第11章　抽样方法与MCMC

2025年11月

本讲介绍

- 本章主要介绍了基本的抽样方法和MCMC抽样方法的原理、步骤及应用场景。
- 在机器学习中，常常需要从复杂的概率分布中抽样以进行参数估计、模型推断等任务。例如，在一般的（非线性非高斯）状态空间模型中，我们需要根据参数的后验分布$\boldsymbol{p}(\boldsymbol{\theta}|\mathbf{y}_{\mathbf{1}:\boldsymbol{T}})$推断参数的性质。
- 然而$\boldsymbol{p}(\boldsymbol{\theta}|\mathbf{y}_{\mathbf{1}:\boldsymbol{T}})$可能是高维的，且包含了隐变量的多重积分变换。传统的抽样方法在面对高维复杂分布时存在诸多局限性，因此需要更强大有效的抽样方法，马尔可夫链蒙特卡罗（MCMC）方法应运而生。
- MCMC 旨在解决高维复杂分布的抽样难题，为贝叶斯推断和生成式机器学习提供有力工具。

本讲介绍

- **传统抽样方法包括简单分布抽样、拒绝采样、重要性采样和SIR等。**
- **MCMC 方法通过构造具有目标平稳分布的马尔可夫链来获取样本，适用于高维复杂分布的抽样和推断。Metropolis - Hastings 算法通过特定的接受概率准则从提议分布中采样并更新状态；Gibbs 采样是Metropolis - Hastings 算法的一个特例，利用条件独立性依次对变量进行采样，保证了100%的接受概率。**
- **本章的重点在于理解各类抽样方法的核心原理和操作流程，尤其是MCMC算法的原理与性质。难点在于学会在实际问题中如何应用 Metropolis - Hastings 算法和 Gibbs 采样，根据目标分布和数据特点选择合适的抽样方法、提议分布和参数设置，优化采样效率。**

本讲关键问题

（1）拒绝采样的原理和操作步骤是怎样的？提议分布起到什么作用？

（2）重要性采样基于什么原理来近似期望结果？提议分布的选择如何影响重要性采样的效果？

（3）SIR方法如何改进拒绝采样？

（4）为何在高维空间中，拒绝采样和重要性采样会面临困难？

（5）MCMC方法的基本思想是什么？

（6）什么是细致平稳条件？马尔可夫链的收敛条件是什么？

（7）Metropolis - Hastings 算法的具体步骤是什么？在提议分布对称和不对称的情况下，算法有何不同？

（8）Gibbs 采样为何能保证接受概率为 1？

关键词： MCMC、 Metropolis - Hastings 算法、抽样、 Gibbs 采样马尔可夫链

一、简单分布抽样

如何利用均匀分布的随机数生成非均匀分布的随机数？

假设可以生成区间$(0,1)$内均匀分布的随机变量z，则可通过$y=F^{-1}(z)$生成密度函数为$p(y)$的随机数，其中$F(\cdot)$是y的累积概率密度函数。这种方法需要$F(\cdot)$具有解析表达式。

若变量 $y\sim\mathcal{N}(0,1)$ ，则变量 $\sigma y+\mu\sim\mathcal{N}(\mu,\sigma^2)$ 。

多维正态分布 $y\sim\mathcal{N}(\mu,\Sigma)$ ，可以使用 Cholesky 分解 $\Sigma=LL^T$。若 $z\sim\mathcal{N}(0,I)$ 是独立标准正态分布向量，则 $y=\mu+Lz\sim\mathcal{N}(\mu,\Sigma)$ 。

二、拒绝采样(Rejection Sampling)

假设我们希望从分布 $p(z)$中采样，但直接从$p(z)$采样相对困难：

$$p(z) = \frac{1}{Z_p}\tilde{p}(z),$$

对任意z，$\tilde{p}(z)$可以计算，但归一化常数Z_p未知。

在拒绝采样中，我们需要一个较简单的分布$q(z)$，称为提议分布（proposal distribution），从中可以直接生成样本。

引入常数 k，使得对于所有z 都有 $kq(z) \geq \tilde{p}(z)$ 。

二、拒绝采样

拒绝采样的步骤涉及生成两个随机数

1. 从分布 $q(z)$ 中采样得到z_0；
2. 生成一个均匀分布的随机数 $u_0 \sim \mathrm{U}[0, kq(z_0)]$。
3. 如果 $u_0 > \tilde{p}(z_0)$，则拒绝 z_0，否则接受 z_0。（如果(z_0, u_0)落在灰色阴影区域内，则被拒绝）。

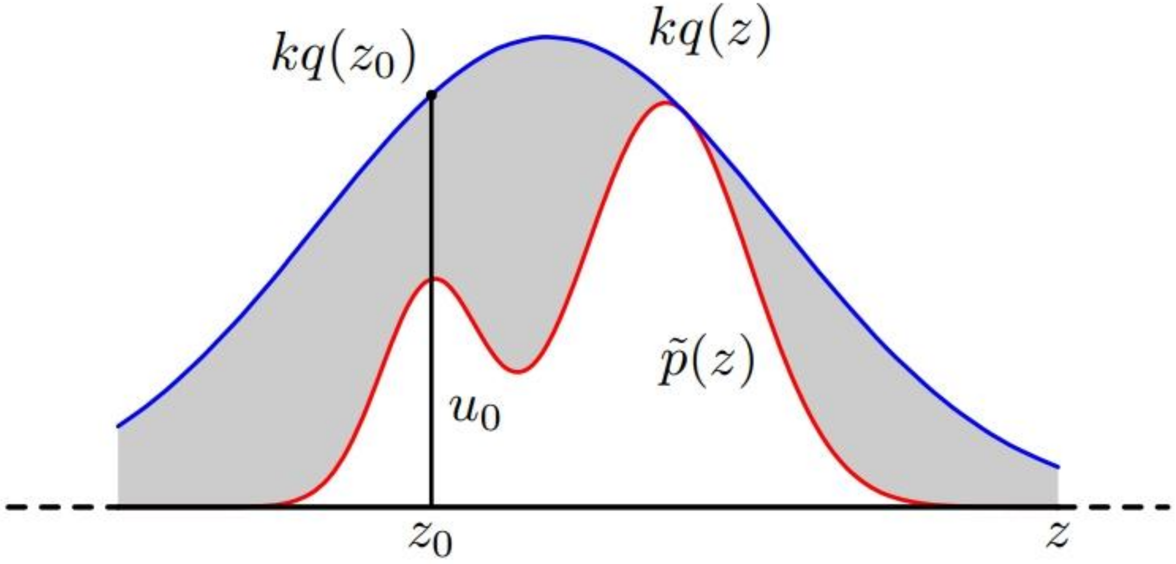

图12.1 拒绝采样

二、拒绝采样

被提出并接受的 z 的概率为：

$$p(\text{propose and accept } z)=\frac{\tilde{p}(z)}{kq(z)}q(z)=\frac{\tilde{p}(z)}{k}.$$

总接受的概率为：

$$p(\text{accept})=\int p(\text{propose and accept } z)dz=\frac{Z_p}{k}.$$

因此，在所有被接受的样本中 z 的分布为：

$$p(z|\text{accept})=\frac{p(\text{propose and accept } z)}{p(\text{accept})}=\frac{\tilde{p}(z)}{Z_p}.$$

接收点的比例取决于 $\tilde{p}(z)$下的面积与 $kq(z)$下的面积之比。常数 k 应该尽可能小，同时确保 $kq(z)\geq\tilde{p}(z)$。

二、拒绝采样

拒绝采样在高维空间中非常难以实现：考 虑 $p(x)=\mathcal{N}(0,\sigma_p^2 I)$, $q(x)=\mathcal{N}(0,\sigma_q^2 I)$, $\sigma_q^2 \geq \sigma_p^2$ 以确保它是一个上界。 在 D 维空间中，最优 $k=\left(\frac{\sigma_q}{\sigma_p}\right)^D$，接受率为 $1/k$。随着 D的增加，接受率呈指数下降。

三、重要性采样(Importance Sampling)

在某些情况下，我们需要估计一个概率的期望值：

$$\mathbb{E}[f] = \int f(z)p(z)dz$$

例如在状态空间模型中，

$$p(z_t|y_{1:t-1}) = \int p(z_t|z_{t-1})p(z_{t-1}|y_{1:t-1})dz_{t-1}$$

重要性采样提供了近似期望结果，但不直接从$p(z)$ 中直接抽样。

三、重要性采样

基于一个提议分布 $q(z)$，期望值 $\mathbb{E}[f]$ 可以近似表示为从$q(z)$中随机样本的有限求和：

$$\mathbb{E}[f] = \int f(z)p(z)dz = \int f(z)\frac{p(z)}{q(z)}q(z)dz \approx \frac{1}{N}\sum_{n=1}^{N}\frac{p(z^{(n)})}{q(z^{(n)})}f(z^{(n)}).$$

权重 $r_n = \frac{p(z^{(n)})}{q(z^{(n)})}$ 被称为重要性权重（importance weight），用以修正从 $q(z)$而非 $p(z)$ 采样所引入的偏差。与拒绝采样不同，生成的所有样本都会被保留。

三、重要性采样

在许多情况下，

$$p(z) = \frac{\tilde{p}(z)}{Z_p}, \quad q(z) = \frac{\tilde{q}(z)}{Z_q}$$

Z_p, Z_q均未知。

在这种情况下，我们有：

$$\mathbb{E}[f] = \frac{Z_q}{Z_p} \int f(z) \frac{\tilde{p}(z)}{\tilde{q}(z)} q(z) dz \approx \frac{Z_q}{Z_p} \frac{1}{N} \sum_{n=1}^{N} r_n f(z^{(n)}),$$

$$\frac{Z_p}{Z_q} = \frac{1}{Z_q} \int \tilde{p}(z) dz = \int \frac{\tilde{p}(z)}{\tilde{q}(z)} q(z) dz \approx \frac{1}{N} \sum_{n=1}^{N} r_n,$$

其中 $r_n = \frac{\tilde{p}(z^{(n)})}{\tilde{q}(z^{(n)})}$ 。该期望值可写为：

$$\mathbb{E}[f] \approx \sum_{n=1}^{N} w_n f(z^{(n)}),$$

$$w_n = \frac{r_n}{\sum_{m=1}^{N} r_m} = \frac{\tilde{p}(z^{(n)}) / \tilde{q}(z^{(n)})}{\sum_{m=1}^{N} \tilde{p}(z^{(m)}) / \tilde{q}(z^{(m)})}. \tag{1}$$

三、重要性采样

重要性采样方法的采样效率在很大程度上取决于采样分布 $q(z)$与目标分布 $p(z)$的匹配程度。

如果 $p(z)$ 波动很大，并且集中于采样分布 $q(z)$中z概率较小区域，那么重要性权重$\{r_n\}$可能被少数具有较大的权重的样本点主导，而其余权重则相对较小，导致有效样本数远小于样本数 N，期望值的估计可能会严重偏离正确值。

这也突出了采样分布 $q(z)$的一个关键要求：**它不应在 $p(z)$可能显著的区域中很小或为零**。

四、采样-重要性-重采样(Sampling-Importance-Resampling, SIR)

拒绝采样依赖于常数 k 的值。足够大的 k 值可以保证分布 $p(z)$ 被 $kq(z)$所覆盖，但会导致接收率极低。

采样-重要性-重采样（SIR）与拒绝采样类似，但无需确定常数 k 。

SIR 遵循如下步骤：

1. 从提议分布 $q(z)$ 中抽取 N 个样本 $z^{(1)},\dots,z^{(N)}$；
2. 使用公式（1）计算权重 $w_1,\dots,w_N$；
3. 根据权重 $(w_1,\dots,w_N)$ 从离散分布 $(z^{(1)},\dots,z^{(N)})$ 中重采样新的一组样本。

四、采样-重要性-重采样

当 $N \to \infty$ 时，得到的样本近似服从目标分布 $p(z)$。对于一维情况，抽样结果 $z \leq a$的概率是

$$P(z \leq a) = \sum_n I\left(z^{(n)} \leq a\right) w_n = \frac{\sum_n I\left(z^{(n)} \leq a\right) \tilde{p}\left(z^{(n)}\right) / q\left(z^{(n)}\right)}{\sum_n \tilde{p}\left(z^{(n)}\right) / q\left(z^{(n)}\right)}$$

其中 $I(\cdot)$ 是示性函数。令$f(z) = I(z \leq a)$看作z的一个函数，当 $N \to \infty$ 时，

$$P(z \leq a) \to \frac{\int I(z \leq a) \tilde{p}(z)/q(z) \times q(z) dz}{\int \tilde{p}(z)/q(z) \times q(z) dz} = \frac{\int I(z \leq a) \tilde{p}(z)\, dz}{Z_p} = \int I(z \leq a) p(z)\, dz$$

表明从 SIR 样本得到的累积分布函数逐渐逼近 $p(z)$的累积分布。

五、马尔可夫链蒙特卡罗方法 (MCMC)

拒绝采样和重要性采样在高维空间中存在局限性。在本节中，我们讨论一个非常通用且强大的方法马尔可夫链蒙特卡罗 (Markov Chain Monte Carlo, MCMC)。

MCMC的思想是构造一条马尔可夫链，其平稳分布即为目标密度 $p^*(x)$。在贝叶斯情境中，目标密度通常是后验分布 $p^*(x) \propto p(x|D)$，但MCMC也可以用于从任何高维分布中生成样本。

在状态空间中生成一个随机过程，使得在每个状态 x 上停留的时间比例与 $p^*(x)$ 成正比。由马尔可夫链所得到的样本序列 $x_0, x_1, x_2, \ldots$ 也可以用于对 p^* 进行蒙特卡罗积分。

初始样本并非来自平稳分布 $p^*(x)$，因此应当予以舍弃。达到平稳状态所需的时间称为混合时间（mixing time）或预烧时间（burn-in time）。减少预烧时间是加速算法的最重要因素之一。

五、马尔可夫链蒙特卡罗方法 (MCMC) – 马尔可夫链

首先我们介绍马尔可夫链：什么条件可以保证一条马尔可夫链收敛到目标分布？

一阶马尔可夫链定义为一系列随机变量 $z^{(1)},\dots,z^{(M)}$，满足条件独立性：

$$p(z^{(m+1)}|z^{(1)},\dots,z^{(m)}) = p(z^{(m+1)}|z^{(m)}),\quad m\in\{1,\dots,M-1\}.$$

一条马尔可夫链可以定义为：

- 变量的初始分布 $p(z^{(0)})$；
- 后续变量的条件概率，即转移概率 $T_m(z^{(m)},z^{(m+1)})\equiv p(z^{(m+1)}|z^{(m)})$ 。

如果所有转移概率都相同，则该马尔可夫链称为**齐次**的。

五、马尔可夫链蒙特卡罗方法 (MCMC) － 马尔可夫链

$z^{(m+1)}$ 的边缘分布可表示为：

$$p(z^{(m+1)}) = \sum_{z^{(m)}} p(z^{(m+1)}|z^{(m)})p(z^{(m)}).$$

如果每一步的边缘分布保持不变，则称该链为**平稳**的。对于具有转移概率 $T(z', z)$的齐次马尔可夫链，其在分布 $p^*(z)$ 下平稳的充要条件为

$$p^*(z) = \sum_{z'} T(z', z)p^*(z').$$

一个给定的马尔可夫链可能具有多个平稳分布。例如，如果转移概率为恒等变换，则 $p^*(z)$可以是任意分布。

五、马尔可夫链蒙特卡罗方法 (MCMC) – 马尔可夫链

保证平稳分布 $p^*(z)$ 的充分但非必要条件是转移概率满足**细致平衡**条件：

$$p^*(z)T(z,z') = p^*(z')T(z',z), \quad \forall z, z'.$$

也即前一状态是 z 而后一状态是 z'的联合概率等于前一状态是z'而后一状态是z的联合概率。因为：

$$\sum_{z'} p^*(z')T(z',z) = \sum_{z'} p^*(z)T(z,z') = p^*(z)\sum_{z'} p(z'|z) = p^*(z).$$

满足细致平衡条件的马尔可夫链称为**可逆**的。

五、马尔可夫链蒙特卡罗方法 (MCMC) – 马尔可夫链

MCMC通过建立一个目标分布是$p^*(z)$且平稳的马尔可夫链来得到$p^*(z)$的随机样本，必须要求当 $m \to \infty$ 时，无论初始分布 $p(z^{(0)})$是什么， $p(z^{(m)})$都会收敛到所需的平稳分布 $p^*(z)$ ，此时该链满足遍历性 (ergodicity)。显然，一个遍历的马尔可夫链只能有一个平稳分布。

将状态转移概率堆积成矩阵A，若存在正整数 n 使得转移概率矩阵 A^n 的所有元素均大于零（不等于0），即所有状态在n步内均可达，则称矩阵 A为**正则**的。

可以证明，**如果一个齐次马尔可夫链满足细致平衡条件，且转移矩阵是正则的，那么该马尔可夫链是遍历的。**

五、马尔可夫链蒙特卡罗方法 (MCMC) – Metropolis-Hastings 算法

目标分布为 $p(x) = \tilde{p}(x) / Z_p$。类似于拒绝采样和重要性采样，MCMC 从提议分布 $q(x)$ 中采样。但在MCMC中，我们需要记录当前状态 $x^{(\tau)}$，且提议分布 $q(x|x^{(\tau)})$ 依赖于 $x^{(\tau)}$ 。

如何生成$x^{(\tau+1)}$：

1. 从 $q(x|x^{(\tau)})$ 中生成候选样本 x'；
2. 根据接受概率 $A(x', x^{(\tau)})$ 决定是否接受 x'；
3. 若 x'被接受，则 $x^{(\tau+1)} = x'$；否则， $x^{(\tau+1)} = x^{(\tau)}$。

样本序列 $\{x^{(1)}, x^{(2)}, x^{(3)}, \ldots\}$ 构成了一条马尔可夫链。

五、马尔可夫链蒙特卡罗方法 (MCMC) － Metropolis-Hastings 算法

在标准的 Metropolis 算法中，$q(x)$ 是对称的，即 $q(x_A|x_B) = q(x_B|x_A)$。候选样本 x' 的接受概率为：

$$A(x', x^{(\tau)}) = \min\left(1, \frac{\tilde{p}(x')}{\tilde{p}(x^{(\tau)})}\right).$$

注意，当 $\tilde{p}(x') \geq \tilde{p}(x^{(\tau)})$ 时，x' 总是被接受。

当 $q(x)$ 不对称时，Metropolis-Hastings 算法的接受概率调整为：

$$A(x', x^{(\tau)}) = \min\left(1, \frac{\tilde{p}(x')q(x^{(\tau)}|x')}{\tilde{p}(x^{(\tau)})q(x'|x^{(\tau)})}\right).$$

对于对称的提议分布，Metropolis-Hastings 算法可简化为标准的 Metropolis 算法。

接受准则的计算不需要知道概率分布 $p(x)$ 的归一化常数 Z_p。

五、马尔可夫链蒙特卡罗方法 (MCMC) – Metropolis-Hastings 算法

只要对于任意 x_A和 x_B 都有 $q(x_A|x_B) > 0$，则 $p(x^{(\tau)})$ 将在 $\tau \to \infty$时收敛到 $p(x)$。

$p(x)$就是由 Metropolis-Hastings 算法构造的马尔可夫链的平稳分布，根据细致平稳条件，

$$\begin{aligned} p(x)q(x'|x)A(x',x) &= \min\left(p(x')q(x|x'), p(x)q(x'|x)\right) \\ &= p(x')q(x|x')A(x,x') \end{aligned} \tag{2}$$

因此，只要对于所有 x' 有 $q(x'|x) > 0$， $p(x)$就是平稳分布。

序列 $x^{(1)}, x^{(2)}, \dots$ 并非独立，因为样本之间具有高度相关性。为了获得独立样本，我们可以舍弃大部分样本，仅保留每隔 M 个样本中的一个。对于足够大的 M，保留的样本在渐近意义下是独立的。

五、马尔可夫链蒙特卡罗方法 (MCMC) – Metropolis-Hastings 算法

- Independence sampler：

$$q(x'|x) = q(x')$$

这类似于重要性采样。$q(x')$ 可以是任意合适的分布，例如高斯分布。由于高斯分布的密度在整个状态空间上都有非零，因此对于任何无约束的连续状态空间都是有效的提议分布。

- Random walk Metropolis (RWM) algorithm：

$$q(x'|x) = N(x'|x, \tau^2 I)$$

Roberts and Rosenthal (2001)证明，如果后验分布是高斯分布，则渐近最优 $\tau^2 = 2.38^2 / D$，D 是 x 的维度，接受率约为0.234，此时随机样本最大覆盖$p(x)$而不被频繁拒绝。

五、马尔可夫链蒙特卡罗方法 (MCMC) – 随机游走 Metropolis 例子

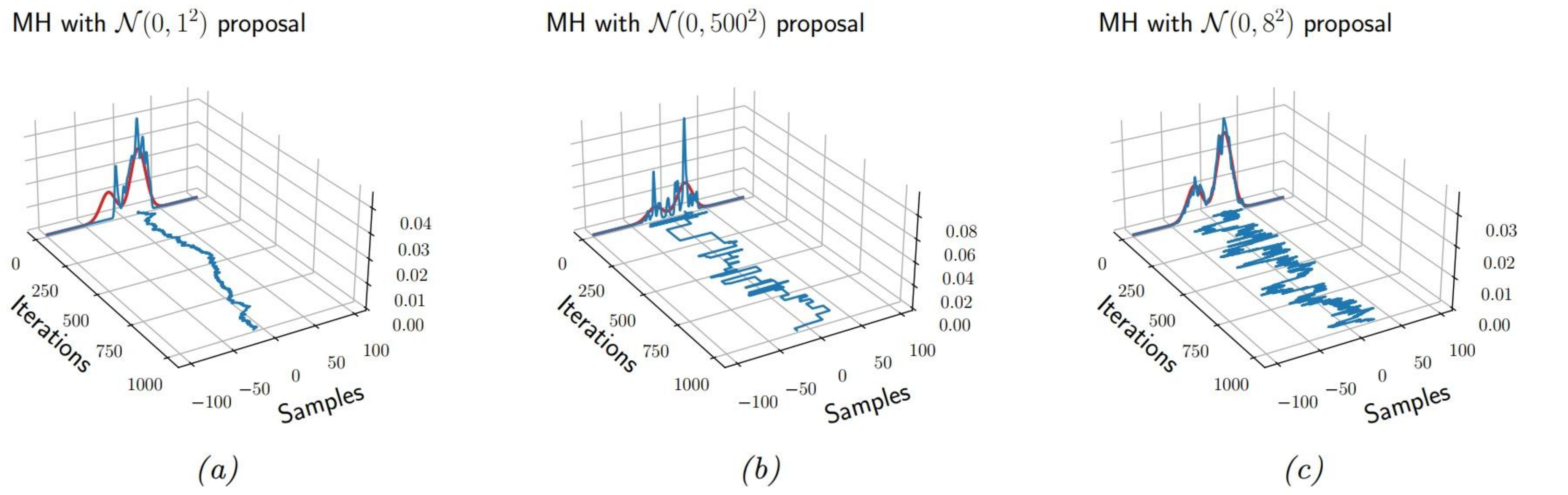

图12.2 RWM 算法

适当地设置方差 τ^2 至关重要。如果方差过小，马尔可夫链将只探索其中一个模态，如图 12.2(a) 所示；如果方差过大，大部分提议的抽样都被拒绝，链将变得非常“黏滞”，长时间停留在同一状态，如图 12.2(b) 所示。将方差设置得恰到好处，则可以对目标分布进行良好探索，如图 12.2 (c) 所示。

五、马尔可夫链蒙特卡罗方法 (MCMC) – Metropolis-Hastings 算法

- Composing proposals：如果存在多个提议分布，则可以将它们组合：

$$q(x'|x) = \sum_{k=1}^{K} w_k q_k(x'|x), \quad \sum_{k=1}^{K} w_k = 1$$

只要每个 $q_k(x'|x)$ 都有效且每个 $w_k > 0$，整体混合分布也是有效的。

提议分布也可以通过串联多个提议分布来组合：

$$q(x'|x) = \sum_{x_1} \cdots \sum_{x_{K-1}} q_1(x_1|x) q_2(x_2|x_1) \cdots q_K(x'|x_{K-1}).$$

一个常见的例子是，每个基础候选 $q_k(x'|x)$ 只更新变量的一个子集。

五、马尔可夫链蒙特卡罗方法 (MCMC) – Gibbs 采样

Metropolis-Hastings 方法的一个主要问题在于如何选择提议分布，而且其接受率可能较低。

Gibbs 采样是 MH 方法的一个特例，它利用条件独立性的性质构造出一个良好的提议分布，使得接受概率为 1。

Gibbs 采样的过程是在给定其他所有变量的条件下依次对每个变量进行采样。例如，如果 $D = 3$，按照如下顺序更新：

$$
\begin{gathered}
x_1^{s+1} \sim p(x_1 | x_2^s, x_3^s) \\
x_2^{s+1} \sim p(x_2 | x_1^{s+1}, x_3^s) \\
x_3^{s+1} \sim p(x_3 | x_1^{s+1}, x_2^{s+1})
\end{gathered}
$$

五、马尔可夫链蒙特卡罗方法 (MCMC) － Gibbs 采样

Gibbs 采样所对应的提议分布是：

$$q(x'|x) = p(x_i'|x_{-i})\mathbb{I}(x_{-i}' = x_{-i}),$$

新状态x'中的$x_i{}'$被采样，而所有其他变量$x_{-i}{}' = x_{-i}$保持不变。因此，新状态 x'的Metropolis-Hastings 的接受概率为：

$$\begin{aligned} A(x', x) &= \frac{p(x')q(x|x')}{p(x)q(x'|x)} = \frac{p(x_i'|x_{-i}')p(x_{-i}')p(x_i|x_{-i}')}{p(x_i|x_{-i})p(x_{-i})p(x_i'|x_{-i})} \\ &= \frac{p(x_i'|x_{-i})p(x_{-i})p(x_i|x_{-i})}{p(x_i|x_{-i})p(x_{-i})p(x_i'|x_{-i})} = 1 \end{aligned}$$

所以，Gibbs 采样每次抽样的接受率为 100 %。

然而，接受率为 100 % 并不必然意味着 Gibbs 采样能迅速收敛，因为它每次只更新一个坐标。将变量分组并同时采样可以显著加快收敛速度。

六、 Stochastic Volatility Model

考虑一个随机波动率模型：

观测过程:

$$y_t|s_t \sim \mathcal{N}(\mu, \exp(s_t)).$$

波动率动态:

$$s_t = \alpha + \phi s_{t-1} + \epsilon_t,$$

其中 $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ 为独立同分布的噪声项。$\theta = (\mu, \alpha, \phi, \sigma^2)$ 为未知参数。

我们使用 Gibbs 采样来模拟 $p(\theta, s_{1:T}|y_{1:T})$。

六、 Stochastic Volatility Model – 后验概率

1. 假设s_t先验为均匀分布，给定参数 $\theta=(\mu,\alpha,\phi,\sigma^2)$ 、 观测数据 $y_{1:T}$ 以及除 s_t 外其他状态 s_{-t}， s_t 的条件分布为：

$$p(s_t|\theta,y_{1:T},s_{-t})\propto p(y_t|s_t,\mu)\cdot p(s_t|s_{t-1},\alpha,\phi,\sigma^2)\cdot p(s_{t+1}|s_t,\alpha,\phi,\sigma^2).$$

$$p(y_t|s_t,\mu)=\frac{1}{\sqrt{2\pi\exp(s_t)}}\exp\left(-\frac{(y_t-\mu)^2}{2\exp(s_t)}\right),$$

$$p(s_t|s_{t-1},\alpha,\phi,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(s_t-(\alpha+\phi s_{t-1}))^2}{2\sigma^2}\right),$$

$$p(s_{t+1}|s_t,\alpha,\phi,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(s_{t+1}-(\alpha+\phi s_t))^2}{2\sigma^2}\right).$$

六、 Stochastic Volatility Model － 后验概率

2. 给定隐含状态 $s_{1:T}$和观测数据 $y_{1:T}$ ，参数 $\mu, \alpha, \phi, \sigma^2$的条件概率为：

对于 μ ，假设先验为均匀分布，则 μ的后验分布为：

$$p(\mu|s_{1:T}, y_{1:T}) \propto p(y_{1:T}|\mu, s_{1:T}) \propto \exp\left(-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t-\mu)^2}{\exp(s_t)}\right).$$

因此， μ的后验分布为：

$$\mu|s_{1:T}, y_{1:T} \sim \mathcal{N}\left(\frac{\sum_{t=1}^{T}\frac{y_t}{\exp(s_t)}}{\sum_{t=1}^{T}\frac{1}{\exp(s_t)}}, \frac{1}{\sum_{t=1}^{T}\frac{1}{\exp(s_t)}}\right).$$

六、 Stochastic Volatility Model － 后验概率

给定 $s_{1:T}$ ，我们有：

$$p(\alpha,\phi,\sigma^2|s_{1:T}) \propto p(s_{1:T}|\alpha,\phi,\sigma^2)$$

因此：

$$\alpha|s_{1:T},\sigma^2,\phi \sim \mathcal{N}\left(\frac{\sum_{t=2}^{T}(s_t-\phi s_{t-1})}{T-1},\frac{\sigma^2}{T-1}\right),$$

$$\phi|s_{1:T},\alpha,\sigma^2 \sim \mathcal{N}\left(\frac{\sum_{t=2}^{T}(s_t-\alpha)\,s_{t-1}}{\sum_{t=2}^{T}s_{t-1}^2},\frac{\sigma^2}{\sum_{t=2}^{T}s_{t-1}^2}\right),$$

$$\sigma^2|s_{1:T},\alpha,\phi \sim \text{Inv-Gamma}\left(\frac{T-3}{2},\frac{1}{2}\sum_{t=2}^{T}(s_t-\alpha-\phi s_{t-1})^2\right).$$

六、 Stochastic Volatility Model － MCMC抽样过程

1. **初始化**：为$\mu,\alpha,\phi,\sigma^2,s_{1:T}$设定初始值。
2. **更新隐藏状态** $s_{1:T}$：给定参数 μ,α,ϕ,σ^2，更新 $s_1,s_2,\ldots,s_T$。
3. **更新参数** μ,α,ϕ,σ^2：根据当前隐含状态 $s_{1:T}$更新参数。
4. **重复步骤 2 和 3**

六、 Stochastic Volatility Model

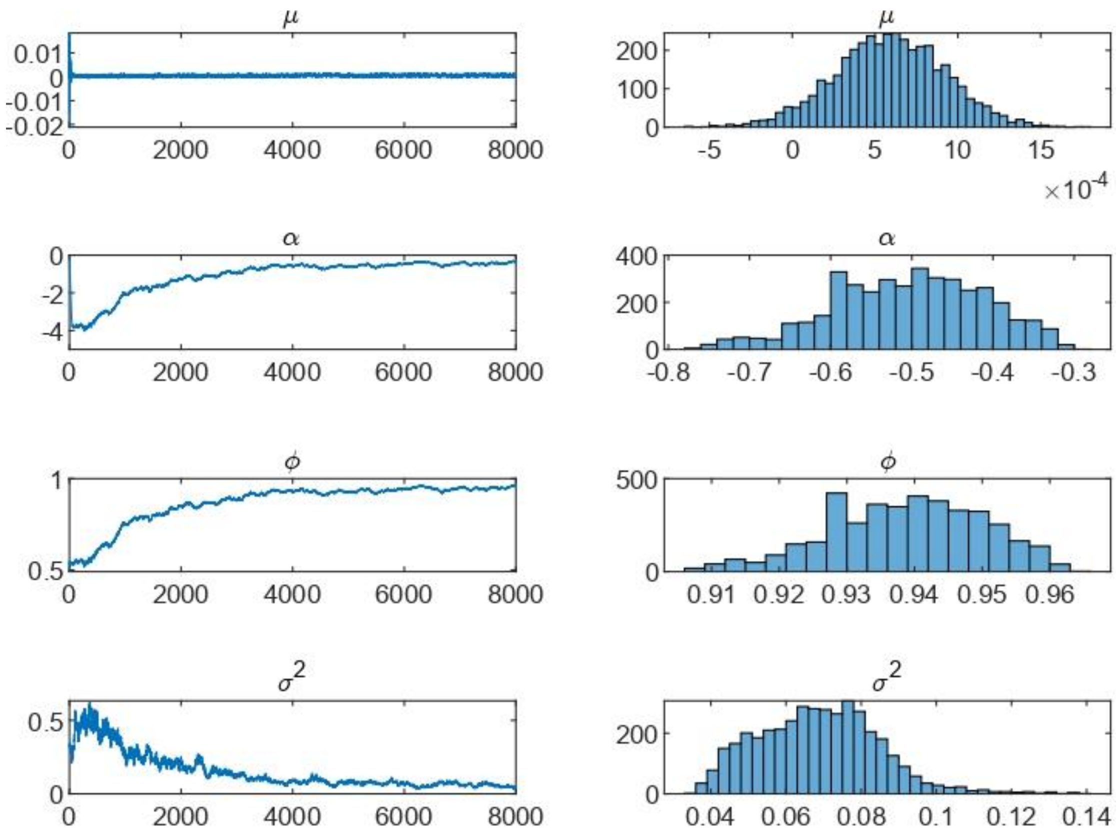

图12.3 抽样轨迹与直方图，一万次

在4000次抽样后，样本分布趋于平稳。我们可以利用后4000个样本拟合估计参数的分布，计算参数的均值和方差。

谢 谢!