Department of Statistics and Data Science at SUSTech

MAT7035: Computational Statistics

# Tutorial 11: Exact IBF Sampling and IBF Sampler

## E. The exact IBF sampling

(a) Assumptions:

- $Z$ is **a discrete r.v. of low dimension**;
- The complete-data posterior distribution $p(\theta|Y_{\rm obs}, Z)$ and the conditional predictive distribution $f(Z|Y_{\rm obs}, \theta)$ are available;
- $\mathcal{S}_{(\theta,Z|Y_{\rm obs})} = \mathcal{S}_{(\theta|Y_{\rm obs})} \times \mathcal{S}_{(Z|Y_{\rm obs})}$.

(b) Goal: To obtain i.i.d. samples from the observed posterior distribution $p(\theta|Y_{\rm obs})$.

(c) Algorithm:

- Identify $\mathcal{S}_{(Z|Y_{\rm obs})} = \{z_1, \dots, z_K\}$ from $f(Z|Y_{\rm obs}, \theta)$, and use the sampling IBF to find

$$p_k = \frac{q_k(\theta_0)}{\sum_{k'=1}^K q_{k'}(\theta_0)}, \quad k = 1, \dots, K,$$

where

$$q_k(\theta_0) = \frac{\Pr\{Z = z_k|Y_{\rm obs}, \theta_0\}}{p(\theta_0|Y_{\rm obs}, z_k)}, \quad k = 1, \dots, K.$$

- Generate i.i.d. samples $\{Z^{(\ell)}\}_{\ell=1}^L$ from $f(Z|Y_{\rm obs})$ with probabilities $\{p_k\}_1^K$ on $\{z_k\}_1^K$;
- Generate $\theta^{(\ell)} \sim p(\theta|Y_{\rm obs}, Z^{(\ell)})$ for $\ell = 1, \dots, L$, then $\{\theta^{(\ell)}\}_{\ell=1}^L$ are i.i.d. samples from the target density $p(\theta|Y_{\rm obs})$.

## F. The IBF sampler

(a) Assumptions:

- Both $p(\theta|Y_{\rm obs}, Z)$ and $f(Z|Y_{\rm obs}, \theta)$ are available;
- The posterior mode $\tilde{\theta}$ has already been obtained via an EM algorithm;

- $\mathcal{S}_{(\theta,Z|Y_{\rm obs})} = \mathcal{S}_{(\theta|Y_{\rm obs})} \times \mathcal{S}_{(Z|Y_{\rm obs})}$.

(b) Goal: To generate i.i.d. samples $\{\theta^{(i)}\}_{i=1}^{I}$ from the observed posterior distribution $p(\theta|Y_{\rm obs})$.

(c) Algorithm:

Based on $f(Z|Y_{\rm obs},\theta)$ and $p(\theta|Y_{\rm obs},Z)$, calculate the posterior mode $\tilde{\theta}$ of $p(\theta|Y_{\rm obs})$ via an EM algorithm and set $\theta_0 = \tilde{\theta}$;

- Draw $J$ i.i.d. samples $\{Z^{(j)}\}_{j=1}^{J}$ from $f(Z|Y_{\rm obs},\theta_0)$ and calculate the weights

$$\omega_j = \frac{p^{-1}\left(\theta_0|Y_{\rm obs}, Z^{(j)}\right)}{\sum_{k=1}^{J} p^{-1}\left(\theta_0|Y_{\rm obs}, Z^{(k)}\right)}, \qquad j = 1,\ldots,J;$$

- Choose a subset $\{Z^{(k_i)}\}_{i=1}^{I}$ with size $I(< J)$ from $\{Z^{(j)}\}_{j=1}^{J}$ via sampling **without replacement** from the discrete distribution on $\{Z^{(j)}\}_{j=1}^{J}$ with probabilities $\{\omega_j\}_{j=1}^{J}$;
- Generate $\theta^{(i)} \sim p(\theta|Y_{\rm obs}, Z^{(k_i)})$ for $i = 1,\ldots,I$, then $\{\theta^{(i)}\}_{i=1}^{I}$ are i.i.d. samples from the observed-data posterior distribution $p(\theta|Y_{\rm obs})$.

**Example T11.1** (Exact IBF sampling). Let $Y_{\rm obs} = \{(n_1,\ldots,n_4);(n_{12},n_{34})\}$ denote the observed frequencies, $\boldsymbol{\theta} = (\theta_1,\ldots,\theta_4)^\top \in \mathbb{T}_4$ denote the cell probability vector. The observed-data likelihood function of $\boldsymbol{\theta}$ is given by

$$L(\boldsymbol{\theta}|Y_{\rm obs}) \propto \left(\prod_{i=1}^{4} \theta_i^{n_i}\right)(\theta_1+\theta_2)^{n_{12}}(\theta_3+\theta_4)^{n_{34}}.$$

Let the prior distribution of $\boldsymbol{\theta}$ be the Dirichlet$(\alpha_1,\ldots,\alpha_4)$, a conjugate prior for $\boldsymbol{\theta}$. Then use the exact IBF sampling to sample from $p(\boldsymbol{\theta}|Y_{\rm obs})$.

**Solution:** By writing $n_{12} = Z_1 + Z_2$ with $Z_2 \equiv n_{12} - Z_1$ and $n_{34} = Z_3 + Z_4$ with $Z_4 \equiv n_{34} - Z_3$, a natural latent vector $Z = (Z_1, Z_3)^\top$ can be introduced so that the likelihood function for the complete-data $Y_{\rm obs}$, $Z$ is proportional to $\prod_{i=1}^{4} \theta_i^{n_i+Z_i}$.

Thus, the complete-data posterior is

$$\boldsymbol{\theta}|(Y_{\rm obs}, Z) \sim \text{Dirichlet}(n_1+Z_1+\alpha_1,\ldots,n_4+Z_4+\alpha_4).$$

Note that given $Y_{\text{obs}}$ and $\boldsymbol{\theta}$, $Z_1$ and $Z_3$ are independently binomially distributed. Thus, the conditional predictive distribution is

$$f(Z|Y_{\text{obs}}, \boldsymbol{\theta}) = \text{Binomial}\left(Z_1 \middle| n_{12}, \frac{\theta_1}{\theta_1 + \theta_2}\right) \times \text{Binomial}\left(Z_3 \middle| n_{34}, \frac{\theta_3}{\theta_3 + \theta_4}\right).$$

To apply the exact IBF sampling, we first need to identify the conditional support of $Z|Y_{\text{obs}}$. We have

$$\mathcal{S}_{(Z|Y_{\text{obs}})} = \mathcal{S}_{(Z|Y_{\text{obs}}, \boldsymbol{\theta})} = \{z_1, \ldots, z_K\} = \left\{ \begin{array}{cccc} (0,0) & (0,1) & \cdots & (0, n_{34}) \\ (1,0) & (1,1) & \cdots & (1, n_{34}) \\ \vdots & \vdots & \ddots & \vdots \\ (n_{12}, 0) & (n_{12}, 1) & \cdots & (n_{12}, n_{34}) \end{array} \right\},$$

where $K = (n_{12}+1)(n_{34}+1)$. Then we calculate $\{p_k\}_{k=1}^{K}$ with $\theta_0 = (0.25, \ldots, 0.25)^{\top}$. Thirdly, we draw $L = 100,000$ i.i.d. samples $\{Z^{(\ell)}\}_{\ell=1}^{L}$ of $Z$ from the pmf $f(Z|Y_{\text{obs}})$ with probabilities $\{p_k\}_{k=1}^{K}$ on $\{z_k\}_{k=1}^{K}$. Finally, we draw $\boldsymbol{\theta}^{(\ell)} \sim p(\boldsymbol{\theta}|Y_{\text{obs}}, Z^{(\ell)})$ for $\ell = 1, \ldots, L$. $\{\boldsymbol{\theta}^{(\ell)}\}_{\ell=1}^{L}$ are i.i.d. samples from the observed posterior distribution $p(\boldsymbol{\theta}|Y_{\text{obs}})$.