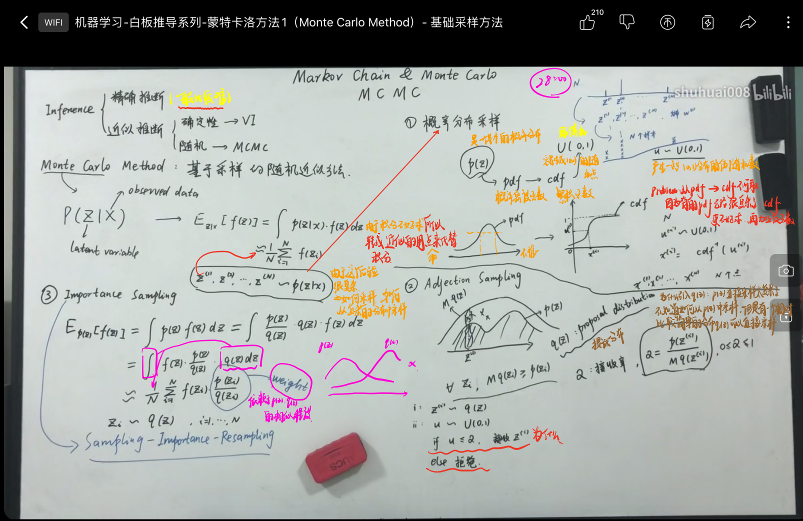
背景. Inference: x → 有隐变量 → 求后验 $p(z|x)$

$\Rightarrow p(z|x)$



Rejection sampling

Importance Sampling

Q: 为什么 $u \le \alpha$ 接收

三种采样的方法

1) 概率分布采样. 求出来 $p(z)$ 的 pdf 再求 cdf.

2) 拒绝采样. $Mq(z)$, $p(z)$, $q(z)$. $\alpha = \frac{p(z)}{Mq(z)}$ $p(z)\uparrow$ $\alpha\uparrow$

① $z^{(i)} \sim q(z)$. 取出 $z^{(i)}$ 这个样本

② 从 $u \sim U(0,1)$ 如果 $u \le \alpha$ 接收 $z^{(i)}$

else. reject.

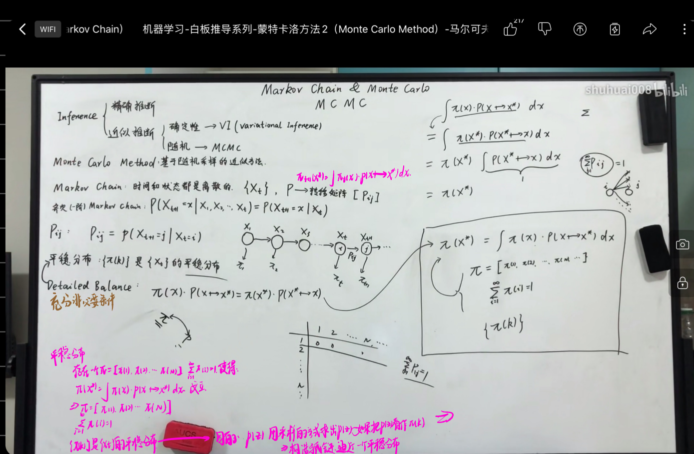
3) Importance sampling

$E_{p(z)}[f(z)] = \int p(z) f(z) dz = \int f(z) \cdot \frac{p(z)}{q(z)} q(z) dz$

$= \frac{1}{N}\sum_{i=1}^{N} f(z) \cdot \frac{p(z)}{q(z)}$ weight

找 $z_i \sim q(z)$ ⇒ E 最大

马氏链.



马氏链是一种随机过程.

要求: 满足马尔科夫性质: 给定当前时刻 未来和过去无关 一阶.

n阶

什么样的马氏链收敛到平稳分布

Detailed Balance.

$\pi(x) \cdot p(x \mapsto x^*) = \pi(x^*) \cdot p(x^* \mapsto x)$

x^* ⇄ x

DB. $\int \pi(x) \cdot p(x \mapsto x^*) dx$

$= \int \pi(x^*) \cdot p(x^* \mapsto x) dx$

$= \pi(x^*) \underbrace{\int p(x^* \mapsto x) dx}_{①}$

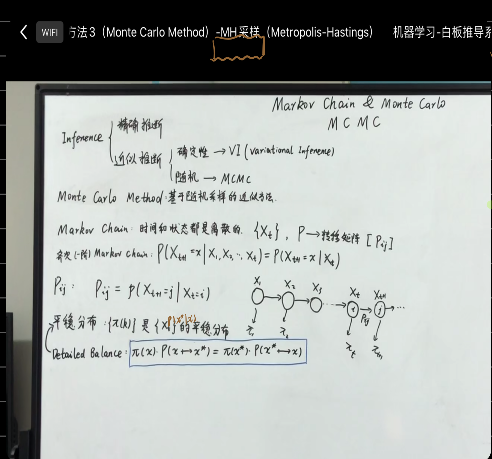
$= \pi(x^*)$

$\sum_{j=1} p_{ij}$ j 取到所有

所以 $\sum_{j=1} p_{ij} = 1$

⇒ 平稳分布.

引出 MCMC



关心的是后验概率 $p(z) \to E_{p(z)}[f(z)]$ $f(z)$ 只是一个函数.

$= \int f(z) p(z) dz$. $z^{(1)}, z^{(2)} \dots z^{(N)} \sim p(z)$. 无法直接得到 $p(z)$

$= \frac{1}{N}\sum_{i=1}^{N} f(z^{(i)})$

⇒ MCMC 来间接得到 N 个样本

把 $\pi(x)$ 看作是 $p(z)$ → 构造一个马氏链. 趋向于平稳分布

⇒ 找到一个满足条件的转移矩阵

前情

$p(z)$. 找一个随机的条件转移矩阵 Q $Q=[Q_{ij}]$ proposal matrix.

$p(z)$ $Q(z \mapsto z^*) \ne p(z^*) Q(z^* \mapsto z)$.

因是随机取的马氏链 ⇒ 找到令上式相等的条件转移矩阵.

随机找到个 Q 显然 Q 不能使 DB 条件成立 所以引入 α

$\underbrace{p(z) Q(z \mapsto z^*) \alpha(z, z^*)}_{p(z) \cdot p(z \mapsto z^*)} = \underbrace{p(z^*) Q(z^* \mapsto z) \alpha(z^*, z)}_{p(z^*) p(z^* \mapsto z)}$

α: 接受率 $\alpha(z, z^*) = \min(1, \frac{p(z^*) Q(z^* \mapsto z)}{p(z) Q(z \mapsto z^*)})$

$Q(z^* \mapsto z) \to Q(z|z^*)$, $Q(z \mapsto z^*) \to Q(z^*|z)$

左边: $p(z) Q(z \mapsto z^*) \alpha(z, z^*)$

$= p(z) Q(z \mapsto z^*) \cdot \min(1, \frac{p(z^*) Q(z^* \mapsto z)}{p(z) Q(z \mapsto z^*)})$

$= \min(p(z) Q(z \mapsto z^*), p(z^*) \cdot Q(z^* \mapsto z))$

这样设置接收率则满足 detailed balance 条件

设 $\alpha = \min(1, \frac{p(z^*) Q(z^* \mapsto z)}{p(z) Q(z \mapsto z^*)})$. 在上式的求取中 $p(z^*)$ 和 $p(z)$ 是如何求得的

⇒ $p(z^*)$ 和 $p(z)$ 直接代入求出来 $p(z)$ 了 实际上这里面有一些省略的

$p(z) = \frac{\hat{p}(z)}{z_p}$ z_p 是归一化因子 未知的 求不出也不用求.

在求 $p^*(z)$ 时 实际上 用的是 $\hat{p}(z)$ 可以求出来的

Metropolis-Hastings:

$u \sim U(0,1)$

$z^* \sim Q(z|z^{(i-1)})$

$\alpha = \min(1, \frac{p(z^*) Q(z^* \mapsto z)}{p(z) Q(z \mapsto z^*)})$ α 的求法有疑问

if $u \le \alpha$. $z^{(i)} = z^*$. else $z^{(i)} = z^{(i-1)}$ 只是会重复

N 次 $z^{(1)}, z^{(2)} \dots z^{(N)}$

⇒ $E_{p(z)}[f(z)] \approx \frac{1}{N}\sum_{i=1}^{N} f(z^{(i)})$

$= p(z^*)Q(z^* \mapsto z)\min(1, \frac{p(z)Q(z \mapsto z^*)}{p(z^*)Q(z^* \mapsto z)})$

$= p(z^*)Q(z^* \mapsto z) \cdot \alpha(z^*, z)$

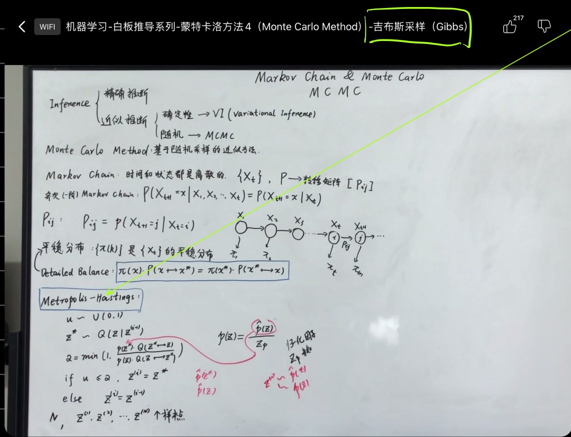
MH算法

likelyhood * prior. $\boxed{\text{似然乘以先验一定可以求}}$???

所以上面的p(z)和p(z)实际上是$\hat{p}(z)$和$\hat{p}(z)$

$z^{(i)} \sim \hat{p}(z)$

$\sim p(z)$



Gibbs $p(z) = p(z_1, z_2 \cdots z_m)$. 固定其中的一维

z_i维 $\sim p(z_i | z_{-i})$. z有3维

$p(z) = p(z_1, z_2, z_3)$ 了例子

i: $z_1^{(0)}, z_2^{(0)}, z_3^{(0)}$ 初始值　在MH中Q是和p不同的

ii) $z_1^{(t+1)} \sim p(z_1 | z_2^{(t)}, z_3^{(t)})$　但是在这里Q→p来代替

$z_2^{(t+1)} \sim p(z_2 | z_1^{(t+1)}, z_3^{(t)})$　不用证明 满足DB条件. 只要证明Gibbs采样是MH就可以了.

$z_3^{(t+1)} \sim p(z_3 | z_1^{(t+1)}, z_2^{(t+1)})$

所以看一下α针对Gibbs是什么　证明Gibbs和MH有关

$p(z^*)Q(z^* \mapsto z)$

$p(z)Q(z \mapsto z^*)$

结果 条件

表示时还用$p(z_i | z^*_{-i})$

了Gibbs采样都代表了什么

$= \frac{p(z_i^* | z_{-i}^*)\, p(z_{-i}^*)\, p(z_i | z_{-i}^*)}{p(z_i | z_{-i})\, p(z_{-i})\, p(z_i^* | z_{-i})}$

因为这里有没有变化了

z_{-i}和z^*_{-i}是一样的东西. z_{-i}和z^*_{-i}表示的剩余分量是一模一样的

$= \frac{p(z_i^* | z_{-i}^*)\, p(z_{-i}^*)\, p(z_i | z_{-i}^*)}{p(z_i | z_{-i}^*)\, p(z_{-i})\, p(z_i^* | z_{-i})} = 1$

$\Rightarrow \alpha = 1$.　Gibbs采样是特殊的MH效率高

根据Gibbs的规定和定义 $\Rightarrow \alpha = 1$.

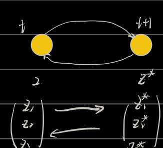
第五节: Gibbs采样即对上节内容的补充

Gibbs采样是特殊的MH采样. ⟺ 接受率为1. 一维一维的采样

假设: $z = (z_1, z_2, z_3)$, $p(z) = p(z_1, z_2, z_3)$

$\alpha = \min(1, \frac{p(z^*)Q(z^* \mapsto z)}{p(z)Q(z \mapsto z^*)})$

$= \frac{p(z_i^* | z_{-i}^*)\, p(z_{-i}^*) \cdot p(z_i | z_{-i}^*)}{p(z_i | z_{-i})\, p(z_{-i})\, p(z_i^* | z_{-i})} = 1$



MH: Q分布 Q是转移概率分布

Gibbs. $p(z_i | z_{-i})$ 把转移矩阵用自己的概率密度分布去表示.

这个要记住

z　t=1. $(z_1^{(1)}, z_2^{(1)}, z_3^{(1)})$

z^*　t=2. $(z_1^{(1)}, z_2^{(1)}, z_3^{(1)})$

t=1 → t=2 只有z_i变 其余都不变

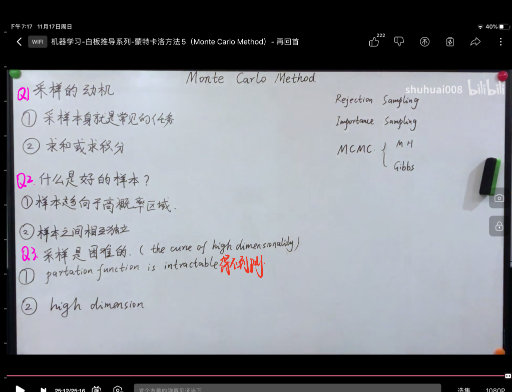
$z_{-i} = z^*_{-i}$

假设现在采第i维. $Q(z \to z^*)$ Gibbs

$= Q(z_i, z_{-i} \to z_i^*, z_{-i}^*)$

$= p(z_i \to z_i^*)$. $= p(z_i | z^*_{-i})$ 固定除i维以外的所有去求第i维

不用书写的Q. 只有原来p这个概率分布 只是固定其中的一维



①回顾MC方法

②根据以前的查缺补漏

Q1: A1: 在机器学习中会有已知的模型 p(x)

有的时候: 从概率模型中产生新的样本. ⇒采样

例如 p(x)是一种生成模型表示的是一堆图片

去生成新的图片 ⇒如果模型学好了之后 ⇒采出一些新的样本

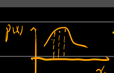
A2: 求和. 求积分问题的简化

在前面的推断学习的过程中会有求不出的积分. 求和

⇒采样方式. 特别是MCMC

$\int p(x) f(x) dx = \frac{1}{N}\sum f(x^{(i)})$. 先在p中采样得到$x^{(i)}$

Q2: A1: 例p(x)中采样 采了$x^{(1)} \dots x^{(N)}$的样本



N=100时. 大部分应该集中在高概率区域

低概率区样本少

人大代表多找的应是大众人民 不是富豪.

A): 样本之间相互独立. 十个人 不在同一个厂 不同一个职业.

相关性太强可以视作是失败了.

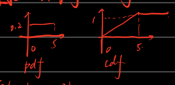
Q): 实操.

① 无向图模型 $p(x)$很复杂.

↓ 令有一个归一化因子. $p(x)=\frac{1}{Z}\hat{p}(x)$ 可求. Z不可求 $Z=\int\hat{p}(x)dx$

$x\in\mathbb{R}^p$ 无法求出来.

① 直接采样 pdf → cdf



很复杂时做不到

② 高维. $x\in\{1,2,\dots,k\}^p$ p个分量. 共有k^p的状态空间. 指数级别 ⇒ intractable.

要通过才可以知道高概率低概率 ⇒ 先有再采样前面

接采样不行

P7. p8没有看