《白板》(13) 采样方法 mcmc 介绍

Inference $\begin{cases} \text{精确推断} \\ \text{近似推断} \begin{cases} \text{确定性} \to VI \\ \text{随机} \to mcmc \end{cases} \end{cases}$

Monte Carlo Method：基于采样的随机近似方法
蒙特卡罗方法

Inference 推断主要任务 求 $p(z|x) \longrightarrow E_{z|x}[f(z)]$

积分并不好求，所以用数值积分 $= \int_z p(z|x) \cdot f(z) dz$

采样 近似 $\approx \frac{1}{N}\sum_{i=1}^{N} f(z_i)$

$z^{(1)}, z^{(2)} \dots z^{(N)} \sim p(z|x)$

如何从复杂分布里面去采样？

③ 重要性采样 Important sample.

不直接对 $p(z)$ 采样，对 $p(z)$ 下的期望采样

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

$$= \int f(z) \cdot \frac{p(z)}{q(z)} \cdot q(z) \cdot dz \simeq \frac{1}{N}\sum_{i=1}^{N} f(z_i) \cdot \boxed{\frac{p(z_i)}{q(z_i)}}$$

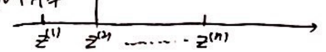
$z_i \sim q(z)$，$i = 1 \cdots\cdots N$

Weight
加权，权重值

重要性采样变种

Sampling - Importance - Resampling

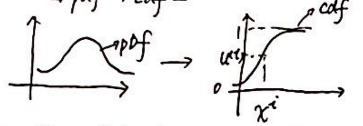
N个样本



采样权重有高有低，那么每个样本点有 weight
再去采样 N个样本点 ⟶ 分两个阶段 ⟹ 注重思想
内容较为复杂

① 概率分布采样

$p(z)$ ↳ pdf → cdf → $u(0,1)$ 均匀分布 $u \sim U(0,1)$



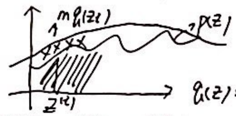
从计算机很容易生成 $u \sim U(0,1)$，通过其生成 x^{i}

$x^{(i)} = cdf^{-1}(u^{(i)})$ ↳ 反函数

采样得到 n个点 $x^{(1)} \dots\dots x^{(n)}$

问题：由 pdf 求 cdf，很可能求不出来，∵ pdf 可能很复杂

② ~~Adjection~~ Rejection Sampling，拒绝采样



$q(z)$：proposal distribution
提议分布

$\forall z_i$，$mq(z_i) \geqslant p(z_i)$
↓
对于任意 z_i，满足，m是常数

前提，$p(z)$ 难采样，有一个分布 $q(z)$，好采样

采样运用接收率 α [落在阴影接收，××处拒绝]

$$\alpha = \frac{p(z^{(i)})}{mq(z^{(i)})}, \quad 0 \leq \alpha \leq 1$$

算法流程
$z^{(i)} \sim q(z)$
$u \sim U(0,1)$
if $u \leq \alpha$，接收 $z^{(i)}$
else 拒绝

为什么乘以m，因为不知道 $p(z)$ 是什么样，仅为了保证 $q(z)$ 大于 $p(z)$，乘以常数 m

白板(13)　　Markov Chain and Monte Carlo

MCMC

Monte Carlo Method：基于随机采样的近似方法　　pdf → cdf（递增且值域在0~1之间）然后在纵轴采样，……参考上张纸

Markov Chain：时间、状态都是离散的.

齐次（一阶）Markov Chain：

↳ $p(x_{t+1}=x \mid x_1 x_2 \cdots x_t) = p(x_{t+1} \mid x_t)$

P → 状态转移矩阵 $[p_{ij}]$

$p_{ij} = p(x_{t+1}=j \mid x_t = i)$

$$\begin{array}{ccccccccc} x_1 & \to & x_2 & \to & x_3 & \cdots\to & x_t\ (i) & \to & x_{t+1}\ (j) \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \pi_1 & & \pi_2 & & \pi_3 & & \pi_t & & \pi_{t+1} \end{array}$$

马氏链

$$\underset{\pi_t(x)}{x} \longrightarrow \underset{t+1}{x^*}$$

表达 → $\pi_{t+1}(x^*) = \int \underbrace{\pi_t(x) \cdot p(x \to x^*)}_{} dx$

$p(x \to x^*)$ ↓ 转移概率

▲ ①

平稳分布，假设存在 $\begin{cases} \pi = [\pi_1, \pi_2 \cdots \pi_n] \\ \sum_{i=1}^{\infty} \pi(i) = 1 \end{cases}$ 使得

$\pi(x^*) = \int \pi(x) \cdot p(x \to x^*) \cdot dx$，那么

$\{\pi(k)\}$ 是 $\{x_t\}$ 的平稳分布. ⇒ 即 π_k 同一分布？

▲ ②

Detail Balance（细致平衡）↓ 来回概率一样

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

若满足②，则推出①平稳分布

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

若一个分布满足②，则一定满足①，
…………①，不一定……②.

▲ 把 π 与 P 联系起来

π ↑ 平稳分布

P ↗ 马氏链的转移矩阵

自报(13)

MCMC - Metropolis-Hastings算法

任务推断后验概率 $p(z) \to E_{p(z)}[f(z)] = \int_z f(z)p(z)dz$

$z^{(1)}, z^{(2)} \cdots z^{(N)} \sim p(z) \Rightarrow$ 通过采样 $\approx \frac{1}{N}\sum_{i=1}^{N} f(z^{(i)})$

由于直接通过 $p(z)$ 采样困难，所以提出 MCMC，构造马氏链，$p(z)$ 看作 $\pi(x)$，让其平稳分布

那么问题变成 $x \to x^*$，$x^* \to x$ 的转移矩阵是什么，才可以满足平稳分布的不等式.

P_{ij} 怎么求. ▲

从马氏链任意取一个 $Q=[Q_{ij}]$，那么 $p(z)\cdot Q(z\to z^*) \neq p(z^*)\cdot Q(z^*\to z)$

定义一个接收率 $\begin{cases}\alpha(z, z^*), \\ \alpha(z^*, z),\end{cases}$ 使得 $p(z)\cdot \underbrace{Q(z\to z^*)\cdot \alpha(z,z^*)}_{P(z\to z^*)} = p(z^*)\cdot \underbrace{Q(z^*\to z)\cdot \alpha(z^*, z)}_{P(z^*\to z)}$

那么 $\alpha(z, z^*) = \min\left(1, \frac{p(z^*)\cdot Q(z^*\to z)}{p(z)\cdot Q(z\to z^*)}\right)$ —— ①

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

$\hookrightarrow Q(z^*|z)$

验证①式，

$$p(z)\cdot Q(z\to z^*)\cdot \alpha(z, z^*) = p(z)\cdot Q(z\to z^*)\cdot \overset{①式}{\min(\cdot\cdot, \cdot\cdot)} = \min\left(p(z)\cdot Q(z\to z^*), p(z^*)\cdot Q(z^*\to z)\right)$$
$$= p(z^*)\cdot Q(z^*\to z)\cdot \underbrace{\min\left(1, \frac{p(z)Q(z\to z^*)}{p(z^*)\cdot Q(z^*\to z)}\right)}_{\alpha(z^*, z)}$$

∴ 只要满足①式，则可以 Detail Balance $\pi(x)\cdot p(x\to x^*) = \pi(x^*)\cdot p(x^*\to x)$

以上就是 Metropolis-Hastings算法 $u \sim U(0,1)$ 均匀分布 $z^* \sim Q(z|z^{(t-1)})$ ↗上一个样本

$\alpha = \min\left(1, \frac{p(z^*)\cdot Q(z^*\to z)}{p(z)\cdot Q(z\to z^*)}\right)$, if $u \le \alpha$, $z^{(t)} = z^*$ (直接接收). else $z^{(t)} = z^{(t-1)}$, (取上一个样本)

MCMC → 吉布斯采样

本质上还是MH采样. 特殊的MH采样。高维度采样，一维一维采(固定其它维度) $z_i \sim p(z_i|z_{-i})$

$\hookrightarrow z_1 z_2 z_3 \cdots$

假定3维. $p(z) = p(z_1, z_2, z_3)$. $z_1^{(0)}, z_2^{(0)}, z_3^{(0)}$，那么 $t+1$，$z_1^{(t+1)} \sim p(z_1|z_2^{(t)}, z_3^{(t)})$

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

↳采样过的 ↳采样过的

验证Gibbs是MH一种

$$\frac{p(z^*)\cdot Q(z^*\to z)}{p(z)\cdot Q(z\to z^*)} = \frac{p(z_i^*|z_{-i}^*)\cdot p(z_{-i}^*)\cdot p(z_i|z_{-i}^*)}{p(z_i|z_{-i})\cdot p(z_{-i})\cdot p(z_i^*|z_{-i})}$$

→ 不论是 z_i 还是 z_i^*，其采样后余下分量一样

∴ $z_{-i}^* = z_{-i}$ △???

框内相等 $= 1$

∴ $\alpha = 1$，是特殊的MH算法

白板(13)

MC MC - 回顾

Monte Carlo Method

正

一.
采样动机(为什么要采样) ①采样本身就是常见的任务 ②求积分或求和，且求和/积分有时很难，一般采用mcmc方法采样近似。

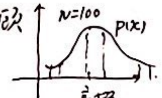
$\int p(x)\cdot f(x)dx = E_{p(x)}[f(x)]$

$x^{(1)}, x^{(2)}, \cdots x^{(N)} \sim p(x)$ 采样.

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

二.什么样的样本才是好的样本？

①样本趋向于高概率区域.兼顾其它区域



②样本之间相互独立，∵相关性强，代表不了全部样本的分布.

$x^{(1)}, x^{(2)} \cdots x^{(N)}$ 样本

三.直接采样是困难的.

① partation function is intractable $p(x)=\frac{1}{Z}\hat{p}(x)$

↳ 归一化因子不可求；高维时

② high dimension 不可行

所以才提出 Rejection Sampling / Importance Sampling. 引入 $q(x)$，绕过 $p(x)$。

mcmc → 马氏链近似 $\begin{cases} MH \\ Gibbs. \end{cases}$

Rejection Sampling / Importance Sampling } $q(x)$与$p(x)$接近且$q(x)$简单易采样

mcmc $\begin{cases} MH \\ Gibbs \end{cases}$

mcmc原理：构建马氏链(性质：经过若干步骤，会收敛到某平稳分布) 证明平稳分布：

状态空间：$\{1,2,\cdots k\}$ 在m之后，是平稳分布

$$\overset{1}{\bigcirc}\to\overset{2}{\bigcirc}\to\cdots\to\overset{t}{\bigcirc}\to\overset{t+1}{\bigcirc}\to\cdots\to\overset{m}{\bigcirc}\to\overset{m+1}{\bigcirc}\to\cdots$$

$q^{(1)}(x)\quad q^{(2)}(x)\quad q^{(t)}(x)\quad q^{(t+1)}(x)\quad q^{(m)}(x)\quad q^{(m+1)}(x)$

$q^{(1)}(x)$

x	$1, 2, 3, \cdots k$
$q^{(1)}(x)$	$q_1^1\ q_1^2\ q_1^3 \cdots q_1^k$

$\sum_{i=1}^{N} q_1^i = 1$

m步之前，其概率分布不同.

m之后 $p\ q^{(m)}(x) = q^{(m+1)}(x) = q^{(m+2)}(x) \cdots$ =相同

其值为 $\frac{1}{k}$

x	$1, 2, \cdots k$
$q^{(m)}(x)$	$\frac{1}{k}\ \frac{1}{k} \cdots \frac{1}{k}$

(均匀分布)

最理想情况，收敛后，其概率分布逼近 $p(x)$

状态转移矩阵：(随机矩阵)

$$Q=\begin{pmatrix} Q_{11} & Q_{12} & \cdots & Q_{1k} \\ Q_{21} & Q_{22} & \cdots & Q_{2k} \\ \vdots & \vdots & & \vdots \\ Q_{k1} & Q_{k2} & \cdots & Q_{kk} \end{pmatrix}_{k\times k}$$

行和=1
列和不一定=1

$$q^{(t+1)}(x=j) = \sum_{i=1}^{k} q^{(t)}(x=i)\, Q_{ij} \quad \text{—①}$$

要求和

令 $q^{(t+1)} = (q^{(t+1)}(x=1)\ \ q^{(t+1)}(x=2) \cdots q^{(t+1)}(x=k))_{1\times k}$

而改写①式.

$\because q^{(t+1)} = \left(\sum_{i=1}^{k} q^{(t)}(x=i)Q_{i1}\ \ \sum_{i=1}^{k} q^{(t)}(x=i)Q_{i2} \cdots \sum_{i=1}^{k} q^{(t)}(x=i)Q_{ik}\right)$

$= q^{(t)} \cdot [Q\ Q\ Q]$ (每一列) $= q^{(t)}\cdot Q$

↳ $q^{(t)} = (q^{(t)}(x=1)\ \ q^{(t)}(x=2) \cdots q^{(t)}(x=k))$ 向一向量

递推 $= q^{(1)}\cdot Q^{t}$ ▲ —②

∵随机矩阵：特征值的绝对值都≤1

$\therefore Q = A\cdot\Lambda\cdot A^{-1}$ [特征分解]

其中 $\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \lambda_3 & \\ & & & \ddots \ \lambda_k \end{pmatrix}$，$|\lambda_i|\le 1$, for $i=1\cdots k$

不妨假设只有一个 $\lambda_i=1$，其它 $\lambda_i<1$

则②式 $= q^{(1)}\cdot(A\Lambda A^{-1})^{t}$ $\underline{\underline{A与A^{-1}约掉}}$ $q^{(1)}\cdot A\Lambda^{t}A^{-1}$

存在足够大的 m，s.t. $\Lambda^{m} = \begin{pmatrix} 0 & & \\ & 1 & \\ & & 0 \end{pmatrix}$

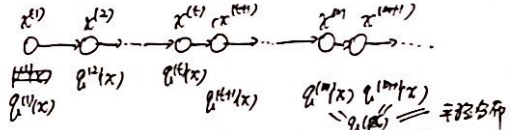
$q^{(m+1)} = q^{(1)}\cdot A\cdot\Lambda^{m}A^{-1}$，$q^{(m+2)} = q^{(1)}\cdot \cancel{A}\cdot\Lambda^{m}\cdot\cancel{A^{-1}}\cdot \cancel{A}\cdot\Lambda \cancel{A^{-1}} = q^{(1)}\cdot\Lambda^{m+1}$

又∵ $\Lambda^{m} = \begin{pmatrix} 0 & & \\ & 1 & \\ & & 0 \end{pmatrix}$ ∴ $\Lambda^{m+1} = \Lambda^{m} \Rightarrow q^{(m+2)} = q^{(m+1)}$

当 $t>m$ 时，$q^{(m+1)} = q^{(m+2)} = \cdots = q^{(m+t)} \Rightarrow$ 平稳分布了

白板（13）　　MCMC — 采样思路及面临困难　　反

状态空间：{1, 2, …, k}
状态转移矩阵：$Q=[Q_{ij}]_{k\cdot k}$　} 参数



假定进入 m 步后，进入平稳分布．

MCMC 是一套方法论．具体方法很多．最基本的是 MH, Gibbs

MCMC：利用 Markov Chain 收敛于平稳分布，设计 Q，使得平稳分布 ≈ 目标分布 $p(x)$，任意分布，包括后验 $p(z|x)$.

拒绝采样和重要性采样在高维空间，很难找到 $q(x)\approx p(x)$.

初始状态 $q^{(1)}(x)$ 可从状态空间任取，只要收敛平稳分布与 Q（转移矩阵）有关，与初始状态 $q^{(1)}(x)$ 无关

从 1 ~ m 步，称之为 burn-in（燃烧期），所花费时间 mixing time.

专注设计 Q，而非在高维空间采样

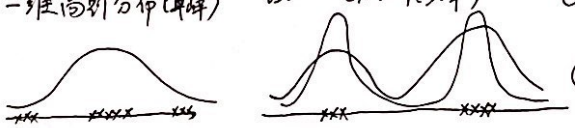
MCMC 问题：① 无法判断收敛于平稳分布的时间，理论只保证了收敛性。
↳ 只能看 xx 步后是否收敛
xx+1 步 … ——

② 混合时间 mixing time 过长 / 不混合 / 不收敛（发散）→ $p(x)$ 太复杂，高维造成．且维度之间相关性强

③ 好的采样：要相互独立．但是 m 与 m+1 一定相关（马尔可夫链）．实际上在进入平稳分布后，相隔 xx 步采样．

~~④ 最大问题：处理多维（Gibbs）~~

一维高斯分布（单峰）　vs.　GMM（多峰）
缓和 vs. 陡峭



① 多峰会卡在一个峰值上，无法越过低概率区域．

② 多峰陡峭采样难度大．混合失败．

有很多 MCMC 会解决上述问题．[高维．陡峭]