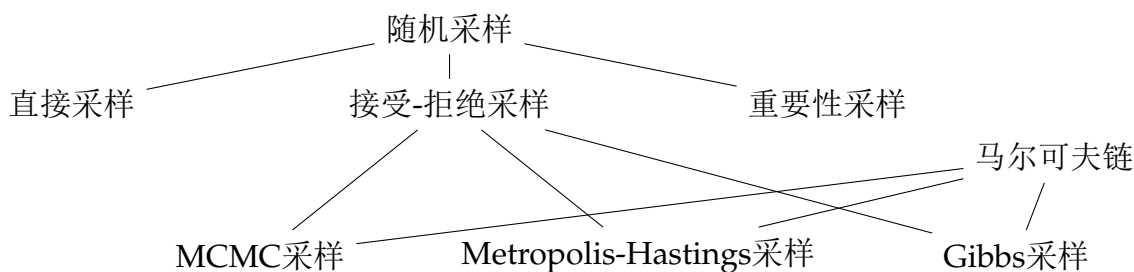


随机采样方法梳理

罗磊

2023 年 6 月 30 日

蒙特卡洛随机采样方法在统计和机器学习等领域具有重要的地位，可用于参数估计、重采样建模等。本文对几种常用的随机采样方法进行简单的介绍和梳理。



本文首先依次对直接采样、接受-拒绝采样和重要性采样进行简要介绍。在此基础上，对接受拒绝采样中更为复杂的马尔可夫链-蒙特卡洛采样方法（如MCMC、Metropolis-Hastings以及Gibbs采样等）进行介绍。

1 简单采样方法

1.1 直接采样

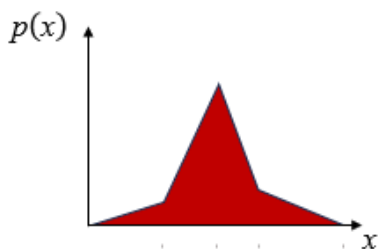
当待采样的概率密度函数PDF已知时：

1. 首先将PDF转换为累积分布密度函数CDF，对应的值域为 $[0, 1]$ ；
2. 从均匀分布 $\text{Uniform}(0, 1)$ 中进行采样，获得 x_i ，然后根据 x_i 从CDF中方向求解出一个样本 y_i ；
3. 对以上第2步进行重复，便获得了一系列 N 个满足对应分布的样本： $\hat{y} = \{y_1, \dots, y_N\}$ 。

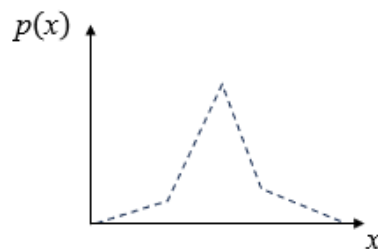
总体流程如图1所示，但需要注意：

1. 直接采样仅适用于待采样PDF已知的情形；
2. 对于高维复杂分布，一般难以获得对应的累积分布密度函数CDF，此时难以使用直接采样。

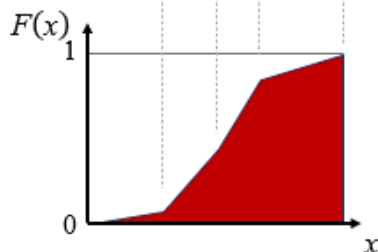
前提: PDF 已知



第三步: 获得样本集 $\hat{x}=\{\hat{x}_1, \dots, \hat{x}_N\}$



第一步: PDF 转换为 CDF:



第二步: 根据CDF反向采样 \hat{x}

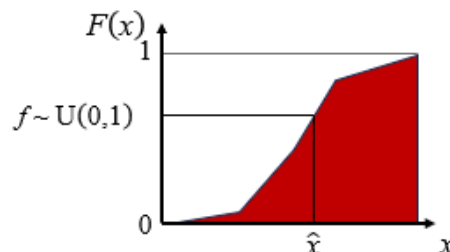


图 1: 直接采样流程图

1.2 接受-拒绝采样

对于某概率分布 $p(X)$, 引入一个简单的提议 (proposal) 分布 $q(X)$, 使得对于 $\forall x_i \in \mathcal{X}$, $M \cdot q(x_i) \geq p(x_i)$, 其中 M 为某常数。这样一来便可通过提议分布 $q(X)$ 实现对 $p(X)$ 的采样:

1. 取 $x_i \sim q(X)$;
2. 计算此时的接受率: $\alpha_i = p(x_i) / (M \cdot q(x_i))$;
3. 从均匀分布中采样 $u \sim \text{Uniform}(0,1)$, 如果 $u \leq \alpha_i$, 则接受 x_i ; 否则, 拒绝 x_i , 不放入样本集;

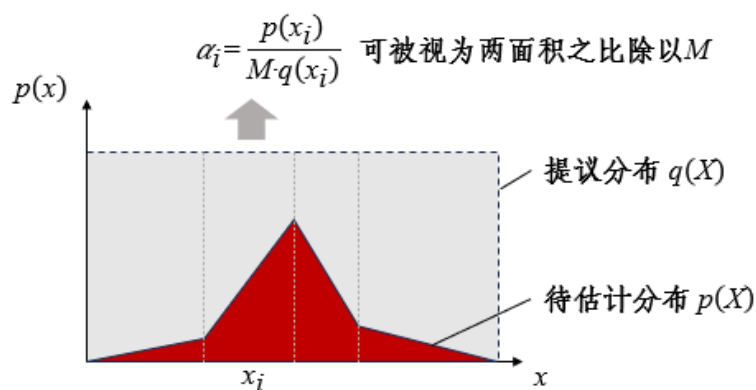


图 2: 接受拒绝采样原理

图2显示了对接受-拒绝采样原理解释, 可以看出:

1. 该方法可与基于蒙特卡洛的积分求解相联系；
2. 在满足上述前提，且不考虑计算效率的情况下，提议分布的选择与最终采样结果分布无关；
3. 常数 M 取值过大或过小对于采样准确性和效率存在影响。当 M 取值过小时，不满足上述前提；当 M 取值过大时，则接受率偏低，导致采样效率降低。因此， M 的最佳设置应该为 $M^* = \max_{x_i \in \mathcal{X}} \frac{p(x_i)}{q(x_i)}$ 。

1.3 重要性采样

以上第一和第二节中介绍的两种方法都是为了对概率分布 p 进行采样，而本节将介绍的重要性采样则是对某符合潜在分布 p 的变量 X 在函数 $f(\cdot)$ 作用下的期望 $\mathbb{E}_{p(x)}[f(x)]$ 进行采样：

$$\begin{aligned}\mathbb{E}_{p(x)}[f(x)] &= \int f(x)p(x)dx \\ &= \int f(x)\frac{p(x)}{q(x)}q(x)dx \\ &\approx \frac{1}{N} \sum_{i=1}^N f(x_i) \frac{p(x_i)}{q(x_i)} \\ x_i &\sim q(X), i = 1, 2, \dots, N\end{aligned}$$

其中， $p(x_i)/q(x_i)$ 被称为重要性值，衡量了不同 x_i 取值所得 $f(x_i)$ 对于 $f(x)$ 整体期望的影响。

2 基于马尔可夫链的复杂采样方法

在工业过程建模、分子动力学模拟等领域，过程高维复杂，难以获得充分的样本反映其全貌。人们可采用基于马尔可夫链的迭代采样逐步获得所感兴趣的参数和变量的近似分布。

2.1 马尔可夫链-蒙特卡洛（MCMC）采样

在该方法中，假设在第 $t-1$ 步获得了样本 \hat{x}_{t-1} ，接下来在第 t 步中，可根据提议分布，按照 $q(X_t|X_{t-1})$ 新采集一个样本 \hat{s} ，然后计算接受率 $\alpha = \min\{1, p(\hat{s})/p(\hat{x}_{t-1})\}$ 。接下来，以概率 α 接受 \hat{s} 为第 t 轮的采样结果，即 $\hat{x}_t = \hat{s}$ ；否则，以上一步结果作为本轮结果， $\hat{x}_t = \hat{x}_{t-1}$ 。

注意，在MCMC采样过程中， \hat{x}_t 取决于 \hat{x}_{t-1} （即马尔可夫过程），其可能与 \hat{x}_{t-1} 以及更早的样本间存在依赖，形成了一条马尔可夫链。因此，相邻的样本间可能不独立，这是基于马尔可夫链的采样方法与1.2节中接受-拒绝采样方法的最大区别。下图3显示了MCMC采样的原理：

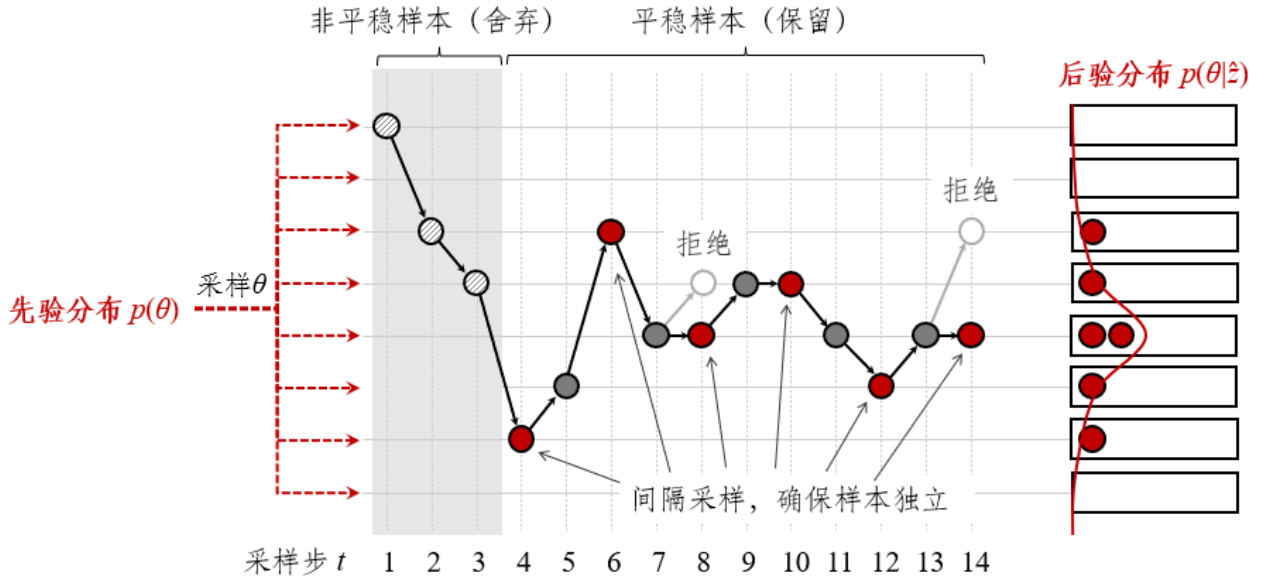


图 3: 直接采样流程图

2.2 Metropolis-Hastings采样

在一些情况下，上述MCMC采样算法中的接受率 α 可能会很小，导致算法需要经历很多次迭代才能到达平稳。因此，Metropolis-Hastings算法考虑把MCMC接受率等式两侧同步放大，将其中的一个接受概率设置为1，这样就能保证每次迭代过程中接收新状态的概率越大，加速算法收敛。

$$\alpha = \min \left\{ 1, \frac{p(\hat{s})Q_{\hat{x}_{t-1} \rightarrow \hat{s}}}{p(\hat{x}_{t-1})Q_{\hat{s} \rightarrow \hat{x}_{t-1}}} \right\}$$

其中， $Q_{\hat{x}_{t-1} \rightarrow \hat{s}}$ 和 $Q_{\hat{s} \rightarrow \hat{x}_{t-1}}$ 为两个状态间相互转移的概率。

2.3 Gibbs采样

对于很多模型，一般难以直接对其中多个参数的高维联合后验分布（joint posterior distribution）进行采样估计，但更易对各参数在低维情形下的全条件分布（full conditional distribution）进行处理。这样便可使用Gibbs采样器以迭代方式进行采样，最终逼近模型参数的联合后验分布。Gibbs采样旨在根据各参数的全条件概率密度进行采样，还原所有模型参数的联合后验概率密度，由此获得模型参数期望、中位数、置信区间等一系列统计指标。Gibbs采样旨在通过以下多步低维采样获得多个参数的联合分布：

$$\begin{cases} p(x_1|x_2, x_3, \dots, x_{k-1}, x_k) \\ p(x_2|x_1, x_3, \dots, x_{k-1}, x_k) \\ \vdots \\ p(x_k|x_1, x_2, \dots, x_{k-1}) \end{cases} \rightarrow p(x_1, x_2, \dots, x_k) \quad (1)$$

具体在第 t 步的采样过程为：

1. $x_1^{(t)} \sim p(x_1 | \hat{x}_2^{(t-1)}, \hat{x}_3^{(t-1)}, \dots, \hat{x}_{k-1}^{(t-1)}, \hat{x}_k^{(t-1)});$
2. $x_2^{(t)} \sim p(x_2 | \hat{x}_1^{(t-1)}, \hat{x}_3^{(t-1)}, \dots, \hat{x}_{k-1}^{(t-1)}, \hat{x}_k^{(t-1)});$
3. $x_3^{(t)} \sim p(x_3 | \hat{x}_1^{(t-1)}, \hat{x}_2^{(t-1)}, \dots, \hat{x}_{k-1}^{(t-1)}, \hat{x}_k^{(t-1)});$
- \vdots

3 基于接受-拒绝采样的重采样回归建模

样本不平衡时机器学习分类或回归建模中常常面临的一个问题，人们可以通过：

1. 调整不同标签（目标为类别值，对应于分类）或不同数值（目标为连续值，对应于回归）上样本的比例；
2. 采用F1分数等适用于不平衡样本情形下的模型评价指标。

等方式来强化模型对于目标分布较少但对于预测准确性影响较大区域的学习。具体算例和代码请见Github代码，此处略。