Bayesian Semi-parametric Logistic and Poisson Regression

Sohee Kang

Department of Public Health Sciences (Biostatistics)
University of Toronto

May 11, 2006

Outline

1. Part I: Bayesian Semi-parametric Logistic regression
   - Introduction to Dirichlet Process
   - Bayesian curve fitting
   - Bayesian Semi-parametric Logistic Regression
   - Simulated examples

2. Part II: Bayesian Semi-parametric Poisson Regression
   - Bayesian Semi-parametric Poisson Regression
   - Simulation study
   - Discussion

Outline

Introduction to Dirichlet Process

Dirichlet Process

What is Dirichlet Process?

- The Dirichlet process is a way of putting a distribution on a class of distributions.
- It is most popular prior process in Non-parametric Bayesian Inference.
- Definition: Consider a space Θ and σ-algebra $\mathcal{B}$ of a subset of Θ. A random probability measure, G on $(\Theta, \mathcal{B})$, follows a Dirichlet process $DP(\alpha, G_0)$, if for any finite measurable partition, $B_1, \ldots, B_k$ of Θ,

$$(G(B_1), \ldots, G(B_k)) \sim Dirichlet(\alpha G_0(B_1), \ldots, \alpha G_0(B_k))$$

- $G \sim DP(\alpha, G_0)$, where α is a precision parameter, which reflects the strength of belif in G_0, the prior expectation of G.

Dirichlet Process

What is Dirichlet Process?

- The Dirichlet process is a way of putting a distribution on a class of distributions.
- It is most popular prior process in Non-parametric Bayesian Inference.
- Definition: Consider a space Θ and σ-algebra $\mathcal{B}$ of a subset of Θ. A random probability measure, G on $(\Theta, \mathcal{B})$, follows a Dirichlet process $DP(\alpha, G_0)$, if for any finite measurable partition, $B_1, \ldots, B_k$ of Θ,

$$(G(B_1), \ldots, G(B_k)) \sim Dirichlet(\alpha G_0(B_1), \ldots, \alpha G_0(B_k))$$

- $G \sim DP(\alpha, G_0)$, where α is a precision parameter, which reflects the strength of belif in G_0, the prior expectation of G.

Dirichlet Process

What is Dirichlet Process?

- The Dirichlet process is a way of putting a distribution on a class of distributions.
- It is most popular prior process in Non-parametric Bayesian Inference.
- Definition: Consider a space Θ and σ-algebra $\mathcal{B}$ of a subset of Θ. A random probability measure, G on $(\Theta, \mathcal{B})$, follows a Dirichlet process $DP(\alpha, G_0)$, if for any finite measurable partition, $B_1, \ldots, B_k$ of Θ,

$$(G(B_1), \ldots, G(B_k)) \sim Dirichlet(\alpha G_0(B_1), \ldots, \alpha G_0(B_k))$$

- $G \sim DP(\alpha, G_0)$, where α is a precision parameter, which reflects the strength of belif in G_0, the prior expectation of G.

Dirichlet Process

What is Dirichlet Process?

- The Dirichlet process is a way of putting a distribution on a class of distributions.
- It is most popular prior process in Non-parametric Bayesian Inference.
- Definition: Consider a space Θ and σ-algebra $\mathcal{B}$ of a subset of Θ. A random probability measure, G on $(\Theta, \mathcal{B})$, follows a Dirichlet process $DP(\alpha, G_0)$, if for any finite measurable partition, $B_1, \ldots, B_k$ of Θ,

$$(G(B_1), \ldots, G(B_k)) \sim Dirichlet(\alpha G_0(B_1), \ldots, \alpha G_0(B_k))$$

- $G \sim DP(\alpha, G_0)$, where α is a precision parameter, which reflects the strength of belif in G_0, the prior expectation of G.

Dirichlet Process

What is Dirichlet Process?

- The Dirichlet process is a way of putting a distribution on a class of distributions.
- It is most popular prior process in Non-parametric Bayesian Inference.
- Definition: Consider a space Θ and σ-algebra $\mathcal{B}$ of a subset of Θ. A random probability measure, G on $(\Theta, \mathcal{B})$, follows a Dirichlet process $DP(\alpha, G_0)$, if for any finite measurable partition, $B_1, \ldots, B_k$ of Θ,

$$(G(B_1), \ldots, G(B_k)) \sim Dirichlet(\alpha G_0(B_1), \ldots, \alpha G_0(B_k))$$

- $G \sim DP(\alpha, G_0)$, where α is a precision parameter, which reflects the strength of belif in G_0, the prior expectation of G.

Dirichlet Process

What is Dirichlet Process?

- The Dirichlet process is a way of putting a distribution on a class of distributions.
- It is most popular prior process in Non-parametric Bayesian Inference.
- Definition: Consider a space Θ and σ-algebra $\mathcal{B}$ of a subset of Θ. A random probability measure, G on $(\Theta, \mathcal{B})$, follows a Dirichlet process $DP(\alpha, G_0)$, if for any finite measurable partition, $B_1, \ldots, B_k$ of Θ,

$$(G(B_1), \ldots, G(B_k)) \sim Dirichlet(\alpha G_0(B_1), \ldots, \alpha G_0(B_k))$$

- $G \sim DP(\alpha, G_0)$, where α is a precision parameter, which reflects the strength of belif in G_0, the prior expectation of G.

Dirichlet Process Representation

Two Representations of Dirichlet Process

(1) Stick-breaking representation (Sethuraman, 1994)

- $G = \sum_{k=1}^{\infty} p_k \delta_{\theta_k}$, here $\theta_k \sim G_0$ and

$$p_1 = V_1 \text{ and } p_k = (1 - V_1)(1 - V_2) \ldots (1 - V_{k-1})V_k, \; k \geq 2,$$

where V_k are independent $Beta(1, \alpha)$ random variables.

Dirichlet Process Representation

Two Representations of Dirichlet Process

(1) Stick-breaking representation (Sethuraman, 1994)

- $G = \sum_{k=1}^{\infty} p_k \delta_{\theta_k}$, here $\theta_k \sim G_0$ and

$$p_1 = V_1 \; and \; p_k = (1 - V_1)(1 - V_2) \ldots (1 - V_{k-1})V_k, \;\; k \geq 2,$$

where V_k are independent $Beta(1, \alpha)$ random variables.

Dirichlet Process Representation

Two Representations of Dirichlet Process

(1) Stick-breaking representation (Sethuraman, 1994)

- $G = \sum_{k=1}^{\infty} p_k \delta_{\theta_k}$, here $\theta_k \sim G_0$ and

$$p_1 = V_1 \text{ and } p_k = (1 - V_1)(1 - V_2) \ldots (1 - V_{k-1})V_k, \ \ k \geq 2,$$

where V_k are independent $Beta(1, \alpha)$ random variables.

Dirichlet Process Representation

Two Representations of Dirichlet Process

(1) Stick-breaking representation (Sethuraman, 1994)

- $G = \sum_{k=1}^{\infty} p_k \delta_{\theta_k}$, here $\theta_k \sim G_0$ and

  $$p_1 = V_1 \; and \; p_k = (1 - V_1)(1 - V_2)\ldots(1 - V_{k-1})V_k, \;\; k \geq 2,$$

  where V_k are independent $Beta(1, \alpha)$ random variables.

Dirichlet Process Representation

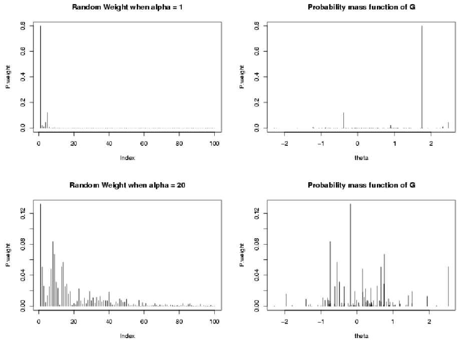
Two Representations of Dirichlet Process

(1) Stick-breaking representation (Sethuraman, 1994)

- $G = \sum_{k=1}^{\infty} p_k \delta_{\theta_k}$, here $\theta_k \sim G_0$ and

$$p_1 = V_1 \textit{ and } p_k = (1 - V_1)(1 - V_2)\ldots(1 - V_{k-1})V_k, \;\; k \geq 2,$$

where V_k are independent $Beta(1, \alpha)$ random variables.



(2) Polya urn representation (Blackwell and MacQueen, 1973)

- $X_1 \sim G_0$
  and
- $X_{n+1} | X_1, \ldots X_n \sim \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^{n} \delta_{X_i}$.

(2) Polya urn representation (Blackwell and MacQueen, 1973)

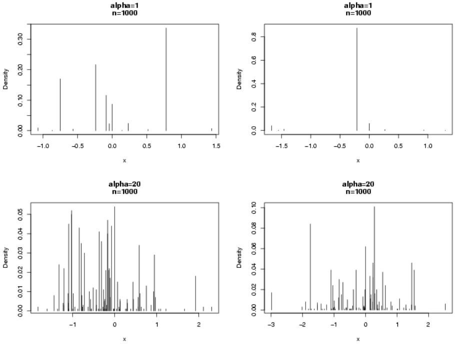
- $X_1 \sim G_0$
  and
- $X_{n+1}|X_1, \ldots X_n \sim \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^{n} \delta_{X_i}$.

(2) Polya urn representation (Blackwell and MacQueen, 1973)

- $X_1 \sim G_0$
  and
- $X_{n+1} | X_1, \ldots X_n \sim \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^{n} \delta_{X_i}.$

(2) Polya urn representation (Blackwell and MacQueen, 1973)

- $X_1 \sim G_0$
  and
- $X_{n+1}|X_1, \dots X_n \sim \frac{\alpha}{\alpha+n} G_0 + \frac{1}{\alpha+n} \sum_{i=1}^{n} \delta_{X_i}$.



Dirichlet Process Mixture models

- The Dirichlet Processes mixtures provide a fomal model to estimate the distribution of random variable Y given a sample $\{y_1, y_2, \ldots, y_n\}$ from an unknown distribution.
-
$$y_i|\theta_i \sim F(\theta_i)$$
$$\theta_i|G \sim G$$
$$G \sim DP(\alpha, G_0)$$
- Escobar (1994, 1995) provided the fundamental developement of the computational technique for the Dirichlet process mixture model for Normal and inverse gamma conjugate distributions.

Dirichlet Process Mixture models

- The Dirichlet Processes mixtures provide a fomal model to estimate the distribution of random variable Y given a sample $\{y_1, y_2, \ldots, y_n\}$ from an unknown distribution.
- $$y_i|\theta_i \sim F(\theta_i)$$ $$\theta_i|G \sim G$$ $$G \sim DP(\alpha, G_0)$$
- Escobar (1994, 1995) provided the fundamental developement of the computational technique for the Dirichlet process mixture model for Normal and inverse gamma conjugate distributions.

Dirichlet Process Mixture models

- The Dirichlet Processes mixtures provide a fomal model to estimate the distribution of random variable Y given a sample $\{y_1, y_2, \ldots, y_n\}$ from an unknown distribution.
-
$$y_i|\theta_i \sim F(\theta_i)$$
$$\theta_i|G \sim G$$
$$G \sim DP(\alpha, G_0)$$

- Escobar (1994, 1995) provided the fundamental developement of the computational technique for the Dirichlet process mixture model for Normal and inverse gamma conjugate distributions.

Dirichlet Process Mixture models

- The Dirichlet Processes mixtures provide a fomal model to estimate the distribution of random variable Y given a sample $\{y_1, y_2, \ldots, y_n\}$ from an unknown distribution.
- $$y_i|\theta_i \sim F(\theta_i)$$
$$\theta_i|G \sim G$$
$$G \sim DP(\alpha, G_0)$$
- Escobar (1994, 1995) provided the fundamental developement of the computational technique for the Dirichlet process mixture model for Normal and inverse gamma conjugate distributions.

Dirichlet Process Mixture Models

How we do a density estimation?

- Posterior predictive density of a future observation y_{n+1} is given by:

$$P(y_{n+1}|D) = \int P(y_{n+1}|\theta) dP(\theta|D)$$

- Computation is possible by MCMC (Markov Chain Monte Carlo) method.
  1. Generating the posterior samples of θ, $\theta^{(r)} = (\theta_1^{(r)}, \dots, \theta_n^{(r)})$
  2. Summing using Monte Carlo, $\hat{P}(y_{n+1}|D) = \frac{1}{R} \sum_{r=1}^{R} P(y_{n+1}|\theta^{(r)})$

How to obtain the Posterior samples of θ?

Dirichlet Process Mixture Models

How we do a density estimation?

- Posterior predictive density of a future observation y_{n+1} is given by:

$$P(y_{n+1}|D) = \int P(y_{n+1}|\theta) dP(\theta|D)$$

- Computation is possible by MCMC (Markov Chain Monte Carlo) method.
  1. Generating the posterior samples of θ, $\theta^{(r)} = (\theta_1^{(r)}, \ldots, \theta_n^{(r)})$
  2. Summing using Monte Carlo, $\hat{P}(y_{n+1}|D) = \frac{1}{R} \sum_{r=1}^{R} P(y_{n+1}|\theta^{(r)})$

How to obtain the Posterior samples of θ?

Dirichlet Process Mixture Models

How we do a density estimation?

- Posterior predictive density of a future observation y_{n+1} is given by:

$$P(y_{n+1}|D) = \int P(y_{n+1}|\theta) dP(\theta|D)$$

- Computation is possible by MCMC (Markov Chain Monte Carlo) method.
  1. Generating the posterior samples of θ, $\theta^{(r)} = (\theta_1^{(r)}, \ldots, \theta_n^{(r)})$
  2. Summing using Monte Carlo, $\hat{P}(y_{n+1}|D) = \frac{1}{R} \sum_{r=1}^{R} P(y_{n+1}|\theta^{(r)})$

How to obtain the Posterior samples of θ?

Dirichlet Process Mixture Models

How we do a density estimation?

- Posterior predictive density of a future observation y_{n+1} is given by:

$$P(y_{n+1}|D) = \int P(y_{n+1}|\theta) dP(\theta|D)$$

- Computation is possible by MCMC (Markov Chain Monte Carlo) method.
  1. Generating the posterior samples of θ, $\theta^{(r)} = (\theta_1^{(r)}, \ldots, \theta_n^{(r)})$
  2. Summing using Monte Carlo, $\hat{P}(y_{n+1}|D) = \frac{1}{R}\sum_{r=1}^{R} P(y_{n+1}|\theta^{(r)})$

How to obtain the Posterior samples of θ?

Dirichlet Process Mixture Models

How we do a density estimation?

- Posterior predictive density of a future observation y_{n+1} is given by:

$$P(y_{n+1}|D) = \int P(y_{n+1}|\theta) dP(\theta|D)$$

- Computation is possible by MCMC (Markov Chain Monte Carlo) method.
  1. Generating the posterior samples of θ, $\theta^{(r)} = (\theta_1^{(r)}, \ldots, \theta_n^{(r)})$
  2. Summing using Monte Carlo, $\hat{P}(y_{n+1}|D) = \frac{1}{R} \sum_{r=1}^{R} P(y_{n+1}|\theta^{(r)})$

How to obtain the Posterior samples of θ?

Dirichlet Process Mixture Models

How we do a density estimation?

- Posterior predictive density of a future observation y_{n+1} is given by:

$$P(y_{n+1}|D) = \int P(y_{n+1}|\theta) dP(\theta|D)$$

- Computation is possible by MCMC (Markov Chain Monte Carlo) method.
  1. Generating the posterior samples of θ, $\theta^{(r)} = (\theta_1^{(r)}, \ldots, \theta_n^{(r)})$
  2. Summing using Monte Carlo, $\hat{P}(y_{n+1}|D) = \frac{1}{R} \sum_{r=1}^{R} P(y_{n+1}|\theta^{(r)})$

How to obtain the Posterior samples of θ?

Dirichlet Process Mixture Models

How we do a density estimation?

- Posterior predictive density of a future observation y_{n+1} is given by:

$$P(y_{n+1}|D) = \int P(y_{n+1}|\theta)dP(\theta|D)$$

- Computation is possible by MCMC (Markov Chain Monte Carlo) method.
  1. Generating the posterior samples of θ, $\theta^{(r)} = (\theta_1^{(r)}, \ldots, \theta_n^{(r)})$
  2. Summing using Monte Carlo, $\hat{P}(y_{n+1}|D) = \frac{1}{R}\sum_{r=1}^{R} P(y_{n+1}|\theta^{(r)})$

How to obtain the Posterior samples of θ?

Dirichlet Process Mixture Models

How we do a density estimation?

- Posterior predictive density of a future observation y_{n+1} is given by:

$$P(y_{n+1}|D) = \int P(y_{n+1}|\theta)dP(\theta|D)$$

- Computation is possible by MCMC (Markov Chain Monte Carlo) method.
  1. Generating the posterior samples of θ, $\theta^{(r)} = (\theta_1^{(r)}, \ldots, \theta_n^{(r)})$
  2. Summing using Monte Carlo, $\hat{P}(y_{n+1}|D) = \frac{1}{R}\sum_{r=1}^{R} P(y_{n+1}|\theta^{(r)})$

How to obtain the Posterior samples of θ?

MDP Computations

- Conditional on $\theta^{(-i)} = \{\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\}$, θ_i has the following mixing distribution:

$$\pi(\theta_i|\theta^{(-i)}, y) \propto q_{i0} g_0(\theta_i) f(y_i|\theta_i) + \sum_{j \neq i} q_{ij} \delta_{\theta_j}.$$

- The mixing weights are:

$$q_{i0} \propto \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i \quad \text{and} \quad q_{ij} = f(y_i|\theta_j).$$

- The Gibbs sampling steps will be as follows:

Step 1. Choose a starting value for θ.

Step 2. Sample an element of θ sequentially by drawing from the distribution. $(\theta_1|\theta^{(-1)}, y)$ then $(\theta_2|\theta^{(-2)}, y)$ and so on up to $(\theta_n|\theta^{(-n)}, y)$.

Step 3. Return step 2 until it converges.

MDP Computations

- Conditional on $\theta^{(-i)} = \{\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\}$, θ_i has the following mixing distribution:

$$\pi(\theta_i|\theta^{(-i)}, y) \propto q_{i0} g_0(\theta_i) f(y_i|\theta_i) + \sum_{j \neq i} q_{ij} \delta_{\theta_j}.$$

- The mixing weights are:

$$q_{i0} \propto \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i \quad \text{and} \quad q_{ij} = f(y_i|\theta_j).$$

- The Gibbs sampling steps will be as follows:

Step 1. Choose a starting value for θ.

Step 2. Sample an element of θ sequentially by drawing from the distribution. $(\theta_1|\theta^{(-1)}, y)$ then $(\theta_2|\theta^{(-2)}, y)$ and so on up to $(\theta_n|\theta^{(-n)}, y)$.

Step 3. Return step 2 until it converges.

MDP Computations

- Conditional on $\theta^{(-i)} = \{\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\}$, θ_i has the following mixing distribution:

$$\pi(\theta_i | \theta^{(-i)}, y) \propto q_{i0} g_0(\theta_i) f(y_i | \theta_i) + \sum_{j \neq i} q_{ij} \delta_{\theta_j}.$$

- The mixing weights are:

$$q_{i0} \propto \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i \quad \text{and} \quad q_{ij} = f(y_i|\theta_j).$$

- The Gibbs sampling steps will be as follows:

Step 1. Choose a starting value for θ.

Step 2. Sample an element of θ sequentially by drawing from the distribution. $(\theta_1|\theta^{(-1)}, y)$ then $(\theta_2|\theta^{(-2)}, y)$ and so on up to $(\theta_n|\theta^{(-n)}, y)$.

Step 3. Return step 2 until it converges.

MDP Computations

- Conditional on $\theta^{(-i)} = \{\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\}$, θ_i has the following mixing distribution:

$$\pi(\theta_i|\theta^{(-i)}, y) \propto q_{i0} g_0(\theta_i) f(y_i|\theta_i) + \sum_{j \neq i} q_{ij} \delta_{\theta_j}.$$

- The mixing weights are:

$$q_{i0} \propto \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i \quad \text{and} \quad q_{ij} = f(y_i|\theta_j).$$

- The Gibbs sampling steps will be as follows:

Step 1. Choose a starting value for θ.

Step 2. Sample an element of θ sequentially by drawing from the distribution. $(\theta_1|\theta^{(-1)}, y)$ then $(\theta_2|\theta^{(-2)}, y)$ and so on up to $(\theta_n|\theta^{(-n)}, y)$.

Step 3. Return step 2 until it converges.

MDP Computations

- Conditional on $\theta^{(-i)} = \{\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\}$, θ_i has the following mixing distribution:

$$\pi(\theta_i|\theta^{(-i)}, y) \propto q_{i0} g_0(\theta_i) f(y_i|\theta_i) + \sum_{j \neq i} q_{ij} \delta_{\theta_j}.$$

- The mixing weights are:

$$q_{i0} \propto \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i \quad \text{and} \quad q_{ij} = f(y_i|\theta_j).$$

- The Gibbs sampling steps will be as follows:

Step 1. Choose a starting value for θ.

Step 2. Sample an element of θ sequentially by drawing from the distribution. $(\theta_1|\theta^{(-1)}, y)$ then $(\theta_2|\theta^{(-2)}, y)$ and so on up to $(\theta_n|\theta^{(-n)}, y)$.

Step 3. Return step 2 until it converges.

MDP Computations

- Conditional on $\theta^{(-i)} = \{\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\}$, θ_i has the following mixing distribution:

$$\pi(\theta_i|\theta^{(-i)}, y) \propto q_{i0} g_0(\theta_i) f(y_i|\theta_i) + \sum_{j \neq i} q_{ij} \delta_{\theta_j}.$$

- The mixing weights are:

$$q_{i0} \propto \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i \quad \text{and} \quad q_{ij} = f(y_i|\theta_j).$$

- The Gibbs sampling steps will be as follows:

Step 1. Choose a starting value for θ.

Step 2. Sample an element of θ sequentially by drawing from the distribution. $(\theta_1|\theta^{(-1)}, y)$ then $(\theta_2|\theta^{(-2)}, y)$ and so on up to $(\theta_n|\theta^{(-n)}, y)$.

Step 3. Return step 2 until it converges.

MDP Computations

- Conditional on $\theta^{(-i)} = \{\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\}$, θ_i has the following mixing distribution:

$$\pi(\theta_i|\theta^{(-i)}, y) \propto q_{i0} g_0(\theta_i) f(y_i|\theta_i) + \sum_{j \neq i} q_{ij} \delta_{\theta_j}.$$

- The mixing weights are:

$$q_{i0} \propto \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i \quad \text{and} \quad q_{ij} = f(y_i|\theta_j).$$

- The Gibbs sampling steps will be as follows:

Step 1. Choose a starting value for θ.

Step 2. Sample an element of θ sequentially by drawing from the distribution. $(\theta_1|\theta^{(-1)}, y)$ then $(\theta_2|\theta^{(-2)}, y)$ and so on up to $(\theta_n|\theta^{(-n)}, y)$.

Step 3. Return step 2 until it converges.

MDP Computations

- Conditional on $\theta^{(-i)} = \{\theta_1, \theta_2, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_n\}$, θ_i has the following mixing distribution:

$$\pi(\theta_i|\theta^{(-i)}, y) \propto q_{i0} g_0(\theta_i) f(y_i|\theta_i) + \sum_{j \neq i} q_{ij} \delta_{\theta_j}.$$

- The mixing weights are:

$$q_{i0} \propto \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i \quad \text{and} \quad q_{ij} = f(y_i|\theta_j).$$

- The Gibbs sampling steps will be as follows:

Step 1. Choose a starting value for θ.

Step 2. Sample an element of θ sequentially by drawing from the distribution. $(\theta_1|\theta^{(-1)}, y)$ then $(\theta_2|\theta^{(-2)}, y)$ and so on up to $(\theta_n|\theta^{(-n)}, y)$.

Step 3. Return step 2 until it converges.

MDP Computations

- Conditional on $\theta^{(-i)} = \{\theta_1, \theta_2, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\}$, θ_i has the following mixing distribution:

$$\pi(\theta_i|\theta^{(-i)}, y) \propto q_{i0} g_0(\theta_i) f(y_i|\theta_i) + \sum_{j \neq i} q_{ij} \delta_{\theta_j}.$$

- The mixing weights are:

$$q_{i0} \propto \alpha \int f(y_i|\theta_i) g_0(\theta_i) d\theta_i \quad \text{and} \quad q_{ij} = f(y_i|\theta_j).$$

- The Gibbs sampling steps will be as follows:

Step 1. Choose a starting value for θ.

Step 2. Sample an element of θ sequentially by drawing from the distribution. $(\theta_1|\theta^{(-1)}, y)$ then $(\theta_2|\theta^{(-2)}, y)$ and so on up to $(\theta_n|\theta^{(-n)}, y)$.

Step 3. Return step 2 until it converges.

Bayesian Curve Fitting using Multivariate Normal Mixtures

Bayesian Curve Fitting using Multivariate Normal Mixture

Peter Muller, Alaattin Erkanli, and Mike West, Biometrika, 1996

- Main idea: instead of modeling the random function g, the nonparametric regression problem can be reduced to a density estimation problem,
  if we consider the joint distribution $(x_i, y_i) \sim F$, and then obtaining $g_F(x) = E_F(y|x)$.
- Model Structure

$$z_i = (y_i, x_{i1}, \ldots, x_{ip}) \sim N(\mu_i, \Sigma_i)$$
$$\theta_i = (\mu_i, \Sigma_i) \sim G$$
$$G \sim DP(\alpha, G_0)$$

- Also, additional level of hierarchy was added on parameter α and G_0.

Bayesian Curve Fitting using Multivariate Normal Mixture

Peter Muller, Alaattin Erkanli, and Mike West, Biometrika, 1996

- Main idea: instead of modeling the random function g, the nonparametric regression problem can be reduced to a density estimation problem,
  if we consider the joint distribution $(x_i, y_i) \sim F$, and then obtaining $g_F(x) = E_F(y|x)$.
- Model Structure

$$z_i = (y_i, x_{i1}, \ldots, x_{ip}) \sim N(\mu_i, \Sigma_i)$$

$$\theta_i = (\mu_i, \Sigma_i) \sim G$$

$$G \sim DP(\alpha, G_0)$$

- Also, additional level of hierarchy was added on parameter α and G_0.

Bayesian Curve Fitting using Multivariate Normal Mixture

Peter Muller, Alaattin Erkanli, and Mike West, Biometrika, 1996

- Main idea: instead of modeling the random function g, the nonparametric regression problem can be reduced to a density estimation problem,
  if we consider the joint distribution $(x_i, y_i) \sim F$, and then obtaining $g_F(x) = E_F(y|x)$.
- Model Structure

$$z_i = (y_i, x_{i1}, \ldots, x_{ip}) \sim N(\mu_i, \Sigma_i)$$

$$\theta_i = (\mu_i, \Sigma_i) \sim G$$

$$G \sim DP(\alpha, G_0)$$

- Also, additional level of hierarchy was added on parameter α and G_0.

Bayesian Curve Fitting using Multivariate Normal Mixture

Peter Muller, Alaattin Erkanli, and Mike West, Biometrika, 1996

- Main idea: instead of modeling the random function g, the nonparametric regression problem can be reduced to a density estimation problem,
  if we consider the joint distribution $(x_i, y_i) \sim F$, and then obtaining $g_F(x) = E_F(y|x)$.
- Model Structure

$$z_i = (y_i, x_{i1}, \ldots, x_{ip}) \sim N(\mu_i, \Sigma_i)$$

$$\theta_i = (\mu_i, \Sigma_i) \sim G$$

$$G \sim DP(\alpha, G_0)$$

- Also, additional level of hierarchy was added on parameter α and G_0.

Bayesian Curve Fitting using Multivariate Normal Mixture

Peter Muller, Alaattin Erkanli, and Mike West, Biometrika, 1996

- Main idea: instead of modeling the random function g, the nonparametric regression problem can be reduced to a density estimation problem,
  if we consider the joint distribution $(x_i, y_i) \sim F$, and then obtaining $g_F(x) = E_F(y|x)$.
- Model Structure

$$z_i = (y_i, x_{i1}, \ldots, x_{ip}) \sim N(\mu_i, \Sigma_i)$$

$$\theta_i = (\mu_i, \Sigma_i) \sim G$$

$$G \sim DP(\alpha, G_0)$$

- Also, additional level of hierarchy was added on parameter α and G_0.

Bayesian Curve Fitting using Multivariate Normal Mixture

Peter Muller, Alaattin Erkanli, and Mike West, Biometrika, 1996

- Main idea: instead of modeling the random function g, the nonparametric regression problem can be reduced to a density estimation problem,
  if we consider the joint distribution $(x_i, y_i) \sim F$, and then obtaining $g_F(x) = E_F(y|x)$.
- Model Structure

$$z_i = (y_i, x_{i1}, \ldots, x_{ip}) \sim N(\mu_i, \Sigma_i)$$

$$\theta_i = (\mu_i, \Sigma_i) \sim G$$

$$G \sim DP(\alpha, G_0)$$

- Also, additional level of hierarchy was added on parameter α and G_0.

Bayesian Curve Fitting using Multivariate Normal Mixture

Peter Muller, Alaattin Erkanli, and Mike West, Biometrika, 1996

- Main idea: instead of modeling the random function g, the nonparametric regression problem can be reduced to a density estimation problem,
  if we consider the joint distribution $(x_i, y_i) \sim F$, and then obtaining $g_F(x) = E_F(y|x)$.
- Model Structure

$$z_i = (y_i, x_{i1}, \ldots, x_{ip}) \sim N(\mu_i, \Sigma_i)$$

$$\theta_i = (\mu_i, \Sigma_i) \sim G$$

$$G \sim DP(\alpha, G_0)$$

- Also, additional level of hierarchy was added on parameter α and G_0.

Regression function Estimation

Suppose $z = (y_{n+1}, x_{n+1})$, and the goal is to estimate the the regression function $E(y|x)$.

$$p(z|\theta) = \frac{\alpha}{\alpha + n} \int f(z|\theta) dG_0(\theta) + \frac{1}{\alpha + n} \sum_{j=1}^{k} n_j f(z|\theta_j^*)$$

Here, $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ are distinct values and n_j is for the number of occurrences $\theta_i = \theta_j^*$.

- Therefore, conditioning on x implies $p(y|x,\ \theta^*) = w_0 p_0(y|x,\ \theta^*) + \sum_{j=1}^{k} w_j(x) f_j(y|x,\ \theta^*)$, where $w_j(x)$ are functions of the marginal densities of x.
- Take an expectation, then $E(y|x, \theta) = \sum_{j=0}^{k} w_j(x) l_j(x)$. ,where $l_j(x)$ is the mean of jth component for y given x.

Regression function Estimation

Suppose $z = (y_{n+1}, x_{n+1})$, and the goal is to estimate the the regression function $E(y|x)$.

$$p(z|\theta) = \frac{\alpha}{\alpha+n}\int f(z|\theta)dG_0(\theta) + \frac{1}{\alpha+n}\sum_{j=1}^{k} n_j f(z|\theta_j^*)$$

Here, $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ are distinct values and n_j is for the number of occurrences $\theta_i = \theta_j^*$.

- Therefore, conditioning on x implies $p(y|x,\ \theta^*) = w_0 p_0(y|x,\ \theta^*) + \sum_{j=1}^{k} w_j(x) f_j(y|x,\ \theta^*)$, where $w_j(x)$ are functions of the marginal densities of x.
- Take an expectation, then $E(y|x,\theta) = \sum_{j=0}^{k} w_j(x) I_j(x)$. ,where $I_j(x)$ is the mean of jth component for y given x.

Regression function Estimation

Suppose $z = (y_{n+1}, x_{n+1})$, and the goal is to estimate the the regression function $E(y|x)$.

$$p(z|\theta) = \frac{\alpha}{\alpha+n}\int f(z|\theta)dG_0(\theta) + \frac{1}{\alpha+n}\sum_{j=1}^{k} n_j f(z|\theta_j^*)$$

Here, $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ are distinct values and n_j is for the number of occurrences $\theta_i = \theta_j^*$.

- Therefore, conditioning on x implies $p(y|x,\ \theta^*) = w_0 p_0(y|x,\ \theta^*) + \sum_{j=1}^{k} w_j(x) f_j(y|x,\ \theta^*)$, where $w_j(x)$ are functions of the marginal densities of x.
- Take an expectation, then $E(y|x,\theta) = \sum_{j=0}^{k} w_j(x) l_j(x)$. ,where $l_j(x)$ is the mean of jth component for y given x.

Regression function Estimation

Suppose $z = (y_{n+1}, x_{n+1})$, and the goal is to estimate the the regression function $E(y|x)$.

$$p(z|\theta) = \frac{\alpha}{\alpha+n}\int f(z|\theta)dG_0(\theta) + \frac{1}{\alpha+n}\sum_{j=1}^{k} n_j f(z|\theta_j^*)$$

Here, $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ are distinct values and n_j is for the number of occurrences $\theta_i = \theta_j^*$.

- Therefore, conditioning on x implies $p(y|x,\ \theta^*) = w_0 p_0(y|x,\ \theta^*) + \sum_{j=1}^{k} w_j(x) f_j(y|x,\ \theta^*)$, where $w_j(x)$ are functions of the marginal densities of x.
- Take an expectation, then $E(y|x,\theta) = \sum_{j=0}^{k} w_j(x) l_j(x)$. ,where $l_j(x)$ is the mean of jth component for y given x.

Regression function Estimation

Suppose $z = (y_{n+1}, x_{n+1})$, and the goal is to estimate the the regression function $E(y|x)$.

$$p(z|\theta) = \frac{\alpha}{\alpha+n}\int f(z|\theta)dG_0(\theta) + \frac{1}{\alpha+n}\sum_{j=1}^{k} n_j f(z|\theta_j^*)$$

Here, $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ are distinct values and n_j is for the number of occurrences $\theta_i = \theta_j^*$.

- Therefore, conditioning on x implies $p(y|x,\ \theta^*) = w_0 p_0(y|x,\ \theta^*) + \sum_{j=1}^{k} w_j(x) f_j(y|x,\ \theta^*)$, where $w_j(x)$ are functions of the marginal densities of x.
- Take an expectation, then $E(y|x,\theta) = \sum_{j=0}^{k} w_j(x) l_j(x)$. ,where $l_j(x)$ is the mean of jth component for y given x.

Regression function Estimation

Suppose $z = (y_{n+1}, x_{n+1})$, and the goal is to estimate the the regression function $E(y|x)$.

$$p(z|\theta) = \frac{\alpha}{\alpha+n}\int f(z|\theta)dG_0(\theta) + \frac{1}{\alpha+n}\sum_{j=1}^{k} n_j f(z|\theta_j^*)$$

Here, $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ are distinct values and n_j is for the number of occurrences $\theta_i = \theta_j^*$.

- Therefore, conditioning on x implies $p(y|x,\ \theta^*) = w_0 p_0(y|x,\ \theta^*) + \sum_{j=1}^{k} w_j(x) f_j(y|x,\ \theta^*)$, where $w_j(x)$ are functions of the marginal densities of x.
- Take an expectation, then $E(y|x,\theta) = \sum_{j=0}^{k} w_j(x) l_j(x)$. ,where $l_j(x)$ is the mean of jth component for y given x.

Regression function Estimation

Suppose $z = (y_{n+1}, x_{n+1})$, and the goal is to estimate the the regression function $E(y|x)$.

$$p(z|\theta) = \frac{\alpha}{\alpha+n}\int f(z|\theta)dG_0(\theta) + \frac{1}{\alpha+n}\sum_{j=1}^{k} n_j f(z|\theta_j^*)$$

Here, $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ are distinct values and n_j is for the number of occurrences $\theta_i = \theta_j^*$.

- Therefore, conditioning on x implies $p(y|x,\ \theta^*) = w_0 p_0(y|x,\ \theta^*) + \sum_{j=1}^{k} w_j(x) f_j(y|x,\ \theta^*)$, where $w_j(x)$ are functions of the marginal densities of x.
- Take an expectation, then $E(y|x,\theta) = \sum_{j=0}^{k} w_j(x) l_j(x)$. ,where $l_j(x)$ is the mean of jth component for y given x.

Regression function Estimation

Suppose $z = (y_{n+1}, x_{n+1})$, and the goal is to estimate the the regression function $E(y|x)$.

$$p(z|\theta) = \frac{\alpha}{\alpha+n}\int f(z|\theta)dG_0(\theta) + \frac{1}{\alpha+n}\sum_{j=1}^{k} n_j f(z|\theta_j^*)$$

Here, $\theta^* = (\theta_1^*, \ldots, \theta_k^*)$ are distinct values and n_j is for the number of occurrences $\theta_i = \theta_j^*$.

- Therefore, conditioning on x implies $p(y|x,\ \theta^*) = w_0 p_0(y|x,\ \theta^*) + \sum_{j=1}^{k} w_j(x) f_j(y|x,\ \theta^*)$, where $w_j(x)$ are functions of the marginal densities of x.
- Take an expectation, then $E(y|x,\theta) = \sum_{j=0}^{k} w_j(x) l_j(x)$. ,where $l_j(x)$ is the mean of jth component for y given x.

Bayesian Semi-parametric Logistic Regression

The model

- Let the response variable y be a Bernoulli random variable, and assume a continuous covariate x.

$$y_i | x_i \sim Bernoulli(H(\beta_{0i} + \beta_{1i} x_i))$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

where $H(u) = \frac{\exp(u)}{(1+\exp(-u))}$.

The model

- Let the response variable y be a Bernoulli random variable, and assume a continuous covariate x.

$$y_i|x_i \sim Bernoulli(H(\beta_{0i} + \beta_{1i}x_i))$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

where $H(u) = \frac{\exp(u)}{(1+\exp(-u))}$.

The model

- Let the response variable y be a Bernoulli random variable, and assume a continuous covariate x.

$$y_i|x_i \sim Bernoulli(H(\beta_{0i} + \beta_{1i}x_i))$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

where $H(u) = \frac{\exp(u)}{(1+\exp(-u))}$.

The model

- Let the response variable y be a Bernoulli random variable, and assume a continuous covariate x.

$$y_i | x_i \sim Bernoulli(H(\beta_{0i} + \beta_{1i} x_i))$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

where $H(u) = \frac{\exp(u)}{(1+\exp(-u))}$.

The model

- Let the response variable y be a Bernoulli random variable, and assume a continuous covariate x.

$$y_i | x_i \sim Bernoulli(H(\beta_{0i} + \beta_{1i} x_i))$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

where $H(u) = \frac{\exp(u)}{(1+\exp(-u))}$.

The model

- Let the response variable y be a Bernoulli random variable, and assume a continuous covariate x.

$$y_i|x_i \sim Bernoulli(H(\beta_{0i} + \beta_{1i}x_i))$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

where $H(u) = \frac{\exp(u)}{(1+\exp(-u))}$.

The model

- Let the response variable y be a Bernoulli random variable, and assume a continuous covariate x.

$$y_i|x_i \sim Bernoulli(H(\beta_{0i} + \beta_{1i}x_i))$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

where $H(u) = \frac{\exp(u)}{(1+\exp(-u))}$.

Specification of Parameters

Specification of prior mean G_0

- $$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix}, \Sigma\right) \text{ here, } \Sigma = \begin{bmatrix} \sigma^2_{\beta_0} & \rho_\beta \sigma_{\beta_0} \sigma_{\beta_1} \\ \rho_\beta \sigma_{\beta_0} \sigma_{\beta_1} & \sigma^2_{\beta_1} \end{bmatrix}$$

$$\mu_x | \tau_x \sim N(\mu_0, \tau_x^{-1} v_0)$$

$$\tau_x \sim Gamma(\frac{a}{2}, \frac{b}{2})$$

Specification of Parameters

Specification of prior mean G_0

- $$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix}, \Sigma \right) \text{ here, } \Sigma = \begin{bmatrix} \sigma^2_{\beta_0} & \rho_\beta \sigma_{\beta_0} \sigma_{\beta_1} \\ \rho_\beta \sigma_{\beta_0} \sigma_{\beta_1} & \sigma^2_{\beta_1} \end{bmatrix}$$

$$\mu_x | \tau_x \sim N(\mu_0, \tau_x^{-1} v_0)$$

$$\tau_x \sim Gamma(\frac{a}{2}, \frac{b}{2})$$

Specification of Parameters

Specification of prior mean G_0

- $$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix}, \Sigma\right) \text{ here, } \Sigma = \begin{bmatrix} \sigma^2_{\beta_0} & \rho_\beta \sigma_{\beta_0}\sigma_{\beta_1} \\ \rho_\beta \sigma_{\beta_0}\sigma_{\beta_1} & \sigma^2_{\beta_1} \end{bmatrix}$$

$$\mu_x | \tau_x \sim N(\mu_0, \tau_x^{-1} v_0)$$

$$\tau_x \sim Gamma(\frac{a}{2}, \frac{b}{2})$$

Specification of Parameters

Specification of prior mean G_0

- $$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix}, \Sigma\right) \text{ here, } \Sigma = \begin{bmatrix} \sigma^2_{\beta_0} & \rho_\beta \sigma_{\beta_0}\sigma_{\beta_1} \\ \rho_\beta \sigma_{\beta_0}\sigma_{\beta_1} & \sigma^2_{\beta_1} \end{bmatrix}$$

$$\mu_x | \tau_x \sim N(\mu_0, \tau_x^{-1} v_0)$$

$$\tau_x \sim Gamma(\frac{a}{2}, \frac{b}{2})$$

Specification of Parameters

Specification of prior mean G_0

- $$\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix}, \Sigma\right) \text{ here, } \Sigma = \begin{bmatrix} \sigma^2_{\beta_0} & \rho_\beta \sigma_{\beta_0} \sigma_{\beta_1} \\ \rho_\beta \sigma_{\beta_0} \sigma_{\beta_1} & \sigma^2_{\beta_1} \end{bmatrix}$$

$$\mu_x | \tau_x \sim N(\mu_0, \tau_x^{-1} v_0)$$

$$\tau_x \sim Gamma(\frac{a}{2}, \frac{b}{2})$$

Specification of Hyper-parameters

Add one more hierarchy to the model:

- $$\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^*_{\beta_0} \\ \mu^*_{\beta_1} \end{pmatrix}, A\right), \quad \text{here, } A = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\Sigma^{-1} \sim W(b, (bB)^{-1})$$

$$\mu_0 \sim N(m, V)$$

$$v_0^{-1} \sim Gamma(\frac{w}{2}, \frac{W}{2}).$$

Here W is a Wishart distribution with degree of freedom b and scale matrix B.

Specification of Hyper-parameters

Add one more hierarchy to the model:

- $$\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^*_{\beta_0} \\ \mu^*_{\beta_1} \end{pmatrix}, A \right), \quad \text{here, } A = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_2 & \sigma_2^2 \end{bmatrix}$$

  $$\Sigma^{-1} \sim W(b, (bB)^{-1})$$

  $$\mu_0 \sim N(m, V)$$

  $$v_0^{-1} \sim Gamma(\frac{w}{2}, \frac{W}{2}).$$

Here W is a Wishart distribution with degree of freedom b and scale matrix B.

Specification of Hyper-parameters

Add one more hierarchy to the model:

- $$\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^*_{\beta_0} \\ \mu^*_{\beta_1} \end{pmatrix}, A\right), \quad \text{here, } A = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\Sigma^{-1} \sim W(b, (bB)^{-1})$$

$$\mu_0 \sim N(m, V)$$

$$v_0^{-1} \sim Gamma(\frac{w}{2}, \frac{W}{2}).$$

Here W is a Wishart distribution with degree of freedom b and scale matrix B.

Specification of Hyper-parameters

Add one more hierarchy to the model:

- $$\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^*_{\beta_0} \\ \mu^*_{\beta_1} \end{pmatrix}, A\right), \quad \text{here, } A = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_2 & \sigma_2^2 \end{bmatrix}$$

  $$\Sigma^{-1} \sim W(b, (bB)^{-1})$$

  $$\mu_0 \sim N(m, V)$$

  $$v_0^{-1} \sim Gamma(\frac{w}{2}, \frac{W}{2}).$$

Here W is a Wishart distribution with degree of freedom b and scale matrix B.

Specification of Hyper-parameters

Add one more hierarchy to the model:

- $$\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^*_{\beta_0} \\ \mu^*_{\beta_1} \end{pmatrix}, A \right), \quad \text{here, } A = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\Sigma^{-1} \sim W(b, (bB)^{-1})$$

$$\mu_0 \sim N(m, V)$$

$$v_0^{-1} \sim Gamma(\frac{w}{2}, \frac{W}{2}).$$

Here W is a Wishart distribution with degree of freedom b and scale matrix B.

Specification of Hyper-parameters

Add one more hierarchy to the model:

- $$\begin{pmatrix} \mu_{\beta_0} \\ \mu_{\beta_1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu^*_{\beta_0} \\ \mu^*_{\beta_1} \end{pmatrix}, A\right), \quad \text{here, } A = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_2\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\Sigma^{-1} \sim W(b, (bB)^{-1})$$

$$\mu_0 \sim N(m, V)$$

$$v_0^{-1} \sim Gamma(\frac{w}{2}, \frac{W}{2}).$$

Here W is a Wishart distribution with degree of freedom b and scale matrix B.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of primary Parameters

- $(\tau_{x_i}|D_i, \mu_0, v_0) \sim Gamma(\frac{a^*}{2}, \frac{b^*}{2})$
- $(\mu_{x_i}|\tau_{x_i}, D_i, \mu_0, v_0) \sim N(\mu_0^*, \tau_x^{-1}\frac{v_0}{(v_0+1)})$
- Here $a^* = (a+1)$, $b^* = (b + \frac{(x_i-\mu_0)^2}{(1+v_0)})$, and $\mu_0^* = \frac{(\mu_0+v_0x_i)}{(1+v_0)}$.
- However, $(\beta_0|\beta_1, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_0|\beta_1, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$, and respectively,
  $(\beta_1|\beta_0, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_1|\beta_0, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$.
  do not have mathematically explicit posterior distributions.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of primary Parameters

- $(\tau_{x_i}|D_i, \mu_0, v_0) \sim Gamma(\frac{a^*}{2}, \frac{b^*}{2})$
- $(\mu_{x_i}|\tau_{x_i}, D_i, \mu_0, v_0) \sim N(\mu_0^*, \tau_x^{-1}\frac{v_0}{(v_0+1)})$
- Here $a^* = (a+1)$, $b^* = (b + \frac{(x_i-\mu_0)^2}{(1+v_0)})$, and $\mu_0^* = \frac{(\mu_0+v_0x_i)}{(1+v_0)}$.
- However, $(\beta_0|\beta_1, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_0|\beta_1, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$, and respectively,
  $(\beta_1|\beta_0, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_1|\beta_0, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$.
  do not have mathematically explicit posterior distributions.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of primary Parameters

- $(\tau_{x_i}|D_i, \mu_0, v_0) \sim Gamma(\frac{a^*}{2}, \frac{b^*}{2})$
- $(\mu_{x_i}|\tau_{x_i}, D_i, \mu_0, v_0) \sim N(\mu_0^*, \tau_x^{-1}\frac{v_0}{(v_0+1)})$
- Here $a^* = (a+1)$, $b^* = (b + \frac{(x_i-\mu_0)^2}{(1+v_0)})$, and $\mu_0^* = \frac{(\mu_0+v_0 x_i)}{(1+v_0)}$.
- However, $(\beta_0|\beta_1, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_0|\beta_1, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$, and respectively, $(\beta_1|\beta_0, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_1|\beta_0, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$. do not have mathematically explicit posterior distributions.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of primary Parameters

- $(\tau_{x_i}|D_i, \mu_0, v_0) \sim Gamma(\frac{a^*}{2}, \frac{b^*}{2})$
- $(\mu_{x_i}|\tau_{x_i}, D_i, \mu_0, v_0) \sim N(\mu_0^*, \tau_x^{-1}\frac{v_0}{(v_0+1)})$
- Here $a^* = (a+1)$, $b^* = (b + \frac{(x_i-\mu_0)^2}{(1+v_0)})$, and $\mu_0^* = \frac{(\mu_0+v_0x_i)}{(1+v_0)}$.
- However, $(\beta_0|\beta_1, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_0|\beta_1, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$, and respectively,
  $(\beta_1|\beta_0, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_1|\beta_0, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$.
  do not have mathematically explicit posterior distributions.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of primary Parameters

- $(\tau_{x_i}|D_i, \mu_0, v_0) \sim Gamma(\frac{a^*}{2}, \frac{b^*}{2})$
- $(\mu_{x_i}|\tau_{x_i}, D_i, \mu_0, v_0) \sim N(\mu_0^*, \tau_x^{-1} \frac{v_0}{(v_0+1)})$
- Here $a^* = (a+1)$, $b^* = (b + \frac{(x_i - \mu_0)^2}{(1+v_0)})$, and $\mu_0^* = \frac{(\mu_0 + v_0 x_i)}{(1+v_0)}$.
- However, $(\beta_0|\beta_1, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_0|\beta_1, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) f(D)$, and respectively, $(\beta_1|\beta_0, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_1|\beta_0, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) f(D)$. do not have mathematically explicit posterior distributions.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of primary Parameters

- $(\tau_{x_i}|D_i, \mu_0, v_0) \sim Gamma(\frac{a^*}{2}, \frac{b^*}{2})$
- $(\mu_{x_i}|\tau_{x_i}, D_i, \mu_0, v_0) \sim N(\mu_0^*, \tau_x^{-1}\frac{v_0}{(v_0+1)})$
- Here $a^* = (a+1)$, $b^* = (b + \frac{(x_i-\mu_0)^2}{(1+v_0)})$, and $\mu_0^* = \frac{(\mu_0+v_0x_i)}{(1+v_0)}$.
- However, $(\beta_0|\beta_1, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_0|\beta_1, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$, and respectively,
  $(\beta_1|\beta_0, D, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma) \propto f(\beta_1|\beta_0, \mu_{\beta_0}, \mu_{\beta_1}, \Sigma)f(D)$.
  do not have mathematically explicit posterior distributions.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of Hyper-Parameters

$\beta = (\beta_0, \beta_1)^T$, $\mu_\beta = (\mu_{\beta_0}, \mu_{\beta_1})^T$ then

- $(\mu_\beta | \beta, \Sigma, A) \sim N(a^*, A^*)$,
  here $A^* = A^{-1} + n\Sigma^{-1}$, $a^* = A^*(A^{-1}\mu_\beta + n\Sigma^{-1}\bar{\beta})$, and
  $\bar{\beta} = (\frac{\sum \beta_0}{n}, \frac{\sum \beta_1}{n})^T$.
- $(\Sigma^{-1} | \beta, \mu_\beta) \sim W(b + n, B^*)$,
  where $B^* = (bB + \sum(\beta - \mu_\beta)(\beta - \mu_\beta)^T)^{-1}$.
- $(\mu_0 | v_0, \mu_x, \tau_x, m, V) \sim N(m^*, V^*)$,
  where $m^* = (1-t)m + t\frac{\sum_{i=1}^n \tau_x \mu_x}{\sum \tau_x}$, $t = \frac{V}{(V + \frac{v_0}{\sum \tau_x})}$, $V^* = \frac{t v_0}{\sum \tau_x}$.
- $(v_0^{-1} | \mu_0, \mu_x, \tau_x, w, W) \sim Gamma(\frac{(w+n)}{2}, \frac{(W+N)}{2})$,
  where $N = \sum \frac{(\mu_x - \mu_0)^2}{\tau_x}$.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of Hyper-Parameters

$\beta = (\beta_0, \beta_1)^T$, $\mu_\beta = (\mu_{\beta_0}, \mu_{\beta_1})^T$ then

- $(\mu_\beta | \beta, \Sigma, A) \sim N(a^*, A^*)$,
  here $A^* = A^{-1} + n\Sigma^{-1}$, $a^* = A^*(A^{-1}\mu_\beta + n\Sigma^{-1}\bar{\beta})$, and $\bar{\beta} = (\frac{\sum \beta_0}{n}, \frac{\sum \beta_1}{n})^T$.
- $(\Sigma^{-1} | \beta, \mu_\beta) \sim W(b + n, B^*)$,
  where $B^* = (bB + \sum(\beta - \mu_\beta)(\beta - \mu_\beta)^T)^{-1}$.
- $(\mu_0 | v_0, \mu_x, \tau_x, m, V) \sim N(m^*, V^*)$,
  where $m^* = (1 - t)m + t\frac{\sum_{i=1}^n \tau_x \mu_x}{\sum \tau_x}$, $t = \frac{V}{\left(V + \frac{v_0}{\sum \tau_x}\right)}$, $V^* = \frac{t v_0}{\sum \tau_x}$.
- $(v_0^{-1} | \mu_0, \mu_x, \tau_x, w, W) \sim Gamma(\frac{(w+n)}{2}, \frac{(W+N)}{2})$,
  where $N = \sum \frac{(\mu_x - \mu_0)^2}{\tau_x}$.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of Hyper-Parameters

$\beta = (\beta_0, \beta_1)^T$, $\mu_\beta = (\mu_{\beta_0}, \mu_{\beta_1})^T$ then

- $(\mu_\beta | \beta, \Sigma, A) \sim N(a^*, A^*)$,
  here $A^* = A^{-1} + n\Sigma^{-1}$, $a^* = A^*(A^{-1}\mu_\beta + n\Sigma^{-1}\bar{\beta})$, and $\bar{\beta} = (\frac{\sum \beta_0}{n}, \frac{\sum \beta_1}{n})^T$.
- $(\Sigma^{-1} | \beta, \mu_\beta) \sim W(b + n, B^*)$,
  where $B^* = (bB + \sum(\beta - \mu_\beta)(\beta - \mu_\beta)^T)^{-1}$.
- $(\mu_0 | v_0, \mu_x, \tau_x, m, V) \sim N(m^*, V^*)$,
  where $m^* = (1-t)m + t\frac{\sum_{i=1}^n \tau_x \mu_x}{\sum \tau_x}$, $t = \frac{V}{(V + \frac{v_0}{\sum \tau_x})}$, $V^* = \frac{t v_0}{\sum \tau_x}$.
- $(v_0^{-1} | \mu_0, \mu_x, \tau_x, w, W) \sim Gamma(\frac{(w+n)}{2}, \frac{(W+N)}{2})$,
  where $N = \sum \frac{(\mu_x - \mu_0)^2}{\tau_x}$.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of Hyper-Parameters

$\beta = (\beta_0, \beta_1)^T$, $\mu_\beta = (\mu_{\beta_0}, \mu_{\beta_1})^T$ then

- $(\mu_\beta | \beta, \Sigma, A) \sim N(a^*, A^*)$,
  here $A^* = A^{-1} + n\Sigma^{-1}$, $a^* = A^*(A^{-1}\mu_\beta + n\Sigma^{-1}\bar{\beta})$, and
  $\bar{\beta} = (\frac{\sum \beta_0}{n}, \frac{\sum \beta_1}{n})^T$.
- $(\Sigma^{-1} | \beta, \mu_\beta) \sim W(b + n, B^*)$,
  where $B^* = (bB + \sum(\beta - \mu_\beta)(\beta - \mu_\beta)^T)^{-1}$.
- $(\mu_0 | v_0, \mu_x, \tau_x, m, V) \sim N(m^*, V^*)$,
  where $m^* = (1 - t)m + t\frac{\sum_{i=1}^n \tau_x \mu_x}{\sum \tau_x}$, $t = \frac{V}{(V + \frac{v_0}{\sum \tau_x})}$, $V^* = \frac{tv_0}{\sum \tau_x}$.
- $(v_0^{-1} | \mu_0, \mu_x, \tau_x, w, W) \sim Gamma(\frac{(w+n)}{2}, \frac{(W+N)}{2})$,
  where $N = \sum \frac{(\mu_x - \mu_0)^2}{\tau_x}$.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of Hyper-Parameters

$\beta = (\beta_0, \beta_1)^T$, $\mu_\beta = (\mu_{\beta_0}, \mu_{\beta_1})^T$ then

- $(\mu_\beta | \beta, \Sigma, A) \sim N(a^*, A^*)$,
  here $A^* = A^{-1} + n\Sigma^{-1}$, $a^* = A^*(A^{-1}\mu_\beta + n\Sigma^{-1}\bar{\beta})$, and
  $\bar{\beta} = (\frac{\sum \beta_0}{n}, \frac{\sum \beta_1}{n})^T$.
- $(\Sigma^{-1} | \beta, \mu_\beta) \sim W(b + n, B^*)$,
  where $B^* = (bB + \sum(\beta - \mu_\beta)(\beta - \mu_\beta)^T)^{-1}$.
- $(\mu_0 | v_0, \mu_x, \tau_x, m, V) \sim N(m^*, V^*)$,
  where $m^* = (1-t)m + t\frac{\sum_{i=1}^n \tau_x \mu_x}{\sum \tau_x}$, $t = \frac{V}{(V + \frac{v_0}{\sum \tau_x})}$, $V^* = \frac{t v_0}{\sum \tau_x}$.
- $(v_0^{-1} | \mu_0, \mu_x, \tau_x, w, W) \sim Gamma(\frac{(w+n)}{2}, \frac{(W+N)}{2})$,
  where $N = \sum \frac{(\mu_x - \mu_0)^2}{\tau_x}$.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of Hyper-Parameters

$\beta = (\beta_0, \beta_1)^T$, $\mu_\beta = (\mu_{\beta_0}, \mu_{\beta_1})^T$ then

- $(\mu_\beta | \beta, \Sigma, A) \sim N(a^*, A^*)$,
  here $A^* = A^{-1} + n\Sigma^{-1}$, $a^* = A^*(A^{-1}\mu_\beta + n\Sigma^{-1}\bar{\beta})$, and
  $\bar{\beta} = (\frac{\sum \beta_0}{n}, \frac{\sum \beta_1}{n})^T$.
- $(\Sigma^{-1} | \beta, \mu_\beta) \sim W(b + n, B^*)$,
  where $B^* = (bB + \sum(\beta - \mu_\beta)(\beta - \mu_\beta)^T)^{-1}$.
- $(\mu_0 | v_0, \mu_x, \tau_x, m, V) \sim N(m^*, V^*)$,
  where $m^* = (1-t)m + t\frac{\sum_{i=1}^n \tau_x \mu_x}{\sum \tau_x}$, $t = \frac{V}{\left(V + \frac{v_0}{\sum \tau_x}\right)}$, $V^* = \frac{t v_0}{\sum \tau_x}$.
- $(v_0^{-1} | \mu_0, \mu_x, \tau_x, w, W) \sim Gamma(\frac{(w+n)}{2}, \frac{(W+N)}{2})$,
  where $N = \sum \frac{(\mu_x - \mu_0)^2}{\tau_x}$.

Conditional distributions of primary Parameters and Hyper-parameters

Conditional distributions of Hyper-Parameters

$\beta = (\beta_0, \beta_1)^T$, $\mu_\beta = (\mu_{\beta_0}, \mu_{\beta_1})^T$ then

- $(\mu_\beta | \beta, \Sigma, A) \sim N(a^*, A^*)$,
  here $A^* = A^{-1} + n\Sigma^{-1}$, $a^* = A^*(A^{-1}\mu_\beta + n\Sigma^{-1}\bar{\beta})$, and $\bar{\beta} = (\frac{\sum \beta_0}{n}, \frac{\sum \beta_1}{n})^T$.
- $(\Sigma^{-1} | \beta, \mu_\beta) \sim W(b+n, B^*)$,
  where $B^* = (bB + \sum(\beta - \mu_\beta)(\beta - \mu_\beta)^T)^{-1}$.
- $(\mu_0 | v_0, \mu_x, \tau_x, m, V) \sim N(m^*, V^*)$,
  where $m^* = (1-t)m + t\frac{\sum_{i=1}^n \tau_x \mu_x}{\sum \tau_x}$, $t = \frac{V}{\left(V + \frac{v_0}{\sum \tau_x}\right)}$, $V^* = \frac{t v_0}{\sum \tau_x}$.
- $(v_0^{-1} | \mu_0, \mu_x, \tau_x, w, W) \sim Gamma(\frac{(w+n)}{2}, \frac{(W+N)}{2})$,
  where $N = \sum \frac{(\mu_x - \mu_0)^2}{\tau_x}$.

Posterior distribution for α

Assume $\alpha \sim Gamma(a_0, b_0)$, then

1. $$(\alpha|D, \beta, \mu_x, \tau_x, k, \eta) \sim \pi_1 \Gamma\{a_0 + k, b_0 - \log(\eta)\} + \pi_2 \Gamma\{a_0 + k - 1, b_0 - \log(\eta)\}$$

   here,

   $$\pi_1 = \frac{(a_0 + k - 1)}{a_0 + k - 1 + n(b_0 - \log(\eta))}, \quad \pi_2 = 1 - \pi_1$$

2. $(\eta|D, \beta, \mu_x \tau_x, k, \alpha) \sim Beta(\alpha + 1, n)$

Here, k is the number of distinct components.

Posterior distribution for α

Assume $\alpha \sim Gamma(a_0, b_0)$, then

1. $$(\alpha | D, \beta, \mu_x, \tau_x, k, \eta) \sim \pi_1 \Gamma\{a_0 + k, b_0 - \log(\eta)\} + \pi_2 \Gamma\{a_0 + k - 1, b_0 - \log(\eta)\}$$

   here,

   $$\pi_1 = \frac{(a_0 + k - 1)}{a_0 + k - 1 + n(b_0 - \log(\eta))}, \ \pi_2 = 1 - \pi_1$$

2. $(\eta | D, \beta, \mu_x \tau_x, k, \alpha) \sim Beta(\alpha + 1, n)$

Here, k is the number of distinct components.

Posterior distribution for α

Assume $\alpha \sim Gamma(a_0, b_0)$, then

1. $$(\alpha|D, \beta, \mu_x, \tau_x, k, \eta) \sim \pi_1 \Gamma\{a_0 + k, b_0 - \log(\eta)\} + \pi_2 \Gamma\{a_0 + k - 1, b_0 - \log(\eta)\}$$

   here,

   $$\pi_1 = \frac{(a_0 + k - 1)}{a_0 + k - 1 + n(b_0 - \log(\eta))}, \quad \pi_2 = 1 - \pi_1$$

2. $(\eta|D, \beta, \mu_x \tau_x, k, \alpha) \sim Beta(\alpha + 1, n)$

Here, k is the number of distinct components.

Posterior distribution for α

Assume $\alpha \sim Gamma(a_0, b_0)$, then

1. $$(\alpha | D, \beta, \mu_x, \tau_x, k, \eta) \sim \pi_1 \Gamma\{a_0 + k, b_0 - \log(\eta)\} + \pi_2 \Gamma\{a_0 + k - 1, b_0 - \log(\eta)\}$$

   here,

   $$\pi_1 = \frac{(a_0 + k - 1)}{a_0 + k - 1 + n(b_0 - \log(\eta))}, \quad \pi_2 = 1 - \pi_1$$

2. $(\eta | D, \beta, \mu_x \tau_x, k, \alpha) \sim Beta(\alpha + 1, n)$

Here, k is the number of distinct components.

Computation of mixing weights

Reall from the Dirichlet Mixture Models:

$$(\theta_i|\theta^{-i}, D_n) \sim q_{0i} G_i(\theta_i|D_i) + \sum_{j=1, j\neq i}^{n} q_{ij}\delta_{\theta_j}(\theta_i)$$

Calculating Mixing weights:

$q_{i0} \propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i)$

$= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i)$

$$= \alpha \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{1-y_i} dF(\beta_0, \beta_1)$$

$$\cdot \int_{-\infty}^{\infty}\int_{0}^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x$$

Computation of mixing weights

Reall from the Dirichlet Mixture Models:

$$(\theta_i|\theta^{-i}, D_n) \sim q_{0i} G_i(\theta_i|D_i) + \sum_{j=1, j\neq i}^{n} q_{ij}\delta_{\theta_j}(\theta_i)$$

Calculating Mixing weights:

$$q_{i0} \propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i)$$

$$= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i)$$

$$= \alpha \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{1-y_i} dF(\beta_0, \beta_1)$$

$$\cdot \int_{-\infty}^{\infty}\int_{0}^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x$$

Computation of mixing weights

Reall from the Dirichlet Mixture Models:

$$(\theta_i|\theta^{-i}, D_n) \sim q_{0i} G_i(\theta_i|D_i) + \sum_{j=1, j\neq i}^{n} q_{ij}\delta_{\theta_j}(\theta_i)$$

Calculating Mixing weights:

$$q_{i0} \propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i)$$

$$= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i)$$

$$= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{1-y_i} dF(\beta_0, \beta_1)$$

$$\cdot \int_{-\infty}^{\infty} \int_{0}^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x$$

Computation of mixing weights

Reall from the Dirichlet Mixture Models:

$$(\theta_i|\theta^{-i}, D_n) \sim q_{0i} G_i(\theta_i|D_i) + \sum_{j=1, j\neq i}^{n} q_{ij}\delta_{\theta_j}(\theta_i)$$

Calculating Mixing weights:

$q_{i0} \propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i)$

$= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i)$

$$= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{1-y_i} dF(\beta_0, \beta_1)$$

$$\cdot \int_{-\infty}^{\infty} \int_{0}^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x$$

Computation of mixing weights

Reall from the Dirichlet Mixture Models:

$$(\theta_i|\theta^{-i}, D_n) \sim q_{0i} G_i(\theta_i|D_i) + \sum_{j=1, j\neq i}^{n} q_{ij}\delta_{\theta_j}(\theta_i)$$

Calculating Mixing weights:

$q_{i0} \propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i)$

$= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i)$

$$= \alpha \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{1-y_i} dF(\beta_0, \beta_1)$$

$$\cdot \int_{-\infty}^{\infty}\int_{0}^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x$$

Computation of mixing weights

Reall from the Dirichlet Mixture Models:

$$(\theta_i|\theta^{-i}, D_n) \sim q_{0i} G_i(\theta_i|D_i) + \sum_{j=1, j\neq i}^{n} q_{ij}\delta_{\theta_j}(\theta_i)$$

Calculating Mixing weights:

$q_{i0} \propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i)$

$= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i)$

$$= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{1-y_i} dF(\beta_0, \beta_1)$$

$$\cdot \int_{-\infty}^{\infty} \int_{0}^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x$$

Computation of mixing weights

Reall from the Dirichlet Mixture Models:

$$(\theta_i|\theta^{-i}, D_n) \sim q_{0i} G_i(\theta_i|D_i) + \sum_{j=1, j\neq i}^{n} q_{ij}\delta_{\theta_j}(\theta_i)$$

Calculating Mixing weights:

$$q_{i0} \propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i)$$

$$= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i)$$

$$= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{1-y_i} dF(\beta_0, \beta_1)$$

$$\cdot \int_{-\infty}^{\infty} \int_{0}^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x$$

Computation of mixing weights

Reall from the Dirichlet Mixture Models:

$$(\theta_i|\theta^{-i}, D_n) \sim q_{0i} G_i(\theta_i|D_i) + \sum_{j=1, j\neq i}^{n} q_{ij}\delta_{\theta_j}(\theta_i)$$

Calculating Mixing weights:

$q_{i0} \propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i)$

$= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i)$

$$= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(\frac{\exp(\beta_0+\beta_1 x_i)}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{y_i} \left(\frac{1}{1+\exp(\beta_0+\beta_1 x_i)}\right)^{1-y_i} dF(\beta_0, \beta_1)$$

$$\cdot \int_{-\infty}^{\infty} \int_{0}^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x$$

Complete Gibbs Sampling Schemes

1. Drawing a new $\theta_i, \quad i = 1, \ldots, n$ from the Dirichlet process. Either it takes old value such as $\theta_j, j \neq i$ or generates new value from the posterior of G_0 depending on the mixing weight q_{0i} and q_{ji}.
   Note that the posterior samples for β_0 and β_1 are obtained by ARS (Adaptive Rejection Sampling).
2. Remixing step: Drawing a new $\theta_j^*, j = 1, \ldots, k$, from its conditional distribution conditioned by the known number of clusters and the set of indices which maps the original data into k distinct groups or clusters.
3. Drawing new hyperparameters based on the latest parameter θ^*.
4. Drawing a Dirichlet process parameter $[\alpha|\theta^*]$ by first sampling $[\eta|\alpha \quad k]$ and then $[\alpha|\eta \quad k]$ where k is the number of distinct values in θ^*.

Complete Gibbs Sampling Schemes

1. Drawing a new $\theta_i, \quad i = 1, \ldots, n$ from the Dirichlet process. Either it takes old value such as $\theta_j, j \neq i$ or generates new value from the posterior of G_0 depending on the mixing weight q_{0i} and q_{ji}.
   Note that the posterior samples for β_0 and β_1 are obtained by ARS (Adaptive Rejection Sampling).
2. Remixing step: Drawing a new $\theta^*_j, j = 1, \ldots, k$, from its conditional distribution conditioned by the known number of clusters and the set of indices which maps the original data into k distinct groups or clusters.
3. Drawing new hyperparameters based on the latest parameter θ^*.
4. Drawing a Dirichlet process parameter $[\alpha|\theta^*]$ by first sampling $[\eta|\alpha \quad k]$ and then $[\alpha|\eta \quad k]$ where k is the number of distinct values in θ^*.

Complete Gibbs Sampling Schemes

1. Drawing a new θ_i, $i = 1, \ldots, n$ from the Dirichlet process. Either it takes old value such as $\theta_j, j \neq i$ or generates new value from the posterior of G_0 depending on the mixing weight q_{0i} and q_{ji}.
   Note that the posterior samples for β_0 and β_1 are obtained by ARS (Adaptive Rejection Sampling).
2. Remixing step: Drawing a new $\theta_j^*, j = 1, \ldots, k$, from its conditional distribution conditioned by the known number of clusters and the set of indices which maps the original data into k distinct groups or clusters.
3. Drawing new hyperparameters based on the latest parameter θ^*.
4. Drawing a Dirichlet process parameter $[\alpha|\theta^*]$ by first sampling $[\eta|\alpha \;\; k]$ and then $[\alpha|\eta \;\; k]$ where k is the number of distinct values in θ^*.

Complete Gibbs Sampling Schemes

1. Drawing a new $\theta_i, \quad i = 1, \ldots, n$ from the Dirichlet process. Either it takes old value such as $\theta_j, j \neq i$ or generates new value from the posterior of G_0 depending on the mixing weight q_{0i} and q_{ji}.
   Note that the posterior samples for β_0 and β_1 are obtained by ARS (Adaptive Rejection Sampling).
2. Remixing step: Drawing a new $\theta_j^*, j = 1, \ldots, k$, from its conditional distribution conditioned by the known number of clusters and the set of indices which maps the original data into k distinct groups or clusters.
3. Drawing new hyperparameters based on the latest parameter θ^*.
4. Drawing a Dirichlet process parameter $[\alpha|\theta^*]$ by first sampling $[\eta|\alpha \;\; k]$ and then $[\alpha|\eta \;\; k]$ where k is the number of distinct values in θ^*.

Complete Gibbs Sampling Schemes

1. Drawing a new $\theta_i, \quad i = 1, \ldots, n$ from the Dirichlet process. Either it takes old value such as $\theta_j, j \neq i$ or generates new value from the posterior of G_0 depending on the mixing weight q_{0i} and q_{ji}.
   Note that the posterior samples for β_0 and β_1 are obtained by ARS (Adaptive Rejection Sampling).
2. Remixing step: Drawing a new $\theta_j^*, j = 1, \ldots, k$, from its conditional distribution conditioned by the known number of clusters and the set of indices which maps the original data into k distinct groups or clusters.
3. Drawing new hyperparameters based on the latest parameter θ^*.
4. Drawing a Dirichlet process parameter $[\alpha|\theta^*]$ by first sampling $[\eta|\alpha \quad k]$ and then $[\alpha|\eta \quad k]$ where k is the number of distinct values in θ^*.

Logistic regression function estimation

Under the assumed structure, $P(D|\theta^*) = P(x|\theta^*)P(y|x,\theta^*)$.

$$P(y|x \ \ \theta^*) = W_0(x)\int P(y|x \ \theta^*)dG_0 + \sum_{j=1}^{k} W_j(x)P_j(y|x \ \theta^*)$$

Here, $W_0(x) = \frac{\int f(x|\theta^*)dG_0}{\int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

and $W_j(x) = \frac{n_j f(x|\theta^*)}{\alpha \int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

Therefore,

$$E(y|x \ \theta^*) = \sum_{j=0}^{k} W_j(x)\frac{\exp(\beta_{0j} + \beta_{1j}x)}{1 + \exp(\beta_{0j} + \beta_{1j}x)}$$

Logistic regression function estimation

Under the assumed structure, $P(D|\theta^*) = P(x|\theta^*)P(y|x, \theta^*)$.

$$P(y|x\ \ \theta^*) = W_0(x)\int P(y|x\ \theta^*)dG_0 + \sum_{j=1}^{k} W_j(x)P_j(y|x\ \theta^*)$$

Here, $W_0(x) = \frac{\int f(x|\theta^*)dG_0}{\int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

and $W_j(x) = \frac{n_j f(x|\theta^*)}{\alpha \int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

Therefore,

$$E(y|x\ \theta^*) = \sum_{j=0}^{k} W_j(x)\frac{\exp(\beta_{0j} + \beta_{1j}x)}{1 + \exp(\beta_{0j} + \beta_{1j}x)}$$

Logistic regression function estimation

Under the assumed structure, $P(D|\theta^*) = P(x|\theta^*)P(y|x, \theta^*)$.

$$P(y|x \;\; \theta^*) = W_0(x) \int P(y|x \; \theta^*) dG_0 + \sum_{j=1}^{k} W_j(x) P_j(y|x \; \theta^*)$$

Here, $W_0(x) = \frac{\int f(x|\theta^*)dG_0}{\int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

and $W_j(x) = \frac{n_j f(x|\theta^*)}{\alpha \int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

Therefore,

$$E(y|x \; \theta^*) = \sum_{j=0}^{k} W_j(x) \frac{\exp(\beta_{0j} + \beta_{1j}x)}{1 + \exp(\beta_{0j} + \beta_{1j}x)}$$

Logistic regression function estimation

Under the assumed structure, $P(D|\theta^*) = P(x|\theta^*)P(y|x,\theta^*)$.

$$P(y|x \;\; \theta^*) = W_0(x)\int P(y|x \; \theta^*)dG_0 + \sum_{j=1}^{k} W_j(x)P_j(y|x \; \theta^*)$$

Here, $W_0(x) = \frac{\int f(x|\theta^*)dG_0}{\int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

and $W_j(x) = \frac{n_j f(x|\theta^*)}{\alpha \int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

Therefore,

$$E(y|x \; \theta^*) = \sum_{j=0}^{k} W_j(x)\frac{\exp(\beta_{0j} + \beta_{1j}x)}{1+\exp(\beta_{0j} + \beta_{1j}x)}$$

Logistic regression function estimation

Under the assumed structure, $P(D|\theta^*) = P(x|\theta^*)P(y|x, \theta^*)$.

$$P(y|x \quad \theta^*) = W_0(x) \int P(y|x \ \theta^*) dG_0 + \sum_{j=1}^{k} W_j(x) P_j(y|x \ \theta^*)$$

Here, $W_0(x) = \frac{\int f(x|\theta^*)dG_0}{\int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

and $W_j(x) = \frac{n_j f(x|\theta^*)}{\alpha \int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

Therefore,

$$E(y|x \ \theta^*) = \sum_{j=0}^{k} W_j(x) \frac{\exp(\beta_{0j} + \beta_{1j}x)}{1 + \exp(\beta_{0j} + \beta_{1j}x)}$$

Logistic regression function estimation

Under the assumed structure, $P(D|\theta^*) = P(x|\theta^*)P(y|x,\theta^*)$.

$$P(y|x \quad \theta^*) = W_0(x)\int P(y|x \ \theta^*)dG_0 + \sum_{j=1}^{k} W_j(x)P_j(y|x \ \theta^*)$$

Here, $W_0(x) = \frac{\int f(x|\theta^*)dG_0}{\int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

and $W_j(x) = \frac{n_j f(x|\theta^*)}{\alpha \int f(x|\theta^*)dG_0 + \sum_{j=1}^{k} n_j f(x|\theta^*)}$

Therefore,

$$E(y|x \ \theta^*) = \sum_{j=0}^{k} W_j(x)\frac{\exp(\beta_{0j} + \beta_{1j}x)}{1 + \exp(\beta_{0j} + \beta_{1j}x)}$$

Simulated Examples

Example 1.

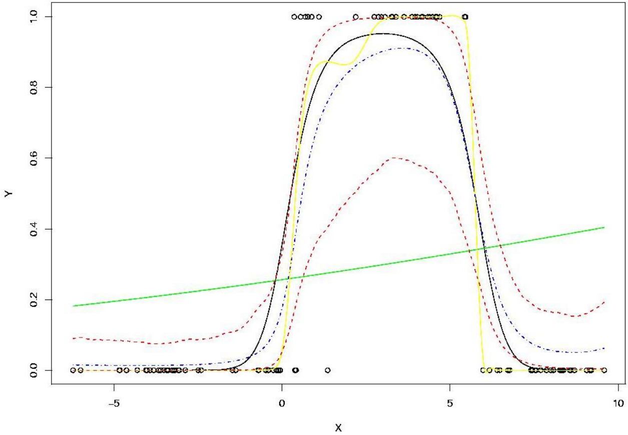
(1) $P(y=1|x) = \dfrac{\exp(-0.4(x-3)^2+3)}{1+\exp(-0.4(x-3)^2+3)}$, n=100

Example 1.

$$(1)\ P(y=1|x)=\frac{\exp(-0.4(x-3)^2+3)}{1+\exp(-0.4(x-3)^2+3)},\ \text{n=100}$$

Example 1.

$$(1)\ P(y=1|x) = \frac{\exp(-0.4(x-3)^2+3)}{1+\exp(-0.4(x-3)^2+3)},\ \text{n=100}$$



Example 2.

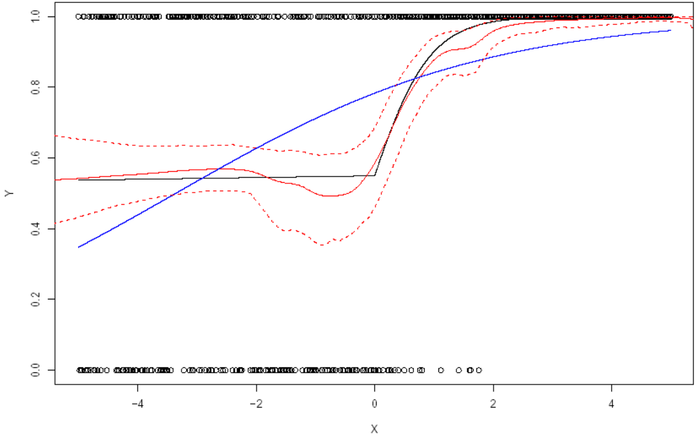
(2) $P(y=1|x)=\frac{\exp(0.2+0.01x)}{1+\exp(0.2+0.01x)}I(x\leq 0)+\frac{\exp(0.2+2x)}{1+\exp(0.2+2x)}I(x>0)$,
n=200

Example 2.

(2) $P(y=1|x) = \frac{\exp(0.2+0.01x)}{1+\exp(0.2+0.01x)} I(x \leq 0) + \frac{\exp(0.2+2x)}{1+\exp(0.2+2x)} I(x > 0)$,
n=200

Example 2.

(2) $P(y=1|x) = \frac{\exp(0.2+0.01x)}{1+\exp(0.2+0.01x)} I(x \le 0) + \frac{\exp(0.2+2x)}{1+\exp(0.2+2x)} I(x > 0)$, n=200



The performance of our method

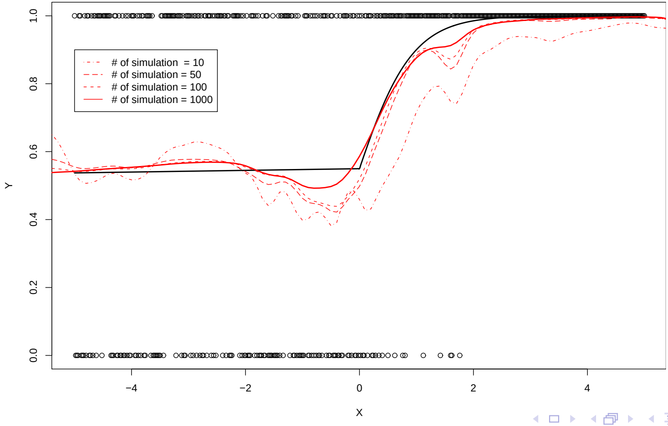
- Our method seems to converge to the target distribution very quickly.

The performance of our method

- Our method seems to converge to the target distribution very quickly.

The performance of our method

- Our method seems to converge to the target distribution very quickly.



Example 3.

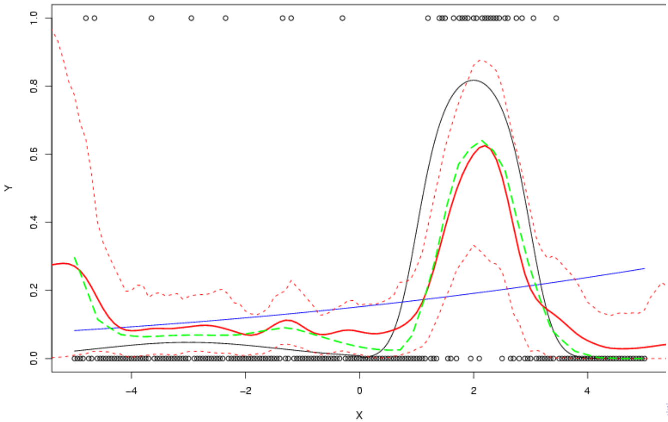
(3) $P(y=1|x) = \frac{\exp(-3-0.2(x+3)^2)}{1+\exp(-3-0.2(x+3)^2)} I(x \leq 0) + \frac{\exp(1.5-2(x-2)^2)}{1+\exp(1.5-2(x-2)^2)} I(x > 0)$, n=200

Example 3.

$$(3)\ P(y=1|x) = \frac{\exp(-3-0.2(x+3)^2)}{1+\exp(-3-0.2(x+3)^2)} I(x \leq 0) + \frac{\exp(1.5-2(x-2)^2)}{1+\exp(1.5-2(x-2)^2)} I(x > 0),\ \text{n=200}$$

Example 3.

$$(3)\ P(y=1|x) = \frac{\exp(-3-0.2(x+3)^2)}{1+\exp(-3-0.2(x+3)^2)} I(x \le 0) + \frac{\exp(1.5-2(x-2)^2)}{1+\exp(1.5-2(x-2)^2)} I(x > 0),\ \text{n=200}$$



Example 3.

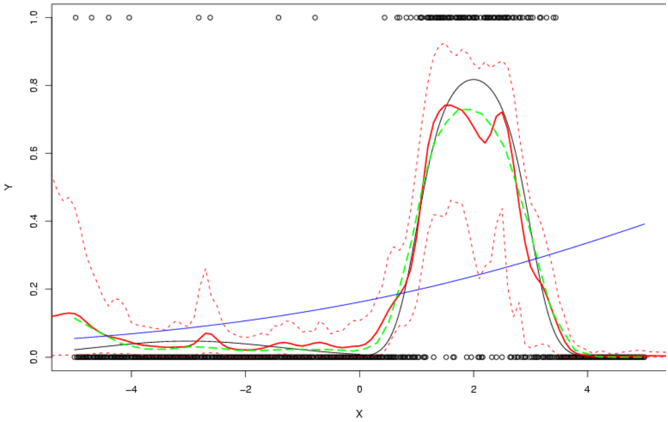
After sample size is increased, $n = 500$.

Example 3.

After sample size is increased, $n = 500$.

Example 3.

After sample size is increased, n = 500.



Simulations

Parameters related to the amount of smoothing

- α: As α values increases Dirichlet process generates more clusters and it gives less smoother smoothing.
- τ: The bigger value of τ gives less smoothing (or more wiggly curve) same as the smaller window size in Kernel Smoothing.
- We are going to illustrate how the different choice of priors for τ affects the amount of smoothing of the estimated curve with the following function:

$$P(y=1|x) = \frac{\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}{1+\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}$$

Simulations

Parameters related to the amount of smoothing

- α: As α values increases Dirichlet process generates more clusters and it gives less smoother smoothing.
- τ: The bigger value of τ gives less smoothing (or more wiggly curve) same as the smaller window size in Kernel Smoothing.
- We are going to illustrate how the different choice of priors for τ affects the amount of smoothing of the estimated curve with the following function:

$$P(y=1|x) = \frac{\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}{1+\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}$$

Simulations

Parameters related to the amount of smoothing

- α: As α values increases Dirichlet process generates more clusters and it gives less smoother smoothing.
- τ: The bigger value of τ gives less smoothing (or more wiggly curve) same as the smaller window size in Kernel Smoothing.
- We are going to illustrate how the different choice of priors for τ affects the amount of smoothing of the estimated curve with the following function:

$$P(y=1|x) = \frac{\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}{1+\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}$$

Simulations

Parameters related to the amount of smoothing

- α: As α values increases Dirichlet process generates more clusters and it gives less smoother smoothing.
- τ: The bigger value of τ gives less smoothing (or more wiggly curve) same as the smaller window size in Kernel Smoothing.
- We are going to illustrate how the different choice of priors for τ affects the amount of smoothing of the estimated curve with the following function:

$$P(y = 1|x) = \frac{\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}{1+\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}$$

Simulations

Parameters related to the amount of smoothing

- α: As α values increases Dirichlet process generates more clusters and it gives less smoother smoothing.
- τ: The bigger value of τ gives less smoothing (or more wiggly curve) same as the smaller window size in Kernel Smoothing.
- We are going to illustrate how the different choice of priors for τ affects the amount of smoothing of the estimated curve with the following function:

$$P(y=1|x) = \frac{\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}{1+\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}$$

Simulations

Parameters related to the amount of smoothing

- α: As α values increases Dirichlet process generates more clusters and it gives less smoother smoothing.
- τ: The bigger value of τ gives less smoothing (or more wiggly curve) same as the smaller window size in Kernel Smoothing.
- We are going to illustrate how the different choice of priors for τ affects the amount of smoothing of the estimated curve with the following function:

$$P(y = 1|x) = \frac{\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}{1+\exp(\exp(-2(x+2.5)^2+2)+\exp(-(x-2.5)^2/8)-2)}$$

Simulated examples

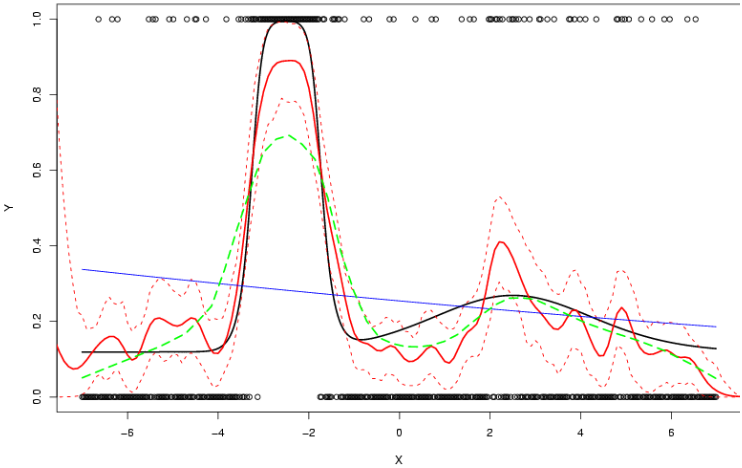
$\tau \sim Gamma(200, 5)$

Simulated examples

$\tau \sim Gamma(200, 5)$

Simulated examples

$\tau \sim Gamma(200, 5)$



Simulated examples

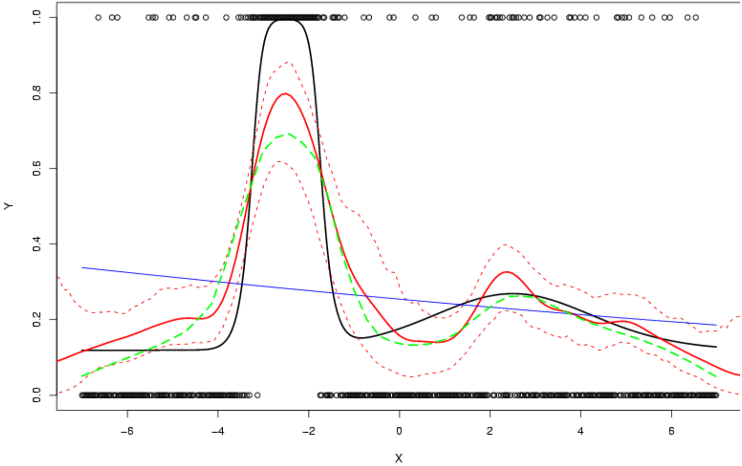
$\tau \sim Gamma(200, 20)$

Simulated examples

$\tau \sim Gamma(200, 20)$

Simulated examples

$\tau \sim Gamma(200, 20)$



Simulated examples

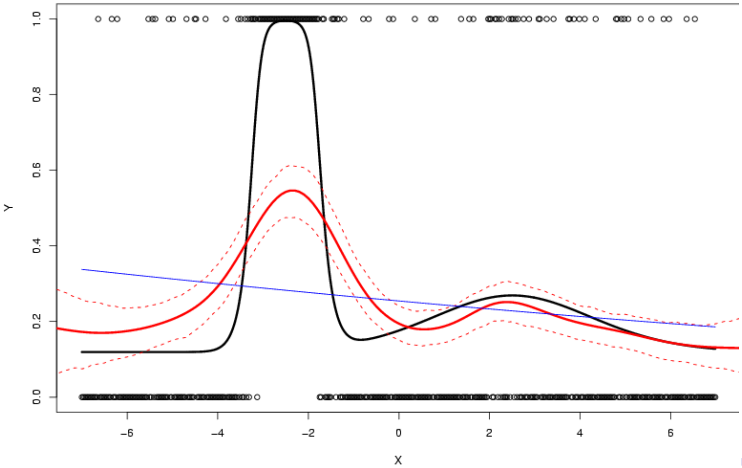
$\tau \sim Gamma(5, 1)$

Simulated examples

$\tau \sim Gamma(5, 1)$

Simulated examples

$\tau \sim Gamma(5, 1)$



The Low birth Weight Data example

The non-smoking groups of mothers

Priors: $\tau \sim Gamma(10, 400)$ and $\alpha \sim Gamma(50, 1)$.

We did the transformation of X: X - mean(X), $n = 115$.

The Low birth Weight Data example

The non-smoking groups of mothers

Priors: $\tau \sim Gamma(10, 400)$ and $\alpha \sim Gamma(50, 1)$.

We did the transformation of X: X - mean(X), $n = 115$.

The Low birth Weight Data example

The non-smoking groups of mothers

Priors: $\tau \sim Gamma(10, 400)$ and $\alpha \sim Gamma(50, 1)$.

We did the transformation of X: X - mean(X), $n = 115$.

The Low birth Weight Data example

The non-smoking groups of mothers

Priors: $\tau \sim Gamma(10, 400)$ and $\alpha \sim Gamma(50, 1)$.

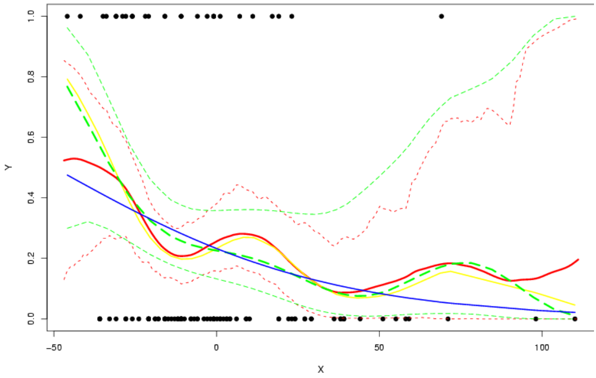
We did the transformation of X: X - mean(X), $n = 115$.

The Low birth Weight Data example

The non-smoking groups of mothers

Priors: $\tau \sim Gamma(10, 400)$ and $\alpha \sim Gamma(50, 1)$.

We did the transformation of X: X - mean(X), $n = 115$.



The Low birth Weight Data example

The smoking groups of mothers

Priors: $\tau \sim Gamma(10, 800)$ and $\alpha \sim Gamma(10, 1)$.

We did the transformation of X: X - mean(X), $n = 74$.

The Low birth Weight Data example

The smoking groups of mothers

Priors: $\tau \sim Gamma(10, 800)$ and $\alpha \sim Gamma(10, 1)$.

We did the transformation of X: X - mean(X), $n = 74$.

The Low birth Weight Data example

The smoking groups of mothers

Priors: $\tau \sim Gamma(10, 800)$ and $\alpha \sim Gamma(10, 1)$.

We did the transformation of X: X - mean(X), $n = 74$.

The Low birth Weight Data example

The smoking groups of mothers

Priors: $\tau \sim Gamma(10, 800)$ and $\alpha \sim Gamma(10, 1)$.

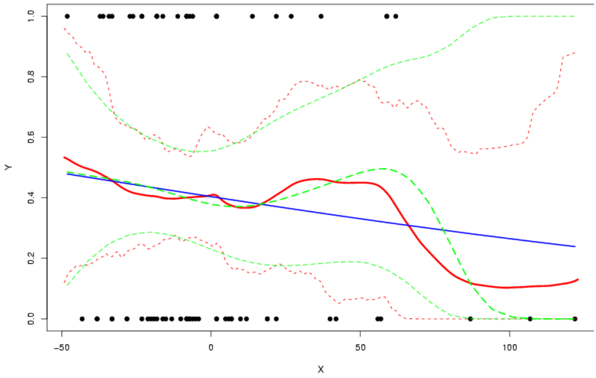
We did the transformation of X: X - mean(X), $n = 74$.

The Low birth Weight Data example

The smoking groups of mothers

Priors: $\tau \sim Gamma(10, 800)$ and $\alpha \sim Gamma(10, 1)$.

We did the transformation of X: X - mean(X), $n = 74$.



Bayesian Semi-parametric Poisson Regression

Bayesian semi-parametric Poisson regression

The model structure for Poisson:

$$y_i|x_i \sim Poi(\lambda_i), \quad \text{where} \quad \lambda_i = \exp(\beta_{0i} + \beta_{1i}x_i)$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

The specification of G_0 and hyper-parameters are same as the Logistic regression case

Bayesian semi-parametric Poisson regression

The model structure for Poisson:

$$y_i|x_i \sim Poi(\lambda_i), \quad \text{where} \quad \lambda_i = \exp(\beta_{0i} + \beta_{1i}x_i)$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

The specification of G_0 and hyper-parameters are same as the Logistic regression case

Bayesian semi-parametric Poisson regression

The model structure for Poisson:

$$y_i|x_i \sim Poi(\lambda_i), \quad \text{where} \quad \lambda_i = \exp(\beta_{0i} + \beta_{1i}x_i)$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

The specification of G_0 and hyper-parameters are same as the Logistic regression case

Bayesian semi-parametric Poisson regression

The model structure for Poisson:

$$y_i|x_i \sim Poi(\lambda_i), \quad \text{where} \quad \lambda_i = \exp(\beta_{0i} + \beta_{1i}x_i)$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

The specification of G_0 and hyper-parameters are same as the Logistic regression case

The mixing weights

The mixing weights are calculated as:

$$\begin{aligned} q_{i0} &\propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i) \\ &= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i) \\ &= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(-\exp(\beta_0 + \beta_1 x_i))(\exp(\beta_0 + \beta_1 x_i))^{y_i}}{y_i!} dF(\beta_0, \beta_1) \\ &\quad \cdot \int_{-\infty}^{\infty} \int_0^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x \end{aligned}$$

- For the first part of integration, again Monte Carlo method is applied.
- $q_{ij} \propto f(D_i|\theta_j) = f(y_i|x_i \ \theta_j) f(x_i|\theta_j)$ are easily evaluated.
- The posterior samples of (β_0, β_1) are again generated by ARS.

The mixing weights

The mixing weights are calculated as:

$$
\begin{aligned}
q_{i0} &\propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(-\exp(\beta_0 + \beta_1 x_i))(\exp(\beta_0 + \beta_1 x_i))^{y_i}}{y_i!} dF(\beta_0, \beta_1) \\
&\quad \cdot \int_{-\infty}^{\infty} \int_0^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x
\end{aligned}
$$

- For the first part of integration, again Monte Carlo method is applied.
- $q_{ij} \propto f(D_i|\theta_j) = f(y_i|x_i\ \theta_j) f(x_i|\theta_j)$ are easily evaluated.
- The posterior samples of (β_0, β_1) are again generated by ARS.

The mixing weights

The mixing weights are calculated as:

$$
\begin{aligned}
q_{i0} &\propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(-\exp(\beta_0 + \beta_1 x_i))(\exp(\beta_0 + \beta_1 x_i))^{y_i}}{y_i!} dF(\beta_0, \beta_1) \\
&\quad \cdot \int_{-\infty}^{\infty} \int_0^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x
\end{aligned}
$$

- For the first part of integration, again Monte Carlo method is applied.
- $q_{ij} \propto f(D_i|\theta_j) = f(y_i|x_i\ \theta_j) f(x_i|\theta_j)$ are easily evaluated.
- The posterior samples of (β_0, β_1) are again generated by ARS.

The mixing weights

The mixing weights are calculated as:

$$
\begin{aligned}
q_{i0} &\propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int f(y_i|x_i,\theta_i) f(x_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\exp(-\exp(\beta_0+\beta_1 x_i))(\exp(\beta_0+\beta_1 x_i))^{y_i}}{y_i!} dF(\beta_0,\beta_1) \\
&\quad \cdot \int_{-\infty}^{\infty}\int_0^{\infty} f(x_i|\mu_x,\tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x
\end{aligned}
$$

- For the first part of integration, again Monte Carlo method is applied.
- $q_{ij} \propto f(D_i|\theta_j) = f(y_i|x_i\ \theta_j) f(x_i|\theta_j)$ are easily evaluated.
- The posterior samples of (β_0,β_1) are again generated by ARS.

The mixing weights

The mixing weights are calculated as:

$$
\begin{aligned}
q_{i0} &\propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int f(y_i|x_i, \theta_i) f(x_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(-\exp(\beta_0 + \beta_1 x_i))(\exp(\beta_0 + \beta_1 x_i))^{y_i}}{y_i!} dF(\beta_0, \beta_1) \\
&\quad \cdot \int_{-\infty}^{\infty} \int_0^{\infty} f(x_i|\mu_x, \tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x
\end{aligned}
$$

- For the first part of integration, again Monte Carlo method is applied.
- $q_{ij} \propto f(D_i|\theta_j) = f(y_i|x_i\ \theta_j) f(x_i|\theta_j)$ are easily evaluated.
- The posterior samples of (β_0, β_1) are again generated by ARS.

The mixing weights

The mixing weights are calculated as:

$$
\begin{aligned}
q_{i0} &\propto \alpha \int f(D_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int f(y_i|x_i,\theta_i) f(x_i|\theta_i) dG_0(\theta_i) \\
&= \alpha \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\exp(-\exp(\beta_0+\beta_1 x_i))(\exp(\beta_0+\beta_1 x_i))^{y_i}}{y_i!} dF(\beta_0,\beta_1) \\
&\quad \cdot \int_{-\infty}^{\infty} \int_0^{\infty} f(x_i|\mu_x,\tau_x) f(\mu_x|\tau_x) f(\tau_x) d\tau_x d\mu_x
\end{aligned}
$$

- For the first part of integration, again Monte Carlo method is applied.
- $q_{ij} \propto f(D_i|\theta_j) = f(y_i|x_i\ \theta_j) f(x_i|\theta_j)$ are easily evaluated.
- The posterior samples of (β_0, β_1) are again generated by ARS.

Example 1.

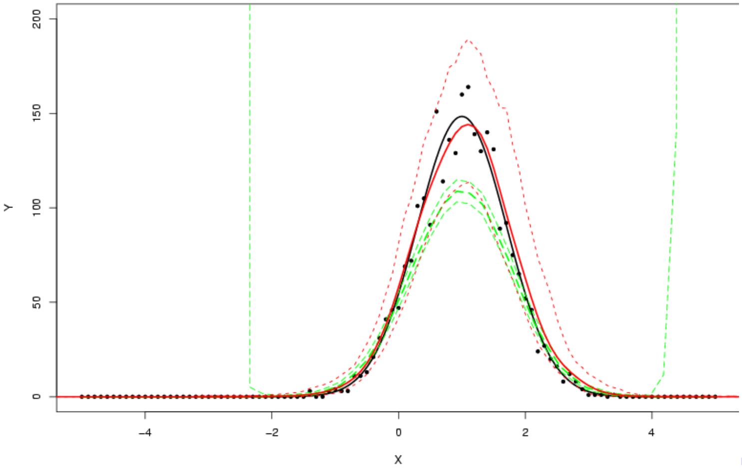
(1) $E(Y|X = x) = \lambda = \exp(-(x-1)^2 + 5)$, n=100

Example 1.

(1) $E(Y|X = x) = \lambda = \exp(-(x-1)^2 + 5)$, n=100

Example 1.

(1) $E(Y|X=x) = \lambda = \exp(-(x-1)^2 + 5)$, n=100



Example 2.

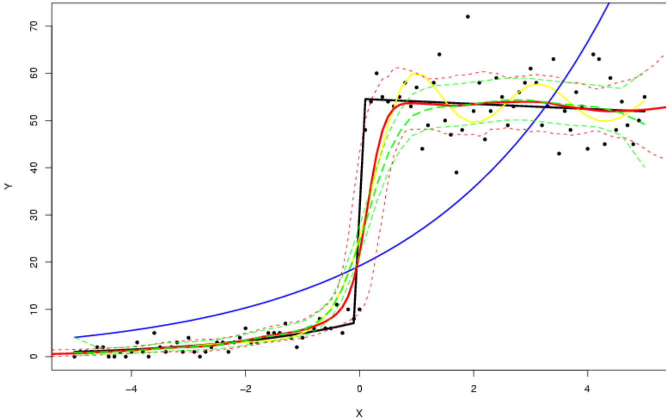
(2) $E(Y|X=x)=\lambda=\exp(0.4x+2)I(x\leq 0)+\exp(-0.01x+4)I(x>0)$, n=100

Example 2.

(2) $E(Y|X=x)=\lambda=\exp(0.4x+2)I(x\leq 0)+$
$\exp(-0.01x+4)I(x>0)$, n=100

Example 2.

(2) $E(Y|X=x) = \lambda = \exp(0.4x+2)I(x \leq 0) + \exp(-0.01x+4)I(x>0)$, n=100



Example 3.

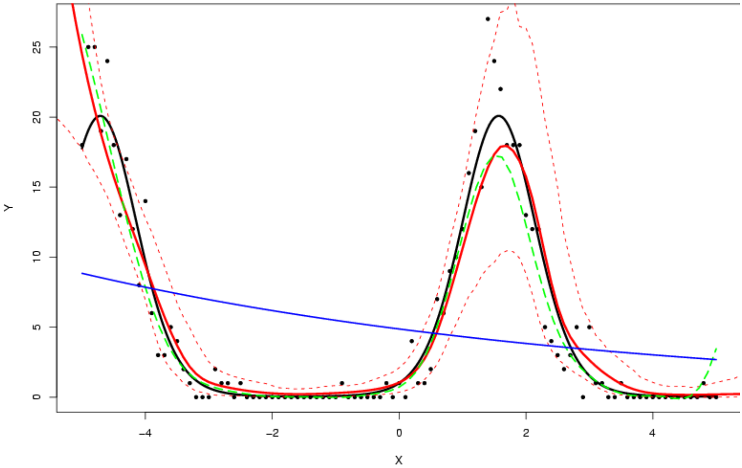
(3) $E(Y|X = x) = \lambda = \exp(3\sin(x))$, n=100

Example 3.

(3) $E(Y|X=x)=\lambda=\exp(3\sin(x))$, n=100

Example 3.

(3) $E(Y|X = x) = \lambda = \exp(3\sin(x))$, n=100



Example 4.

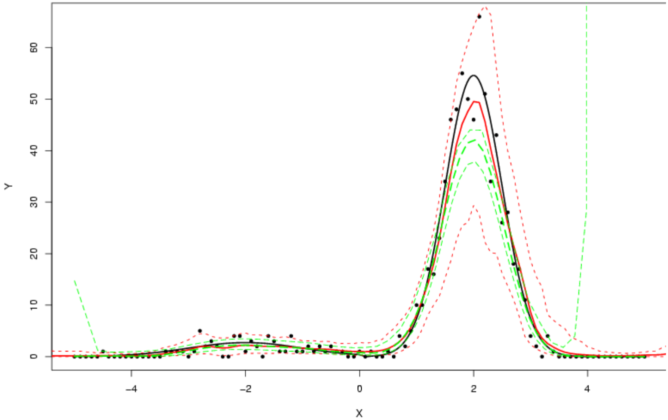
(4) $E(Y|X = x) = \lambda = \exp(-0.5(x + 2)^2 + 1)I(x \le 0) + \exp(-2(x - 2)^2 + 4)I(x > 0)$, n=100

Example 4.

(4) $E(Y|X = x) = \lambda = \exp(-0.5(x + 2)^2 + 1)I(x \leq 0) + \exp(-2(x - 2)^2 + 4)I(x > 0)$, n=100

Example 4.

$$(4)\ E(Y|X=x)=\lambda=\exp(-0.5(x+2)^2+1)I(x\leq 0)+\exp(-2(x-2)^2+4)I(x>0),\ \text{n=100}$$



Poisson regression function

How our method works?

(1) $E(Y|X=x) = \lambda = \exp(-(x-1)^2 + 5)$, n=100

We took the sampled parameters after 83th iteration: $k = 5$.

Poisson regression function

How our method works?

(1) $E(Y|X=x)=\lambda=\exp(-(x-1)^2+5)$, n=100

We took the sampled parameters after 83th iteration: $k=5$.

Poisson regression function

How our method works?

(1) $E(Y|X=x)=\lambda=\exp(-(x-1)^2+5)$, n=100

We took the sampled parameters after 83th iteration: $k=5$.

Poisson regression function

How our method works?

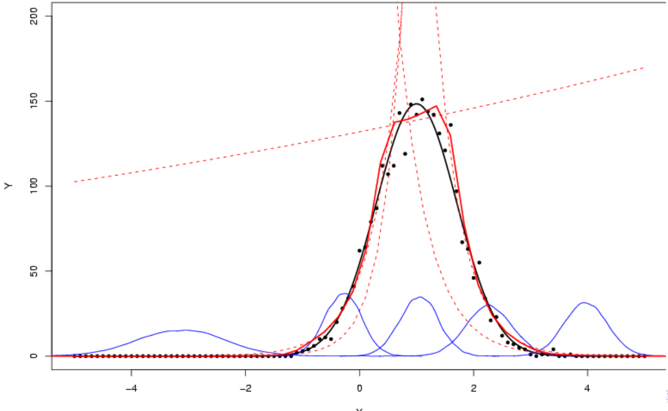
(1) $E(Y|X=x) = \lambda = \exp(-(x-1)^2 + 5)$, n=100
We took the sampled parameters after 83th iteration: $k = 5$.

Poisson regression function

How our method works?

(1) $E(Y|X=x) = \lambda = \exp(-(x-1)^2+5)$, n=100

We took the sampled parameters after 83th iteration: $k=5$.



Poisson regression function

(2) $E(Y|X = x) = \lambda = \exp(0.4x + 2)I(x \leq 0)+$
$\exp(-0.01x + 4)I(x > 0)$, n=100

We took the sampled parameters after 709th iteration: $k = 5$.

Poisson regression function

(2) $E(Y|X = x) = \lambda = \exp(0.4x + 2)I(x \leq 0)+$
$\exp(-0.01x + 4)I(x > 0)$, n=100

We took the sampled parameters after 709th iteration: $k = 5$.

Poisson regression function

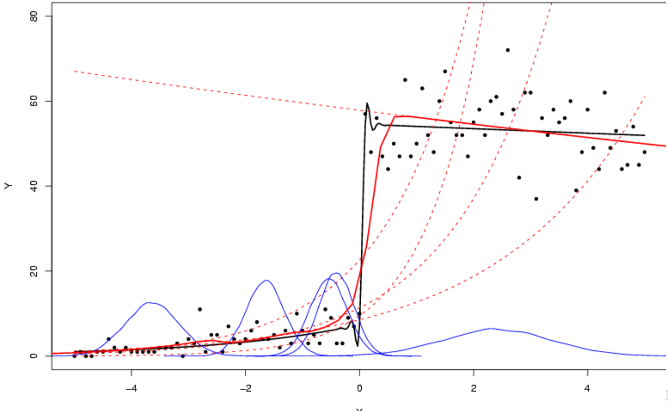
(2) $E(Y|X = x) = \lambda = \exp(0.4x + 2)I(x \leq 0) + \exp(-0.01x + 4)I(x > 0)$, n=100

We took the sampled parameters after 709th iteration: $k = 5$.

Poisson regression function

(2) $E(Y|X=x)=\lambda=\exp(0.4x+2)I(x\leq 0)+\exp(-0.01x+4)I(x>0)$, n=100

We took the sampled parameters after 709th iteration: $k=5$.



Posterior distribution of β_0 and β_1

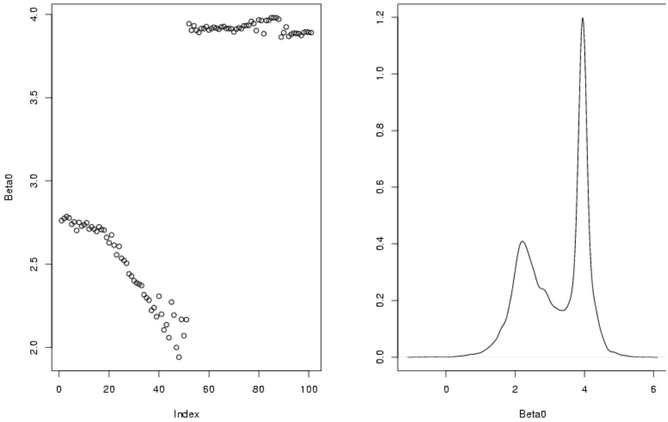
Posterior distribution of β_0

Posterior distribution of β_0 and β_1

Posterior distribution of β_0

Posterior distribution of β_0 and β_1

Posterior distribution of β_0



Posterior distribution of β_0 and β_1

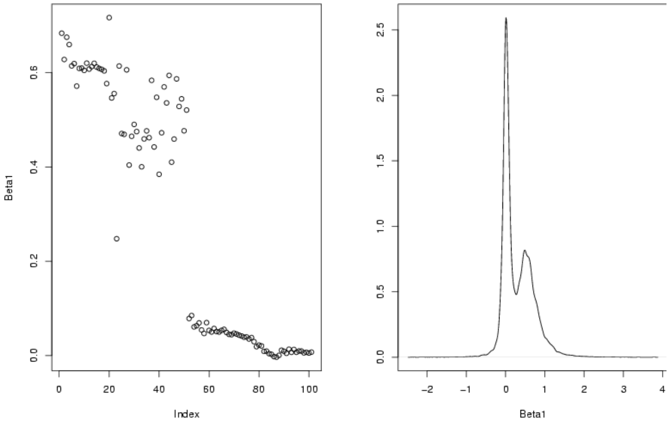
Posterior distribution of β_1

Posterior distribution of β_0 and β_1

Posterior distribution of β_1

Posterior distribution of β_0 and β_1

Posterior distribution of β_1



Discussion and Future work

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Discussion

Benefit of our method

- It is an effective way of estimating the true Logistic and Poisson regression functions, especially when the functions are spatially heterogenous.
- It is a new way of doing semi-parametric regressions with Bayesian perspective
- It is conceptually easy and provides easy-to-implement simulation environment.
- We can directly obtain the distributions of the primary parameters of interests

Concerns:

- How sensitive our method to the choice of priors?
- How good approximation of our Dirichlet mixing weight, q_0, which requires a numerical integration for each iteration?
- How heavy is the cost of computing to implement the method?

Future work

Future work

- The multivariate extension of our work should be studied further.
- This method can also be applied to estimate the hazard function in Survival Analysis.

How is it applied to survival analysis?

- Let $y = (y_1, y_2, \ldots, y_n)'$ be a survival time, and each having an exponential distribution with parameter λ_i.
- Let $\delta = (\delta_1, \delta_2, \ldots, \delta_n)'$ be a censoring indicator, where $\delta_i = 0$ if y_i is right censored and $\delta_i = 1$ if y_i is a true failure time.

Future work

Future work

- The multivariate extension of our work should be studied further.
- This method can also be applied to estimate the hazard function in Survival Analysis.

How is it applied to survival analysis?

- Let $y = (y_1, y_2, \ldots, y_n)'$ be a survival time, and each having an exponential distribution with parameter λ_i.
- Let $\delta = (\delta_1, \delta_2, \ldots, \delta_n)'$ be a censoring indicator, where $\delta_i = 0$ if y_i is right censored and $\delta_i = 1$ if y_i is a true failure time.

Future work

Future work

- The multivariate extension of our work should be studied further.
- This method can also be applied to estimate the hazard function in Survival Analysis.

How is it applied to survival analysis?

- Let $y = (y_1, y_2, \ldots, y_n)'$ be a survival time, and each having an exponential distribution with parameter λ_i.
- Let $\delta = (\delta_1, \delta_2, \ldots, \delta_n)'$ be a censoring indicator, where $\delta_i = 0$ if y_i is right censored and $\delta_i = 1$ if y_i is a true failure time.

Future work

Future work

- The multivariate extension of our work should be studied further.
- This method can also be applied to estimate the hazard function in Survival Analysis.

How is it applied to survival analysis?

- Let $y = (y_1, y_2, \ldots, y_n)'$ be a survival time, and each having an exponential distribution with parameter λ_i.
- Let $\delta = (\delta_1, \delta_2, \ldots, \delta_n)'$ be a censoring indicator, where $\delta_i = 0$ if y_i is right censored and $\delta_i = 1$ if y_i is a true failure time.

Future work

Future work

- The multivariate extension of our work should be studied further.
- This method can also be applied to estimate the hazard function in Survival Analysis.

How is it applied to survival analysis?

- Let $y = (y_1, y_2, \ldots, y_n)'$ be a survival time, and each having an exponential distribution with parameter λ_i.
- Let $\delta = (\delta_1, \delta_2, \ldots, \delta_n)'$ be a censoring indicator, where $\delta_i = 0$ if y_i is right censored and $\delta_i = 1$ if y_i is a true failure time.

Future work

Future work

- The multivariate extension of our work should be studied further.
- This method can also be applied to estimate the hazard function in Survival Analysis.

How is it applied to survival analysis?

- Let $y = (y_1, y_2, \ldots, y_n)'$ be a survival time, and each having an exponential distribution with parameter λ_i.
- Let $\delta = (\delta_1, \delta_2, \ldots, \delta_n)'$ be a censoring indicator, where $\delta_i = 0$ if y_i is right censored and $\delta_i = 1$ if y_i is a true failure time.

Future work

Future work

- The multivariate extension of our work should be studied further.
- This method can also be applied to estimate the hazard function in Survival Analysis.

How is it applied to survival analysis?

- Let $y = (y_1, y_2, \ldots, y_n)'$ be a survival time, and each having an exponential distribution with parameter λ_i.
- Let $\delta = (\delta_1, \delta_2, \ldots, \delta_n)'$ be a censoring indicator, where $\delta_i = 0$ if y_i is right censored and $\delta_i = 1$ if y_i is a true failure time.

Survival Analysis

- Our model can be written as:

$$y_i|x_i \sim Exp(\lambda_i), \quad \text{where} \quad \lambda_i = \beta_{0i} + \beta_{1i}x_i$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

The specification of G_0 and hyper-parameters are same as Logistic and Poisson case.

- The likelihood function with the censoring information is:

$$\begin{aligned} L(\beta|D) &= \prod_{i=1}^{n} f(y_i|\lambda_i)^{\delta_i} S(y_i|\lambda_i)^{(1-\delta_i)} \\ &= \exp\{\sum_{i=1}^{n} \delta_i(\beta_{0i} + \beta_{1i}x_i)\} \exp\{-\sum_{i=1}^{n} y_i \exp(\beta_{0i} + \beta_{1i}x_i)\} \end{aligned}$$

Survival Analysis

- Our model can be written as:

$$y_i|x_i \sim Exp(\lambda_i), \quad \text{where} \quad \lambda_i = \beta_{0i} + \beta_{1i}x_i$$
$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$
$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$
$$G \sim D(G_0, \alpha)$$

The specification of G_0 and hyper-parameters are same as Logistic and Poisson case.

- The likelihood function with the censoring information is:

$$\begin{aligned} L(\beta|D) &= \prod_{i=1}^{n} f(y_i|\lambda_i)^{\delta_i} S(y_i|\lambda_i)^{(1-\delta_i)} \\ &= \exp\{\sum_{i=1}^{n} \delta_i(\beta_{0i} + \beta_{1i}x_i)\} \exp\{-\sum_{i=1}^{n} y_i \exp(\beta_{0i} + \beta_{1i}x_i)\} \end{aligned}$$

Survival Analysis

- Our model can be written as:

$$y_i | x_i \sim Exp(\lambda_i), \quad \text{where} \quad \lambda_i = \beta_{0i} + \beta_{1i} x_i$$

$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$

$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$

$$G \sim D(G_0, \alpha)$$

  The specification of G_0 and hyper-parameters are same as Logistic and Poisson case.

- The likelihood function with the censoring information is:

$$\begin{aligned} L(\beta|D) &= \prod_{i=1}^{n} f(y_i|\lambda_i)^{\delta_i} S(y_i|\lambda_i)^{(1-\delta_i)} \\ &= \exp\{\sum_{i=1}^{n} \delta_i(\beta_{0i} + \beta_{1i} x_i)\} \exp\{-\sum_{i=1}^{n} y_i \exp(\beta_{0i} + \beta_{1i} x_i)\} \end{aligned}$$

Survival Analysis

- Our model can be written as:

$$y_i|x_i \sim Exp(\lambda_i), \quad \text{where} \quad \lambda_i = \beta_{0i} + \beta_{1i}x_i$$
$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$
$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$
$$G \sim D(G_0, \alpha)$$

The specification of G_0 and hyper-parameters are same as Logistic and Poisson case.

- The likelihood function with the censoring information is:

$$L(\beta|D) = \prod_{i=1}^{n} f(y_i|\lambda_i)^{\delta_i} S(y_i|\lambda_i)^{(1-\delta_i)}$$
$$= \exp\{\sum_{i=1}^{n} \delta_i(\beta_{0i} + \beta_{1i}x_i)\} \exp\{-\sum_{i=1}^{n} y_i \exp(\beta_{0i} + \beta_{1i}x_i)\}$$

Survival Analysis

- Our model can be written as:

$$y_i|x_i \sim Exp(\lambda_i), \quad \text{where} \quad \lambda_i = \beta_{0i} + \beta_{1i}x_i$$
$$x_i \sim N(\mu_{xi}, \tau_{xi}^{-1})$$
$$\theta_i = (\beta_{0i}, \beta_{1i}, \mu_{xi}, \tau_{xi}) \sim G$$
$$G \sim D(G_0, \alpha)$$

  The specification of G_0 and hyper-parameters are same as Logistic and Poisson case.

- The likelihood function with the censoring information is:

$$L(\beta|D) = \prod_{i=1}^{n} f(y_i|\lambda_i)^{\delta_i} S(y_i|\lambda_i)^{(1-\delta_i)}$$
$$= \exp\{\sum_{i=1}^{n} \delta_i(\beta_{0i} + \beta_{1i}x_i)\} \exp\{-\sum_{i=1}^{n} y_i \exp(\beta_{0i} + \beta_{1i}x_i)\}$$

Survival Analysis

- After we are able to obtain all mixing weights and posterior samples for the parameters, then we can estimate the hazard function as:

$$h(x) = E(y|x\ \theta^*) = \sum_{j=0}^{k} W_j(x)(\beta_{0j} + \beta_{1j}x)$$

- It would be deduced to a few mixtures of linear functions weighted by functions of marginal distributions of X.
- Now we have a hazard function which is a mixtures of linear functions, and it would provide to fitting the wide range of spatially heterogenous hazard functions.

Survival Analysis

- After we are able to obtain all mixing weights and posterior samples for the parameters, then we can estimate the hazard function as:

$$h(x) = E(y|x\ \theta^*) = \sum_{j=0}^{k} W_j(x)(\beta_{0j} + \beta_{1j}x)$$

- It would be deduced to a few mixtures of linear functions weighted by functions of marginal distributions of X.
- Now we have a hazard function which is a mixtures of linear functions, and it would provide to fitting the wide range of spatially heterogenous hazard functions.

Survival Analysis

- After we are able to obtain all mixing weights and posterior samples for the parameters, then we can estimate the hazard function as:

$$h(x) = E(y|x\ \theta^*) = \sum_{j=0}^{k} W_j(x)(\beta_{0j} + \beta_{1j}x)$$

- It would be deduced to a few mixtures of linear functions weighted by functions of marginal distributions of X.
- Now we have a hazard function which is a mixtures of linear functions, and it would provide to fitting the wide range of spatially heterogenous hazard functions.

Survival Analysis

- After we are able to obtain all mixing weights and posterior samples for the parameters, then we can estimate the hazard function as:

$$h(x) = E(y|x\ \theta^*) = \sum_{j=0}^{k} W_j(x)(\beta_{0j} + \beta_{1j}x)$$

- It would be deduced to a few mixtures of linear functions weighted by functions of marginal distributions of X.
- Now we have a hazard function which is a mixtures of linear functions, and it would provide to fitting the wide range of spatially heterogenous hazard functions.

Survival Analysis

- After we are able to obtain all mixing weights and posterior samples for the parameters, then we can estimate the hazard function as:

$$h(x) = E(y|x\ \theta^*) = \sum_{j=0}^{k} W_j(x)(\beta_{0j} + \beta_{1j}x)$$

- It would be deduced to a few mixtures of linear functions weighted by functions of marginal distributions of X.
- Now we have a hazard function which is a mixtures of linear functions, and it would provide to fitting the wide range of spatially heterogenous hazard functions.

The End