

刘建平Pinard

十年码农，对数学统计学，数据挖掘，机器学习，大数据平台，大数据平台应用开发，大数据可视化感兴趣。

[博客园](#) [首页](#) [新随笔](#) [联系](#) [订阅](#) [管理](#)

## MCMC(一)蒙特卡罗方法

MCMC(一)蒙特卡罗方法

[MCMC\(二\)马尔科夫链](#)

[MCMC\(三\)MCMC采样和M-H采样](#)

[MCMC\(四\)Gibbs采样](#)

作为一种随机采样方法，马尔科夫链蒙特卡罗（Markov Chain Monte Carlo，以下简称MCMC）在机器学习、深度学习以及自然语言处理等领域都有广泛的应用，是很多复杂算法求解的基础。比如我们前面讲到的[分解机\(Factorization Machines\)](#)、[推荐算法](#)，还有前面讲到的[受限玻尔兹曼机（RBM）原理总结](#)，都用到了MCMC来做一些复杂运算的近似求解。下面我们就对MCMC的原理做一个总结。

## 1. MCMC概述

从名字我们可以看出，MCMC由两个MC组成，即蒙特卡罗方法（Monte Carlo Simulation，简称MC）和马尔科夫链（Markov Chain，也简称MC）。要弄懂MCMC的原理我们首先得搞清楚蒙特卡罗方法和马尔科夫链的原理。我们将用三篇来完整学习MCMC。在本篇，我们关注于蒙特卡罗方法。

## 2. 蒙特卡罗方法引入

蒙特卡罗原来是一个赌场的名称，用它作为名字大概是因为蒙特卡罗方法是一种随机模拟的方法，这很像赌场里面的扔骰子的过程。最早的蒙特卡罗方法都是为了求解一些不太好求解的求和或者积分问题。比如积分：

### 公告

★珠江追梦，饮岭南茶，恋鄂北家★

昵称：刘建平Pinard

园龄：1年5个月

粉丝：1090

关注：13

[+加关注](#)

<	2018年3月						>
日	一	二	三	四	五	六	
25	26	27	28	1	2	3	
4	5	6	7	8	9	10	
11	12	13	14	15	16	17	
18	19	20	21	22	23	24	
25	26	27	28	29	30	31	
1	2	3	4	5	6	7	

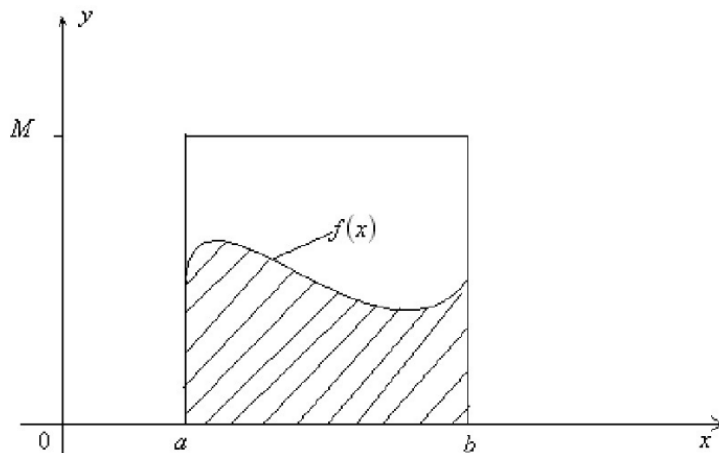
### 常用链接

[我的随笔](#)

[我的评论](#)

$$\theta = \int_a^b f(x)dx$$

如果我们很难求解出 $f(x)$ 的原函数，那么这个积分比较难求解。当然我们可以通过蒙特卡罗方法来模拟求解近似值。如何模拟呢？假设我们函数图像如下图：



则一个简单的近似求解方法是在 $[a, b]$ 之间随机的采样一个点。比如 $x_0$ ，然后用 $f(x_0)$ 代表在 $[a, b]$ 区间上所有的 $f(x)$ 的值。那么上面的定积分的近似求解为：

$$(b-a)f(x_0)$$

当然，用一个值代表 $[a, b]$ 区间上所有的 $f(x)$ 的值，这个假设太粗糙。那么我们可以采样 $[a, b]$ 区间的 $n$ 个值： $x_0, x_1, \dots, x_{n-1}$ ，用它们的均值来代表 $[a, b]$ 区间上所有的 $f(x)$ 的值。这样我们上面的定积分的近似求解为：

$$\frac{b-a}{n} \sum_{i=0}^{n-1} f(x_i)$$

虽然上面的方法可以一定程度上求解出近似的解，但是它隐含了一个假定，即 $x$ 在 $[a, b]$ 之间是均匀分布的，而绝大部分情况， $x$ 在 $[a, b]$ 之间不是均匀分布的。如果我们用上面的方法，则模拟求出的结果很可能和真实值相差甚远。

怎么解决这个问题呢？如果我们可以得到 $x$ 在 $[a, b]$ 的概率分布函数 $p(x)$ ，那么我们的定积分求和可以这样进行：

$$\theta = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{p(x)} p(x)dx \approx \frac{1}{n} \sum_{i=0}^{n-1} \frac{f(x_i)}{p(x_i)}$$

我的参与

最新评论

我的标签

#### 随笔分类(101)

- 0040. 数学统计学(4)
- 0081. 机器学习(62)
- 0082. 深度学习(10)
- 0083. 自然语言处理(23)
- 0121. 大数据挖掘(1)
- 0122. 大数据平台(1)
- 0123. 大数据可视化

#### 随笔档案(101)

- 2017年8月 (1)
- 2017年7月 (3)
- 2017年6月 (8)
- 2017年5月 (7)
- 2017年4月 (5)
- 2017年3月 (10)
- 2017年2月 (7)
- 2017年1月 (13)
- 2016年12月 (17)
- 2016年11月 (22)
- 2016年10月 (8)

#### 常去的机器学习网站

52 NLP

上式最右边的这个形式就是蒙特卡罗方法的一般形式。当然这里是连续函数形式的蒙特卡罗方法，但是在离散时一样成立。

可以看出，最上面我们假设 $x$ 在 $[a, b]$ 之间是均匀分布的时候， $p(x_i) = 1/(b-a)$ ，带入我们有概率分布的蒙特卡罗积分的上式，可以得到：

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{f(x_i)}{1/(b-a)} = \frac{b-a}{n} \sum_{i=0}^{n-1} f(x_i)$$

也就是说，我们最上面的均匀分布也可以作为一般概率分布函数 $p(x)$ 在均匀分布时候的特例。那么我们现在的问题转到了如何求出 $x$ 的分布 $p(x)$ 的若干和样本上来。

### 3. 概率分布采样

上一节我们讲到蒙特卡罗方法的关键是得到 $x$ 的概率分布。如果求出了 $x$ 的概率分布，我们可以基于概率分布去采样基于这个概率分布的 $n$ 个 $x$ 的样本集，带入蒙特卡罗求和的式子即可求解。但是还有一个关键的问题需要解决，即如何基于概率分布去采样基于这个概率分布的 $n$ 个 $x$ 的样本集。

对于常见的均匀分布 $uniform(0, 1)$ 是很容易采样样本的，一般通过线性同余发生器可以很方便的生成 $(0, 1)$ 之间的伪随机数样本。而其他常见的概率分布，无论是离散的分布还是连续的分布，它们的样本都可以通过 $uniform(0, 1)$ 的样本转换而得。比如二维正态分布的样本 $(Z_1, Z_2)$ 可以通过通过独立采样得到的 $uniform(0, 1)$ 样本对 $(X_1, X_2)$ 通过如下的式子转换而得：

$$Z_1 = \sqrt{-2\ln X_1} \cos(2\pi X_2)$$

$$Z_2 = \sqrt{-2\ln X_1} \sin(2\pi X_2)$$

其他一些常见的连续分布，比如 $t$ 分布， $F$ 分布， $Beta$ 分布， $Gamma$ 分布等，都可以通过类似的方式从 $uniform(0, 1)$ 得到的采样样本转化得到。在python的numpy，scikit-learn等类库中，都有生成这些常用分布样本的函数可以使用。

不过很多时候，我们的 $x$ 的概率分布不是常见的分布，这意味着我们没法方便的得到这些非常见的概率分布的样本集。那这个问题怎么解决呢？

### 4. 接受-拒绝采样

Analytics Vidhya  
机器学习库  
机器学习路线图  
深度学习进阶书  
深度学习入门书

#### 积分与排名

积分 - 300539  
排名 - 609

#### 阅读排行榜

1. 梯度下降 ( Gradient Descent ) 小结(98432)
2. 梯度提升树(GBDT)原理小结(47788)
3. 线性判别分析LDA原理总结(32128)
4. scikit-learn决策树算法类库使用小结(28391)
5. 谱聚类 ( spectral clustering ) 原理总结(21369)

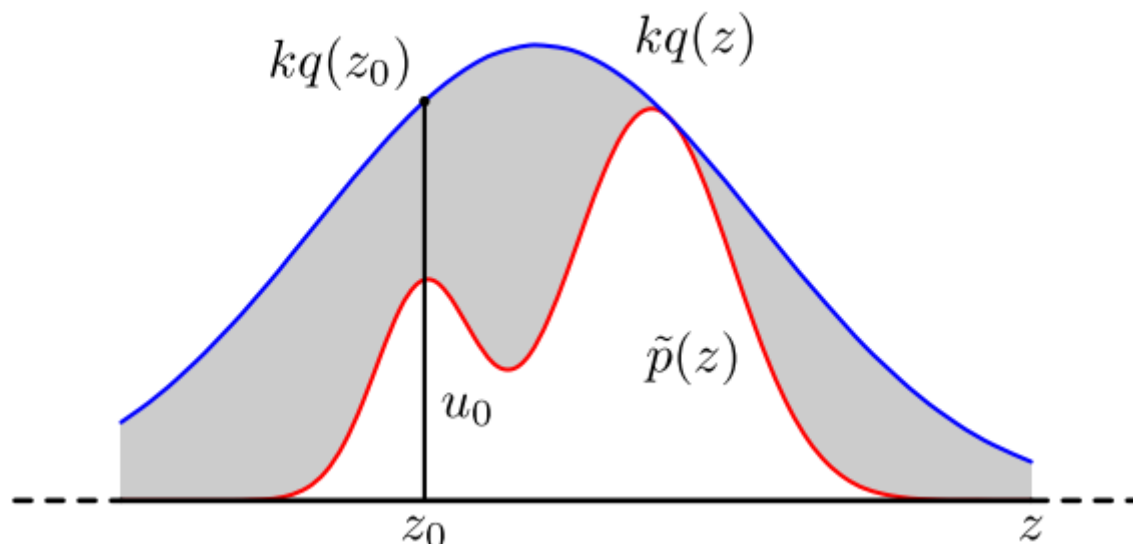
#### 评论排行榜

1. 梯度提升树(GBDT)原理小结(81)
2. 谱聚类 ( spectral clustering ) 原理总结(73)
3. 梯度下降 ( Gradient Descent ) 小结(62)
4. 卷积神经网络(CNN)反向传播算法(57)
5. 集成学习之Adaboost算法原理小结(50)

#### 推荐排行榜

1. 梯度下降 ( Gradient Descent ) 小结(41)
2. 集成学习原理小结(15)
3. 卷积神经网络(CNN)反向传播算法(14)
4. 集成学习之Adaboost算法原理小结(13)

对于概率分布不是常见的分布，一个可行的办法是采用接受-拒绝采样来得到该分布的样本。既然  $p(x)$  太复杂在程序中没法直接采样，那么我设定一个程序可采样的分布  $q(x)$  比如高斯分布，然后按照一定的方法拒绝某些样本，以达到接近  $p(x)$  分布的目的，其中  $q(x)$  叫做 proposal distribution。



具体采用过程如下，设定一个方便采样的常用概率分布函数  $q(x)$ ，以及一个常量  $k$ ，使得  $p(x)$  总在  $kq(x)$  的下方。如上图。

首先，采样得到  $q(x)$  的一个样本  $z_0$ ，采样方法如第三节。然后，从均匀分布  $(0, kq(z_0))$  中采样得到一个值  $u$ 。如果  $u$  落在了上图中的灰色区域，则拒绝这次抽样，否则接受这个样本  $z_0$ 。重复以上过程得到  $n$  个接受的样本  $z_0, z_1, \dots, z_{n-1}$  则最后的蒙特卡罗方法求解结果为：

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{f(z_i)}{p(z_i)}$$

整个过程中，我们通过一系列的接受拒绝决策来达到用  $q(x)$  模拟  $p(x)$  概率分布的目的。

## 5. 蒙特卡罗方法小结

使用接受-拒绝采样，我们可以解决一些概率分布不是常见的分布的时候，得到其采样集并用蒙特卡罗方法求和的目的。但是接受-拒绝采样也只能部分满足我们的需求，在很多时候我们还是很难得到我们的概率分布的样本集。比如：

1) 对于一些二维分布 $p(x, y)$ , 有时候我们只能得到条件分布 $p(x|y)$ 和 $p(y|x)$ 和, 却很难得到二维分布 $p(x, y)$ 一般形式, 这时我们无法用接受-拒绝采样得到其样本集。

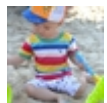
2) 对于一些高维的复杂非常见分布 $p(x_1, x_2, \dots, x_n)$ , 我们要找到一个合适的 $q(x)$ 和 $k$ 非常困难。

从上面可以看出, 要想将蒙特卡罗方法作为一个通用的采样模拟求和的方法, 必须解决如何方便得到各种复杂概率分布的对应的采样样本集的问题。而我们下一篇要讲到的马尔科夫链就是帮助找到这些复杂概率分布的对应的采样样本集的白衣骑士。下一篇我们来总结马尔科夫链的原理。

( 欢迎转载, 转载请注明出处。欢迎沟通交流: pinard.liu@ericsson.com )

分类: [0040. 数学统计学](#)

标签: [机器学习中的数学](#)



刘建平Pinard

关注 - 13

粉丝 - 1090

[+加关注](#)

6

0

« 上一篇: [受限玻尔兹曼机 \(RBM\) 原理总结](#)

» 下一篇: [MCMC\(二\)马尔科夫链](#)

---

posted @ 2017-03-27 15:08 刘建平Pinard 阅读(3647) 评论(8) 编辑 收藏

## 评论列表

---

#1楼 2017-03-27 17:13 codesnippet.info

加油!

支持(0) 反对(0)

---

#2楼 2017-03-27 20:05 xulu1352

大神又开更了，，大神后期能不能讲讲特征工程的东东，想看看老司机怎么做的

支持(0) 反对(0)

---

#3楼[楼主] 2017-03-28 10:11 刘建平Pinard

@ xulu1352

写作提纲里有特征工程部分，应该在写完自然语言处理部分就会开始。

支持(0) 反对(0)

---

#4楼 2017-03-28 19:04 桂。

感谢分享，学到了许多干货！

支持(0) 反对(0)

---

#5楼 2017-10-31 09:54 开拓者亮仔

@刘建平Pinard 怎么解决这个问题呢？如果我们可以得到 $x$ 在 $[a,b]$ 的概率分布函数 $p(x)$ ，那么我们的定积分求和可以这样进行。那面的式子的得出不是很明白，可否详细说下。

支持(0) 反对(0)

---

#6楼[楼主] 2017-10-31 10:37 刘建平Pinard

@ 开拓者亮仔

你说的是这个式子吧。

$$\theta = \int_a^b f(x)dx = \int_a^b \frac{f(x)}{p(x)}p(x)dx \approx \frac{1}{n} \sum_{i=0}^{n-1} \frac{f(x_i)}{p(x_i)}$$

猜你的问题在最后一步转换。由于 $\int_a^b \frac{f(x)}{p(x)}p(x)dx$ 可以看做是 $\frac{f(x)}{p(x)}$ 基于概率分布 $p(x)$ 的期望，那么我们可以用期望的方法来求这个式子的值。而计算期望的一个近似方法是取 $\frac{f(x)}{p(x)}$ 的若干个基于分布 $p(x)$ 的采样点，然后求平均值得到。

支持(5) 反对(0)

---

#7楼 2018-02-08 11:29 stillriver

@ 刘建平Pinard

引用

@开拓者亮仔

你说的是这个式子吧。

$$\int_{-\infty}^{\infty} f(x) p(x) dx \approx \frac{1}{n} \sum_{i=0}^{n-1} f(x_i) p(x_i)$$
  
$$\theta = \int_{-\infty}^{\infty} f(x) p(x) dx = \int_{-\infty}^{\infty} f(x) p(x) dx \approx \frac{1}{n} \sum_{i=0}^{n-1} f(x_i) p(x_i)$$

大数定理

支持(0) 反对(0)

#8楼 2018-02-10 15:39 扼杀

是的，基于分布 $p(x)$ 的实际采样时，样本的密度不一样， $f(x)/p(x)$ 刚好是对应的值。当采样的数量足够多，就大数定理，基本符合 $p(x)$ 的分布了。采样，是一个从分布函数到 $n$ 个采样点的映射。刚一看，真是不理解，说明以前看的也没有仔细问题

支持(0) 反对(0)

[刷新评论](#) [刷新页面](#) [返回顶部](#)

注册用户登录后才能发表评论，请 [登录](#) 或 [注册](#)，[访问网站首页](#)。

【推荐】超50万VC++源码: 大型组态工控、电力仿真CAD与GIS源码库！

【缅怀】传奇谢幕，回顾霍金76载传奇人生

【推荐】业界最快速.NET数据可视化图表组件

【腾讯云】买域名送解析+SSL证书+建站

【活动】2050 科技公益大会 - 年青人因科技而团聚



#### 最新IT新闻:

- 详解Uber自动驾驶汽车传感器系统，怎样的配置才能避免撞人事件
  - 华尔街日报：腾讯音乐估值飙升，预计将在今年公开上市
  - MIT开发出全新成像系统，让自动驾驶汽车在浓雾中也能“看见”路况
  - 分析：数据泄露事件对Facebook影响究竟有多大？
  - 谷歌CEO皮猜：在AI时代 各界都要支持终生学习
- » 更多新闻...



#### 最新知识库文章:

- 写给自学者的入门指南
  - 和程序员谈恋爱
  - 学会学习
  - 优秀技术人的管理陷阱
  - 作为一个程序员，数学对你到底有多重要
- » 更多知识库文章...