

Markov Chain Monte Carlo 05 Sampling

Cheh Gong

03 January 2020

在前面的章节中，我们已经基本介绍了 Markov Chain Monte Carlo Sampling 的基本概念，基本思路 and 主要方法。那么这一小节中，我们将主要来介绍一下，什么是采样？我们为什么而采样？什么样的样本是好的样本？以及我们采样中主要会遇到哪些困难？

1 采样的动机

这一小节的目的就是我们要知道什么是采样的动机，我们为什么而采样？

1. 首先第一点很简单，采样本身就是发出常见的任务，我们机器学习中经常需要进行采样来完成各种各样的任务。如果从一个 $P(X)$ 中采出一堆样本。

2. 求和求积分。包括大名鼎鼎的 Monte Carlo 算法。我们求 $P(X)$ 主要是为了求在 $P(X)$ 概率分布下的一个相关函数的期望，也就是：

$$\int P(x)f(x)dx = \mathbb{E}_{P(X)}[f(X)] \approx \frac{1}{N} \sum_{i=1}^N f(x^{(i)}) \quad (1)$$

而我们是通过采样来得到 $P(X) \sim \{x^{(1)}, x^{(2)}, \dots, x^{(N)}\}$ 样本点。

2 什么样的样本

既然，我们知道了采样的目的和动机，下一个问题就自然是，同样是采样，什么样的样本就是好样本呢？或者说是采样的效率更高一些。

1. 首先样本的分布肯定是要趋向于原始的目标分布吧，也就是说样本要趋向于高概率选择区域。或者是说，采出来的样本出现的概率和实际的目标分布的概率保持一致。

2. 样本和样本之间是相互独立的。这个就没有那么直观了。大家想一想就知道了，如果我采出来的一堆样本之间都差不多，那么就算采出来了趋向于高概率选择区域的样本，那采样效率太低了，样本中反映的信息量太有限了。

3 实际采样中的困难

实际采样中，采样时困难的，为什么呢？我们这里主要介绍两点：

1. **Partation function is intractable.** 我们的后验分布往往被写成 $P(X) = \frac{1}{Z} \hat{P}(X)$ ，上面这个 $\hat{P}(X)$ 都比较好求，就是等于 Likelihood \times Prior。而 Z 就是我们要求的归一化常数，它非常的难

以计算, $Z = \int \hat{P}(X)dX$, 这几乎就是不可计算的。所以, 有很多采样方法就是想要跳过求 $P(X)$ 的过程, 来从一个近似的分布中进行采样, 当然这个近似的分布采样要比原分布简单。比如: Rejection Sampling 和 Importance Sampling。

2. **The curse of high dimension.** 如果样本空间 $\mathcal{X} \in \mathbb{R}^p$, 每个维度都有 K 个状态的话。那么总的样本空间就有 K^p 的状态。要知道那个状态的概率高, 就必须要遍历整个样本空间, 不然就不知道哪个样本的概率高, 如果状态的数量是这样指数型增长的话, 全看一遍之后进行采样时不可能的。所以, 直接采样的方法是不可行的。

4 采样方法

Rejection Sampling 和 Importance Sampling, 都是借助一个 $Q(x)$ 去逼近目标分布 $P(x)$, 通过从 $Q(x)$ 中进行采样来达到在 $P(x)$ 中采样的目的, 而且在 $Q(x)$ 中采样比较简单。当时如果 $Q(x)$ 和 $P(x)$ 直接的差距太大的话, 采样效率会变得很低。

而 MCMC 方法, 我们主要介绍了 MH Sampling 和 Gibbs Sampling, 我们主要是通过构建一个马氏链去逼近目标分布, 具体的描述将在下一节中展开描述。