} \\
&= E_q[\log p(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y},\boldsymbol{x},\boldsymbol{a})] - E_q[\log q(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y})] \\
&= -H(q|p) + H(q)
\end{aligned}
$$

Next we simplify each term in the bound separately. An important difference between these derivations and those we did for the mean-field updates is that we must evaluate this function exactly and therefore cannot drop any constants. Therefore when we plug in the log joint distribution we use the full, unsimplified form of each distribution rather than the simplified form in Equation 11. The reasoning behind each step here is similar enough to those in the update derivations above that no additional explanation is provided.

The first term of the lower bound $-H(q|p)$ is

$$
\begin{aligned}
&E_q[\log p(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\phi},\boldsymbol{y},\boldsymbol{x},\boldsymbol{a})] \\
&= E_q\Big[\log p(\boldsymbol{\theta}) + \sum_j\sum_k \log p(\boldsymbol{\gamma}_{jk}) + \sum_k \log p(\boldsymbol{\phi}_k) + \sum_i \log p(y_i|\boldsymbol{\theta}) \\
&\quad + \sum_i\sum_j \log p(\boldsymbol{a}_{ij}|y_i,\boldsymbol{\gamma}_{jy_i}) + \sum_i \log p(\boldsymbol{x}_i|\boldsymbol{\phi}_{y_i})\Big] \\
&= E_q\Big[-\log B(\boldsymbol{b}^{(\theta)}) + \sum_k (b_k^{(\theta)} - 1)\log\theta_k \\
&\quad + \sum_j\sum_k -\log B(\boldsymbol{b}_{jk}^{(\gamma)}) + \sum_{k'} (b_{jkk'}^{(\gamma)}\gamma_{jk} - 1)\log\gamma_{jkk'}
\end{aligned}
$$

$$+\sum_k -\log B(\boldsymbol{b}_k^{(\phi)}) + \sum_f (b_{kf}^{(\phi)}\phi_k - 1)\log \phi_{kf}$$

$$+\sum_i \log \theta_{y_i}$$

$$+\sum_i \sum_j \left(\log |\boldsymbol{a}_{ij}|_1! - \sum_k \log a_{ijk}!\right) + \sum_k a_{ijk}\log \gamma_{jy_i}$$

$$+\sum_i \left(\log |\boldsymbol{x}_i|_1! - \sum_f \log x_{if}!\right) + \sum_f x_{if}\log \phi_{y_{if}}\Big]$$

$$= E_q\Big[-\log B(\boldsymbol{b}^{(\theta)}) + \sum_k (b_k^{(\theta)} - 1)\log \theta_k$$

$$+\sum_j \sum_k -\log B(\boldsymbol{b}_{jk}^{(\gamma)}) + \sum_{k'} (b_{jkk'}^{(\gamma)}\gamma_{jk} - 1)\log \gamma_{jkk'}$$

$$+\sum_k -\log B(\boldsymbol{b}_k^{(\phi)}) + \sum_f (b_{kf}^{(\phi)}\phi_k - 1)\log \phi_{kf}$$

$$+\sum_k n_k^{(\theta)}$$

$$+\sum_i \sum_j \left(\log |\boldsymbol{a}_{ij}|_1! - \sum_k \log a_{ijk}!\right) + \sum_i \sum_j \sum_{k'} n_{jkk'}^{(\gamma)}\log \gamma_{jkk'}$$

$$+\sum_i \left(\log |\boldsymbol{x}_i|_1! - \sum_f \log x_{if}!\right) + \sum_k \sum_f n_{kf}^{(\phi)}\log \phi_{kf}\Big]$$

$$= E_q\Big[-\log B(\boldsymbol{b}^{(\theta)}) + \sum_k (b_k^{(\theta)} + n_k^{(\theta)} - 1)\log \theta_k$$

$$+\sum_j \sum_k -\log B(\boldsymbol{b}_{jk}^{(\gamma)}) + \sum_{k'} (b_{jkk'}^{(\gamma)} + n_{jkk'}^{(\gamma)} - 1)\log \gamma_{jkk'}$$

$$+\sum_k -\log B(\boldsymbol{b}_k^{(\phi)}) + \sum_f (b_{kf}^{(\phi)} + n_{kf}^{(\phi)} - 1)\log \phi_{kf}$$

$$+\sum_i \sum_j \left(\log |\boldsymbol{a}_{ij}|_1! - \sum_k \log a_{ijk}!\right) + \sum_i \left(\log |\boldsymbol{x}_i|_1! - \sum_f \log x_{if}!\right)\Big]$$

$$= -\log B(\boldsymbol{b}^{(\theta)}) + \sum_k E_{\pi,\boldsymbol{g}}[(b_k^{(\theta)} + n_k^{(\theta)} - 1)\log \theta_k]$$

$$+\sum_j \sum_k -\log B(\boldsymbol{b}_{jk}^{(\gamma)}) + \sum_{k'} E_{v_{jk},\boldsymbol{g}}[(b_{jkk'}^{(\gamma)} + n_{jkk'}^{(\gamma)} - 1)\log \gamma_{jkk'}]$$

$$+\sum_k -\log B(\boldsymbol{b}_k^{(\phi)}) + \sum_f E_{\lambda_k,\boldsymbol{h}}[(b_{kf}^{(\phi)} + n_{kf}^{(\phi)} - 1)\log \phi_{kf}]$$

$$+\sum_i \sum_j \left(\log |\boldsymbol{a}_{ij}|_1! - \sum_k \log a_{ijk}!\right) + \sum_i \left(\log |\boldsymbol{x}_i|_1! - \sum_f \log x_{if}!\right)$$

$$= -\log B(\boldsymbol{b}^{(\theta)}) + \sum_k (b_k^{(\theta)} + \sum_i g_i(k) - 1)E_\pi[\log \theta_k]$$

$$+\sum_j \sum_k -\log B(\boldsymbol{b}_{jk}^{(\gamma)}) + \sum_{k'} (b_{jkk'}^{(\gamma)} + \sum_i a_{ijk'}g_i(k) - 1)E_{v_{jk}}[\log \gamma_{jkk'}]$$

$$+\sum_k -\log B(\boldsymbol{b}_k^{(\phi)}) + \sum_f (b_{kf}^{(\phi)} + \sum_i x_{if} h_i(k) - 1) E_{\lambda_k}[\log \phi_{kf}]$$

$$+\sum_i \sum_j \left(\log |\boldsymbol{a}_{ij}|_1! - \sum_k \log a_{ijk}!\right) + \sum_i \left(\log |\boldsymbol{x}_i|_1! - \sum_f \log x_{if}!\right)$$

where $B(\cdot)$ is the multivariate Beta function that normalizes a Dirichlet distribution.

The second term of the lower bound $H(q)$ is

$$E_q[\log q(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\phi}, \boldsymbol{y})]$$

$$= E_q\left[\log \pi(\boldsymbol{\theta}) + \sum_j \sum_k \nu_{jk}(\boldsymbol{\gamma}_{jk}) + \sum_k \lambda_k(\boldsymbol{\phi}_k) + \sum_i g_i(y_i)\right]$$

$$= E_q\left[-B(\boldsymbol{\alpha}^{(\theta)}) + \sum_k (\alpha_k^{(\theta)} - 1)\log \theta_k\right.$$

$$+\sum_j \sum_k -\log B(\boldsymbol{\alpha}_{jk}^{(\gamma)}) + \sum_{k'} (\alpha_{jk}^{(\gamma)} - 1)\log \gamma_{jkk'}$$

$$+\sum_k -\log B(\boldsymbol{\alpha}_k^{(\phi)}) + \sum_f (\alpha_{kf}^{(\phi)} - 1)\log \phi_{kf}$$

$$\left.+\sum_i \log \alpha_{y_i}^{(y)}\right]$$

$$= -B(\boldsymbol{\alpha}^{(\theta)}) + \sum_k (\alpha_k^{(\theta)} - 1) E_\pi[\log \theta_k]$$

$$+\sum_j \sum_k -\log B(\boldsymbol{\alpha}_{jk}^{(\gamma)}) + \sum_{k'} (\alpha_{jk}^{(\gamma)} - 1) E_{\nu_{jk}}[\log \gamma_{jkk'}]$$

$$+\sum_k -\log B(\boldsymbol{\alpha}_k^{(\phi)}) + \sum_f (\alpha_{kf}^{(\phi)} - 1) E_{\lambda_k}[\log \phi_{kf}]$$

$$+\sum_i E_{g_i}[\log \alpha_{y_i}^{(y)}]$$

$$= -B(\boldsymbol{\alpha}^{(\theta)}) + \sum_k (\alpha_k^{(\theta)} - 1) E_\pi[\log \theta_k]$$

$$+\sum_j \sum_k -\log B(\boldsymbol{\alpha}_{jk}^{(\gamma)}) + \sum_{k'} (\alpha_{jk}^{(\gamma)} - 1) E_{\nu_{jk}}[\log \gamma_{jkk'}]$$

$$+\sum_k -\log B(\boldsymbol{\alpha}_k^{(\phi)}) + \sum_f (\alpha_{kf}^{(\phi)} - 1) E_{\lambda_k}[\log \phi_{kf}]$$

$$+\sum_i \sum_k \alpha_k^{(y)} \log \alpha_k^{(y)}$$