作业要求

一、考虑以下由三个高斯混合成分构成的一元概率分布

$$p(x)=0.2\mathcal{N}\left(x\mid\mu_1,\sigma_1^2\right)+0.5\mathcal{N}\left(x\mid\mu_2,\sigma_2^2\right)+0.3\mathcal{N}\left(x\mid\mu_3,\sigma_3^2\right)$$

其中 $\mu_1=-5$，$\mu_2=1$，$\mu_3=6$，$\sigma_1^2=\sigma_2^2=\sigma_3^2=1$。使用 MH 算法采样 10000 个满足上述分布的样本，并绘制样本序列的折线图、样本的频率直方图以及样本序列自相关图(plot_acf)。

二、考虑以下仅有截距项 μ 的回归模型

$$y_i=\mu+e_i,\ e_i\sim\mathcal{N}\left(0,\sigma^2\right)$$

使用贝叶斯方法和 MH 算法学习参数 μ 和 σ^2 的分布。具体要求如下：

（1）(**数据生成**)以均值为 5 方差为 1 的一元正态分布随机生成 1000 个随机数作为数据。

（2）假设截距项 μ 的先验是均值为 $\eta=0$ 方差为 $\tau^2=100$ 的正态分布。

（3）对于方差 σ^2，通过变换 $\alpha=\log\sigma$，假设 α 的先验是均值为 $s=0$ 方差为 $\omega^2=5$ 的正态分布。

（4）使用 MH 算法从参数 μ 和 σ^2 的后验分布中采样得到 1000 个样本。

（5）检查 MH 算法中的马尔可夫链是否到达平稳分布，绘制样本序列的折线图、样本的频率直方图以及样本序列自相关图(plot_acf)。

三、考虑以下回归模型

$$y_i=\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i+e_i,\ e_i\sim\mathcal{N}\left(0,\sigma^2\right)$$

使用贝叶斯方法和 Gibbs 算法学习参数 $\boldsymbol{\beta}$ 和 σ^2 的分布。具体要求如下：

（1）(**数据生成**)假设数据维度为 20 维，即 $\mathbf{x}_i\in\mathbb{R}^{20}$。以[0,1]区间上的均匀分布随机生成 $n=5000$ 个数据 $\mathbf{x}_i$。以[-1,1]区间上的均匀分布随机生成回归系数 $\boldsymbol{\beta}$。将 $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i$ 加上 $e_i\sim\mathcal{N}(0,1)$ 作为 y_i。

（2）假设参数 $\boldsymbol{\beta}$ 的共轭先验分布是均值向量为 $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ 协方差矩阵为 $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ 的多元正态分布 $\mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\beta}},\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right)$，其中 $\boldsymbol{\mu}_{\boldsymbol{\beta}}=\mathbf{0}$，$\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ 是单位矩阵。理论上可以证明：当参数 σ^2 已知时，参数 $\boldsymbol{\beta}$ 的后验分布（即满条件分布）是以下多元正态分布

$$p\left(\boldsymbol{\beta}\mid\{y_i\}_{i=1}^n,\sigma^2\right)=\mathcal{N}\left(\boldsymbol{\mu}_{\boldsymbol{\beta}},\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\right)$$

其中均值向量 $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ 和协方差矩阵 $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ 满足

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}=\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}+\sigma^{-2}\mathbf{X}\mathbf{X}^{\mathrm{T}}$$

$$\boldsymbol{\mu}_{\boldsymbol{\beta}}=\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\mu}_{\boldsymbol{\beta}}+\sigma^{-2}\mathbf{X}\mathbf{y}\right)$$

其中 $\mathbf{X}=\left[\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_n\right]$，$\mathbf{y}=\left(y_1,y_2,\ldots,y_n\right)^{\mathrm{T}}$。

（3）假设参数 σ^2 的共轭先验分布是参数 $a=3,b=1$ 的逆伽马分布 $\mathrm{IG}(a,b)$。查表可知，当参数 $\boldsymbol{\beta}$ 已知时，参数 σ^2 的后验分布（即满条件分布）是以下逆伽马分布

$$p\left(\sigma^2\mid\{y_i\}_{i=1}^n,\boldsymbol{\beta}\right)=\mathrm{IG}\left(a_1,b_1\right)$$

其中

$$a_1 = a + n/2 \quad b_1 = b + \frac{1}{2}\sum_{i=1}^{n}\left(y_i - \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\right)^2$$

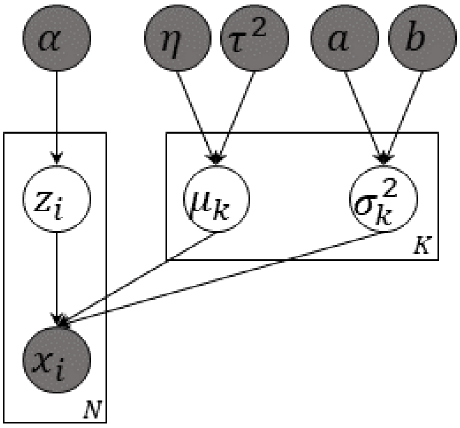
（4）使用 Gibbs 算法从$\boldsymbol{\beta}$和σ^2的后验分布中采样得到 10000 个样本。

（5）检查 Gibbs 算法中的马尔可夫链是否到达平稳分布，绘制样本序列的折线图、样本的频率直方图以及样本序列自相关图(plot_acf)。

四、考虑以下一维的贝叶斯高斯混合模型：

$$\begin{aligned}
&x_i \sim \mathcal{N}\left(\mu_{z_i}, \sigma_{z_i}^2\right), \quad i = 1, \ldots, N, \\
&z_i \sim \mathrm{Mult}(\boldsymbol{\alpha}), \quad i = 1, \ldots, N, \\
&\mu_k \sim \mathcal{N}\left(\eta, \tau^2\right), \quad k = 1, \ldots, K, \\
&\sigma_k^2 \sim \mathrm{IG}\left(a, b\right), \quad k = 1, \ldots, K,
\end{aligned}$$

其中N是样本数，x_i是第i个样本，K是混合成分的个数，z_i是样本x_i对应的混合成分的编号，$\boldsymbol{\alpha}, \eta, \tau^2, a, b$为超参数。上述模型如下图所示：



使用贝叶斯方法和 Gibbs 算法学习参数$\{\mu_k\}_{k=1}^K$，$\{\sigma_k^2\}_{k=1}^K$，$\{z_i\}_{i=1}^N$的分布。具体要求如下：

（1）(**数据生成**)假设数据是按照以下三个高斯混合成分构成的一元概率分布生成的$N = 1000$个样本$D = \{x_1, \ldots x_N\}$：

$$p(x) = 0.2\mathcal{N}\left(x \mid \mu_1, \sigma_1^2\right) + 0.5\mathcal{N}\left(x \mid \mu_2, \sigma_2^2\right) + 0.3\mathcal{N}\left(x \mid \mu_3, \sigma_3^2\right)$$

其中$\mu_1 = -5$，$\mu_2 = 1$，$\mu_3 = 6$，$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$。

（2）分别推导参数$\{\mu_k\}_{k=1}^K$，$\{\sigma_k^2\}_{k=1}^K$，$\{z_i\}_{i=1}^N$的满条件分布。

（3）假设超参数的取值分别为$\boldsymbol{\alpha} = (1/K, \ldots, 1/K)^{\mathrm{T}}$，$\eta = 0$，$\tau^2 = 100$，$a = 3$，$b = 1$。

（4）使用 Gibbs 算法从参数$\{\mu_k\}_{k=1}^K$，$\{\sigma_k^2\}_{k=1}^K$，$\{z_i\}_{i=1}^N$的满条件分布中采样得到 10000 个样本。

（5）通过采样得到的样本计算参数$\{\mu_k\}_{k=1}^K$，$\{\sigma_k^2\}_{k=1}^K$的估计值，并与真实的参数值进行比较。

（6）通过采样得到的样本确定各样本x_i对应的混合成分的编号，并与真实的编号比较，计算预测精度。