Gibbs Sampling for parameter estimation of Unidimensional GMM

Hrishikesh Pable

February 2023

Contents

1 Introduction **2**

2 Notations **3**

3 Graphical Model of Gaussian Mixture Model **4**

4 Gibbs Sampling Derivation **6**

5 Collapsed Gibbs Sampling Derivation **10**

Chapter 1

Introduction

This paper contains the derivation of the conditional posteriors for Gibbs Sampling of a unidimensional Gaussian Mixture Model with k components. We consider a Gaussian-Gamma prior over the mean and variance of each gaussian, and a dirichlet prior over the mixture proportions π.

Chapter 2

Notations

- N : Total number of data points.
- K : Total number of components of the Mixture Model
- μ : $\mu_1, \mu_2, \ldots, \mu_K$
- ν : $\nu_1, \nu_2, \ldots, \nu_K$
- μ_k : The mean of the kth Mixture Component.
- ν_k : The precision (i.e. the inverse of the variance) of the kth Mixture Component.
- X : Vector of n data points.
- x_i : ith data point.
- m_k : The mean of the gaussian part of the Gaussian-Gamma prior over the mean and variance of the kth mixture component (μ_i).
- g_k : The precision of the gaussian part of the Gaussian-Gamma prior over the mean and variance of the kth mixture component (μ_k, ν_k). $g_k = p_k * \nu_k$.
- a_k, b_k : Hyperparameters of the Gamma part of the Gaussian-Gamma prior over the mean and variance of the ith component (μ_k, ν_k).
- Z : Vector consisting of cluster assignments of the n data points.
- z_i : The mixture component to which the ith data point belongs.
- $z_{i,k}$: The indicator variable $\mathrm{I}(z_i = k)$
- n_k : The number of data points belonging to the kth cluster. $n_k = \sum_{i=1}^{n} z_{i,k}$.
- π : $\pi_1, \pi_2, \ldots, \pi_K$ The mixture proportions corresponding each of the K clusters.
- β : The hyperparameter of the Dirichlet prior over the mixing proportions π.

Chapter 3

Graphical Model of Gaussian Mixture Model

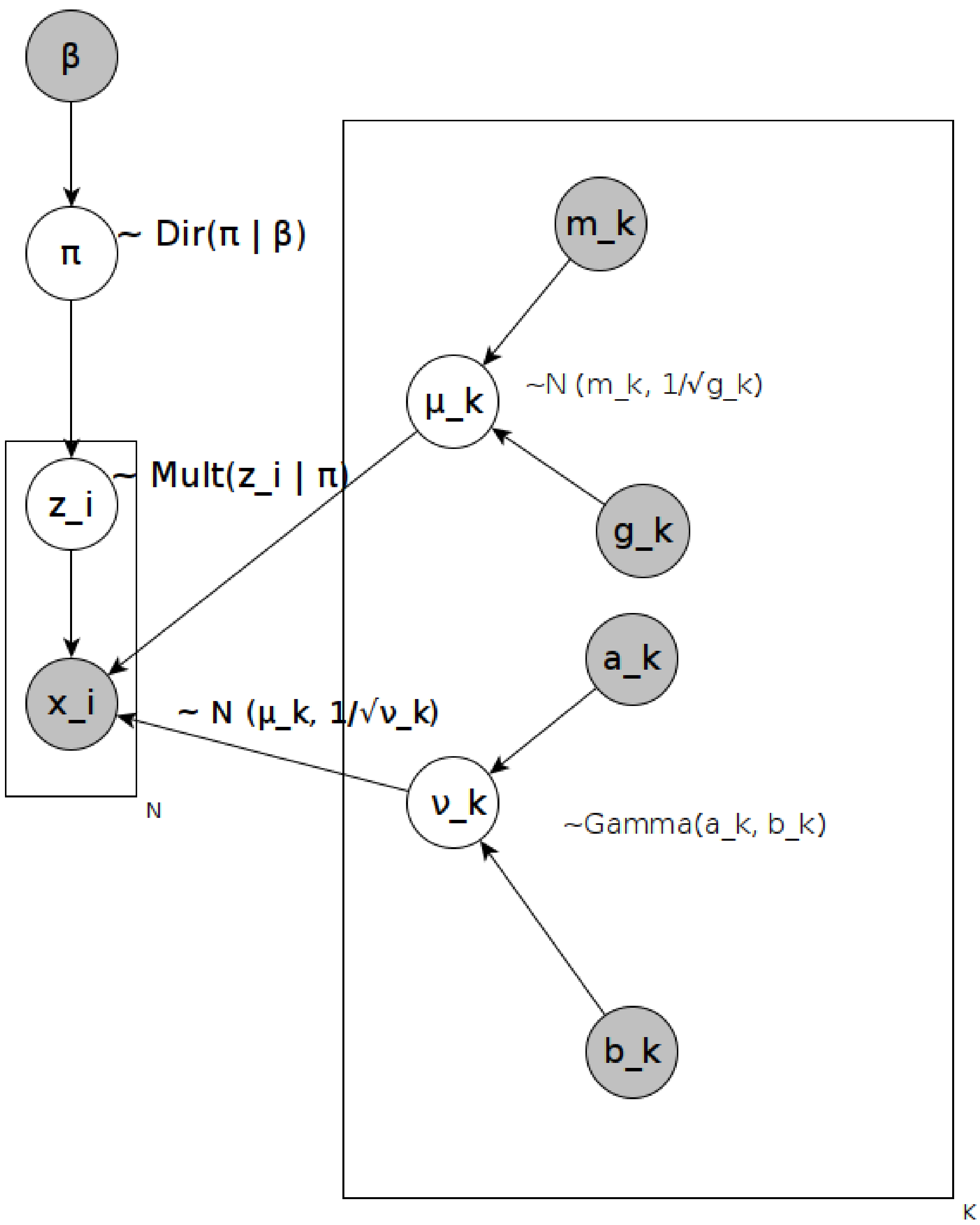

Figure 3.1: Gaussian Mixture Model

Chapter 4

Gibbs Sampling Derivation

We need the following 4 conditional distributions:

1. $P(Z|X, \mu, \nu, \pi)$
2. $P(\mu|X, Z, \nu, \pi)$
3. $P(\nu|X, Z, \mu, \pi)$
4. $P(\pi|X, Z, \mu, \nu)$

First, we will find $P(Z|X, \mu, \nu, \pi)$.

$$P(Z|X, \mu, \nu, \pi) = \frac{P(Z, X, \mu, \nu, \pi)}{P(X, \mu, \nu, \pi)}$$

where $P(X, \mu, \nu, \pi) = \sum_{Z} P(X, Z, \mu, \nu, \pi)$
Now, we can factorise the joint distribution by using the above bayesian network as follows:

$$\begin{aligned}
P(\pi, Z, X, \mu, \nu) &= P(\pi|\beta)P(Z|\pi)P(X|Z, \mu, \nu)P(\mu, \nu) \\
&= P(\pi|\beta)P(Z|\pi)P(X|Z, \mu, \nu)P(\mu|m, g)P(\nu|a, b) \\
&= Dir(\pi|\beta)\prod_{i=1}^{N} Mult(z_i|\pi)\prod_{i=1}^{N}\mathcal{N}\left(x_i|\mu_{z_i}, 1/\sqrt{\nu_{z_i}}\right)\prod_{k=1}^{K}\mathcal{N}\left(\mu_k|m_k, 1/\sqrt{g_k}\right)\prod_{k=1}^{K}\mathcal{G}(\nu_k|a_k, b_k)
\end{aligned} \tag{4.1}$$

We can simplify these terms seperately as follows:

$$\begin{aligned}
P(Z|\pi)P(\pi|\beta) &= \left(\prod_{i=1}^{N} Mult(z_i|\pi)\right)\left(Dir(\pi|\beta)\right) \\
&\propto \left(\prod_{i=1}^{N}\prod_{k=1}^{K}(\pi_k)^{z_{i,k}}\right)\left(\prod_{k=1}^{K}{\pi_k}^{\beta_k-1}\right) \\
&\propto \left(\prod_{k=1}^{K}{\pi_k}^{\left(\sum_{i=1}^{N} z_{i,k}\right)}\right)\left(\prod_{k=1}^{K}{\pi_k}^{(\beta_k-1)}\right) \\
&\propto \prod_{k=1}^{K}{\pi_k}^{\left(\sum_{i=1}^{N} z_{i,k}+\beta_k-1\right)} \\
&\propto \prod_{k=1}^{K}{\pi_k}^{(n_k+\beta_k-1)}
\end{aligned}$$

$$
\begin{aligned}
P(X|Z,\mu,\nu) &= \prod_{i=1}^{N} P(x_i|\mu_{z_i},\nu_{z_i}) \\
&\propto \prod_{i=1}^{N} (\nu_{z_i})^{1/2} \exp\left(\frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right) \\
&\propto \left(\prod_{i=1}^{N} (\nu_{z_i})^{1/2}\right)\left(\prod_{i=1}^{N} \exp\left(\frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right)\right) \\
&\propto \left(\prod_{k=1}^{K} (\nu_k)^{n_k/2}\right)\left(\prod_{k=1}^{K}\prod_{i:z_{i,k}=1} \exp\left(\frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right)\right) \\
&\propto \left(\prod_{k=1}^{K} (\nu_k)^{n_k/2}\right)\left(\prod_{k=1}^{K} \exp\left(\sum_{i:z_{i,k}=1} \frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right)\right) \\
&\propto \left(\prod_{k=1}^{K} (\nu_k)^{n_k/2} \exp\left(\sum_{i:z_{i,k}=1} \frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
P(\mu,\nu) &= \prod_{k=1}^{K} P(\mu_k,\nu_k|m_k,g_k,a_k,b_k) \\
&= \prod_{k=1}^{K} P(\mu_k|m_k,g_k)P(\nu_k|a_k,b_k) \\
&\propto \prod_{k=1}^{K} (\nu_k p_k)^{1/2} \exp\left(\frac{-p_k\nu_k}{2}(\mu_k-m_k)^2\right)(\nu_k)^{a_k-1}\exp(-b_k\nu_k) \\
&\propto \prod_{k=1}^{K} (\nu_k)^{a_k-1/2} \exp\left(\frac{-p_k\nu_k}{2}(\mu_k-m_k)^2 - b_k\nu_k\right) \\
&\propto \prod_{k=1}^{K} (\nu_k)^{a_k-1/2} \exp\left(-\nu_k\left(b_k+\frac{p_k}{2}(\mu_k-m_k)^2\right)\right)
\end{aligned}
$$

We can find $P(Z|\mu,\nu,\pi)$ by dividing the joint distribution in equation 4.1 by the same joint distribution marginalised over z. This would lead to the $Dir(\pi|\beta)$, $\mathcal{N}(\mu_k|m_k,g_k)$ and $\mathcal{G}(\nu_k|a_k,b_k)$ terms getting cancelled from the numerator and the denominator. So, we get,

$$
\begin{aligned}
P(Z|X,\mu,\nu,\pi) &= \frac{P(Z,X,\mu,\nu,\pi)}{\sum\limits_{Z} P(Z,X,\mu,\nu,\pi)} \\
&= \frac{\prod\limits_{i=1}^{N} Mult(z_i|\pi)\prod\limits_{i=1}^{N}\mathcal{N}(x_i|\mu_{z_i},\nu_{z_i})}{\sum\limits_{Z}\prod\limits_{i=1}^{N} Mult(z_i|\pi)\prod\limits_{i=1}^{N}\mathcal{N}(x_i|\mu_{z_i},\nu_{z_i})} \\
&\propto \frac{\left(\prod\limits_{i=1}^{N}\prod\limits_{k=1}^{K}(\pi_k)^{z_{i,k}}\right)\left(\prod_{i=1}^{N}(\nu_{z_i})^{1/2}\exp\left(\frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right)\right)}{\sum_{Z}\left(\prod\limits_{i=1}^{N}\prod\limits_{k=1}^{K}(\pi_k)^{z_{i,k}}\right)\left(\prod_{i=1}^{N}(\nu_{z_i})^{1/2}\exp\left(\frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right)\right)} \\
&\propto \prod_{i=1}^{N}\frac{\left(\prod\limits_{k=1}^{K}(\pi_k)^{z_{i,k}}\right)(\nu_{z_i})^{1/2}\exp\left(\frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right)}{\sum\limits_{z_i=1}^{K}\left(\prod\limits_{k=1}^{K}(\pi_k)^{z_{i,k}}\right)(\nu_{z_i})^{1/2}\exp\left(\frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right)} \\
&\propto \prod_{i=1}^{N}\frac{\left(\prod\limits_{k=1}^{K}(\pi_k)^{z_{i,k}}\right)(\nu_{z_i})^{1/2}\exp\left(\frac{-\nu_{z_i}}{2}(x_i-\mu_{z_i})^2\right)}{\sum\limits_{k'=1}^{K}(\pi_k')(\nu_{k'})^{1/2}\exp\left(\frac{-\nu_{k'}}{2}(x_i-\mu_{k'})^2\right)}
\end{aligned}
$$

Therefore, we have,

$$
P(z_i=k|x_i,\mu,\nu,\pi) \propto \frac{(\pi_k)(\nu_k)^{1/2}\exp\left(\frac{-\nu_k}{2}(x_i-\mu_k)^2\right)}{\sum\limits_{k'=1}^{K}(\pi_k')(\nu_{k'})^{1/2}\exp\left(\frac{-\nu_{k'}}{2}(x_i-\mu_{k'})^2\right)}
$$

Now, we will find the other three posterior conditionals. For that, we need to simplify the joint probability expression in equation 4.1 by putting together the individual simplifications that we made to the parts of the joint probability as follows:

$$
\begin{aligned}
P(\pi, Z, X, \mu, \nu) &= P(\pi|\beta)P(Z|\pi)P(X|\mu,\nu)P(\mu,\nu) \\
&= P(\pi|\beta)P(Z|\pi)P(X|Z,\mu,\nu)P(\mu|m,g)P(\nu|a,b) \\
&= Dir(\pi|\beta)\prod_{i=1}^{N} Mult(z_i|\pi)\prod_{i=1}^{N}\mathcal{N}\left(x_i|\mu_{z_i},\nu_{z_i}\right)\prod_{k=1}^{K}\mathcal{N}\left(\mu_k|m_k,g_k\right)\prod_{k=1}^{K}\mathcal{G}(\nu_k|a_k,b_k) \\
&\propto \left(\prod_{k=1}^{K}\pi_k{}^{(n_k+\beta_k-1)}\right)\left(\prod_{k=1}^{K}(\nu_k)^{n_k/2}\exp\left(\sum_{i:z_{i,k}=1}\frac{-\nu_{z_i}}{2}\left(x_i-\mu_{z_i}\right)^2\right)\right) \\
&\quad \left(\prod_{k=1}^{K}(\nu_k)^{a_k-1/2}\exp\left(-\nu_k\left(b_k+\frac{p_k}{2}(\mu_k-m_k)^2\right)\right)\right) \\
&\propto \left(\prod_{k=1}^{K}\pi_k{}^{(n_k+\beta_k-1)}\nu_k^{n_k/2+a_k-1/2}\exp\left(-\nu_k\left(b_k+\frac{p_k}{2}(\mu_k-m_k)^2+\frac{1}{2}\sum_{i:z_{i,k}=1}(x_i-\mu_k)^2\right)\right)\right) \\
&\propto \left(\prod_{k=1}^{K}\pi_k{}^{(n_k+\beta_k-1)}\nu_k^{n_k/2+a_k-1/2}\exp\left(-\nu_k\left(b_k+\frac{p_k}{2}(\mu_k-m_k)^2+\frac{1}{2}\sum_{i:z_{i,k}=1}(x_i-\mu_k)^2\right)\right)\right)
\end{aligned}
$$

Now, let $I = b_k + \frac{p_k}{2}(\mu_k - m_k)^2 + \frac{1}{2}\sum_{i:z_{i,k}=1}(x_i-\mu_k)^2$

we need to bring it into the form $I = b_k^* + \frac{p_k^*}{2}(\mu_k - m_k^*)^2$, to get the parameters for the posterior gaussian-gamma distribution.

$$
\begin{aligned}
I &= b_k + \frac{p_k}{2}(\mu_k-m_k)^2 + \frac{1}{2}\sum_{i:z_{i,k}=1}(x_i-\mu_k)^2 \\
&= b_k + \frac{p_k}{2}\left(\mu_k^2 - 2\mu_k m_k + m_k^2\right) + \frac{1}{2}\left(\sum_{i:z_{i,k}=1}\left(x_i^2 - 2x_i\mu_k + \mu_k^2\right)\right) \\
&= b_k + \frac{1}{2}\left(p_k\mu_k^2 - 2p_k m_k\mu_k + p_k m_k^2 + \left(\sum_{i:z_{i,k}=1}(x_i^2)\right) - 2\mu_k\left(\sum_{i:z_{i,k}=1}x_i\right) + n_k\mu_k^2\right) \\
&= b_k + \frac{1}{2}\left(\left(p_k+n_k\right)\mu_k^2 - 2\left(p_k m_k + \bar{x}_k n_k\right)\mu_k + \left(p_k m_k^2 + \sum_{i:z_{i,k}=1}x_i^2\right)\right) \\
&= b_k + \frac{1}{2}\left(p_k m_k^2 + \sum_{i:z_{i,k}=1}x_i^2\right) + \frac{(p_k+n_k)}{2}\left(\mu_k^2 - 2\left(\frac{p_k m_k + \bar{x}_k n_k}{p_k+n_k}\right)\mu_k\right) \\
&= b_k^* + \frac{p_k^*}{2}(\mu_k - m_k^*)^2
\end{aligned}
$$

Here,

$b_k^* = b_k + \frac{1}{2}\left(p_k m_k^2 + \sum_{i:z_{i,k}=1}x_i^2\right) - p_k^* m_k^*$

$p_k^* = (p_k + n_k)$

$m_k^* = \left(\frac{p_k m_k + \bar{x}_k n_k}{p_k + n_k}\right)$

So the joint probability can now be written as,

$$
\begin{aligned}
P(\pi, Z, X, \mu, \nu) &\propto \left(\prod_{k=1}^{K} {\pi_k}^{(n_k+\beta_k-1)} \nu_k^{(n_k/2+a_k-1/2)} \exp\left(-\nu_k\left(b_k^* + \frac{p_k^*}{2}(\mu_k - m_k^*)^2\right)\right)\right) \\
&\propto \left(\prod_{k=1}^{K} {\pi_k}^{(\beta_k^*-1)} \nu_k^{(a_k^*-1/2)} \exp\left(-\nu_k\left(b_k^* + \frac{p_k^*}{2}(\mu_k - m_k^*)^2\right)\right)\right) \ [\beta_k^* = n_k + \beta_k, a_k^* = a_k + n_k/2] \\
&\propto \left(\prod_{k=1}^{K} {\pi_k}^{(\beta_k^*-1)} \times \nu_k^{1/2} \exp\left(\frac{-\nu_k p_k^*}{2}(\mu_k - m_k^*)^2\right) \times \nu_k^{(a_k^*-1)} \exp\left(-\nu_k b_k^*\right)\right) \\
&\propto \left(\prod_{k=1}^{K} {\pi_k}^{(\beta_k^*-1)}\right) \times \left(\prod_{k=1}^{K} (\nu_k p_k^*)^{1/2} \exp\left(\frac{-\nu_k p_k^*}{2}(\mu_k - m_k^*)^2\right)\right) \times \left(\prod_{k=1}^{K} \nu_k^{(a_k^*-1)} \exp\left(-\nu_k b_k^*\right)\right) \\
&\propto \left(Dir(\pi|\beta^*)\right)\left(\prod_{k=1}^{K} \mathcal{N}(\mu_k|m_k^*, 1/\sqrt{\nu_k p_k^*})\right)\left(\prod_{k=1}^{K} \mathcal{G}(\nu_k|a_k^*, b_k^*)\right)
\end{aligned}
$$

So, we finally have,

$$
\begin{aligned}
P(\pi|Z, X, \mu, \nu) &= \frac{P(\pi, Z, X, \mu, \nu)}{P(Z, X, \mu, \nu)} \\
&= \frac{P(\pi, Z, X, \mu, \nu)}{\sum\limits_{\pi} P(\pi, Z, X, \mu, \nu)} \\
&= Dir(\pi|\beta^*)
\end{aligned}
$$

$$
\begin{aligned}
P(\mu|Z, X, \nu, \pi) &= \frac{P(\pi, Z, X, \mu, \nu)}{P(\pi, Z, X, \nu)} \\
&= \frac{P(\pi, Z, X, \mu, \nu)}{\sum\limits_{\mu} P(\pi, Z, X, \mu, \nu)} \\
&= \prod_{k=1}^{K} \mathcal{N}(\mu_k|m_k^*, 1/\sqrt{\nu_k p_k^*})
\end{aligned}
$$

$$
\begin{aligned}
P(\nu|Z, X, \mu, \pi) &= \frac{P(\pi, Z, X, \nu, \mu)}{P(\pi, Z, X, \mu)} \\
&= \frac{P(\pi, Z, X, \nu, \mu)}{\sum\limits_{\nu} P(\pi, Z, X, \nu, \mu)} \\
&= \prod_{k=1}^{K} \mathcal{G}(\nu_k|a_k^*, b_k^*)
\end{aligned}
$$

Chapter 5

Collapsed Gibbs Sampling Derivation

Here we directly want to find the probability of Z given the data X. Since $\mu, \nu, and \pi$ are intermediate statistics, they can be obtained from Z.

So, we want,

$$P(Z|X) = \frac{P(Z,X)}{P(W)} = \frac{\prod\limits_{i=1}^{N} P(z_i, x_i)}{\prod_{i=1}^{N} \sum_{k=1}^{K} P(z_i = k, x_i)}$$

The numerator can be factored as:

$$P(X, Z|\beta, m, p, a, b) = P(X|Z, m, p, a, b)P(Z, \beta) \tag{5.1}$$

For finding the first term in equation 5.1, we begin with,

$$P(X|Z, \mu, \nu) = \prod_{i=1}^{N} (\nu_{z_i})^{1/2} \exp\left(\frac{-\nu_{z_i}}{2} (x_i - \mu_{z_i})^2\right)$$

The target $P(X|Z, m, p, a, b)$ is obtained by integrating over μ and ν as follows:

$$
\begin{aligned}
P(X|Z,m,p,a,b) &= \int_\nu \int_\mu P(X|Z,\mu,\nu)P(\mu,\nu|m,p,a,b)d\mu_k\ d\nu_k \\
&= \int_\nu \int_\mu \left(\prod_{k=1}^K \left(\frac{\nu_k}{2\pi}\right)^{n_k/2} \exp\left(\sum_{i:z_{i,k}=1} \frac{-\nu_k}{2}\left(x_i-\mu_k\right)^2 \right)\right) \\
&\quad \left(\prod_{k=1}^K \left(\frac{\nu_k p_k}{2\pi}\right)^{1/2} \exp\left(\frac{-p_k\nu_k}{2}(\mu_k-m_k)^2\right) \times \frac{b_k^{a_k}}{\Gamma(a_k)}(\nu_k)^{a_k-1}\exp(-b_k\nu_k)\right) d\mu_k\ d\nu_k \\
&= \prod_{k=1}^K \int_\nu \int_\mu \left(\frac{\nu_k}{2\pi}\right)^{n_k/2} \exp\left(\sum_{i:z_{i,k}=1} \frac{-\nu_k}{2}\left(x_i-\mu_k\right)^2 \right) \\
&\quad \left(\frac{\nu_k p_k}{2\pi}\right)^{1/2} \exp\left(\frac{-p_k\nu_k}{2}(\mu_k-m_k)^2\right) \times \frac{b_k^{a_k}}{\Gamma(a_k)}(\nu_k)^{a_k-1}\exp(-b_k\nu_k) d\mu_k\ d\nu_k \\
&= \prod_{k=1}^K \int_\nu \frac{b_k^{a_k}}{\Gamma(a_k)}(\nu_k)^{a_k-1}\exp(-b_k\nu_k)\left(\frac{\nu_k}{2\pi}\right)^{n_k/2} \int_\mu \exp\left(\sum_{i:z_{i,k}=1} \frac{-\nu_k}{2}\left(x_i-\mu_k\right)^2 \right) \\
&\quad \left(\frac{\nu_k p_k}{2\pi}\right)^{1/2} \exp\left(\frac{-p_k\nu_k}{2}(\mu_k-m_k)^2\right) \times d\mu_k\ d\nu_k \\
&= \prod_{k=1}^K \int_\nu \frac{b_k^{a_k}}{\Gamma(a_k)}(\nu_k)^{a_k-1}\exp(-b_k\nu_k)\left(\frac{\nu_k}{2\pi}\right)^{n_k/2} \int_\mu \left(\frac{\nu_k p_k}{2\pi}\right)^{1/2} \\
&\quad \exp\left(\sum_{i:z_{i,k}=1} \frac{-\nu_k}{2}\left(x_i-\mu_k\right)^2 - \frac{p_k\nu_k}{2}(\mu_k-m_k)^2 \right) \times d\mu_k\ d\nu_k \\
&= \prod_{k=1}^K \int_\nu \frac{b_k^{a_k}}{\Gamma(a_k)}(\nu_k)^{a_k-1}\exp(-b_k\nu_k)\left(\frac{\nu_k}{2\pi}\right)^{n_k/2} \int_\mu \left(\frac{\nu_k p_k}{2\pi}\right)^{1/2} \\
&\quad \exp\left((b_k^*-b_k)+\frac{p_k^*}{2}(\mu_k-m_k^*)^2\right) d\mu_k\ d\nu_k \\
&= \prod_{k=1}^K \int_\nu \frac{b_k^{a_k}}{\Gamma(a_k)}(\nu_k)^{a_k-1}\exp(-b_k\nu_k)\left(\frac{\nu_k}{2\pi}\right)^{n_k/2} \left(\frac{p_k}{p_k^*}\right)^{1/2} \exp(b_k^*-b_k) \int_\mu \left(\frac{\nu_k p_k^*}{2\pi}\right)^{1/2} \\
&\quad \exp\left(\frac{p_k^*}{2}(\mu_k-m_k^*)^2\right) d\mu_k\ d\nu_k \\
&= \prod_{k=1}^K \int_\nu \frac{b_k^{a_k}}{\Gamma(a_k)}(\nu_k)^{a_k-1}\exp(-b_k\nu_k)\left(\frac{\nu_k}{2\pi}\right)^{n_k/2} \left(\frac{p_k}{p_k^*}\right)^{1/2} \exp(b_k^*-b_k) \int_\mu \mathcal{N}(\mu|m_k^*,\nu_k^* p_k^*) d\mu_k\ d\nu_k \\
&= \prod_{k=1}^K \left(\frac{p_k}{p_k^*}\right)^{1/2} \exp(b_k^*-b_k) \int_\nu \frac{b_k^{a_k}}{\Gamma(a_k)}(\nu_k)^{a_k-1}\exp(-b_k\nu_k)\left(\frac{\nu_k}{2\pi}\right)^{n_k/2} d\nu_k \\
&= \prod_{k=1}^K \left(\frac{p_k}{p_k^*}\right)^{1/2} \frac{\exp(b_k^*-b_k)}{(2\pi)^{n_k/2}} \frac{b_k^{a_k}}{\Gamma(a_k)} \frac{\Gamma(a_k^*)}{b_k^{a_k^*}} \int_\nu \frac{b_k^{a_k^*}}{\Gamma(a_k^*)}(\nu_k)^{a_k^*-1}\exp(-b_k\nu_k)\ d\nu_k \\
&= \prod_{k=1}^K \left(\frac{p_k}{p_k^*}\right)^{1/2} \frac{\exp(b_k^*-b_k)}{(2\pi b_k)^{n_k/2}} \frac{\Gamma(a_k^*)}{\Gamma(a_k)}
\end{aligned}
$$

Hence, $P(X|Z,m,p,a,b) = \prod_{k=1}^K \left(\frac{p_k}{p_k^*}\right)^{1/2} \frac{\exp(b_k^*-b_k)}{(2\pi b_k)^{n_k/2}} \frac{\Gamma(a_k^*)}{\Gamma(a_k)}$

Analogous to $P(X|Z,m,p,a,b)$, we can find $P(Z|\beta)$. We start with $P(Z|\pi,\beta)$ and then integrate out π.

$$
\begin{aligned}
P(Z|\beta) &= \int_\pi P(Z|\pi)P(\pi|\beta)\ d\pi \\
&= \int_\pi \prod_{k=1}^K \frac{1}{B(\beta_k)} \pi_k^{(n_k+\beta_k-1)}\ d\pi_k \\
&= \prod_{k=1}^K \frac{B(\beta_k^*)}{B(\beta_k)} \int_\pi \frac{1}{B(\beta_k^*)} \pi_k^{(\beta_k^*-1)}\ d\pi_k \\
&= \prod_{k=1}^K \frac{B(\beta_k^*)}{B(\beta_k)}
\end{aligned}
$$

Now, we will derive the conditional distribution for each z_i given the other variables, from this joint distribution of Z and W.

$$
\begin{aligned}
P(z_i = k | Z_{\neg i}, X) &= \frac{P(X, Z)}{P(X, Z_{\neg i})} \\
&= \frac{P(X|Z)}{P(X_{\neg i}|Z_{\neg i})P(x_i)} \frac{P(Z)}{P(Z_{\neg i})} \\
&\propto \frac{P(X|Z)}{P(X_{\neg i}|Z_{\neg i})} \frac{P(Z)}{P(Z_{\neg i})} \\
&= \prod_{k=1}^{K} \frac{\left[\left(\frac{p_k}{p_k^*} \right)^{1/2} \frac{\exp(b_k^* - b_k)}{(2\pi b_k)^{n_k/2}} \frac{\Gamma(a_k^*)}{\Gamma(a_k)} \frac{B(\beta_k^*)}{B(\beta_k)} \right]}{\left[\left(\frac{p_k}{p_k^*} \right)^{1/2} \frac{\exp(b_k^* - b_k)}{(2\pi b_k)^{n_k/2}} \frac{\Gamma(a_k^*)}{\Gamma(a_k)} \frac{B(\beta_k^*)}{B(\beta_k)} \right]_{\neg i}}
\end{aligned}
$$