Lab 8: Data Augmentation

Olivier Binette

Friday October 9, 2020

Agenda

- Problem statement
- Go through the lab's tasks
- Office hours

Problem statement

Problem statement

Data points $Y_1, Y_2, \ldots, Y_n$ coming from a **mixture model**:

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2).$$

What does this mean?

For every data point:

- let Z_i be a random variable such that $\mathbb{P}(Z_i = j) = w_j$ for $j = 1, 2, 3$,
- let $Y_i \mid Z_i \sim N(\mu_{Z_i}, \varepsilon^2)$.

Problem statement

Data points $Y_1, Y_2, \ldots, Y_n$ coming from a **mixture model**:

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2).$$

What does this mean?

For every data point:

- let Z_i be a random variable such that $\mathbb{P}(Z_i = j) = w_j$ for $j = 1, 2, 3,$
- let $Y_i \mid Z_i \sim N(\mu_{Z_i}, \varepsilon^2)$.

Problem statement

Data points $Y_1, Y_2, \ldots, Y_n$ coming from a **mixture model**:

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2).$$

What does this mean?

For every data point:

- let Z_i be a random variable such that $\mathbb{P}(Z_i = j) = w_j$ for $j = 1, 2, 3,$
- let $Y_i \mid Z_i \sim N(\mu_{Z_i}, \varepsilon^2)$.

Problem statement

Data points $Y_1, Y_2, \ldots, Y_n$ coming from a **mixture model**:

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2).$$

What does this mean?

For every data point:

- let Z_i be a random variable such that $\mathbb{P}(Z_i = j) = w_j$ for $j = 1, 2, 3,$
- let $Y_i \mid Z_i \sim N(\mu_{Z_i}, \varepsilon^2)$.

Problem statement

Data points $Y_1, Y_2, \dots, Y_n$ coming from a **mixture model**:

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2).$$

What does this mean?

For every data point:

- let Z_i be a random variable such that $\mathbb{P}(Z_i = j) = w_j$ for $j = 1, 2, 3,$
- let $Y_i \mid Z_i \sim N(\mu_{Z_i}, \varepsilon^2)$.

Problem statement

Let's see what this mixture model could look like in an example.

Let $\mu_1 = -5$, $\mu_2 = 0$ and $\mu_3 = 5$, and let $\varepsilon = 1$. Let $w_j = 1/3$ for $j = 1, 2, 3$.

Now let's generate data from the mixture model:

```
n = 100
mu = c(-5, 0, 5)

Z = sample(1:3, size=n, replace=TRUE)
Y = rnorm(n, mean=mu[Z], sd=1)

hist(Y, breaks=20)
```

Problem statement

Let's see what this mixture model could look like in an example.

Let $\mu_1 = -5$, $\mu_2 = 0$ and $\mu_3 = 5$, and let $\varepsilon = 1$. Let $w_j = 1/3$ for $j = 1, 2, 3$.

Now let's generate data from the mixture model:

```
n = 100
mu = c(-5, 0, 5)

Z = sample(1:3, size=n, replace=TRUE)
Y = rnorm(n, mean=mu[Z], sd=1)

hist(Y, breaks=20)
```

Problem statement

Let's see what this mixture model could look like in an example.

Let $\mu_1 = -5$, $\mu_2 = 0$ and $\mu_3 = 5$, and let $\varepsilon = 1$. Let $w_j = 1/3$ for $j = 1, 2, 3$.

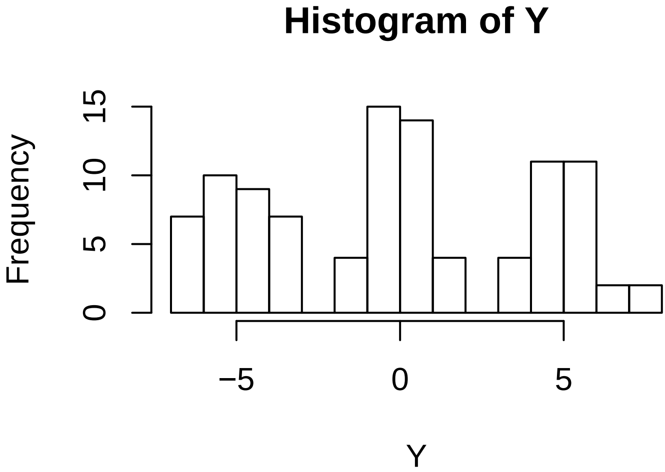
Now let's generate data from the mixture model:

```
n = 100
mu = c(-5, 0, 5)

Z = sample(1:3, size=n, replace=TRUE)
Y = rnorm(n, mean=mu[Z], sd=1)

hist(Y, breaks=20)
```

Problem statement



Problem statement

We have the model

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2),$$

now we need priors on the unknown parameters μ_j, w_j and ε.

Priors

For the means:

$$\begin{aligned} \mu_j \mid \mu_0, \sigma_0 &\sim N(\mu_0, \sigma_0^2) \\ \mu_0 &\sim N(0, 3) \\ \sigma_0^2 &\sim IG(2, 2) \end{aligned}$$

and recall that $\sigma_0^2 \sim IG(2, 2)$ means that $\phi_0 = 1/\sigma_0^2 \sim \text{Gamma}(2, 2)$.

Problem statement

We have the model

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2),$$

now we need priors on the unknown parameters μ_j, w_j and ε.

Priors

For the means:

$$\begin{aligned} \mu_j \mid \mu_0, \sigma_0 &\sim N(\mu_0, \sigma_0^2) \\ \mu_0 &\sim N(0, 3) \\ \sigma_0^2 &\sim IG(2, 2) \end{aligned}$$

and recall that $\sigma_0^2 \sim IG(2, 2)$ means that $\phi_0 = 1/\sigma_0^2 \sim \text{Gamma}(2, 2)$.

Problem statement

We have the model

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2),$$

now we need priors on the unknown parameters μ_j, w_j and ε.

Priors

For the means:

$$\mu_j \mid \mu_0, \sigma_0 \sim N(\mu_0, \sigma_0^2)$$
$$\mu_0 \sim N(0, 3)$$
$$\sigma_0^2 \sim IG(2, 2)$$

and recall that $\sigma_0^2 \sim IG(2, 2)$ means that $\phi_0 = 1/\sigma_0^2 \sim \text{Gamma}(2, 2)$.

Problem statement

We have the model

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2),$$

now we need priors on the unknown parameters μ_j, w_j and ε.

Priors

For the means:

$$\begin{aligned} \mu_j \mid \mu_0, \sigma_0 &\sim N(\mu_0, \sigma_0^2) \\ \mu_0 &\sim N(0, 3) \\ \sigma_0^2 &\sim IG(2, 2) \end{aligned}$$

and recall that $\sigma_0^2 \sim IG(2, 2)$ means that $\phi_0 = 1/\sigma_0^2 \sim \text{Gamma}(2, 2)$.

Problem statement

We have the model

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2),$$

now we need priors on the unknown parameters μ_j, w_j and ε.

Priors

For the means:

$$\begin{aligned} \mu_j \mid \mu_0, \sigma_0 &\sim N(\mu_0, \sigma_0^2) \\ \mu_0 &\sim N(0, 3) \\ \sigma_0^2 &\sim IG(2, 2) \end{aligned}$$

and recall that $\sigma_0^2 \sim IG(2, 2)$ means that $\phi_0 = 1/\sigma_0^2 \sim \text{Gamma}(2, 2)$.

Problem statement

We have the model

$$Y_i \sim \sum_{j=1}^{n} w_j N(\mu_j, \varepsilon^2),$$

now we need priors on the unknown parameters μ_j, w_j and ε.

Priors

For the means:

$$\begin{aligned} \mu_j \mid \mu_0, \sigma_0 &\sim N(\mu_0, \sigma_0^2) \\ \mu_0 &\sim N(0, 3) \\ \sigma_0^2 &\sim IG(2, 2) \end{aligned}$$

and recall that $\sigma_0^2 \sim IG(2,2)$ means that
$\phi_0 = 1/\sigma_0^2 \sim \text{Gamma}(2, 2)$.

Problem statement

Priors

For the mixture weights:

$$(w_1, w_2, w_3) \sim \text{Dirichlet}(\mathbf{1})$$

which means that $p(w_1, w_2, w_3) \propto 1$.

Recall that, in general,

$$\begin{aligned}&(w_1, w_2, w_3) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \alpha_3)\\ \Rightarrow &p(w_1, w_2, w_3) \propto w_1^{\alpha_1 - 1} w_2^{\alpha_2 - 1} w_3^{\alpha_3 - 1}.\end{aligned}$$

Problem statement

Priors

For the mixture weights:

$$(w_1, w_2, w_3) \sim \text{Dirichlet}(\mathbf{1})$$

which means that $p(w_1, w_2, w_3) \propto 1$.

Recall that, in general,

$$(w_1, w_2, w_3) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \alpha_3)$$
$$\Rightarrow p(w_1, w_2, w_3) \propto w_1^{\alpha_1 - 1} w_2^{\alpha_2 - 1} w_3^{\alpha_3 - 1}.$$

Problem statement

Priors

For the mixture weights:

$$(w_1, w_2, w_3) \sim \text{Dirichlet}(\mathbf{1})$$

which means that $p(w_1, w_2, w_3) \propto 1$.

Recall that, in general,

$$\begin{aligned}
&(w_1, w_2, w_3) \sim \text{Dirichlet}(\alpha_1, \alpha_2, \alpha_3) \\
\Rightarrow &p(w_1, w_2, w_3) \propto w_1^{\alpha_1 - 1} w_2^{\alpha_2 - 1} w_3^{\alpha_3 - 1}.
\end{aligned}$$

Problem statement

Priors

For the variance:

$$\varepsilon^2 \sim IG(2, 2)$$

This means that

$$\tau = 1/\varepsilon^2 \sim \text{Gamma}(2, 2)$$

Problem statement

Priors

For the variance:

$$\varepsilon^2 \sim IG(2,2)$$

This means that

$$\tau = 1/\varepsilon^2 \sim \text{Gamma}(2,2)$$

Problem statement

In summary,

$$
\begin{aligned}
p(Y_i|\mu_1,\mu_2,\mu_3,w_1,w_2,w_3,\tau) &= \sum_{j=1}^{3} w_i N(\mu_j,\tau^{-1}) \\
\mu_j|\mu_0,\sigma_0^2 &\sim N(\mu_0,\phi_0^{-1}) \\
\mu_0 &\sim N(0,3) \\
\phi_0 &\sim \text{Gamma}(2,2) \\
(w_1,w_2,w_3) &\sim Dirichlet(\mathbf{1}) \\
\tau &\sim \text{Gamma}(2,2),
\end{aligned}
$$

for $i = 1,\ldots n$.

Task 1

Derive the joint posterior $p(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_{1:N})$ up to a normalizing constant.

Let's do the derivations using $\tau = 1/\varepsilon^2$ and $\phi_0 = 1/\sigma_0^2$.

The posterior distribution is always proportional to the full joint distribution:

$$
\begin{aligned}
&p(Y_{1:n}, \mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(w_{1:3}) p(\tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(\tau) \\
=&\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0) p(\phi_0) p(\tau).
\end{aligned}
$$

Task 1

Derive the joint posterior $p(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_{1:N})$ up to a normalizing constant.

Let's do the derivations using $\tau = 1/\varepsilon^2$ and $\phi_0 = 1/\sigma_0^2$.

The posterior distribution is always proportional to the full joint distribution:

$$
\begin{aligned}
&p(Y_{1:n}, \mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(w_{1:3}) p(\tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(\tau) \\
=&\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0) p(\phi_0) p(\tau).
\end{aligned}
$$

Task 1

Derive the joint posterior $p(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_{1:N})$ up to a normalizing constant.

Let's do the derivations using $\tau = 1/\varepsilon^2$ and $\phi_0 = 1/\sigma_0^2$.

The posterior distribution is always proportional to the full joint distribution:

$$
\begin{aligned}
& p(Y_{1:n}, \mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=& p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=& p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(w_{1:3}) p(\tau) \\
=& p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(\tau) \\
=& \left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau) \right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0) \right) p(\mu_0) p(\phi_0) p(\tau).
\end{aligned}
$$

Task 1

Derive the joint posterior $p(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_{1:N})$ up to a normalizing constant.

Let's do the derivations using $\tau = 1/\varepsilon^2$ and $\phi_0 = 1/\sigma_0^2$.

The posterior distribution is always proportional to the full joint distribution:

$$
\begin{aligned}
&p(Y_{1:n}, \mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau)p(\mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau)p(\mu_{1:3} \mid \mu_0, \phi_0)p(\mu_0)p(\phi_0)p(w_{1:3})p(\tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau)p(\mu_{1:3} \mid \mu_0, \phi_0)p(\mu_0)p(\phi_0)p(\tau) \\
=&\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right)\left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right)p(\mu_0)p(\phi_0)p(\tau).
\end{aligned}
$$

Task 1

Derive the joint posterior $p(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_{1:N})$ up to a normalizing constant.

Let's do the derivations using $\tau = 1/\varepsilon^2$ and $\phi_0 = 1/\sigma_0^2$.

The posterior distribution is always proportional to the full joint distribution:

$$
\begin{aligned}
& p(Y_{1:n}, \mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=& p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=& p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(w_{1:3}) p(\tau) \\
=& p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(\tau) \\
=& \left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau) \right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0) \right) p(\mu_0) p(\phi_0) p(\tau).
\end{aligned}
$$

Task 1

Derive the joint posterior $p(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_{1:N})$ up to a normalizing constant.

Let's do the derivations using $\tau = 1/\varepsilon^2$ and $\phi_0 = 1/\sigma_0^2$.

The posterior distribution is always proportional to the full joint distribution:

$$
\begin{aligned}
&p(Y_{1:n}, \mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(w_{1:3}) p(\tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(\tau) \\
=&\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0) p(\phi_0) p(\tau).
\end{aligned}
$$

Task 1

Derive the joint posterior $p(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_{1:N})$ up to a normalizing constant.

Let's do the derivations using $\tau = 1/\varepsilon^2$ and $\phi_0 = 1/\sigma_0^2$.

The posterior distribution is always proportional to the full joint distribution:

$$
\begin{aligned}
& p(Y_{1:n}, \mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
= & p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
= & p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(w_{1:3}) p(\tau) \\
= & p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(\tau) \\
= & \left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau) \right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0) \right) p(\mu_0) p(\phi_0) p(\tau).
\end{aligned}
$$

Task 1

Derive the joint posterior $p(w_1, w_2, w_3, \mu_1, \mu_2, \mu_3, \varepsilon^2, \mu_0, \sigma_0^2 | Y_{1:N})$ up to a normalizing constant.

Let's do the derivations using $\tau = 1/\varepsilon^2$ and $\phi_0 = 1/\sigma_0^2$.

The posterior distribution is always proportional to the full joint distribution:

$$
\begin{aligned}
&p(Y_{1:n}, \mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3}, \mu_0, \phi_0, w_{1:3}, \tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(w_{1:3}) p(\tau) \\
=&p(Y_{1:n} \mid \mu_{1:3}, w_{1:3}, \tau) p(\mu_{1:3} \mid \mu_0, \phi_0) p(\mu_0) p(\phi_0) p(\tau) \\
=&\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0) p(\phi_0) p(\tau).
\end{aligned}
$$

Task 1

The full joint:

$$\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0)p(\phi_0)p(\tau).$$

Also:

$$p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau) = \sum_{j=1}^{3} w_j N(Y_i; \mu_j, \tau),$$

$$p(\mu_j \mid \mu_0, \phi_0) = N(\mu_j; \mu_0, \phi_0^{-1}),$$

$$p(\mu_0) = N(\mu_0; 0, 3),$$

$$p(\phi_0) = \text{Gamma}(\phi_0; 2, 2),$$

$$p(\tau) = \text{Gamma}(\tau; 2, 2).$$

Task 1

The full joint:

$$\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0)p(\phi_0)p(\tau).$$

Also:

$$p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau) = \sum_{j=1}^{3} w_j N(Y_i; \mu_j, \tau),$$

$$p(\mu_j \mid \mu_0, \phi_0) = N(\mu_j; \mu_0, \phi_0^{-1}),$$

$$p(\mu_0) = N(\mu_0; 0, 3),$$

$$p(\phi_0) = \mathrm{Gamma}(\phi_0; 2, 2),$$

$$p(\tau) = \mathrm{Gamma}(\tau; 2, 2).$$

Task 1

The full joint:

$$\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0)p(\phi_0)p(\tau).$$

Also:

$$p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau) = \sum_{j=1}^{3} w_j N(Y_i; \mu_j, \tau),$$

$$p(\mu_j \mid \mu_0, \phi_0) = N(\mu_j; \mu_0, \phi_0^{-1}),$$

$$p(\mu_0) = N(\mu_0; 0, 3),$$

$$p(\phi_0) = \text{Gamma}(\phi_0; 2, 2),$$

$$p(\tau) = \text{Gamma}(\tau; 2, 2).$$

Task 1

The full joint:

$$\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right)\left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0)p(\phi_0)p(\tau).$$

Also:

$$p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau) = \sum_{j=1}^{3} w_j N(Y_i; \mu_j, \tau),$$

$$p(\mu_j \mid \mu_0, \phi_0) = N(\mu_j; \mu_0, \phi_0^{-1}),$$

$$p(\mu_0) = N(\mu_0; 0, 3),$$

$$p(\phi_0) = \text{Gamma}(\phi_0; 2, 2),$$

$$p(\tau) = \text{Gamma}(\tau; 2, 2).$$

Task 1

The full joint:

$$\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right)\left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0)p(\phi_0)p(\tau).$$

Also:

$$p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau) = \sum_{j=1}^{3} w_j N(Y_i; \mu_j, \tau),$$

$$p(\mu_j \mid \mu_0, \phi_0) = N(\mu_j; \mu_0, \phi_0^{-1}),$$

$$p(\mu_0) = N(\mu_0; 0, 3),$$

$$p(\phi_0) = \mathrm{Gamma}(\phi_0; 2, 2),$$

$$p(\tau) = \mathrm{Gamma}(\tau; 2, 2).$$

Task 1

The full joint:

$$\left(\prod_{i=1}^{n} p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau)\right) \left(\prod_{j=1}^{3} p(\mu_j \mid \mu_0, \phi_0)\right) p(\mu_0)p(\phi_0)p(\tau).$$

Also:

$$p(Y_i \mid \mu_{1:3}, w_{1:3}, \tau) = \sum_{j=1}^{3} w_j N(Y_i; \mu_j, \tau),$$

$$p(\mu_j \mid \mu_0, \phi_0) = N(\mu_j; \mu_0, \phi_0^{-1}),$$

$$p(\mu_0) = N(\mu_0; 0, 3),$$

$$p(\phi_0) = \text{Gamma}(\phi_0; 2, 2),$$

$$p(\tau) = \text{Gamma}(\tau; 2, 2).$$