Supplementary Material for "Estimating Accuracy from Unlabeled Data: A Bayesian Approach"

Emmanouil Antonios Platanios E.A.PLATANIOS@CS.CMU.EDU
Avinava Dubey AKDUBEY@CS.CMU.EDU
Tom Mitchell TOM.MITCHELL@CS.CMU.EDU
Carnegie Mellon University, Pittsburgh, PA 15213, USA

1. Gibbs Sampling Equations

In this section we provide the equations necessary to perform Gibbs sampling over the models defined in our paper.

1.1. Coupled Bayesian Error Estimation

The conditional probabilities we use during the first, collapsed sampling phase are as follows:

$$P(p^d \mid \cdot) = \text{Beta}(\alpha_p + \sigma_{\boldsymbol{\ell}}^d, \beta_p + S^d - \sigma_{\boldsymbol{\ell}}^d), \tag{1}$$

$$P(\ell_i^d \mid \cdot) \propto (p^d)^{\ell_i^d}(1-p^d)^{1-\ell_i^d}\mathcal{B}(A_i^d, B_i^d), \tag{2}$$

$$P(z^d = k \mid \cdot) \propto \begin{cases} Z_k^d \mathcal{P}_k^d & \text{, if } Z_k^d > 0, \\ \alpha \mathcal{P}_{\text{new}}^d & \text{, otherwise,} \end{cases} \tag{3}$$

where:

$$\sigma_{\boldsymbol{\ell}}^d = \sum_{i=1}^{S^d} \ell_i^d, \tag{4}$$

$$A_i^d = \alpha_{z^d}^d + \sum_{\hat{i}=1}^{N}\sum_{j=1}^{M} \mathbb{1}_{\{\hat{f}_{\hat{i}j}^d \neq \ell_{\hat{i}}^d\}}, \tag{5}$$

$$B_i^d = \beta_{z^d}^d + \sum_{\hat{i}=1}^{N}\sum_{j=1}^{M} \mathbb{1}_{\{\hat{f}_{\hat{i}j}^d = \ell_{\hat{i}}^d\}}, \tag{6}$$

$$\alpha_k^d = \alpha_e + \sum_{\substack{\hat{d}=1 \\ \hat{d}\neq d}}^{D} \mathbb{1}_{\{z^{\hat{d}}=k\}}\sigma^{\hat{d}}, \tag{7}$$

$$\beta_k^d = \beta_e + \sum_{\substack{\hat{d}=1 \\ \hat{d}\neq d}}^{D} \mathbb{1}_{\{z^{\hat{d}}=k\}}(S^{\hat{d}} - \sigma^{\hat{d}}), \tag{8}$$

$$\sigma_j^d = \sum_{i=1}^{S^d} \mathbb{1}_{\{\hat{f}_{ij}^d \neq \ell_i^d\}}, \quad \sigma^d = \sum_{j=1} \sigma_j^d, \tag{9}$$

$$Z_k^d = \sum_{\substack{\hat{d}=1 \\ \hat{d}\neq d}}^{D} \mathbb{1}_{\{z^{\hat{d}}=k\}}, \tag{10}$$

$$\mathcal{P}_k^d = \frac{\mathcal{B}(\alpha_k^d + \sigma^d, \beta_k^d + S^d - \sigma^d)}{\mathcal{B}(\alpha_e + \alpha_k^d, \beta_e + \beta_k^d)}, \tag{11}$$

$$\mathcal{P}_{\text{new}}^d = \frac{\mathcal{B}(\alpha_e + \sigma^d, \beta_e + S^d - \sigma^d)}{\mathcal{B}(\alpha_e, \beta_e)}, \tag{12}$$

$\mathbb{1}_{\{\cdot\}}$ evaluates to one if its subscript's argument statement is true and to zero otherwise, and $\mathcal{B}(\cdot,\cdot)$ is the Beta function.

After that phase, we start sampling the error rates along with the rest of the variables and store the samples we obtain. During that second phase, we use the following conditional probabilities:

$$P(p^d \mid \cdot) = \text{Beta}(\alpha_p + \sigma_{\boldsymbol{\ell}}^d, \beta_p + S^d - \sigma_{\boldsymbol{\ell}}^d), \tag{13}$$

$$P(\ell_i^d \mid \cdot) \propto (p^d)^{\ell_i^d}(1-p^d)^{1-\ell_i^d}\pi_i^d \tag{14}$$

$$P(z^d = k \mid \cdot) \propto \begin{cases} Z_k^d \mathcal{R}_k^d & \text{, if } Z_k^d > 0, \\ \alpha \mathcal{P}_{\text{new}}^d & \text{, otherwise,} \end{cases} \tag{15}$$

$$P([\phi_k]_j \mid \cdot) = \text{Beta}(\Phi_j^\alpha, \Phi_j^\beta), \tag{16}$$

where:

$$\pi_i^d = \prod_{j=1}^{N} (e_j^d)^{\mathbb{1}_{\{\hat{f}_{ij}^d \neq l_i^d\}}}(1-e_j^d)^{\mathbb{1}_{\{\hat{f}_{ij}^d = l_i^d\}}}, \tag{17}$$

$$\mathcal{R}_k^d = \prod_{j=1}^{N} (e_j^d)^{\sigma_j^d}(1-e_j^d)^{S^d - \sigma_j^d}, \tag{18}$$

$$\Phi_j^\alpha = \alpha_e + \sum_{d=1}^{D} \mathbb{1}_{\{z^d=k\}}\sigma_j^d, \tag{19}$$

$$\Phi_j^\beta = \beta_e + \sum_{d=1}^{D} \mathbb{1}_{\{z^d=k\}}(S^d - \sigma_j^d). \tag{20}$$

In the case of missing data the conditional probability of the simple error estimation model can be used, which is provided in our paper.

1.2. Hierarchical Coupled Bayesian Error Estimation

The conditional probabilities we use during the first, collapsed sampling phase are as follows:

$$P(p^d \mid \cdot) = \text{Beta}(\alpha_p + \sigma^d_{\boldsymbol{\ell}}, \beta_p + S^d - \sigma^d_{\boldsymbol{\ell}}), \quad (21)$$

$$P(\ell^d_i \mid \cdot) \propto (p^d)^{\ell^d_i}(1-p^d)^{1-\ell^d_i} L^d_i, \quad (22)$$

$$P(z^d_j = t \mid k^d_t = k, \cdot) \propto Z^d_{kt}\mathcal{P}^d_{jk}, \quad (23)$$

$$P(k^d_t = k \mid \cdot) \propto Z^d_k \mathcal{P}^d_{kt}, \quad (24)$$

where:

$$L^d_i = \prod_{k=1}^{K} \mathcal{B}(A^d_{ik}, B^d_{ik}), \quad (25)$$

$$A^d_{ik} = \alpha^d_k + \sum_{j=1}^{M}\sum_{\hat{i}=1}^{N} \mathbb{1}_{\{k^d_{z^d_{\hat{j}}}=k\}} \mathbb{1}_{\{\hat{f}^d_{\hat{i}j} \neq \ell^d_{\hat{i}}\}}, \quad (26)$$

$$B^d_{ik} = \beta^d_k + \sum_{\hat{i}=1}^{N}\sum_{j=1}^{M} \mathbb{1}_{\{k^d_{z^d_{\hat{j}}}=k\}} \mathbb{1}_{\{\hat{f}^d_{\hat{i}j} = \ell^d_{\hat{i}}\}}, \quad (27)$$

$$\alpha^d_k = \alpha_e + \sum_{\substack{\hat{d}=1\\ \hat{d}\neq d}}^{D}\sum_{\hat{j}=1}^{N} \mathbb{1}_{\{k^{\hat{d}}_{z^{\hat{d}}_{\hat{j}}}=k\}} \sigma^{\hat{d}}_{\hat{j}}, \quad (28)$$

$$\beta^d_k = \beta_e + + \sum_{\substack{\hat{d}=1\\ \hat{d}\neq d}}^{D}\sum_{\hat{j}=1}^{N} \mathbb{1}_{\{k^{\hat{d}}_{z^{\hat{d}}_{\hat{j}}}=k\}} (S^{\hat{d}} - \sigma^{\hat{d}}_{\hat{j}}), \quad (29)$$

$$\alpha^d_{jk} = \alpha^d_k + \sum_{\substack{\hat{j}=1\\ \hat{j}\neq j}}^{N} \mathbb{1}_{\{k^d_{z^d_{\hat{j}}}=k\}} \sigma^d_{\hat{j}}, \quad (30)$$

$$\beta^d_{jk} = \beta^d_k + \sum_{\substack{\hat{j}=1\\ \hat{j}\neq j}}^{N} \mathbb{1}_{\{k^d_{z^d_{\hat{j}}}=k\}} (S^d - \sigma^d_{\hat{j}}), \quad (31)$$

$$Z^d_{kt} = \begin{cases} \sum_{\substack{\hat{j}=1\\ \hat{j}\neq j}}^{N} \mathbb{1}_{\{z^d_{\hat{j}}=t\}} & \text{, if } t \text{ occupied,} \\ \alpha \sum_{\substack{\hat{d}=1\\ \hat{d}\neq d}}^{D} \sum_{\hat{t}} \mathbb{1}_{\{k^{\hat{d}}_{\hat{t}}=k\}} & \text{, if } t \text{ unoccupied and } k \text{ exists,} \\ \alpha\gamma & \text{, if } t \text{ unoccupied and } k \text{ is new,} \end{cases} \quad (32)$$

$$\mathcal{P}^d_{jk} = \frac{\mathcal{B}(\alpha^d_{jk} + \sigma^d_j, \beta^d_{jk} + S^d - \sigma^d_j)}{\mathcal{B}(\alpha^d_{jk}, \beta^d_{jk})}, \quad (33)$$

$$Z^d_k = \begin{cases} \sum_{\substack{\hat{d}=1\\ \hat{d}\neq d}}^{D} \sum_{\hat{t}} \mathbb{1}_{\{k^{\hat{d}}_{\hat{t}}=k\}} & \text{, if } k \text{ exists,} \\ \gamma & \text{, if } k \text{ is new,} \end{cases} \quad (34)$$

where, noting that our previous definitions for α^d_{jk} and β^d_{jk} can also apply to sets over functions, $\boldsymbol{j}$, in the following way:

$$\alpha^d_{\boldsymbol{j}k} = \alpha^d_k + \sum_{\substack{\hat{j}=1\\ \hat{j}\notin \boldsymbol{j}}}^{N} \mathbb{1}_{\{k^d_{z^d_{\hat{j}}}=k\}} \sigma^d_{\hat{j}}, \quad (35)$$

$$\beta^d_{\boldsymbol{j}k} = \beta^d_k + \sum_{\substack{\hat{j}=1\\ \hat{j}\notin \boldsymbol{j}}}^{N} \mathbb{1}_{\{k^d_{z^d_{\hat{j}}}=k\}} (S^d - \sigma^d_{\hat{j}}), \quad (36)$$

we have that:

$$\mathcal{P}^d_{kt} = \frac{\mathcal{B}(\alpha^d_{\boldsymbol{j}k} + \sum_{j\in\boldsymbol{j}} \sigma^d_j, \beta^d_{\boldsymbol{j}k} + \sum_{j\in\boldsymbol{j}} (S^d - \sigma^d_j)}{\mathcal{B}(\alpha^d_{\boldsymbol{j}k}, \beta^d_{\boldsymbol{j}k})}, \quad (37)$$

where $\boldsymbol{j} = \{j : z^d_j = t\}$.

After that phase, we start sampling the error rates along with the rest of the variables and store the samples we obtain. During that second phase, we use the following conditional probabilities:

$$P(p^d \mid \cdot) = \text{Beta}(\alpha_p + \sigma^d_{\boldsymbol{\ell}}, \beta_p + S^d - \sigma^d_{\boldsymbol{\ell}}), \quad (38)$$

$$P(\ell^d_i \mid \cdot) \propto (p^d)^{\ell^d_i}(1-p^d)^{1-\ell^d_i} \pi^d_i \quad (39)$$

$$P(z^d_j = t \mid k^d_t = k, \cdot) \propto Z^d_{kt}\mathcal{R}^d_{jk}, \quad (40)$$

$$P(\phi_k \mid \cdot) = \text{Beta}(\Phi^\alpha, \Phi^\beta), \quad (41)$$

$$P(k^d_t = k \mid \cdot) \propto Z^d_k \mathcal{R}^d_{kt}, \quad (42)$$

where:

$$\pi^d_i = \prod_{j=1}^{N} (e^d_j)^{\mathbb{1}_{\{\hat{f}^d_{ij} \neq l^d_i\}}} (1-e^d_j)^{\mathbb{1}_{\{\hat{f}^d_{ij} = l^d_i\}}}, \quad (43)$$

$$\mathcal{R}^d_{jk} = (e^d_j)^{\sigma^d_j}(1-e^d_j)^{S^d - \sigma^d_j}, \quad (44)$$

$$\mathcal{R}^d_{kt} = \prod_{j\in\boldsymbol{j}} (e^d_j)^{\sigma^d_j}(1-e^d_j)^{S^d - \sigma^d_j}, \quad (45)$$

$$\Phi^\alpha = \alpha_e + \sum_{d=1}^{D}\sum_{j=1}^{N} \mathbb{1}_{\{k^d_{z^d_j}=k\}} \sigma^d_j, \quad (46)$$

$$\Phi^\beta = \beta_e + \sum_{d=1}^{D}\sum_{j=1}^{N} \mathbb{1}_{\{k^d_{z^d_j}=k\}} (S^d - \sigma^d_j). \quad (47)$$

where $\boldsymbol{j} = \{j : z^d_j = t\}$. In the case of missing data the conditional probability of the simple error estimation model can be used, which is provided in our paper.