基于贝叶斯有限混合模型的多元混合样本的质量控制

姚可辉

目录

摘要

在统计质量控制中，假设产品是由一条生产线上的多个机器制造的。在某一个或多个时间点，过程失控，不同机器生产的产品发生了不同的均值漂移。再过一段时间，控制图报警信号被发出。我们把从受控状态和失控状态下的所有机器生产的产品看作一组随机样本。基于这些样本，本文提出了贝叶斯有限混合模型，它能够在受控参数已知的情况下，用变点控制图理论，估计各机器发生过程失控的变点。同时估计变点后各机器新的中心。在随机模拟阶段，我们会用 MCMC 和 Gibbs 抽样等方法，并使用 JAGS 软件来辅助计算。

关键词: 贝叶斯，有限混合模型，MCMC，Gibbs，JAGS

1 介绍

1.1 模型概述

在目前的统计过程控制研究中，我们一般假设所有产品都产生于同一个过程，产品质量变化都来源于同一个原因，即过程（包括机器、原材料、环境等因素）的改变。但是，在实际的生产过程中，也会出现混合样本的情况，即产品是由多个机器制造的。各机器在正常工作状态下生产的产品可以认为是由一个中心产生的。但当各机器由于外界原因发生故障，导致过程失控时，我们假定发生故障后的机器生产的样本不是服从同一中心。在此前提下，我们希望有效地估计在过程中不同机器的过程失控的时间点及过程失控后的中心。由于在这一领域的研究还很局限，实际的数据集也很少。因此本文将采用大量参数不同的随机模拟数据集验证该算法的有效性。

1.2 模型假设

在很多情形下，产品的总体质量取决于监控多个有相关性的特征变量。对此，常见的方法就是多元控制图。本文将会在多元情况下建立模型。多元统计控制主要任务是两个：第一是利用某种方法来监测测量值的分布是否发生了漂移；第二则是监测到底是哪些分量发生了漂移。当一些经典的多元控制图，例如 T^2 控制图 Mason et al. (1995)，或是多元 EWMA 控制图 Prabhu and Runger (1997) 发出报警信号后，我们都会对过程变化何时产生，哪些分量发生了漂移，漂移后新的分布是什么感兴趣。我们很自然的引入基于变点模型的控制图来分析这些样本。变点控制图假设在变点前的所有样本服从一个特定的分布，变点后的所有样本服从另一个特定分布。我们会使用贝叶斯的方法来分析变点模型里我们感兴趣的参数。

传统质量控制可以分为两个阶段。在第一阶段，我们的主要目标是检验历史数据是否受控，并用受控的数据估计过程参数。在第二阶段，利用第一阶段对受控过程参数的估计，计算控制线，对过程进行监控。在控制图给出过程失控警报的条件下，检查过程的失控原因。本文提出的方法的适用背景是在质量控制的 *PhaseⅡ*，即假设受控过程的参数已知当控制图给出报警信号后，分析报警信号发出前的所有样本，进行统计推断。

1.3 文章结构

本文往后可以分为三个部分，第一部分主要介绍提出贝叶斯有限混合模型的动机和模型建立的过程，包括模型中各参数的似然函数和后验概率密度函数的推导，各机器变点（故障发生时点）相同或不同时的模型假设等等；第二部分是随机模拟，主要是测试该方法在不同情况的数据集下的稳健性；第三部分是该方法和极大似然估计方法的对比研究，由于极大似然估计方法不适用于混合样本的情况，所以在这一部分，我们对贝叶斯有限混合模型作了简化，使其适用于非混合样本的情况。

2 贝叶斯有限混合模型

2.1 背景

Barnard (1959) 和 Chernoff and Zacks (1964) 是最早提出用贝叶斯方法分析变点问题的。一个世纪以后，Smith 又给出了贝叶斯方法推断运用于变点模型的更详尽解释。比起传统的 MLE 方法 Pignatiello Jr and Samuel (2001)，贝叶斯方法不仅能估计变点发生的位置，还能估计受控过程中的参数和漂移量的大小，而 MLE 方法只能估计变点的位置。当样本来自混合总体时，传统的贝叶斯变点控制图方法也不适用，对此，本文基于贝叶斯有限混合模型的框架 Diebolt and Robert (1994)，提出了适用于混合样本的变点控制图的方法。

2.2 模型假设

假设混合样本 $X_1,\cdots,X_m$ 是由 K 个机器生产的，所有机器生产的产品在受控时都服从一个相同的多元正态分布 $N(\mu_0,\Sigma_0)$。第 k 个机器在总过程产生第 τ_k 各样本的时间点上，从受控状态进入失控状态，失控后机器 k 生产的产品服从 $N(\mu_k,\Sigma_k)$ 。定义隐变量 $z_j \in 1,\cdots,K$, z_j 表示产生第 j 个样本的机器编号。定义 $\boldsymbol{\pi}_j=(\pi_{j1},\cdots,\pi_{jk})$ 表示第 j 个样本是第 k 个机器生产的概率。π_{jk} 满足 $0\le\pi_{jk}\le 1$ ，且 $\Sigma_{k=1}^{K}\pi_{jk}=1$ 。不难发现，z_j 服从一个广义的伯努利分布, $Bernoulli(\boldsymbol{\pi}_j)$, 它的概率分布函数可以写作 $f(z_j|\pi_j,K)=\prod_{k=1}^{K}\pi_{jk}^{\delta_k(z_j))}=\Sigma_{k=1}^{K}\pi_{jk}\delta_k(z_j)$。这里 $\delta_k(z_j)=(z_j=k)$。对 $\boldsymbol{\pi}_j$ 设置一个狄利克雷分布 $Dir(\alpha_0/K*1_K)$，α_0/K 是狄利克雷分布的中心化参数，1_K 是一个所有元素都为 1 的长度为 K 的列向量。$\alpha_0\sim Gamma(\frac{1}{2},\frac{1}{2})$。由于狄利克雷分布是多项分布的共轭先验分布，如果我们把隐变量 z_j 写成向量形式 $\boldsymbol{z}_j=(z_{j1},\cdots,z_{jk})'$, 那么 $z_{j1},\cdots,z_{jk}$ 中满足只有一个元素为 1，其余都为 0。此时 $\boldsymbol{z}_j$ 服从多项分布 $multinomial(1,\boldsymbol{\pi}_j)$。最后，我们假设 $\tau_k,k=1,\cdots,K$ 独立同分布于 K 个离散的均匀分布，$f(\tau_k=i)=\frac{1}{n-1},i=1,\cdots,n-1$。各机器变点后的均值 $\boldsymbol{\mu}_k$ 的先验分布为 $N(\boldsymbol{\mu}_0,\phi)$, ϕ 的先验分布为均匀分布 $Unif(0,100)$。最后假定 Σ_k 满足 $\Sigma_k=c_k\cdot\Sigma_0$, 且 $c_k\sim Unif(0,1)$,。

2.3 似然函数

对于单个样本 j, 它的似然函数可以写作

$$\begin{aligned}&L_j(\boldsymbol{\tau},\boldsymbol{\mu_0},\cdots,\boldsymbol{\mu_K},\Sigma_0,c_1,\cdots,c_K,\boldsymbol{\pi_j},\alpha_0|\boldsymbol{z}_j,X_j,K)=p(x_j|\boldsymbol{z}_j,\cdot)\cdot p(\boldsymbol{z}_j\\&|\boldsymbol{\pi_j})=\prod_{k=1}^{K}p(x_j|\boldsymbol{\mu_0},\Sigma_0)^{1(j\le\tau_k)z_{jk}}\cdot p(x_j|\boldsymbol{\mu_k},\Sigma_k)^{1(j>\tau_k)\cdot z_{jk}}\cdot\pi_{jk}^{z_{jk}}\end{aligned}\tag{1}$$

对所有混合样本 $X_1,\cdots,X_m$, 似然函数可以写作:

$$L_{all}(\boldsymbol{\tau},\boldsymbol{\mu_0},\cdots,\boldsymbol{\mu_K},\Sigma_0,c_1,\cdots,c_K,\boldsymbol{\pi_j},\alpha_0|\boldsymbol{z},\boldsymbol{X},K)=\prod_{j=1}^{m}\prod_{k=1}^{K}p(x_j|\boldsymbol{\mu_0},\Sigma_0)^{1(j\le\tau_k)z_{jk}}\cdot p(x_j|\boldsymbol{\mu_k},\Sigma_k)^{1(j>\tau_k)\cdot z_{jk}}\cdot\pi_{jk}^{z_{jk}}\tag{2}$$

这里

$$p(X_j|\boldsymbol{\mu_k},\Sigma_k)=\frac{1}{\sqrt{(2\pi)^k|c_k\Sigma_0|}}exp(-\frac{1}{2}(X_j-\boldsymbol{\mu_k})^T\Sigma^{-1}(X_j-\boldsymbol{\mu_k}))$$

2.4 变点相同情况下参数的后验分布

我们先来看一种简单的情况。假设混合样本受控和失控状态下都服从二元正态分布，并假设 K 个机器的变点都相同，也就是说，所有机器在同一时点失控，失控的时点即为总体过程产生第 τ 个样本的时点。满足 $\tau_1=\cdots,\tau_k=\tau$。根据模型假设, μ_0 和 Σ_0 是在 *Phase*Ⅱ 阶段估计得到的值，可以认为是已知的。这里我们假定 $\boldsymbol{\mu_0}=(0,0)',\Sigma_0=I_2$, 那么模型中各个参数的全条件概率密度函数可以写为：

$$p(\boldsymbol{\pi_j})\propto f(\boldsymbol{\pi_j})L_j(\boldsymbol{\pi_j}|\cdot)\propto\prod_{k=1}^{K}p(x|\boldsymbol{\mu_k},\Sigma_k)^{z_{jk}}\cdot\pi_{jk}^{z_{jk}}$$

$$\boldsymbol{\pi_j}\propto Dir(\alpha_0/K+z_{j1},\cdots,\alpha_0/K+z_{jk}) \tag{3}$$

$$p(\alpha_0|\cdot)\propto f(\alpha_0)L_{all}(\alpha_0|\cdot) \tag{4}$$

$$p(\tau_k|\cdot)\propto f(\tau_k)L_{all}(\tau_k|\cdot)\propto L_{all}(\tau_k|\cdot) \tag{5}$$

$$f(c_k|\cdot)\propto f(c_k)L_{all}(c_k|\cdot)\propto L_{all}(c_k|\cdot) \tag{6}$$

$$\begin{aligned}&p(\boldsymbol{\mu_k}|\cdot)\propto f(\boldsymbol{\mu_k})L_{all}(\boldsymbol{\mu_k}|\cdot)\\&\propto exp(-\frac{1}{2}(\boldsymbol{\mu_k}-\boldsymbol{\mu_0})^T\Lambda_0^{-1}(\boldsymbol{\mu_k}-\boldsymbol{\mu_0})+\Sigma_{j=1}^{m}(X_j-\boldsymbol{\mu_k})^T(c_k\Sigma_0)^{-1}(X_j-\boldsymbol{\mu_k}))\end{aligned} \tag{7}$$

这里，$\boldsymbol{\mu_0}$ 和 Λ_0 分别是 $\boldsymbol{\mu_k}$ 先验分布的均值和协方差矩阵。令

$$\boldsymbol{\mu_n}=(\Lambda_0^{-1}+n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\boldsymbol{\mu_0}+n\Sigma^{-1}\frac{\Sigma_{j=1}^{m}X_j}{m})$$

$$\Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}$$

我们可以推得 $\boldsymbol{\mu_k}$ 的后验分布服从 $N(\boldsymbol{\mu_n}, \Lambda_n)$。
$\boldsymbol{z_j}$ 的后验分布为

$$\begin{aligned} p(\boldsymbol{z_j}) &\propto f(\boldsymbol{z_j}) L_j(\boldsymbol{z_j}|\cdot) \\ &= \frac{p(X_j|\boldsymbol{z_j},\cdot)p(\boldsymbol{z_j}|\boldsymbol{\pi_j})}{\Sigma_{k=1}^{K} p(z_j=k)p(X_j|z_j=k,\cdot)} \\ &\propto \prod_{k=1}^{K} p(X_j|\boldsymbol{\mu_k},\Sigma_k)\pi_{jk}^{z_{jk}} \end{aligned} \tag{8}$$

2.5 变点不同情况下参数的后验分布

当 $\tau_1, \cdots, \tau_K$ 不全相同时：

$$\boldsymbol{\pi_j} \propto Dir(\alpha_0/K + z_{j1}, \cdots, \alpha_0/K + z_{jk}) \tag{9}$$

$$p(\tau_k|\cdot) \propto \prod_{j=1}^{m} p(X_j|\boldsymbol{\mu_0},\Sigma_0)^{\mathbf{1}(j\le\tau_k)z_{jk}} p(X_j|\boldsymbol{\mu_k},\Sigma_k)^{\mathbf{1}(j>\tau_k)z_{jk}} \tag{10}$$

$$p(\alpha_0|\cdot) \propto f(\alpha_0)\prod_{j=1}^{m} p(\boldsymbol{\pi_j}|\alpha_0,K) \tag{11}$$

$$p(z_j=k|\cdot) \propto p(X_j|\boldsymbol{\mu_0})^{\mathbf{1}(j\le\tau_k)} p(X_j|\boldsymbol{\mu_k})^{\mathbf{1}(j>\tau_k)}\pi_{jk} \tag{12}$$

$$\begin{aligned} p(\boldsymbol{\mu_k}|\cdot) &\propto f(\boldsymbol{\mu_k})\prod_{j=1}^{m} p(X_j|\boldsymbol{\mu_k},\cdot) \\ &\propto exp(-\frac{1}{2}(\boldsymbol{\mu_k}-\boldsymbol{\mu_0})^T\Lambda_0^{-1}(\boldsymbol{\mu_k}-\boldsymbol{\mu_0}) + \Sigma_{j=1}^{m}(X_j-\boldsymbol{\mu_k})^T(c_k\Sigma_0)^{-1}(X_j-\boldsymbol{\mu})\mathbf{1}(j\ge\tau_k)) \end{aligned} \tag{13}$$

$$f(c_k|\cdot) \propto \prod_{j=1}^{m} p(X_j|c_k\Sigma_0,\cdot)\mathbf{1}(j>\tau_k) \tag{14}$$

2.6 贝叶斯后验分布抽样

变点模型的计算非常复杂，即使是在各机器变点都相同的情况下，一些参数的后验分布仍然没有闭合解。如果没有 MCMC 算法，我们必须假设上述的二元过程均值和方差都已知，且变点前后的分布都是共轭先验的。很明显，这些条件都满足是不现实的。为此，我们用 MCMC 算法中比较著名的 Gibbs sampling，从各参数后验分布中随机抽样。实现的过程运用了 JAGS 和 R2JAGS 接口。我们对 MCMC 算法的初值作如下设定: 根据 τ_k 的先验分布抽取它的初值 $\tau_k^{(0)}, k=1,\cdots,K$。从 $\boldsymbol{\mu_k}$ 的先验分布中抽取它的初值 $\boldsymbol{\mu_k^{(0)}}, k=1,\cdots,K$。从 c_k 的先验分布 $Unif(0,1)$ 中抽取 c_k 的初值 $c_k^{(0)}, k=1,\cdots,K$。从 α_0 的先验分布 $Gamma(\frac{1}{2},\frac{1}{2})$ 中抽取它的初值 $\alpha_0^{(0)}$。在给定 $\alpha_0^{(0)}$ 的情况下, 从 $Dir(\alpha_0^{(0)}/K)$ 中抽取 $\boldsymbol{\pi_j}$ 的初值 $\boldsymbol{\pi_j^{(0)}}, j=1,\cdots,m$。再根据 Eq.8和 Eq.12, 抽取隐变量 $\boldsymbol{z_j}$ 的初值 $\boldsymbol{z_j^{(0)}}$。至此，所有参数的初值都已经产生。

下面是算法的迭代更新过程。

Step(1) 根据 Eq.3和 Eq.9更新 $\boldsymbol{\pi_j}$ 产生 $\boldsymbol{\pi_j^{(t)}}, j=1,2,\cdots,m$。

Step(2) 根据 Eq.4和 Eq.11，更新 α_0 产生新的 $\alpha_0^{(t)}$。

Step(3) 根据 Eq.7和 Eq.13产生新的 $\boldsymbol{\mu_k}, k=1,\cdots,K$。

Step(4) 根据 Eq.6和 Eq.14产生新的 $c_k, k=1,\cdots,K$。

Step(5) 根据 Eq.5和 Eq.10 产生新的 $\tau_k, k=1,\cdots,K$。

Step(6) 根据 Eq.8和 Eq.12产生新的 $z_j, j=1,\cdots,m$。

注:假设 $x_1,\cdots,x_n$ 狄利克雷分布 $Dir(\alpha_1,\cdots,\alpha_n)$, 我们可以先从 n 个独立的伽马分布 $Gamma(\alpha_i,1)$ 中抽取分别抽取一个样本，记作 $y_1,\cdots,y_k$, 然后取 $x_i=\frac{y_i}{\Sigma_{i=1}^n y_i}, i=1,\cdots,n$，此时 $x_1,\cdots,x_n$ 就是狄利克雷分布 $Dir(\alpha_1,\cdots,\alpha_n)$ 的一个样本。

我们在随机模拟中用 JAGS 辅助运算。初值不同的五条链会一直迭代更新直到 Gelman-Rubin 统计量收敛到 1。各参数的点估计是它们除去 burn-in 阶段的后验均值。由于总机器数是事先给定的，所以贝叶斯有限混合模型的参数数量始终都是恒定的，所以在理论上这些 MCMC 样本会收敛到一个预期的联合分布。当然在实现的过程中，我们经常需要检查 MCMC 链的收敛与否。比较流行的两种诊断方法是 Raftery and Lewis (1996) 和 Gelman and Rubin (1992)。R-L 方法是基于监控每一条单链的自相关性，它给出了一条 MCMC 链理论上最短的迭代次数和建议的 burn-in 阶段的长度。G-R 方法则需要运行每条单链许多次，每次选取不同的初值，并通过计算 Gelman-Rubin 统计量判断收敛性。Nylander et al. (2007) 讨论了 MCMC 算法估计后验均值的准确性，这里不再赘述。

表 1: 在变点相同情况下的四个场景

场景	描述
1	总量为 70 的混合样本产生于两个机器，混合样本的比例相同受控状态下服从均值为 $(0,0)$，协方差为 I_2 的二元正态分布。变点位置设置在获得 30 个观测的位置。变点后机器一的中心为 $(0,1)$, 机器二的中心为 $(1,0)$，方差不变。
2	总量为 90 的混合样本产生于三个机器，混合样本的比例相同。受控状态下服从均值为 $(0,0)$，协方差为 I_2 的二元正态分布。变点位置设置在获得 30 个观测的位置。变点后机器一的中心为 $(0,1)$, 机器二的中心为 $(1,0)$，机器三的中心为 $(1,1)$，方差不变。这里设置小漂移的目的是尽可能多地采集失控状态下的样本。
3	总量为 70 的混合样本产生于两个机器，混合样本的比例不同，其余条件都于场景 1 相同。
4	总量为 70 的混合样本产生于两个机器，混合样本的比例相同受控状态下服从均值为 $(0,0)$，协方差为 I_2 的二元正态分布。变点位置设置在获得 30 个观测的位置。变点后产品的中心不变，机器 1 生产产品的方差变为 $0.5\cdot I_2$, 而机器 2 生产产品的方差变为 $2\cdot I_2$。

3 随机模拟

3.1 场景设置

我们分别从把随机模拟分为两个部分。第一个部分所有机器发生故障的时间点均相同，在这个前提假设下，我们考虑产生数据的时变点的位置，变点后各机器新的均值，以及各机器产生样本的混合比例等情况，设置了四个场景。第二部分考虑各机器变点不一定相同，此时同样设置四个场景。在这两个部分，我们都假设受控分布已知，且样本都来自于控制图发出报警信号之前。场景设置的具体参数可以见表格 1和 2。

3.2 混合样本的产生

产生变点相同的混合样本时，首先确定变点的位置 τ。然后依次产生 τ 个服从 $N(\mathbf{0},I_2)$ 的二元随机样本，再等概率地从 K 个失控分布中产生 N 个样本，依次排列在 τ 个服从受控分布的二元随机

表 2: 在变点不同情况下的四个场景

场景	描述
1	总量为 70 的混合样本产生于两个机器，混合样本的比例相同受控状态下都服从均值为 $(0,0)$，协方差为 I_2 的二元正态分布。机器一的变点设置在获得 20 个观测的位置，机器二的变点设置在获得 40 个观测的位置。变点后机器一的中心为 $(0,1)$,，变点后机器二的中心为 $(1,0)$，方差不变。
2	总量为 90 的混合样本产生于三个机器，混合样本的比例相同。受控状态下服从均值为 $(0,0)$，协方差为 I_2 的二元正态分布。机器一的断点设置在获得 10 个观测的位置，机器二的断点设置在获得 20 个观测的位置，机器三的断点设置在获得 30 个观测的位置。变点后机器一的中心为 $(0,0.5)$，机器二的中心为 $(0.5,0)$，机器三的中心为 $(0.50.5)$，方差不变。设置小漂移的目的是尽可能多地采集失控状态下的样本。
3	总量为 70 的混合样本产生于两个机器，混合样本的比例不同，其余条件都于场景 1 相同。
4	总量为 70 的混合样本产生于两个机器，混合样本的比例相同受控状态下服从均值为 $(0,0)$，协方差为 I_2 的二元正态分布。在受控分布下，机器一产生了 10 个样本，机器二产生了 20 个样本。变点后产品的中心不变，机器 1 生产产品的方差变为 $0.5\cdot I_2$，而机器 2 生产产品的方差变为 $2\cdot I_2$。

表 3: 场景一：MEWMA 控制图报警前共收集 48 个样本,h=8.455,$ARL_0 = 200$

	真实值	估计值
变点位置	30	29
μ_{11}	0	0.087
μ_{12}	1	1.493
μ_{21}	1	0.732
μ_{22}	0	0.108

表 4: 场景三：MEWMA 控制图报警前共收集 52 个样本,h=8.455,$ARL_0 = 200$

	真实值	估计值
变点位置	30	33
μ_{11}	0	0.013
μ_{12}	0.5	0.378
μ_{21}	0.5	0.447
μ_{22}	0	0.102
μ_{31}	0.5	0.391
μ_{32}	0.5	0.324

样本之后。

产生变点不同的混合样本时，先确定 K 个变点的位置 $\tau_1,\cdots,\tau_K$。不失一般性，我们可以假设 $\tau_1 \leq \cdots \leq \tau_K$。为了方便说明，我们用表 2下的场景一来举例说明产生样本的过程。依次产生 20 个服从 $N(\mathbf{0},I_2)$ 的受控样本; 依次等概率产生 20 个服从 $N((0,1),I_2)$ 和 $N(\mathbf{0},I_2)$ 的样本；依次等概率产生 20 个服从 $N((0,1),I_2)$ 和 $N(10,I_2)$ 的样本。三组样本依次排列。

3.3 MEWMA 控制图

我们用 MEWMA 控制图来监测多元生产过程。假设 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots$ 在受控状态下服从一个 p 元正态分布 $N_2(\boldsymbol{\mu_0},\Sigma_0)$，这里 $\boldsymbol{\mu_0},\Sigma_0$ 设为已知。定义统计量 $\boldsymbol{E}_n = \Lambda(\boldsymbol{X}_n - \boldsymbol{\mu_0}) + (I_{pxp} - \Lambda)\boldsymbol{E}_{n-1}$, 这里 $\boldsymbol{E}_0 = 0,\Lambda = diag(\lambda_1,\cdots,\lambda_p)$。当 $V_n^2 = \boldsymbol{E}_n'\Sigma_{\boldsymbol{E}_n}^{-1}\boldsymbol{E}_n > h$ 时，控制图给出报警警报。这里，$\lambda_1 = \cdots = \lambda_p = \lambda, \Sigma_{\boldsymbol{E}_n} = \frac{\lambda}{2-\lambda}[1-(1-\lambda)^{2n}]\Sigma_0$.。$h$ 的选择满足 $ARL_0 = c$,Lowry et al. (1992),c 是我们预先定好的值。

这里，我们取 $h = 200, \lambda = 0.1$。用 MEWMA 控制图监控每组场景下的生产过程，并取控制图报警前的样本作为输入数据，用 JAGS 进行运算。

3.4 模拟结果

在每组场景设置下，随机模拟 200 次，得到 200 组各参数的估计值，取均值，结果可参见表??。

图 1是表 1场景 1 下变点位置的后验概率密度函数图。

图 2是表 2场景 1 下两个变点位置的后验概率密度函数图。

图 3是表 1场景一下机器一均值第一个分量的后验分布。

图 4该场景下机器一均值第二个分量的后验分布。

图 5该场景下机器二均值第一个分量的后验分布。

表 5: 场景三：MEWMA 控制图报警前共收集 46 个样本,h=8.455,$ARL_0 = 200$

	真实值	估计值
变点位置	30	30
μ_{11}	0	0.1217
μ_{12}	1	0.870
μ_{21}	1	0.898
μ_{22}	0	0.133

表 6: 场景四：MEWMA 控制图报警前共收集 41 个样本,h=8.455,$ARL_0 = 200$

	真实值	估计值
变点位置	30	27
c_1	0.5	0.326
c_2	2	1.520

表 7: 变点相同情况下混合样本的聚类准确性 %

场景 1	场景 2	场景 3	场景 4
82%	78%	80%	75%

表 8: 场景一：MEWMA 控制图报警前共收集 45 个样本,h=8.455,$ARL_0 = 200$

	真实值	估计值
机器一变点位置	20	22
机器二变点位置	40	38
μ_{11}	0	0.139
μ_{12}	1	0.844
μ_{21}	1	0.747
μ_{22}	0	0.102

表 9: 场景二：MEWMA 控制图报警前共收集 54 个样本,h=8.455,$ARL_0 = 200$

	真实值	估计值
机器一变点位置	10	14
机器二变点位置	20	19
机器三变点位置	30	36
μ_{11}	0	0.009
μ_{12}	0.5	0.420
μ_{21}	0.5	0.237
μ_{22}	0	0.202
μ_{31}	0.5	0.491
μ_{32}	0.5	0.224

表 10: 场景三：MEWMA 控制图报警前共收集 46 个样本,h=8.455,$ARL_0 = 200$

	真实值	估计值
变点位置	30	34
μ_{11}	0	0.207
μ_{12}	1	0.802
μ_{21}	1	0.901
μ_{22}	0	0.323

表 11: 场景四：MEWMA 控制图报警前共收集 41 个样本,h=8.455,$ARL_0 = 200$

	真实值	估计值
变点位置	30	27
c_1	0.5	0.326
c_2	2	1.520

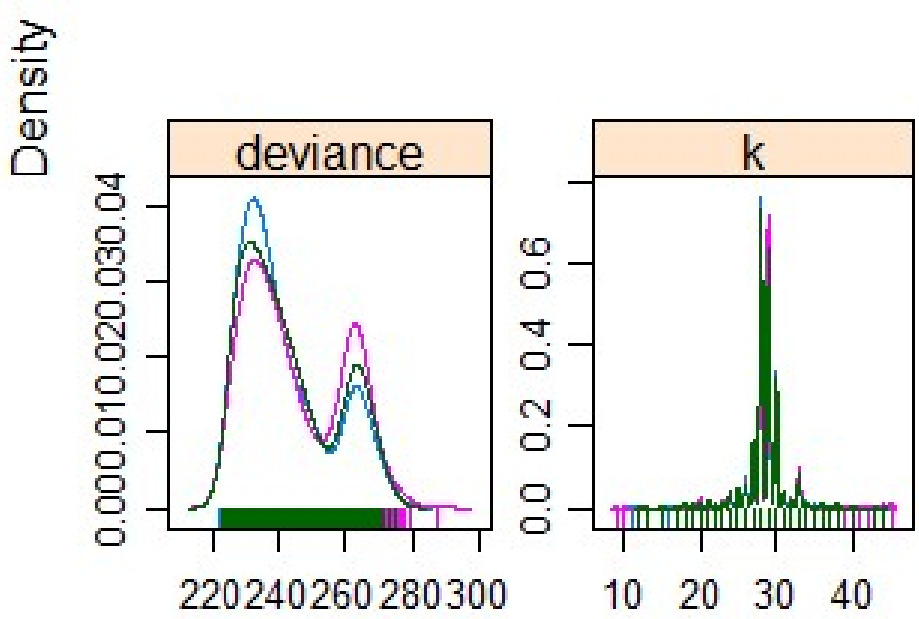

图 1: 变点相同时场景 1 下变点位置的后验概率密度函数图

图 6该场景下机器二均值第二个分量的后验分布

4 贝叶斯方法与极大似然估计方法的对比研究

Hawkins et al. (2003) 给出了仅过程均值发生变化时的变点检测方法，之后 Hawkins and Zamba (2005) 给出了过程方差都发生变化时的变点检测方法。但两者都不适用于混合样本的情况。对于贝叶斯有限混合模型而言，样本产生于同一总体却可以看作是混合样本的一个特例。要比较两种方法，首先产生一个 100 个服从二元过程的样本。这些样本在受控前每个分量均值都为 0，方差为 1，两个变量之间的相关系数为 0.5. 我们假设过程在产生 50 个样本之后出现了均值漂移，我们用 MEWMA 控制图监测该过程，直到控制图报警终止采样。我们设置 MEWMA 控制图的参数 $\lambda = 0.1$,$ARL_0 = 200$, 并由此计算出控制图的控制限为 8., 从图 7可以看出，在控制图在产生了 76 个样本之后发出了报警信号.

为了方便比较贝叶斯方法和 MLE 方法估计的结果，我们计算估计值与真值的误差,$bias(\%) = \frac{(\hat{\tau}-\tau)}{\tau}$，通过多次随机模拟再取均值，误差越小，则估计结果越理想。表 12是 200 次随机模拟后的结果。

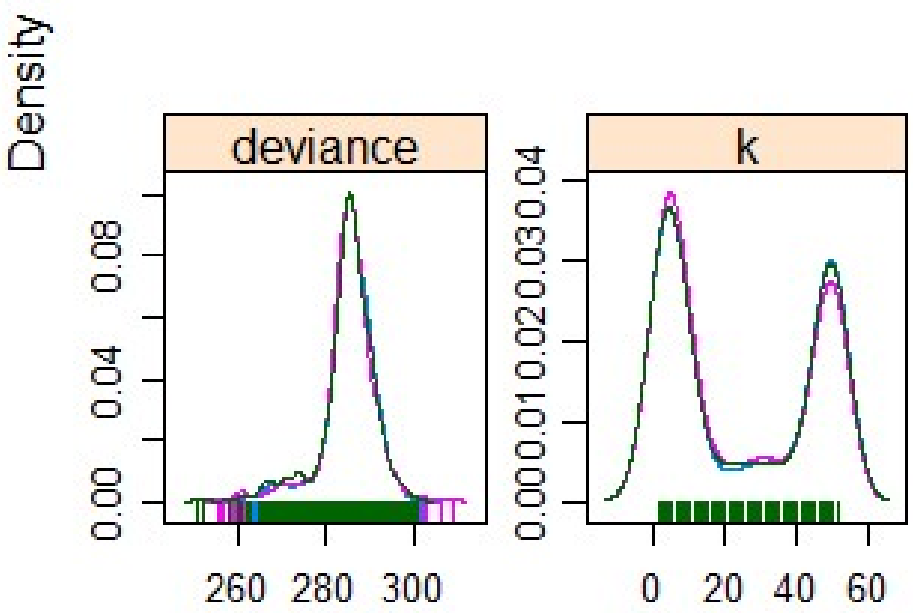

图 2: 变点不同时场景 1 下两个变点位置的后验概率密度函数图

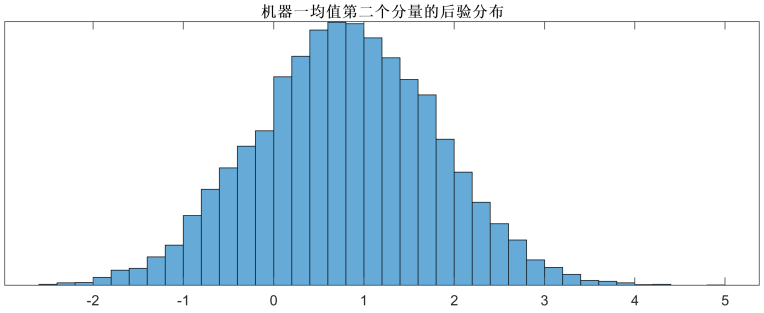

图 3: 机器一均值第一个分量的后验分布

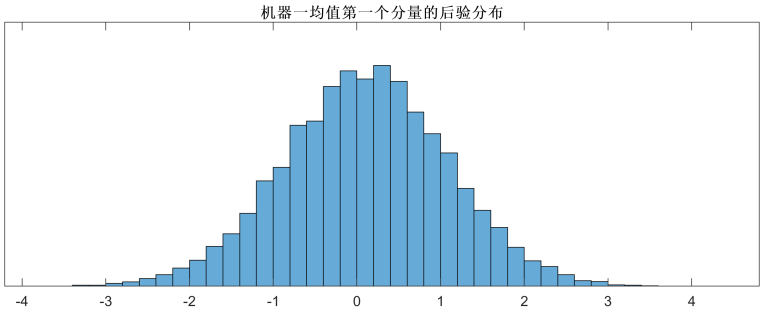

图 4: 机器一均值第二个分量的后验分布

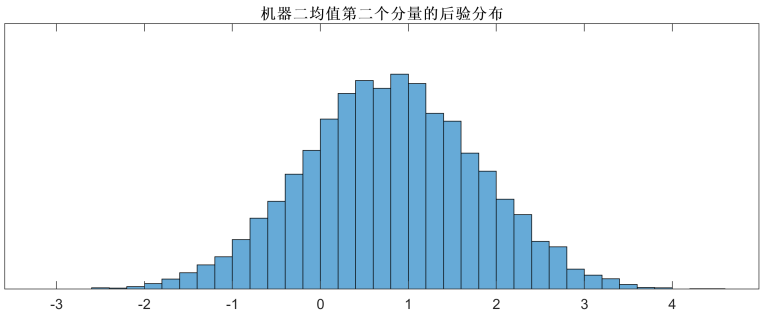

图 5: 机器二均值第一个分量的后验分布

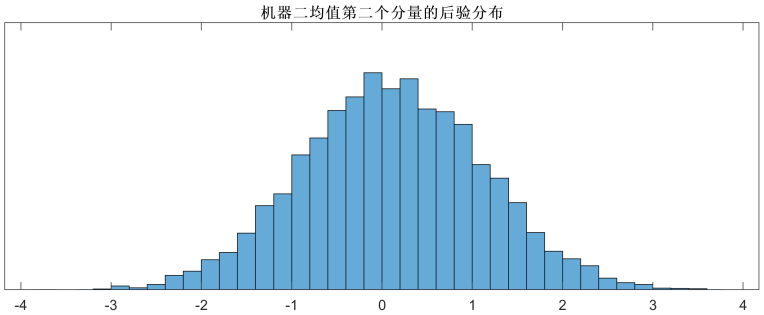

图 6: 机器二均值第二个分量的后验分布

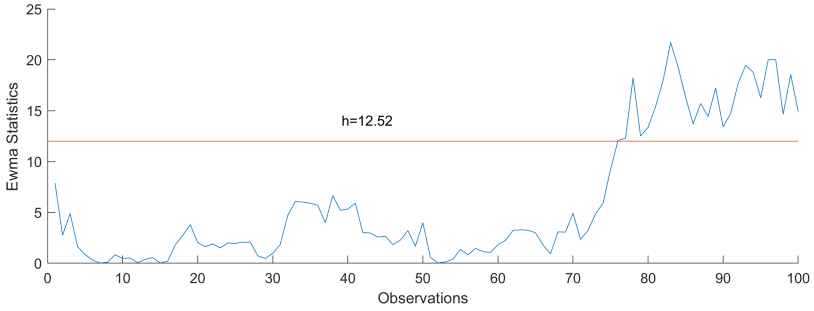

图 7: 随机二元过程的 MEWMA 控制图

表 12: 贝叶斯方法和 MLE 方法的误差率

方法	均值漂移量为 0.5σ	均值漂移量为 1σ
贝叶斯	23.6	12.1
MLE	37.0	18.2

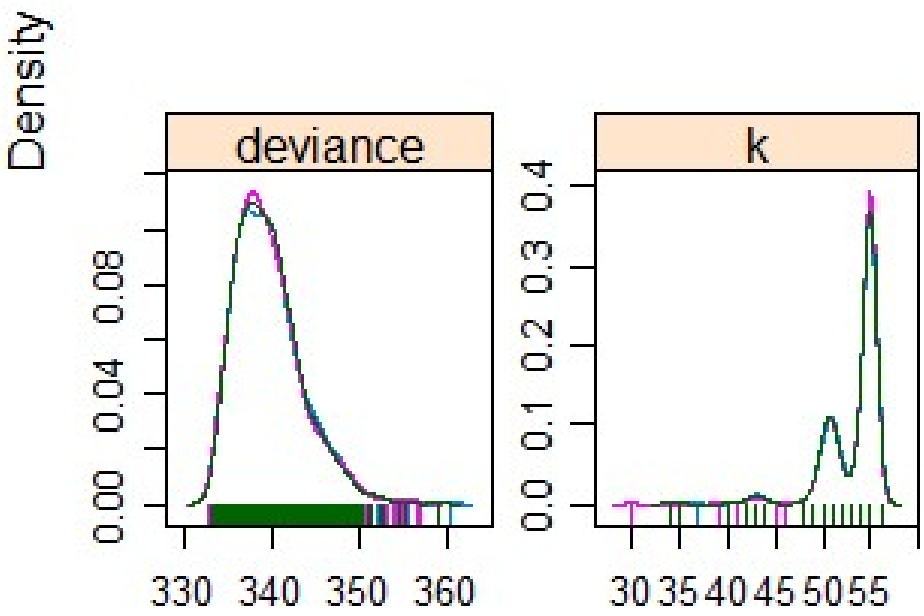

图 8: 贝叶斯方法对非混合样本的变点估计

图 8描述了其中一次随机模拟中变点位置的后验概率密度函数图。

5 结论

通过随机模拟，我们可以看到，当混合样本的变点相同时，贝叶斯有限混合模型能够较为精确地检测混合样本的变点，但由于过程从失控到控制图报警这段时间内产生的样本量有限，所以在估计变点后各机器产品的均值时准确性不够高。为了解决这个问题，我们可以通过调高控制限，从而获得失控状态下的样本，从而更好地估计上述参数。用贝叶斯方法的一个优点是它可以包括许多实际的先验信息，从而提高估计的准确率。比如，如果我们认为过程至少在 10 个观测之后失控，那么变点的先验分布满足 $P(\tau_k \in \{1,\cdots,10\}) = 0$, 这样在后验推断中就不会出现有变点落在 $1-10$ 的区间内的情况，从而提高了估计结果的准确性。

6 讨论和展望

在本文中，我们展示了基于贝叶斯有限混合模型对混合样本过程变点的检测及变点前后均值的估计。我们在模型中通常假定混合样本来自 K 个机器，K 是已知的。在大多数情况下，这个假设是可以被满足的。当如果我们不知道机器的总数，即 K 是一个变量时，我们也需要对 K 作贝叶斯后验推断。Li et al. (2018) 假设 K 的先验分布是一个左闭右开的狄利克雷分布，在最后选择模型的时候用到了 DIC 准则。这种方法虽然简单，但不能很好地估计潜在的机器个数。我会在后续研究的过程

中使用贝叶斯因子的方法 Berger and Pericchi (1996) 进行模型选择，最终目的都是对未知总体混合样本过程的统计分析。

7 致谢

参考文献

Barnard, G. A. (1959). Control charts and stochastic processes. Journal of the Royal Statistical Society. Series B (Methodological), pages 239–271.

Berger, J. O. and Pericchi, L. R. (1996). The intrinsic bayes factor for model selection and prediction. Journal of the American Statistical Association, 91(433):109–122.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. The Annals of Mathematical Statistics, 35(3):999–1018.

Diebolt, J. and Robert, C. P. (1994). Estimation of finite mixture distributions through bayesian sampling. Journal of the Royal Statistical Society. Series B (Methodological), pages 363–375.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical science, pages 457–472.

Hawkins, D. M., Qiu, P., and Kang, C. W. (2003). The changepoint model for statistical process control. Journal of quality technology, 35(4):355.

Hawkins, D. M. and Zamba, K. (2005). Statistical process control for shifts in mean or variance using a changepoint formulation. Technometrics, 47(2):164–173.

Li, Q., Guo, F., Kim, I., Klauer, S. G., and Simons-Morton, B. G. (2018). A bayesian finite mixture change-point model for assessing the risk of novice teenage drivers. Journal of applied statistics, 45(4):604–625.

Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. Technometrics, 34(1):46–53.

Mason, R. L., Tracy, N. D., and Young, J. C. (1995). Decomposition of t 2 for multivariate control chart interpretation. Journal of quality technology, 27(2):99–108.

Nylander, J. A., Wilgenbusch, J. C., Warren, D. L., and Swofford, D. L. (2007). Awty (are we there yet?): a system for graphical exploration of mcmc convergence in bayesian phylogenetics. Bioinformatics, 24(4):581–583.

Pignatiello Jr, J. J. and Samuel, T. R. (2001). Estimation of the change point of a normal process mean in spc applications. Journal of Quality technology, 33(1):82–95.

Prabhu, S. S. and Runger, G. C. (1997). Designing a multivariate ewma control chart. Journal of Quality Technology, 29(1):8.

Raftery, A. E. and Lewis, S. M. (1996). Implementing mcmc. Markov chain Monte Carlo in practice, pages 115–130.