贝叶斯统计

贝叶斯计算

陈垚翰

安徽大学 大数据与统计学院

后验分布的计算和抽样

- 后验分布可以一般表示为

$$\pi(\theta \mid \mathbf{x}) = \frac{\pi(\theta)p(\mathbf{x} \mid \theta)}{\int_{\Theta} \pi(\theta)p(\mathbf{x} \mid \theta)d\theta}$$

- 分母中的积分 $\int_{\Theta} \pi(\theta)p(\mathbf{x} \mid \theta)d\theta$ 难以计算
- 从 $\pi(\theta \mid \mathbf{x})$ 中抽样 $\left\{\theta^{(i)}\right\}_{i=1}^{m}$，并根据大数定律计算

$$\mathrm{E}\left[g(\theta) \mid \mathbf{x}\right] \approx \frac{1}{m} \sum_{i=1}^{m} g\left(\theta^{(i)}\right)$$

马氏链

- **马氏性:** 链 $\left\{\theta^{(0)}, \theta^{(1)}, \ldots\right\}$ 具有马氏性是指 $\theta^{(t+1)}$ 只依赖于 $\theta^{(t)}$，而不依赖于 $\left\{\theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(t-1)}\right\}$ $(t \geqslant 1)$。
- **马氏链的生成:** 给定初始值 $\theta^{(0)}$，通过条件分布

$$p\left(\theta \mid \theta^{(t)}\right), t=1,2,\cdots$$

产生新的随机数 $\theta^{(t+1)}$，其中 $p(\cdot \mid \cdot)$ 又称为转移核。

马氏链

有限状态离散形式的马氏链（Discrete Markov Chain with finite state space）：

- 状态空间有 K 个状态，记为 $1,\ldots,K$。转移核构成如下转移概率矩阵

$$\mathbf{P}^{[t-1,t]}=\begin{bmatrix} p_{1,1}^{[t-1,t]} & \cdots & p_{1,K}^{[t-1,t]} \\ \vdots & \cdots & \vdots \\ p_{K,1}^{[t-1,t]} & \cdots & p_{K,K}^{[t-1,t]} \end{bmatrix} \text{简记为 } \mathbf{P}=\begin{bmatrix} p_{1,1} & \cdots & p_{1,K} \\ \vdots & \vdots & \vdots \\ p_{K,1} & \cdots & p_{K,K} \end{bmatrix}$$

  其中，$p_{i,j}^{[t-1,t]}=\mathrm{P}\left(\theta^{(t)}=j \mid \theta^{(t-1)}=i\right)$。

- $\sum_j p_{i,j}^{[t-1,t]}=1.$

马氏链

令

$$\mathbf{P}^{[2]} := \begin{bmatrix} p_{1,1}^{[2]} & \cdots & p_{1,K}^{[2]} \\ \vdots & \vdots & \vdots \\ p_{K,1}^{[2]} & \cdots & p_{K,K}^{[2]} \end{bmatrix}$$

其中, $p_{i,j}^{[2]} = \mathrm{P}\left(\theta^{(t+2)} = j \mid \theta^{(t)} = i\right)$ 状态 i 经 2 步转移至 j 的概率。

- 注意到

$$\left(\theta^{(2)} = j \mid \theta^{(0)} = i\right) = \bigcup_k \left(\theta^{(2)} = j, \theta^{(1)} = k \mid \theta^{(0)} = i\right).$$

马氏链

由条件概率公式和马氏性可知

$$
\begin{aligned}
P\left(\theta^{(2)}=j \mid \theta^{(0)}=i\right) &= \sum_{k=1}^{K} P\left(\theta^{(2)}=j, \theta^{(1)}=k \mid \theta^{(0)}=i\right) \\
&= \sum_{k=1}^{K} P\left(\theta^{(2)}=j \mid \theta^{(1)}=k, \theta^{(0)}=i\right) \\
&\quad \times P\left(\theta^{(0)}=i\right) \\
&= \sum_{k=1}^{K} P\left(\theta^{(2)}=j \mid \theta^{(1)}=k\right) \\
&\quad \times P\left(\theta^{(0)}=i\right).
\end{aligned}
$$

所以 $p_{i,j}^{[2]}=\sum_{k=1}^{K} p_{i,k} \times p_{k,j}$，即 $\mathbf{P}^{[2]}=\mathbf{P} \times \mathbf{P}=\mathbf{P}^{2}$。

马氏链

可以推广至一般情形：

$$\left(\theta^{(m+n)} = j \mid \theta^{(0)} = i\right) = \bigcup_k \left(\theta^{(m+n)} = j, \theta^{(n)} = k \mid \theta^{(0)} = i\right)$$

$$\mathrm{P}\left(\theta^{(m+n)} = j \mid \theta^{(0)} = i\right) = \sum_{k=1}^{K} \mathrm{P}\left(\theta^{(m+n)} = j, \theta^{(n)} = k \mid \theta^{(0)} = i\right).$$

$$\begin{aligned}\mathrm{P}\left(\theta^{(m+n)} = j \mid \theta^{(0)} = i\right) = \sum_k \Big\{ & \mathrm{P}\left(\theta^{(m+n)} = j \mid \theta^{(n)} = k\right) \\ & \times \mathrm{P}\left(\theta^{(n)} = k \mid \theta^{(0)} = i\right)\Big\}.\end{aligned}$$

马氏链

$$p_{i,j}^{[m+n]} = \sum_{k=1}^{K} p_{i,k}^{[m]} \times p_{k,j}^{[n]}$$

$$\mathbf{P}^{[m+n]} = \mathbf{P}^{[m]} \times \mathbf{P}^{[n]}$$

$$\mathbf{P}^{[n]} = \mathbf{P}^{n}$$

如果以下极限存在,则存在一种稳定的状态

$$\begin{aligned}\lim_{t\to\infty} \mathbf{I} \times \mathbf{P}^{t} &= \lim_{t\to\infty} \mathbf{P}^{t} \\ &= \begin{bmatrix} \pi_1 & \dots & \pi_K \\ \vdots & \vdots & \vdots \\ \pi_1 & \dots & \pi_K \end{bmatrix}.\end{aligned}$$

马氏链

如果长期稳定的状态存在

$$\mathbf{P}^{[t+1]} = \mathbf{P}^{[t]} \times \mathbf{P}$$

则同时在等式两边取极限有

$$\begin{bmatrix} \pi_1 & \dots & \pi_K \\ \vdots & \vdots & \vdots \\ \pi_1 & \dots & \pi_K \end{bmatrix} = \begin{bmatrix} \pi_1 & \dots & \pi_K \\ \vdots & \vdots & \vdots \\ \pi_1 & \dots & \pi_K \end{bmatrix} \times \mathbf{P}.$$

每一行可以表示为

$$\boldsymbol{\pi} = \boldsymbol{\pi} \times \mathbf{P}$$

也称之为**稳态方程**（steady state equation）。

马氏链

- **平稳性（收敛性）**：链的平稳性是指，当 $t \to \infty$ 时，马氏链 $\{\theta^0, \theta^{(1)}, \theta^{(2)}, \cdots,\}$ 依分布收敛到某个概率分布，此分布称为**平稳分布**（stationary distribution）。
- **正常返性**: 状态 i 是正常返的，如果

$$M_i = \mathrm{E}\left[T_i\right] < \infty,$$

其中 $T_i = \inf\left\{n \geqslant 1 : \theta^{(n)} = i \mid \theta^{(0)} = i\right\}$。

马氏链

- **非周期性**：令

$$k = \gcd\left\{n : \mathrm{P}\left(\theta^{(n)} = i \mid \theta^{(0)} = i\right) > 0\right\},$$

  其中，gcd 表示最大公约数。如果 $\forall\, i$ 都有 $k = 1$，则称马氏链是非周期的（aperiodic）。
- **遍历性**：如果一个马氏链的状态 i 是非周期的且正常返的，则称 i 是遍历的，如果马氏链的所有状态都是遍历的，则称马氏链是遍历的。
- **不可约性**：遍历的马氏链称为不可约的（irreducible）。

马氏链

定理

设 $\{\theta^{(0)}, g \geqslant 0\}$ 为不可约的非周期的马氏链，π 为其平稳分布，π_0 为初始分布，则

$$\pi_t \to \pi, t \to \infty$$

其中 π_t 为马氏链在时刻 t 的分布。

例子:不可约马氏链的平稳分布

$$\mathbf{P}=\begin{bmatrix} 0.35 & 0.35 & 0.10 & 0.10 & 0.10 \\ 0.15 & 0.55 & 0.10 & 0.10 & 0.10 \\ 0.15 & 0.15 & 0.10 & 0.20 & 0.40 \\ 0.15 & 0.15 & 0.40 & 0.10 & 0.20 \\ 0.15 & 0.15 & 0.20 & 0.40 & 0.10 \end{bmatrix}$$

$$\mathbf{P}^{[2^5]}=\begin{bmatrix} 0.187500 & 0.312500 & 0.166667 & 0.166667 & 0.166667 \\ 0.187500 & 0.312500 & 0.166667 & 0.166667 & 0.166667 \\ 0.187500 & 0.312500 & 0.166667 & 0.166667 & 0.166667 \\ 0.187500 & 0.312500 & 0.166667 & 0.166667 & 0.166667 \\ 0.187500 & 0.312500 & 0.166667 & 0.166667 & 0.166667 \end{bmatrix}$$

定理（马氏链大数定律）

设 $\left\{\theta^{(1)},\theta^{(2)},\cdots,\theta^{(n)},\cdots\right\}$ 为遍历的平稳分布为 π 的马氏链，则 $\theta^{(n)}$ 依分布收敛到 π 的随机变量 θ，且对任意连续函数 g，当 $\mathrm{E}_\pi\left[g(\theta)\right]$ 存在时，有

$$\frac{1}{n}\sum_{i=1}^{n} g\left(\theta^{(i)}\right) \to \mathrm{E}_\pi\left[g(\theta)\right], \quad n \to \infty.$$

马氏链

不可约的马氏链有平稳分布;而对于平稳分布,是否存在相应的不可约马氏链?或者相应的转移核?

马氏链

不可约的马氏链有平稳分布;而对于平稳分布,是否存在相应的不可约马氏链?或者相应的转移核?

$$
\begin{aligned}
q_{i,j} &= P\left(\theta^{(t)} = j \mid \theta^{(t+1)} = i\right) \\
&= \frac{P\left(\theta^{(t)} = j, \theta^{(t+1)} = i\right)}{P\left(\theta^{(t+1)} = i\right)} \\
&= \frac{P\left(\theta^{(t)} = j\right) P\left(\theta^{(t+1)} = i \mid \theta^{(t)} = j\right)}{P\left(\theta^{(t+1)} = i\right)}
\end{aligned}
$$

马氏链

不可约的马氏链有平稳分布;而对于平稳分布,是否存在相应的不可约马氏链?或者相应的转移核?

$$
\begin{aligned}
q_{i,j} &= P\left(\theta^{(t)} = j \mid \theta^{(t+1)} = i\right) \\
&= \frac{P\left(\theta^{(t)} = j, \theta^{(t+1)} = i\right)}{P\left(\theta^{(t+1)} = i\right)} \\
&= \frac{P\left(\theta^{(t)} = j\right) P\left(\theta^{(t+1)} = i \mid \theta^{(t)} = j\right)}{P\left(\theta^{(t+1)} = i\right)}
\end{aligned}
$$

在稳态时,

$$
q_{i,j} = \frac{\pi_j p_{j,i}}{\pi_i}
$$

马氏链

- 当 $\forall\, i,j$ 都有 $p_{i,j} = q_{i,j}$, 则称相应的马氏链是时序可逆的（time reversible）,且

$$\pi_i p_{i,j} = \pi_j p_{j,i}.$$

上述等式也称之为 “detailed balance” 条件。

定理

满足 $\pi_i p_{i,j} = \pi_j p_{j,i}$ 的转移核 $p_{i,j}$ 有平稳分布 $\boldsymbol{\pi}$。

$$
\begin{aligned}
\sum_i \pi_i p_{i,j} &= \sum_i \pi_j p_{j,i} \\
&= \pi_j \sum_i p_{j,i} \\
&= \pi_j \sum_j p_{i,j} \\
&= \pi_j
\end{aligned}
$$

Gibbs 抽样

- 从后验分布 $\pi(\boldsymbol{\theta} \mid \mathbf{x})$ 中对 $\boldsymbol{\theta}$ 分块抽样，$\boldsymbol{\theta} = \left(\theta_1, \theta_2, \ldots, \theta_p\right)$。

$$\pi\left(\theta_j \mid \theta_{-j}, \mathbf{x}\right) = \frac{\pi\left(\theta_1, \cdots, \theta_{j-1}, \theta_j, \theta_{j+1}, \cdots, \theta_p \mid \mathbf{x}\right)}{\int \pi\left(\theta_1, \cdots, \theta_{j-1}, \theta_j, \theta_{j+1}, \cdots, \theta_p \mid \mathbf{x}\right) d\theta_j}$$

  称之为 full conditional distribution，
  $\theta_{-j} = \left(\theta_1, \cdots, \theta_{j-1}, \theta_{j+1}, \cdots, \theta_p\right)$

- Gibbs 抽样：根据 full conditional distribution 构造马氏链。

Gibbs 抽样

1. 给定初始值 $\theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_p^{(0)}$
2. 对 $t = 0, 1, 2, \cdots$
   a) 从条件分布 $\pi\left(\theta_1 \mid \theta_2^{(t)}, \cdots, \theta_p^{(t)}, \mathbf{x}\right)$ 中产生 $\theta_1^{(t+1)}$
   b) 从条件分布 $\pi\left(\theta_2 \mid \theta_1^{(t+1)}, \theta_3^{(t)}, \cdots, \theta_p^{(t)}, \mathbf{x}\right)$ 产生 $\theta_2^{(t+1)}$
   ……
   c) 从条件分布 $\pi\left(\theta_p \mid \theta_1^{(t+1)}, \theta_2^{(t+1)}, \cdots, \theta_p^{(t+1)}, \mathbf{x}\right)$ 产生 $\theta_p^{(t+1)}$

重复 2. 中的 *a*), *b*) 和 *c*) 生成马氏链 $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(t)}, \cdots$

由马氏链大数定律

$$\frac{1}{n}\sum_{t=1}^{n} g\left(\boldsymbol{\theta}^{(t)}\right) \to \int g(\boldsymbol{\theta})\, \pi(\boldsymbol{\theta} \mid \mathbf{x})\, d\boldsymbol{\theta}$$

Gibbs 抽样

一个例子：

目标是从 $p(\theta_1, \theta_2) \propto e^{-\theta_1\theta_2}\mathbf{1}\{\theta_1, \theta_2 \in (0, c)\}$ 中抽样

- $p(\theta_1, \theta_2)$ 具有对称结构，且

$$p(\theta_1 \mid \theta_2) \propto p(\theta_1, \theta_2) \propto e^{-\theta_1\theta_2}\mathbf{1}(0 < \theta_1 < c) \propto \mathrm{Exp}(\theta_1 \mid \theta_2)\mathbf{1}(\theta_1 < c)$$

即 $p(\theta_1 \mid \theta_2)$ 在 $(0, c)$ 上是截尾的指数分布，记作 $\mathrm{TExp}(\theta_2, (0, c))$。
根据对称性，$p(\theta_2 \mid \theta_1)$ 也是在 $(0, c)$ 上是截尾的指数分布，
记作 $\mathrm{TExp}(\theta_1, (0, c))$。

Gibbs 抽样

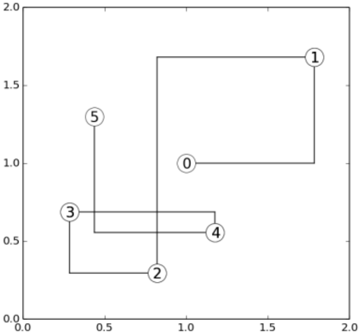
在上述例子中，Gibbs 抽样步骤如下：

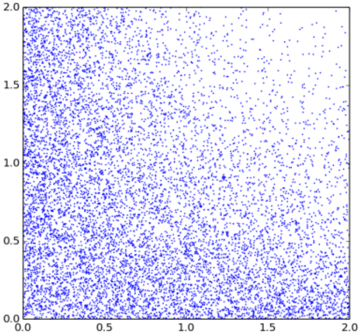
0. 初始化 $\theta_1^{(0)}, \theta_2^{(0)} \in (0, c)$
1. 从 $\mathrm{TExp}\left(\theta_2^{(0)}, (0, c)\right)$ 中抽 $\theta_1^{(1)}$, 然后从 $\mathrm{TExp}\left(\theta_1^{(1)}, (0, c)\right)$ 中抽 $\theta_2^{(1)}$
2. 从 $\mathrm{TExp}\left(\theta_2^{(1)}, (0, c)\right)$ 中抽 $\theta_1^{(2)}$, 然后从 $\mathrm{TExp}\left(\theta_1^{(2)}, (0, c)\right)$ 中抽 $\theta_2^{(2)}$

   $\cdots$

n. 从 $\mathrm{TExp}\left(\theta_2^{(n-1)}, (0, c)\right)$ 中抽 $\theta_1^{(n)}$, 然后从 $\mathrm{TExp}\left(\theta_1^{(n)}, (0, c)\right)$ 中抽 $\theta_2^{(n)}$.

Gibbs 抽样





Metropolis-Hastings 算法

当在 Gibbs 抽样中，不易从条件分布中抽样时怎么办？

Metropolis-Hastings 算法

当在 Gibbs 抽样中,不易从条件分布中抽样时怎么办?

- 选择合适的转移核构造马氏链 $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(t)}, \cdots$,使得该马氏链有相应的平稳分布 $\pi(\boldsymbol{\theta})$ 或者 $\pi(\boldsymbol{\theta} \mid \mathbf{x})$。
- 转移核用以由 $\boldsymbol{\theta}^{(t)}$ 生成 $\boldsymbol{\theta}^{(t+1)}$。
- 稳态方程 $\boldsymbol{\pi} = \boldsymbol{\pi} \times \mathbf{P}$ 在连续状态空间 Θ 的形式

$$\int_A \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int \pi(\boldsymbol{\theta}) \, \mathrm{P}(\boldsymbol{\theta}, A) d\boldsymbol{\theta}$$

  其中 $\boldsymbol{\theta} \in \Theta$ 且 $A \subset \Theta$ 是 Θ 的可测子集,$\mathrm{P}(\boldsymbol{\theta}, A)$ 是转移核。
- 记 $q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ 为由 $\boldsymbol{\theta}$ 生成 $\boldsymbol{\theta}'$ 的建议分布(proposal distribution)。

Metropolis-Hastings 算法

定理

如果 proposal distribution $q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ 满足 “detailed balance” 条件，

$$\pi(\boldsymbol{\theta}) \times q(\boldsymbol{\theta}, \boldsymbol{\theta}') = \pi(\boldsymbol{\theta}') \times q(\boldsymbol{\theta}', \boldsymbol{\theta}) \text{ 对任意 } \boldsymbol{\theta}, \boldsymbol{\theta}'$$

则

$$\mathrm{P}(\boldsymbol{\theta}, A) = \int_{\{A \cap \{\boldsymbol{\theta}' \neq \boldsymbol{\theta}\}\}} q(\boldsymbol{\theta}, \boldsymbol{\theta}')\, d\boldsymbol{\theta}' + r(\boldsymbol{\theta}) \boldsymbol{\delta}_A(\boldsymbol{\theta})$$

且 $\pi(\boldsymbol{\theta})$ 是以 $\mathrm{P}(\boldsymbol{\theta}, A)$ 为转移核的马氏链的平稳分布。

其中, $r(\boldsymbol{\theta}) = 1 - \int_{\{\boldsymbol{\theta}' \neq \boldsymbol{\theta}\}} q(\boldsymbol{\theta}, \boldsymbol{\theta}') d\boldsymbol{\theta}'$ 是停留在 $\boldsymbol{\theta}$ 的概率

$$\delta_A(\boldsymbol{\theta}) = \begin{cases} 1 \text{ if } \ \boldsymbol{\theta} \in A \\ 0 \text{ if } \ \boldsymbol{\theta} \notin A \end{cases}$$

证明

$$
\begin{aligned}
&\int \pi(\boldsymbol{\theta}) \mathrm{P}(\boldsymbol{\theta}, A) d\boldsymbol{\theta} \\
= &\iint_{\{A \cap \{\boldsymbol{\theta}' \neq \boldsymbol{\theta}\}\}} \pi(\boldsymbol{\theta}) q(\boldsymbol{\theta}, \boldsymbol{\theta}') d\boldsymbol{\theta}' d\boldsymbol{\theta} + \int \pi(\boldsymbol{\theta}) r(\boldsymbol{\theta}) \delta_A(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
= &\int_A \int_{\{\boldsymbol{\theta} \neq \boldsymbol{\theta}'\}} \pi(\boldsymbol{\theta}) q(\boldsymbol{\theta}, \boldsymbol{\theta}') d\boldsymbol{\theta} d\boldsymbol{\theta}' + \int_A \pi(\boldsymbol{\theta}) r(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
= &\int_A \int_{\{\boldsymbol{\theta} \neq \boldsymbol{\theta}'\}} \pi(\boldsymbol{\theta}') q(\boldsymbol{\theta}', \boldsymbol{\theta}) d\boldsymbol{\theta} d\boldsymbol{\theta}' + \int_A \pi(\boldsymbol{\theta}) r(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
= &\int_A \pi(\boldsymbol{\theta}') (1 - r(\boldsymbol{\theta}')) d\boldsymbol{\theta}' + \int_A \pi(\boldsymbol{\theta}) r(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
= &\int_A \pi(\boldsymbol{\theta}') d\boldsymbol{\theta}' = \int_A \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}
\end{aligned}
$$

Metropolis-Hastings 算法

- “detailed balance” 条件，

$$\pi(\boldsymbol{\theta}) \times q(\boldsymbol{\theta}, \boldsymbol{\theta}') = \pi(\boldsymbol{\theta}') \times q(\boldsymbol{\theta}', \boldsymbol{\theta}) \text{ 对任意 } \boldsymbol{\theta}, \boldsymbol{\theta}'$$

  是 Metropolis-Hastings 算法的关键。
- 但不是所有的 proposal distribution $q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ 都满足 “detailed balance” 条件，如何找到 proposal distribution 满足 “detailed balance” 条件。
- 令

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left[1, \frac{\pi(\boldsymbol{\theta}')\, q(\boldsymbol{\theta}', \boldsymbol{\theta})}{\pi(\boldsymbol{\theta}) q(\boldsymbol{\theta}, \boldsymbol{\theta}')}\right]$$

  则 $\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') \times q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ 一定满足 “detailed balance” 条件。

证明

$$
\begin{aligned}
\pi(\boldsymbol{\theta}) \times \alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') \times q(\boldsymbol{\theta}, \boldsymbol{\theta}') &= \pi(\boldsymbol{\theta}) \times \min\left[1, \frac{\pi(\boldsymbol{\theta}')\, q(\boldsymbol{\theta}', \boldsymbol{\theta})}{\pi(\boldsymbol{\theta}) q(\boldsymbol{\theta}, \boldsymbol{\theta}')}\right] \times q(\boldsymbol{\theta}, \boldsymbol{\theta}') \\
&= \min\left[\pi(\boldsymbol{\theta})\, q(\boldsymbol{\theta}, \boldsymbol{\theta}'), \pi(\boldsymbol{\theta}')\, q(\boldsymbol{\theta}', \boldsymbol{\theta})\right] \\
&= \pi(\boldsymbol{\theta}') \times \min\left[1, \frac{\pi(\boldsymbol{\theta})\, q(\boldsymbol{\theta}, \boldsymbol{\theta}')}{\pi(\boldsymbol{\theta}') q(\boldsymbol{\theta}', \boldsymbol{\theta})}\right] \times q(\boldsymbol{\theta}', \boldsymbol{\theta}) \\
&= \pi(\boldsymbol{\theta}') \times \alpha(\boldsymbol{\theta}', \boldsymbol{\theta}) \times q(\boldsymbol{\theta}', \boldsymbol{\theta})
\end{aligned}
$$

Metropolis-Hastings 算法

当状态空间有 K 个状态，记为 $1,\ldots,K$，

- 马氏链有转移核 $p_{i,j}$
- 平稳分布为 $\boldsymbol{\pi}=(\pi_1,\ldots,\pi_K)$
- 对每个状态 i 和 状态 j，定义

$$\alpha_{i,j}=\min\left[1,\frac{\pi_j p_{j,i}}{\pi_i p_{i,j}}\right].$$

- 对每个状态 i 和 $j\neq i$，定义 $p'_{i,j}=\alpha_{i,j}p_{i,j}$
- 定义

$$p'_{i,i}=1-\sum_{j\neq i}p'_{i,j}$$

则 $\boldsymbol{\pi}$ 是以 $p'_{i,j}$ 转移核的平稳分布。

证明

注意到

$$\begin{aligned}\pi_i p'_{i,j} &= \pi_i \alpha_{i,j} p_{i,j} \\ &= \pi_i \times \min\left[1, \frac{\pi_j p_{j,i}}{\pi_i p_{i,j}}\right] \times p_{i,j} \\ &= \min\left[\pi_j p_{j,i}, \pi_i p_{i,j}\right]\end{aligned}$$

同时

$$\begin{aligned}\pi_j p'_{j,i} &= \pi_j \alpha_{j,i} p_{j,i} \\ &= \pi_j \times \min\left[1, \frac{\pi_i p_{i,j}}{\pi_j p_{j,i}}\right] \times p_{j,i} \\ &= \min\left[\pi_j p_{j,i}, \pi_i p_{i,j}\right]\end{aligned}$$

所以 $p'_{i,j}$ 满足 “detailed balance” 条件，即 $\boldsymbol{\pi}$ 是以 $p'_{i,j}$ 为转移核的平稳分布。

Metropolis-Hastings 算法

Metropolis-Hastings 算法的主要步骤如下：

1. 初始化 $\boldsymbol{\theta}^{(0)}$
2. 对于 $t = 1, 2, \cdots$
   a) 由 $q\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\theta}'\right)$ 产生 $\boldsymbol{\theta}'$
   b) 计算接受概率 $\alpha\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\theta}'\right)$
   c) 从均匀分布 $U(0, 1)$ 中生成 u
   d) 如果 $u < \alpha\left(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\theta}'\right)$ 则令 $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}'$，
   否则令 $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$

重复 2. 中的 $a), b), c)$ 和 $d)$ 生成马氏链 $\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \cdots, \boldsymbol{\theta}^{(t)}, \cdots$

Metropolis-Hastings 算法

如果将 $\pi(\boldsymbol{\theta} \mid \mathbf{x})$ 作为平稳分布,则

$$\alpha\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\prime}\right)=\min \left[1, \frac{\pi\left(\boldsymbol{\theta}^{\prime} \mid \mathbf{x}\right) q\left(\boldsymbol{\theta}^{\prime}, \boldsymbol{\theta}\right)}{\pi(\boldsymbol{\theta} \mid \mathbf{x}) q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\prime}\right)}\right]$$

因此无需计算 $\int_{\Theta} \pi(\boldsymbol{\theta}) p(\mathbf{x} \mid \boldsymbol{\theta}) d \boldsymbol{\theta}$,其中 $\pi(\boldsymbol{\theta})$ 表示先验分布。

Metropolis-Hastings 算法：一个例子

考虑如下平稳分布

$$\pi(\theta) = \frac{\exp\left(-\theta^2\right)(2+\sin(5\theta)+\sin(2\theta))}{\int_{-\infty}^{\infty}\exp\left(-\theta^2\right)(2+\sin(5\theta)+\sin(2\theta))d\theta}$$

利用 Metropolis-Hastings 算法，可以在不用计算

$$\int_{-\infty}^{\infty}\exp\left(-\theta^2\right)(2+\sin(5\theta)+\sin(2\theta))d\theta$$

的条件下从 $\pi(\theta)$ 中抽样。

Metropolis-Hastings 算法：一个例子

- 选用正态分布作为 proposal distribution

$$\theta' \mid \theta \sim \text{Normal}\left(\theta, \sigma^2\right)$$

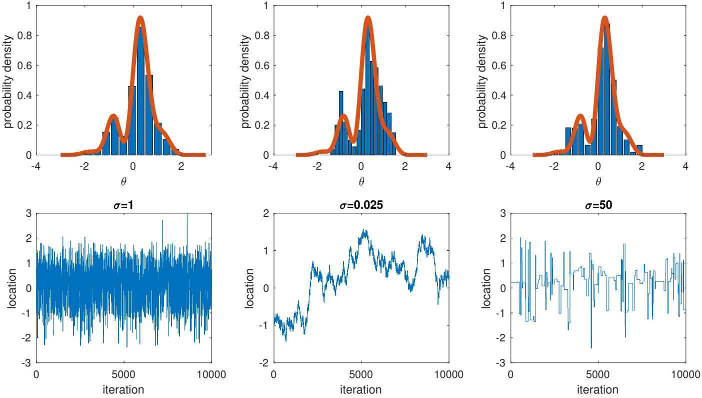
- 计算接受概率

$$\begin{aligned}
\alpha(\theta, \theta') &= \min\left[1, \frac{\pi(\theta')}{\pi(\theta)} \times \frac{q(\theta', \theta)}{q(\theta, \theta')}\right] \\
&= \min\left[1, \frac{\pi(\theta')}{\pi(\theta)}\right]
\end{aligned}$$

其中，

$$\frac{\pi(\theta')}{\pi(\theta)} = \frac{\exp\left(-\theta'^2\right)(2 + \sin(5\theta') + \sin(2\theta'))}{\exp\left(-\theta^2\right)(2 + \sin(5\theta) + \sin(2\theta))}$$

Metropolis-Hastings 算法：一个例子



Proposal Distribution 的选取

- 随机游走形式的 proposal distribution，$q(\boldsymbol{\theta}, \boldsymbol{\theta}') = q_1(\boldsymbol{\theta}' - \boldsymbol{\theta})$，其中 $q_1(\cdot)$ 是关于 0 的对称函数。

$$
\begin{aligned}
\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') &= \min\left[1, \frac{\pi(\boldsymbol{\theta}')\, q(\boldsymbol{\theta}', \boldsymbol{\theta})}{\pi(\boldsymbol{\theta}) q(\boldsymbol{\theta}, \boldsymbol{\theta}')}\right] \\
&= \min\left[1, \frac{\pi(\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})}\right]
\end{aligned}
$$

Proposal Distribution 的选取

- 独立 proposal distribution，$q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\prime}\right)=q_{2}\left(\boldsymbol{\theta}^{\prime}\right)$，其中 $q_{2}(\boldsymbol{\theta}^{\prime})$ 与当前状态 $\boldsymbol{\theta}$ 无关。

$$
\begin{aligned}
\alpha\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\prime}\right) & =\min \left[1, \frac{\pi\left(\boldsymbol{\theta}^{\prime}\right) q\left(\boldsymbol{\theta}^{\prime}, \boldsymbol{\theta}\right)}{\pi(\boldsymbol{\theta}) q\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{\prime}\right)}\right] \\
& =\min \left[1, \frac{\pi\left(\boldsymbol{\theta}^{\prime}\right)}{\pi(\boldsymbol{\theta})} \times \frac{q_{2}(\boldsymbol{\theta})}{q_{2}\left(\boldsymbol{\theta}^{\prime}\right)}\right]
\end{aligned}
$$

贝叶斯计算的应用：Real Business Cycle 模型的估计

- Real Business Cycle（RBC）模型由 Kydland 和 Prescott (1982) 提出，旨在新古典增长模型（neoclassical growth model）中对不确定性建模，RBC 模型是后续的动态随机一般均衡（ Dynamic Stochastic General Equilibrium，DSGE ）的基础。
- 以 King 和 Rebelo (1999) 中的 RBC 模型为例。
- Social Planner（政府）考虑如下效用最大化问题

$$\max_{\{C_t,N_t,I_t\}} \mathrm{E}\left\{\sum_{t=0}^{\infty}\beta^t\left[\ln C_t+\chi\ln(1-N_t)\right]\right\},$$

满足约束条件

$$C_t + I_t = z_t K_t^{\alpha} N_t^{1-\alpha} = Y_t$$

$$K_{t+1} = (1-\delta)K_t + I_t$$

$$\ln z_t = \rho \ln z_{t-1} + \sigma\varepsilon_t, \quad \varepsilon_t \sim \text{IID}\, N(0,1),$$

其中

- C_t 表示社会总消费需求（aggregate consumption）；
- N_t 表示社会总劳动需求（aggregate labor demand）；
- K_t 表示社会资本总量（aggregate capital）；
- I_t 表示社会投资总量（aggregate invetment）；
- z_t 表示技术进步对全要素生产率的冲击（total factor productivity (TFP) shocks），服从 AR(1) 过程；
- δ 表示资本折旧率。

注意到在约束条件下

$$\max_{\{C_t,N_t,I_t\}} \mathrm{E}\left\{\sum_{t=0}^{\infty}\beta^t\left[\ln C_t+\chi\ln\left(1-N_t\right)\right]\right\}$$

$$\Leftrightarrow$$

$$\max_{\{K_{t+1},N_t\}} \mathrm{E}\left\{\sum_{t=0}^{\infty}\beta^t\left[\ln\left(z_tK_t^{\alpha}N_t^{1-\alpha}+(1-\delta)K_t-K_{t+1}\right)+\chi\ln\left(1-N_t\right)\right]\right\}$$

关于 K_{t+1} 的一阶偏导数

$$-\frac{1}{C_t}+\mathrm{E}_t\left\{\frac{\beta}{C_{t+1}}\left[\alpha z_{t+1}K_{t+1}^{\alpha-1}N_{t+1}^{1-\alpha}+1-\delta\right]\right\}$$

关于 N_t 的一阶偏导数

$$\frac{(1-\alpha)z_tK_t^{\alpha}N_t^{-\alpha}}{C_t}-\frac{\chi}{1-N_t}$$

令关于 K_{t+1} 的一阶偏导数和关于 N_t 的一阶偏导数分别等于 0,则有最优一阶条件（欧拉方程）:

$$\frac{1}{C_t} = \mathrm{E}_t\left\{\frac{\beta}{C_{t+1}}\left[\alpha z_{t+1}K_{t+1}^{\alpha-1}N_{t+1}^{1-\alpha}+1-\delta\right]\right\} \tag{1}$$

$$\frac{\chi C_t}{1-N_t} = (1-\alpha)z_tK_t^{\alpha}N_t^{-\alpha} \tag{2}$$

假设全社会拥有资本,则实际利率和劳动工资分别为

$$R_{kt} = \alpha z_tK_t^{\alpha-1}N_t^{1-\alpha} \tag{3}$$

和

$$w_t = (1-\alpha)z_tK_t^{\alpha}N_t^{-\alpha} \tag{4}$$

贝叶斯计算的应用：Real Business Cycle 模型的估计

- $\boldsymbol{\theta} = (\alpha, \beta, \delta, \chi, \rho, \sigma)$
- Steady State（稳态）：所有变量没有时序上的动态特征，$X_t = X$。
- 在稳态下令 $z_t = 1$，则欧拉方程 (1) 为

$$1 = \beta \left[\alpha K^{\alpha-1} N^{1-\alpha} + 1 - \delta \right]. \tag{5}$$

并由 (5) 式可以解出

$$\frac{K}{N} = \left(\frac{\alpha}{1/\beta - (1-\delta)} \right)^{\frac{1}{1-\alpha}}.$$

因此，在稳态下

$$R_{kt} = \alpha \frac{Y}{K} = \alpha \left(\frac{K}{N}\right)^{\alpha-1}, w_t = (1-\alpha)\left(\frac{K}{N}\right)^{\alpha}.$$

由 $K_{t+1} = (1-\delta)K_t + I_t$ 可知，在稳态下

$$I = \delta K = \frac{\delta K}{N} N,$$

且

$$Y = \left(\frac{K}{N}\right)^{\alpha} N.$$

有欧拉方程 (2) 可以解得在稳态下

$$\chi = (1-\alpha)\left(\frac{K}{N}\right)^{\alpha} \frac{1-N}{C}.$$

做对数线性化，令 $\hat{X}_t = \ln X_t - \ln X$。在 (2) 两边同时取对数并在稳态取全微分

$$(\ln C_t - \ln C) + \frac{e^{\ln N}}{1 - e^{\ln N}} (\ln N_t - \ln N) = (\ln z_t - \ln z) + \alpha (\ln K_t - \ln K) - \alpha (\ln N_t - \ln N)$$

即

$$\hat{C}_t + \frac{N}{1 - N}\hat{N}_t = \hat{z}_t + \alpha\hat{K}_t - \alpha\hat{N}_t.$$

并可以解得

$$\hat{N}_t = \frac{1}{\alpha + \zeta}\hat{z}_t + \frac{\alpha}{\alpha + \zeta}\hat{K}_t - \frac{1}{\alpha + \zeta}\hat{C}_t.$$

其中，$\zeta \equiv N/(1 - N)$

将 $K_{t+1}=(1-\delta)K_t+I_t$ 代入 $C_t+I_t=z_tK_t^{\alpha}N_t^{1-\alpha}=Y_t$ 并在稳态取全微分

$$
\begin{aligned}
& dC_t+dK_{t+1}-(1-\delta)dK_t \\
=& K^{\alpha}N^{1-\alpha}dz_t+\alpha K^{\alpha-1}N^{1-\alpha}dK_t+(1-\alpha)K^{\alpha}N^{-\alpha}dN_t.
\end{aligned}
$$

注意到 $d\ln x=\frac{1}{x}dx$,进而有

$$
\begin{aligned}
& C\hat{C}_t+K\hat{K}_{t+1}-(1-\delta)K\hat{K}_t \\
=& K^{\alpha}N^{1-\alpha}\hat{z}_t+\alpha K^{\alpha}N^{1-\alpha}\hat{K}_t+(1-\alpha)K^{\alpha}N^{1-\alpha}\hat{N}_t.
\end{aligned}
$$

应用稳态下的欧拉方程 (5),可以解得

$$
\begin{aligned}
\hat{K}_{t+1} &= \frac{Y}{K}\left[\hat{z}_t+\alpha\hat{K}_t+(1-\alpha)\hat{N}_t\right]-\frac{C}{K}\hat{C}_t+(1-\delta)\hat{K}_t \\
&= \frac{1}{\beta}\hat{K}_t+\frac{Y}{K}\left[\hat{z}_t+(1-\alpha)\hat{N}_t\right]-\frac{C}{K}\hat{C}_t.
\end{aligned}
$$

将 $\hat{N}_t = \frac{1}{\alpha+\zeta}\hat{z}_t + \frac{\alpha}{\alpha+\zeta}\hat{K}_t - \frac{1}{\alpha+\zeta}\hat{C}_t$ 代入:

$$\hat{K}_{t+1} = \left(\frac{1}{\beta} + \frac{1-\alpha}{\alpha+\zeta}\alpha\frac{Y}{K}\right)\hat{K}_t + \frac{Y}{K}\left(\frac{1+\zeta}{\zeta+\alpha}\right)\hat{z}_t - \left(\frac{C}{K} + \frac{1-\alpha}{\alpha+\zeta}\frac{Y}{K}\right)\hat{C}_t. \tag{6}$$

对欧拉方程 (1) 在稳态取全微分

$$\begin{aligned}
-\frac{dC_t}{C^2} =& -\mathrm{E}_t\left\{\frac{\beta dC_{t+1}}{C^2}\left[\alpha K^{\alpha-1}N^{1-\alpha} + 1 - \delta\right]\right\} \\
&+ \mathrm{E}_t\left\{\frac{\beta}{C}\left[\alpha K^{\alpha-1}N^{1-\alpha}dz_{t+1} + \alpha(\alpha-1)K^{\alpha-2}N^{1-\alpha}dK_{t+1}\right]\right\} \\
&+ \mathrm{E}_t\left\{\frac{\beta}{C}\alpha(1-\alpha)K^{\alpha-1}N^{-\alpha}dN_{t+1}\right\}.
\end{aligned}$$

注意到 $dC_t = C\hat{C}_t, dz_t = \hat{z}_{t+1}, dK_{t+1} = K\hat{K}_{t+1}, dN_{t+1} = N\hat{N}_{t+1}$,并且 $Y = K^{\alpha}N^{1-\alpha}$,进一步有

$$-\hat{C}_t = -\mathrm{E}_t\left[\hat{C}_{t+1}\right] + \beta\alpha\frac{Y}{K}\mathrm{E}_t\left[\hat{z}_{t+1} + (\alpha-1)\hat{K}_{t+1} + (1-\alpha)\hat{N}_{t+1}\right]$$

并将 $\hat{N}_t = \frac{1}{\alpha+\zeta}\hat{z}_t + \frac{\alpha}{\alpha+\zeta}\hat{K}_t - \frac{1}{\alpha+\zeta}\hat{C}_t$ 代入:

$$\begin{aligned}-\hat{C}_t = &-\left[1 + (1-\alpha)\beta\alpha\frac{Y}{K}\frac{1}{\alpha+\zeta}\right]\mathrm{E}_t\left[\hat{C}_{t+1}\right] + \beta\alpha\frac{Y}{K}\frac{1+\zeta}{\alpha+\zeta}\mathrm{E}_t\left[\hat{z}_{t+1}\right]\\ &-\beta\alpha(1-\alpha)\frac{Y}{K}\frac{\zeta}{\alpha+\zeta}\mathrm{E}_t\left[\hat{K}_{t+1}\right].\end{aligned} \tag{7}$$

在劳动力市场, 由欧拉方程 (2) 和实际劳动工资 (4) 可知劳动力市场的供给方程为

$$\frac{\chi C_t}{1 - N_t} = w_t. \tag{8}$$

以及对数线性化后的劳动力市场供给方程

$$\frac{N}{1 - N}\hat{N}_t = -\hat{C}_t + \hat{w}_t. \tag{9}$$

同时对 (4) 作对数线性化可得劳动力市场的需求方程

$$\hat{w}_t = \hat{z}_t + \alpha \hat{K}_t - \alpha \hat{N}_t. \tag{10}$$

对 $Y_t = z_t K_t^{\alpha} N_t^{1-\alpha}$ 做对数线性化有

$$\hat{Y}_t = \alpha \hat{K}_t + (1-\alpha)\hat{N}_t + \hat{z}_t. \tag{11}$$

对 $R_{kt} = \alpha z_t K_t^{\alpha-1} N_t^{1-\alpha}$ 做对数线性化有

$$\hat{R}_{kt} = \hat{Y}_t - \hat{K}_t. \tag{12}$$

综合 (6),(7),(9),(10),(11),(12):

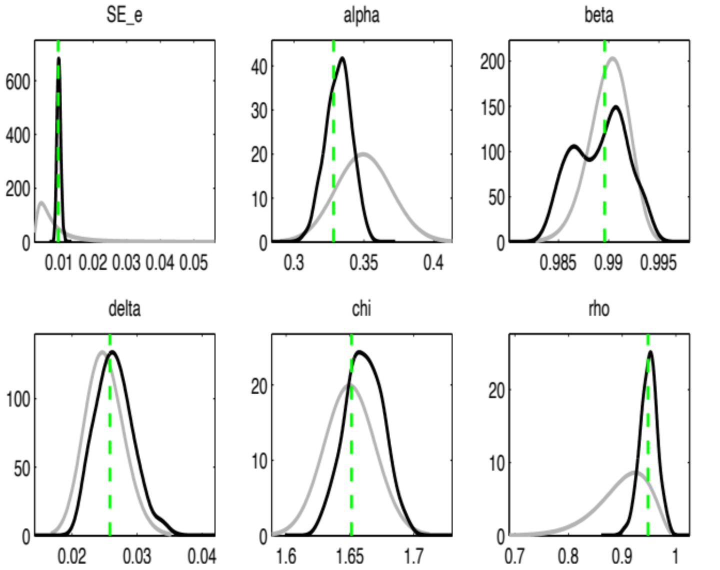
$$
\begin{aligned}
\hat{K}_{t+1} &= \left(\frac{1}{\beta} + \frac{1-\alpha}{\alpha+\zeta}\alpha\frac{Y}{K}\right)\hat{K}_t + \frac{Y}{K}\left(\frac{1+\zeta}{\zeta+\alpha}\right)\hat{z}_t - \left(\frac{C}{K} + \frac{1-\alpha}{\alpha+\zeta}\frac{Y}{K}\right)\hat{C}_t \\
-\hat{C}_t &= -\left[1+(1-\alpha)\beta\alpha\frac{Y}{K}\frac{1}{\alpha+\zeta}\right]\mathrm{E}_t\left[\hat{C}_{t+1}\right] + \beta\alpha\frac{Y}{K}\frac{1+\zeta}{\alpha+\zeta}\mathrm{E}_t\left[\hat{z}_{t+1}\right] \\
&\quad - \beta\alpha(1-\alpha)\frac{Y}{K}\frac{\zeta}{\alpha+\zeta}\mathrm{E}_t\left[\hat{K}_{t+1}\right] \\
\frac{N}{1-N}\hat{N}_t &= -\hat{C}_t + \hat{w}_t. \\
\hat{w}_t &= \hat{z}_t + \alpha\hat{K}_t - \alpha\hat{N}_t. \\
\hat{Y}_t &= \alpha\hat{K}_t + (1-\alpha)\hat{N}_t + \hat{z}_t. \\
\hat{R}_{kt} &= \hat{Y}_t - \hat{K}_t \\
\ln z_t &= \rho\ln z_{t-1} + \sigma\varepsilon_t
\end{aligned}
$$

贝叶斯计算的应用：Real Business Cycle 模型的估计

- 上述是状态空间模型（state space model）的方程组，数据是 $\hat{Y}_t$，参数是 $\boldsymbol{\theta} = (\alpha, \beta, \delta, \chi, \rho, \sigma)$，$\{\hat{K}_t, \hat{C}_t, \hat{N}_t, \hat{z}_t\}$ 是隐变量（latent variable ）
- 对于 $\boldsymbol{\theta}$ 选择合适的先验:

参数	先验分布	先验均值	先验标准差
α	Beta	0.35	0.02
β	Beta	0.99	0.002
δ	Beta	0.025	0.003
χ	Gamma	1.65	0.02
ρ	Beta	0.9	0.05
σ^2	Inv-Gamma	0.01	∞

- MCMC 获得 $\boldsymbol{\theta}$ 的后验分布。


SE_e
alpha
beta
delta
chi
rho
600
400
200
0
0.01 0.02 0.03 0.04 0.05
40
30
20
10
0
0.3
0.35
0.4
200
150
100
50
0
0.985
0.99
0.995
100
50
0
0.02
0.03
0.04
20
10
0
1.6
1.65
1.7
20
10
0
0.7
0.8
0.9
1