



Xianling Mao的专栏

Explore the world with my head and feet.

目录视图

摘要视图

RSS 订阅

个人资料



xianlingmao

访问：900031次

积分：4578

等级：

排名：第6289名

原创：27篇 转载：13篇

译文：0篇 评论：159条

文章搜索

文章分类

- Emacs (3)
- mathmatics (16)
- other (8)
- QA (0)
- 信息抽取 (5)
- 信息检索 (2)
- 图模型 (5)
- 机器学习 (14)
- 模拟与采样 (5)
- topic model (6)
- lda (2)
- gibbs sampling (3)
- latex (1)
- Deep Learning (1)

文章存档

- 2013年01月 (1)
- 2012年09月 (1)
- 2012年08月 (2)
- 2012年07月 (5)
- 2012年06月 (3)

展开

阅读排行

- 深入理解拉格朗日乘法 (231386)
- 随机模拟的基本思想和常 (64191)
- 浅谈深度学习(Deep Leai

赠书 | **AI**专栏（**AI**圣经！《深度学习》中文版）

机器学习&数据挖掘 系统实训

【获奖公布】征文 | 你会为 **AI** 转型么？

随机模拟的基本思想和常用采样方法（sampling）

2012-07-23 15:27

64202人阅读

评论(23)

收藏

举报

分类：

gibbs sampling（2）

机器学习（13）

模拟与采样（4）

mathmatics（15）

版权声明：本文为博主原创文章，未经博主允许不得转载。

通常，我们会遇到很多问题无法用分析的方法来求得精确解，例如由于式子特别，真的解不出来；

一般遇到这种情况，人们经常会采用一些方法去得到近似解（越逼近精确解越好，当然如果一个近似算法与精确解的接近程度能够通过一个式子来衡量或者有上下界，那么这种近似算法比较好，因为人们可以知道接近程度，换个说法，一般一个近似算法被提出后，人们通常都会去考察或寻求刻划近似程度的式子）。

本文要谈的随机模拟就是一类近似求解的方法，这种方法非常的牛逼哦，它的诞生虽然最早可以追溯到18xx年法国数学家蒲松的投针问题（用模拟的方法来求解pi的问题），但是真正的大规模应用还是被用来解决二战时候美国佬生产原子弹所碰到的各种难以解决的问题而提出的蒙特卡洛方法（Monte Carlo），从此一发不可收拾。

本文将分为两个大类来分别叙述，首先我们先谈谈随机模拟的基本思想和基本思路，然后我们考察随机模拟的核心：对一个分布进行抽样。我们将介绍常用的抽样方法，1. 接受-拒绝抽样；2）重要性抽样；3）MCMC（马尔科夫链蒙特卡洛方法）方法，主要介绍MCMC的两个非常著名的采样算法（metropolis-hasting算法和它的特例Gibbs采样算法）。

一. 随机模拟的基本思想

我们先看一个例子：现在假设我们有一个矩形的区域R（大小已知），在这个区域中有一个不规则的区域M（即不能通过公式直接计算出来），现在要求取M的面积？怎么求？近似的方法很多，例如：把这个不规则的区域M划分为很多很多个小的规则区域，用这些规则区域的面积求和来近似M，另外一个近似的方法就是采样的方法，我们抓一把黄豆，把它们均匀地铺在矩形区域，如果我们知道黄豆的总个数S，那么只要我们数数位于不规则区域M中的黄豆个数S1，那么我们就可以求出M的面积：M=S1*R/S。

另外一个例子，在**机器学习**或统计计算领域，我们常常遇到这样一类问题：即如何求取一个定积分： $\int_a^b f(x) dx$ ，如归一化因子等。

如何来求解这类问题呢？当然如果定积分可以解析求出，直接求出即可，如果不能解析求出，只能求取近似解了，常用的近似方法是采用蒙特卡洛积分，即把上述式子改写为：

$$\int_a^b f(x) \cdot g(x) / g(x) \, dx = \int_a^b (1/g(x)) \cdot f(x) \cdot g(x) \, dx$$

那么把f(x)/g(x)作为一个函数，而把g(x)看做是[a,b]上的一个概率分布，抽取n个样本之后，上述式子可以继续写为： $\frac{1}{n} \sum_{i=1}^n [f(x_i)/g(x_i)]$ ，当n趋向无穷大的时候，根据大数定理，上述式子是和要求的定积分式子是相等的，因此可以用抽样的方法来得到近似解。

通过上述两个例子，我们大概能够理解抽样方法解决问题的基本思想，其基本思路就是要把待解决的问题转化为一种可以通过某种采样方法可以解决的问题，至于怎么转化，还是挺有创造性，没有定法。因此随机模拟方法的核心就是如何对一个概率分布得到样本，即抽样（sampling）。因此下一节，我们将介绍常用的抽样方法。

（pku，sewm，shinning）

二. 常见的抽样方法

2.0 直接抽样法

模型选择的几种方法：AI

(62319)

狄利克雷过程（dirichlet |

(56202)

核方法(kernel method)的

(49746)

什么叫共轭先验或者共轭

(43730)

俄罗斯的数学太牛逼了，

(43482)

深入理解模拟退火算法（

(42146)

话题模型（topic model）

(42073)

(34547)

评论排行

深入理解拉格朗日乘子法

(66)

随机模拟的基本思想和常

(23)

核方法(kernel method)的

(16)

浅谈深度学习(Deep Lear

(13)

一. 图模型（graphical m

(9)

看懂信息检索和网络数据

(6)

Topic Model的分类和设

(3)

话题模型（topic model）

(3)

狄利克雷过程（dirichlet |

(3)

EM算法

(3)

推荐文章

* CSDN日报20170725——《新的开始，从研究生到入职亚马逊》

* 深入剖析基于并发AQS的重入锁(ReentrantLock)及其Condition实现原理

* Android版本的"Wannacry"文件加密病毒样本分析(附带锁机)

* 工作与生活真的可以平衡吗？

* 《Real-Time Rendering 3rd》提炼总结——高级着色：BRDF及相关技术

* 《三体》读后思考-泰勒展开/维度打击/黑暗森林

最新评论

随机模拟的基本思想和常用采样: task_force: 楼主，有一事请教，对g(x)分布进行采样得到样本值x_i后，和式为什么不是\sum_1^n？

随机模拟的基本思想和常用采样: task_force: 楼主，有一点不明白希望请教一下，n个样本按照g(x)的分布进行采样，则积分为什么变为和式{\sum_...

核方法(kernel method)的主要思: roguesir: 赞

什么叫共轭先验或者共轭分布？qq_30330939: 不太对哦，请看一下维基百科，那里明明是说先验和后验共轭

深入理解拉格朗日乘子法（Lagrangian Multiplier）: MaRinQ: 解释得非常不错 赞一个！

核方法(kernel method)的主要思: jason-刘: 请教楼主一个问题，证明一个函数满足Mercer定理时，训练数据集是怎么选取的？大小有规定吗？还有...

测度论--长度是怎样炼成的[zz]death_ray: 谢谢，学实分析感觉怀疑自己的数学体系，看完了之后豁然开朗的感觉。

略，因为较为简单，而且只能解决很简单的问题，一般是一维分布的问题。

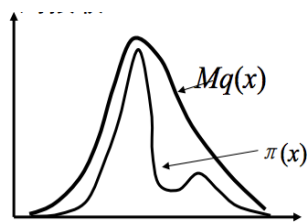
2.1 接受-拒绝抽样（Acceptance-Rejection sampling）

又简称拒绝抽样，直观地理解，为了得到一个分布的样本，我们通过某种机制得到了很多的初步样本，然后其中一部分初步样本会被作为有效的样本（即要抽取的分布的样本），一部分初步样本会被认为是无效样本舍弃掉。这个算法的基本思想是：我们需要对一个分布f(x)进行采样，但是却很难直接进行采样，所以我们想通过另外一个容易采样的分布g(x)的样本，用某种机制去除掉一些样本，从而使得剩下的样本就是来自与所求分布f(x)的样本。

它有几个条件：1）对于任何一个x，有f(x)<=M*g(x); 2) g(x)容易采样；3) g(x)最好在形状上比较接近f(x)。具体的采样过程如下：

1. 对于g(x)进行采样得到一个样本xi, xi ~ g(x);
2. 对于均匀分布采样 ui ~ U(a,b);
3. 如果ui<= f(x)/[M*g(x)], 那么认为xi是有效的样本；否则舍弃该样本；（# 这个步骤充分体现了这·接受-拒绝）
4. 反复重复步骤1~3，直到所需样本达到要求为止。

这个方法可以如图所示：



(说明：这是从其他地方弄来的图，不是自己画的，符号有些和文中不一致，其中\pi(x) 就是文中的f(x)，q(x)就是文中的g(x))

2.2 重要性抽样(Importance sampling)

重要性采样和蒙特卡洛积分密切相关，看积分：

\int f(x) dx = \int f(x)*(1/g(x))*g(x) dx, 如果g(x)是一个概率分布，从g(x)中抽取N个样本，上述的式子就约等于(1/N)*\sum f(xi)*(1/g(xi))。这相当于给每个样本赋予了一个权重，g(xi)大意味着概率大，那么N里面含有这样的样本xi就多，即这些样本的权重大，所以称为重要性抽样。

抽样的步骤如下：

1. 选择一个容易抽样的分布g(x), 从g(x)中抽取N个样本；
2. 计算(1/N)* \sum f(xi)*(1/g(xi))，从而得到近似解。

（pku，sewm，shinning）

2.3 MCMC抽样方法

无论是拒绝抽样还是重要性采样，都是属于独立采样，即样本与样本之间是独立无关的，这样的采样效率比较低，如拒绝采样，所抽取的样本中有很大部分是无效的，这样效率就比较低，MCMC方法是关联采样，即下一个样本与这个样本有关系，从而使得采样效率高。MCMC方法的基本思想是：通过构建一个markov chain使得该markov chain的稳定分布是我们所要采样的分布f(x)。如果这个markov chain达到稳定状态，那么来自这个chain的每个样本都是f(x)的样本，从而实现抽样的目的。这里存在一个核心问题，如何构建满足要求的markov chain？（什么是markov chain，什么是稳定分布，请查资料，这里假设读者已知。）

在本文，我们介绍两个著名MCMC抽样方法，它们能够方便地构建满足要求的markov chain。

A). Metropolis-Hasting算法

这个算法是两个作者的合称，但不是同一篇论文的，一个是1953年，另外一个197x年对1953年的工作进行了一些扩展，所以以这两位作者的名字来命名这个算法。

假设要采样的概率分布是\pi(x)，现在假设有一个概率分布p(y|x)，使得\pi(x)*p(y|x) = \pi(y)*p(x|y)成立，称细致平衡公式，这个细致平衡公式是markov chain能达到稳定分布的必要条件。因此关键是构建出一个概率分布p(y|x)使得它满足细致平衡。现在假设我们有一个容易采样的分布q(y|x)（称为建议分布），对于目前的样本x，它能够通过q(y|x)得到下一个建议样本y，这个建议样本y按照一定的概率被接受或者不被接受，称为比率\alpha(x, y) = min{1,

关闭

看懂信息检索和网络数据挖掘领：任蒙蒙-Angel: 认真琢磨，对于提高写作能力很有帮助

深入理解拉格朗日乘子法（Lagrange multiplier）: yandoudoujiayou: 请问博主，在求解带有1范数和核范数的模型中使用增广拉格朗日函数，等式约束为矩阵Z3=Z1,Z3=Z2...

随机模拟的基本思想和常用采样方法: friskit: 博主你好，公式中的inf貌似应该换成int。\\int是积分，\\inf是下确界

$q(x|y) \cdot \pi(y) / [q(y|x) \cdot \pi(x)]$ 。即如果知道样本 x_i ，如何知道下一个样本 x_{i+1} 是什么呢？就是通过 $q(y|x_i)$ 得到一个建议样本 y ，然后根据 $\alpha(x_i, y)$ 决定 $x_{i+1}=y$ 还是 $x_{i+1}=x_i$ 。可以证明分布 $q(y|x) \cdot \alpha(x,y)$ 满足细致平衡，同时可以证明这样抽取得到的样本是分布 $\pi(x)$ 的样本。具体的步骤如下：

1. 给定一个起始样本 x_0 和一个建议分布 $q(y|x)$ ；
2. 对于第 i 个样本 x_i ，通过 $q(y|x_i)$ 得到一个建议样本 y ；计算比率 $\alpha(x_i, y) = \min\{1, q(x_i|y) \cdot \pi(y) / [q(y|x_i) \cdot \pi(x_i)]\}$;
3. 抽取一个均匀分布样本 $u_i \sim U(0,1)$ ，如果 $u_i \leq \alpha(x_i,y)$ ，则 $x_{i+1} = y$ ；否则 $x_{i+1} = x_i$ ；
4. 重复步骤2~3，直到抽取到想要的样本数量为止。

如果，建议分布 $q(y|x)$ 满足： $q(y|x) = q(x|y)$ ，即对称，这个时候比率 $\alpha(x, y) = \min\{1, \pi(y) / \pi(x)\}$ 就是1953年最原始的算法，后来hasting把这个算法扩展了，不要求建议分布式对称的，从而得到了上述的算法。然而这个算法有一个缺点，就是抽样的效率不高，有些样本会被舍弃掉。从而产生了Gibbs算法。

B). Gibbs采样算法

Gibbs算法，很简单，就是用条件分布的抽样来替代全概率分布的抽样。例如， $X=\{x_1,x_2,...x_n\}$ 满足对 $p(X)$ 进行抽样呢？如果我们知道它的条件分布 $p(x_1|X_{-1}),...,p(x_i|X_{-i}),....$ ，其中 X_{-i} 表示除了 x_i 以外的所有变量。如果这些条件分布都是很容易抽样的，那么我们就可以通过对条件分布的抽样来对全概率分布 $p(x_1,x_2,...,x_n)$ 进行抽样。

Gibbs采样算法的步骤：

1. 给定一个初始样本 $X_0=\{x_{10},x_{20},...,x_{n0}\}$
- 2.已知一个样本 $X_i=\{x_{1i},x_{2i},...,x_{ni}\}$ ，对于 $x_{1_{i+1}}$ 进行抽样， $x_{1_{i+1}} \sim p(x_1|X_{i,-1})$
3. 对于 $x_{2_{i+1}}$ 进行抽样， $x_{2_{i+1}} \sim p(x_2|x_{1_{i+1}}, x_{3i},...x_{ni})$
-
- 4.对于 $x_{n_{i+1}}$ 进行抽样， $x_{n_{i+1}} \sim p(x_n|x_{1_{i+1}}, x_{2_{i+1}},...x_{n-1_{i+1}})$
- 5.步骤2~4可以得到 X 的一个样本，然后重复步骤2~4可以不断地得到 X 的样本。

当然无论是metropolis-hasting算法还是gibbs算法，都有一个burn in的过程，所谓burn in的过程就是因为这个两个算法本身都是markov chain的算法，要达到稳定状态需要一定的步骤才能达到，所以需要有一个burn in过程，只有在达到平衡状态时候得到的样本才能是平衡状态时候的目标分布的样本，因此，在burn in过程中产生的样本都需要被舍弃。如何判断一个过程是否达到了平衡状态还没有一个成熟的方法来解决，目前常见的方法是看是否状态已经平稳（例如画一个图，如果在较长的过程中，变化已经不大，说明很有可能已经平衡）当然这个方法并不能肯定一个状态是否平衡，你可以举出反例，但是却是实际中没有办法的办法。

可以证明Gibbs算法是metropolis-hasting算法的一个特例,即比率 $\alpha(x,y) = 1$ 的一个特例。具体证明，此处略。

（pku，sewm，shinning）

顶

16

踩

0

关闭

相关文章推荐

- 机器学习方法(八)：随机采样方法整理（MCMC、...
- MC, MCMC, Gibbs采样 原理&实现（in R）
- 随机采样和随机模拟：吉布斯采样Gibbs Sampling
- word2vec基于负采样的模型原理介绍
- word2vec 中的数学原理详解（六）若干源码细节

- word2vec的负采样算法
- 深入理解拉格朗日乘法（Lagrange Multiplier）...
- 拉格朗日乘法
- 拉格朗日乘法
- 拉格朗日乘法




猜你在找

- 【直播】机器学习&数据挖掘7周实训--韦玮
 - 【直播】3小时掌握Docker最佳实战-徐西宁
 - 【直播】计算机视觉原理及实战--屈教授
 - 【直播】机器学习之矩阵--黄博士
 - 【直播】机器学习之凸优化--马博士
- 【套餐】系统集成项目管理工程师顺利通关--徐朋
 - 【套餐】机器学习系列套餐（算法+实战）--唐宇迪
 - 【套餐】微信订阅号+服务号Java版 v2.0--翟东平
 - 【套餐】微信订阅号+服务号Java版 v2.0--翟东平
 - 【套餐】Javascript 设计模式实战--曾亮


查看评论

16楼 [task_force](#) 2017-07-28 10:48发表




楼主，有一事请教，对g(x)分布进行采样得到样本值x_i后，和式为什么不是 $\sum_1^n [P(X=x_i)*f(x_i)/g(x_i)]$ ？

15楼 [task_force](#) 2017-07-28 10:44发表




楼主，有一点不明白希望请教一下，n个样本按照g(x)的分布进行采样，则积分为什么变为和式 $\sum_1^n [f(x_i)/g(x_i)]/n$ ，为什么不是 $\sum_1^n [f(x_i)/g(x_i)*P(X=x_i)]$ ？样本如果是按照g(x)采样，那么样本应该有个分布P(X=x_i)才对。

14楼 [friskit](#) 2017-05-19 09:53发表




博主你好，公式中的inf貌似应该换成int。 \int 是积分， \inf 是下确界

13楼 [pussy_foot](#) 2016-08-29 17:15发表




细致平稳条件是马氏链收敛的充分不必要条件吧。

12楼 [Jia_11](#) 2016-08-13 16:26发表




博主你好，我刚刚学习这个，对于采样方法的思想还没弄清楚。我想问一下：在拒绝采样中，对于任意z，如果我们可以判断接受的概率，相当于我们知道了p(z)的大致取值，那为什么还要采样呢？

11楼 [Solomon1588](#) 2016-05-28 22:51发表




产生一个服从U[0,1]分布随机数的用意是：将[0,1]区间划分为[0,a)和[a,1],若随机数落到[0,a)，就按照概率a接受这个状态转移。

10楼 [yangmei3668038](#) 2014-12-26 10:48发表




讲的很细，理解的很透彻了我。谢谢了。

9楼 [wyk4322129](#) 2014-03-14 10:14发表




您好，向您请教一下，在接受-拒绝抽样中，为什么要和均匀分布作比较，来决定是接受还是拒绝呢？

8楼 [Particlefilter](#) 2014-03-06 09:13发表



随机采样问题很重要，学习了


7楼 [billbargen](#) 2013-04-25 17:04发表




请问教一下，在接受-拒绝采样的讲解中：“它有几个条件：1）对于任何一个x，有 $f(x) \leq M * g(x)$ ；”里面M代表什么？

关闭


6楼 [kira423336903](#) 2013-03-25 20:15发表

 你好，关于Gibbs抽样的描述：第三步， 对于 $x_{2\{i+1\}}$ 进行抽样， $x_{2\{i+1\}} \sim p(x_2|x_{1\{i+1\}}, x_{2i}, \dots, x_{ni}), \dots$ 这个分布应该是 $x_{2\{i+1\}} \sim p(x_2|x_{1\{i+1\}}, x_{3i}, \dots, x_{ni})$ 吧？没接触过，不太理解。


Re: [xianlingmao](#) 2013-03-26 15:22发表

 回复kira423336903：你是对的，我已经改正了，谢谢！

5楼 [dendily](#) 2013-03-22 14:28发表


 很不错，通俗易懂，受教了

4楼 [duskwaitor](#) 2012-10-22 08:27发表


 引用“xianlingmao”的评论：
回复duskwaitor：思想是这样的： $f(x)/g(x)$ 看为一个整体函...

楼主回复的很详细啊，感谢楼主，祝福！


Re: [xfortius](#) 2013-03-09 09:58发表

 回复duskwaitor：这里 $g(x)$ 是均匀分布，这个蒙特卡洛模拟法这样就把不具随机性的事件可以通过事件来模拟或逼近。


3楼 [duskwaitor](#) 2012-10-21 11:18发表

 另外也没看懂第一节“随机模拟的基本思想”第五段“ $\frac{1}{n} \sum [f(x_i)/g(x_i)]$ ”是什么意思，我QQ25565853，楼主能加赐教一下吗，谢谢！


Re: [xianlingmao](#) 2012-10-21 13:59发表

 回复duskwaitor：思想是这样的： $f(x)/g(x)$ 看为一个整体函数 $F(x)$ ，即 $F(x) = f(x)/g(x)$ ， $g(x)$ 是其 x 的概率分布，那么从 $g(x)$ 中抽取 n 个样本， $F(x)$ 的期望就可以通过求在这 n 个样本上的平均值来近似，因为 n 个样本的出现是按照 $g(x)$ 的概率的，这就是 $\frac{1}{n} \sum [f(x_i)/g(x_i)]$ 的意思。或者你换个角度，假设 x 的取值有 m 个， n 个样本中对应这 m 个值得个数是 n_1, n_2, \dots, n_m ，这个时候求取 $F(x)$ 的期望，也就是平均值，可以写为： $E(F(x)) = [n_1 * F(x_1) + \dots + n_m * F(x_m)]/n$ ，这个式子与 $\frac{1}{n} \sum [f(x_i)/g(x_i)]$ 是等价的。qq就不加了，我很少用，谢谢！


2楼 [duskwaitor](#) 2012-10-21 11:03发表

 感谢楼主，另外 $\int_a^b f(x) dx$ 是什么意思？是 $\int_a^b f(x) dx$ 的意思么？


Re: [xianlingmao](#) 2012-10-21 13:50发表

 回复duskwaitor：是的 :)

1楼 [bneliao](#) 2012-07-24 01:57发表

 讲解深入浅出，通俗易懂。lz强悍
就是公式不能用数学符号表示，不是很直观


Re: [xianlingmao](#) 2012-07-24 13:44发表

 回复bneliao：主要是csdn不能输入公式，每个公式都要弄成图片放入，花费时间会较多，不过写为 $latex$ 形式，也不影响阅读。

发表评论

用户名：[haijunz](#)

评论内容：



提交

* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

关闭

关闭