Lecture 13 - Markov chain Monte Carlo

赵尉辰

南开大学 统计与数据科学学院

目录

1. 背景：Monte Carlo 方法与采样问题
2. Markov Chain Monte Carlo
   - Gibbs Sampling
   - Metropolis-Hastings 算法
3. 模拟退火

目录

近似计算积分

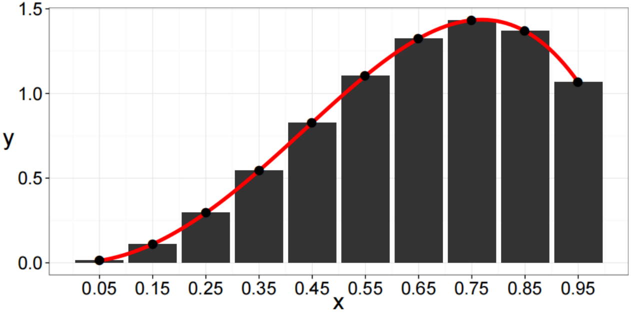
考虑一个积分计算的问题，对于$f : S \to \mathbb{R}$，定义

$$I = \int_S f(x)\, dx.$$

近似计算积分

若$S=[0,1]$，则可以近似计算I：

$$\widehat{I}_n=\frac{1}{n}\sum_{i=0}^{n-1}f\left(\frac{i+1/2}{n}\right).$$



如果 $\sup_{x\in[0,1]}|f'(x)|<M<\infty$，则近似误差为 $\mathcal{O}\left(n^{-1}\right)$.

近似计算积分

进一步，若$S = [0,1] \times [0,1]$，则可以近似计算I：

$$\widehat{I}_n = \frac{1}{m^2} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} f\left(\frac{i+1/2}{m}, \frac{j+1/2}{m}\right)$$

此时，$n = m^2$，近似误差为 $\mathcal{O}\left(n^{-1/2}\right)$. 更一般地，对于$S = [0,1]^d$，近似误差为 $\mathcal{O}\left(n^{-1/d}\right)$.

这说明随着维度的增长，近似计算积分越来越困难，这经常被称为“维度诅咒 (curse of dimensionality)”。

然而在统计物理、机器学习等领域中，这样高维积分计算问题非常普遍。

贝叶斯推断

推断任务

在机器学习中，所有未知的量，无论是对未来的预测，系统的隐藏状态，还是模型的参数，都被视为随机变量，并被赋予概率分布。推断任务是指根据已知数据，计算这些随机变量的后验分布。

设ϕ是未知变量，$\mathcal{D}$是已知变量，给定先验$p(\phi)$和似然$p(\mathcal{D}|\phi)$，我们可以通过贝叶斯定理计算后验：

$$p(\phi|\mathcal{D}) = \frac{p(\phi)p(\mathcal{D}|\phi)}{p(\mathcal{D})}$$

计算的瓶颈在于归一化系数

$$p(\mathcal{D}) = \int p(\mathcal{D}|\phi)p(\phi)d\phi$$

尤其是在高维情形。

Monte Carlo 方法

对于$f : S \to \mathbb{R}$，积分可以改写为

$$I = \int_S f(x)\,dx = \int_S \varphi(x)\pi(x)dx = \mathbb{E}_\pi[\varphi].$$

其中π是S上的概率分布，$\varphi : x \mapsto f(x)/\pi(x)$.

Monte Carlo 方法

- 得到n个独立的服从π的样本$X_1, X_2, \ldots, X_n$
- 计算

$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i)$$

Monte Carlo 方法的近似误差

$$
\begin{aligned}
(I-\widehat{I}_n)^2 &= I^2 - 2I\widehat{I}_n + \widehat{I}_n^2 \\
&= I^2 - \frac{2I}{n}\sum_{i=1}^n \varphi(X_i) + \frac{1}{n^2}\sum_{i=1}^n \varphi(X_i)^2 + \frac{1}{n^2}\sum_{i\neq j}\varphi(X_i)\varphi(X_j).
\end{aligned}
$$

由于X_i i.i.d.且 $I = \mathbb{E}_\pi[\varphi(X)]$, 那么

$$
\begin{aligned}
\mathbb{E}_\pi[(I-\widehat{I}_n)^2] &= I^2 - 2I^2 + \frac{1}{n}\mathbb{E}_\pi[\varphi(X_1)^2] + \frac{1}{n^2}n(n-1)I^2 \\
&= \frac{\mathbb{E}_\pi[\varphi(X_1)^2] - I^2}{n} = \frac{\mathbb{V}_\pi(\varphi(X_1))}{n}
\end{aligned}
$$

于是如果 test function 满足$|\varphi(x)| \le 1,\ \forall x$, 那么

$$
\sqrt{\mathbb{E}_\pi[(I-\widehat{I}_n)^2]} = \frac{\sqrt{\mathrm{Var}_\pi(\varphi(X_1))}}{\sqrt{n}} \le \frac{1}{\sqrt{n}}
$$

这说明 Monte Carlo 方法的近似误差是维度无关的!

集中不等式[1][2][3]

Michel Talagrand awarded the 2024 Abel Prize

«for his groundbreaking contributions to probability theory and functional analysis, with outstanding applications in mathematical physics and statistics.»

---

[1] Wainwright M J. High-dimensional statistics: A non-asymptotic viewpoint[M]. Cambridge university press, 2019.

[2] Dubhashi D P, Panconesi A. Concentration of measure for the analysis of randomized algorithms[M]. Cambridge University Press, 2009.

[3] Boucheron, Stéphane, Gábor Lugosi, and Pascal Massart, Concentration Inequalities: A Nonasymptotic Theory of Independence, Oxford Academic, 2013.

机器学习

监督学习的目标

学习一个映射函数 $f : \mathcal{X} \to \mathcal{Y}$，使得预测结果尽可能准确。

- 损失函数 $\ell(f(x), y)$ 衡量预测值与真实值的差异
- 常用损失函数：平方损失、交叉熵、0-1损失等

机器学习

机器学习主流是概率模型[4][5]，一般认为数据是随机变量，服从一定的概率分布。

定义 1 (期望风险)

期望风险 *(Expected Risk)* 定义为

$$R_{\exp}(f) = \mathbb{E}_{(x,y)\sim p(x,y)}\left[\ell(f(x), y)\right]$$

- 反映模型在 全体数据分布 上的表现
- 理论上的理想优化目标
- 关键问题：真实数据分布 $p(x,y)$ 未知!

[4] Murphy K P. Probabilistic machine learning: an introduction[M]. MIT press, 2022.

[5] Murphy K P. Probabilistic machine learning: Advanced topics[M]. MIT press, 2023.

经验风险

定义 2 (经验风险)

经验风险 *(Empirical Risk)* 定义为

$$R_{\mathrm{emp}}(f) = \frac{1}{n}\sum_{i=1}^{n} \ell(f(x_i), y_i)$$

- 基于 训练数据集 $\{(x_i, y_i)\}_{i=1}^{n}$ 的近似
- 实际优化目标

Monte Carlo 方法[6]

Monte Carlo 方法近似 $\mathbb{E}_\pi[\varphi(X)] \Leftrightarrow$ simulation method to sample π

[6] Liu Jun. Monte Carlo strategies in scientific computing[M]. New York: springer, 2001.

采样

问题 1 (采样)

设 π 是一个概率分布，采样问题*(Sampling)*是指：如何获得随机样本x，使得x的分布为π。

问题 2 (采样)

给定光滑函数$V : \mathbb{R}^d \to \mathbb{R}$, 如何获得$\mathbb{R}^d$上的随机样本服从概率分布

$$\pi = \frac{1}{Z}e^{-V}, \qquad \pi \propto e^{-V}$$

我们一般称V为位势函数*(potential function)*

采样

问题 1 (采样)

设 π 是一个概率分布，采样问题*(Sampling)*是指：如何获得随机样本x，使得x的分布为π。

问题 2 (采样)

给定光滑函数$V:\mathbb{R}^d\to\mathbb{R}$, 如何获得$\mathbb{R}^d$上的随机样本服从概率分布

$$\pi=\frac{1}{Z}e^{-V},\qquad \pi\propto e^{-V}$$

我们一般称V为位势函数*(potential function)*

生成

生成任务

在机器学习中，**生成任务** 是指模型的目标为生成新的数据实例，这些实例与已有数据(训练数据)具有相似的特征或模式，常见的生成任务包括文本生成、图像生成、视频生成等。

处理生成任务通常包含两个部分：

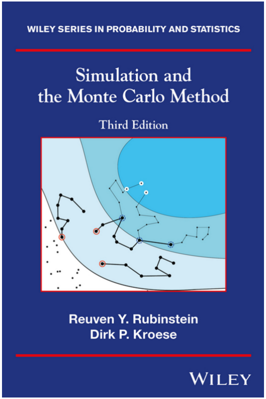
(1) 学习数据分布

学习通常利用深度神经网络实现，神经网络是一个参数化的模型，需要通过数学优化更新模型参数。

(2) 根据数据分布生成实例

生成新的数据样本即就是从数据分布中**采样**，需要通过采样算法实现。

随机模拟



- 随机数/基本随机变量/随机过程/随机向量的模拟
- Rejection Sampling
- Importance Sampling and Variance Reduction Methods

模拟高斯分布的 Galton's machine

目录

Markov Chain Monte Carlo (MCMC)

定义 3 (Markov链)

一个*Markov链*是一个随机过程 $\{X_n\}_{n=0}^{\infty}$，满足：未来状态只依赖于当前状态，而与过去状态无关。即，对于任意的 n 和状态 i, j，有：

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = P(X_{n+1} = j \mid X_n = i)$$

进一步，如果转移概率不随时间变化，即：

$$P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i) \quad \forall n$$

那么我们称*Markov*链是时齐的。

Markov Chain Monte Carlo (MCMC)

定义 4 (不变分布)

设 $\{X_n\}_{n=0}^{\infty}$ 是一个*Markov*链，其状态空间为 S。如果存在一个概率分布 π 满足以下条件：

$$\pi(j) = \sum_{i \in S} \pi(i) P_{ij} \quad \forall j \in S$$

其中 P_{ij} 是从状态 i 转移到状态 j 的概率，则称 π 为该*Markov*链的不变分布*(Invariant Distribution)*。

Markov Chain Monte Carlo (MCMC)

MCMC的思想即是：构造一个Markov链，使得它的平稳分布是我们采样的目标分布 π。那么从任意状态分布(容易获得样本的分布)出发，经过充分的状态转移，就能获得目标分布π的样本。

给定目标分布π，构造Markov链X_n，使得，$n \to \infty$，$X_n \sim \pi$，那么随着$N \to \infty$

$$\frac{1}{N}\sum_{n=1}^{N}\varphi\left(X_n\right) \to \int \varphi\left(x\right)\pi\left(x\right)dx$$

Gibbs Sampling

设目标分布为

$$\pi(x) = \pi(x_1, x_2, ..., x_d).$$

记$x_{-i} := (x_1, ..., x_{i-1}, x_{i+1}, ..., x_d)$.

Systematic scan Gibbs sampler

1. 选择初始状态 $X^{(0)} = (X_1^{(0)}, ..., X_d^{(0)})$。
2. 对于每一步 $t = 1, 2, \ldots, N$:
   - 从提议分布 $\pi_{X_1|X_{-1}}\left(\cdot|X_2^{(t-1)}, ..., X_d^{(t-1)}\right)$ 中生成候选状态 $X_1^{(t)}$。
   - ...
   - 采样$X_j^{(t)} \sim \pi_{X_j|X_{-j}}\left(\cdot|X_1^{(t)}, ..., X_{j-1}^{(t)}, X_{j+1}^{(t-1)}, ..., X_d^{(t-1)}\right)$
   - ...
   - 采样$X_d^{(t)} \sim \pi_{X_d|X_{-d}}\left(\cdot|X_1^{(t)}, ..., X_{d-1}^{(t)}\right)$

Gibbs Sampling

Random scan Gibbs sampler

1. 选择初始状态 $X^{(0)} = (X_1^{(0)}, ..., X_d^{(0)})$。
2. 对于每一步 $t = 1, 2, \ldots, N$:
   - 从维度指标集中$\{1, 2, \ldots, d\}$采样指标J，一般可以均匀的采样；
   - 采样$X_J^{(t)} \sim \pi_{X_J|X_{-J}}\left(\cdot|X_1^{(t-1)}, ..., X_{J-1}^{(t-1)}, X_{J+1}^{(t-1)}, ..., X_d^{(t-1)}\right)$

Gibbs Sampling

- 联合分布 π 是否能由条件分布$\pi_{X_i|X_{-i}}$唯一确定？
- Gibbs Sampler 是否以目标分布 π 为不变分布？
- Gibbs Sampler 是否能收敛到不变分布 π ？

Gibbs Sampling

定理 1 (Hammersley-Clifford)

设概率密度 $\pi\left(x_1, x_2, \ldots, x_d\right)$满足正定性条件*(positivity condition)*，即如果对于所有$x_1, \ldots, x_d$，边缘密度$\pi_{X_i}\left(x_i\right) > 0$，就有

$$\pi\left(x_1, x_2, \ldots, x_d\right) > 0$$

成立，那么对于所有$(z_1, \ldots, z_d) \in supp(\pi)$，即$\pi\left(z_1, \ldots, z_d\right) > 0$，

$$\pi\left(x_1, x_2, \ldots, x_d\right) \propto \prod_{j=1}^{d} \frac{\pi_{X_j|X_{-j}}\left(x_j | x_1, \ldots, x_{j-1}, z_{j+1}, \ldots, z_d\right)}{\pi_{X_j|X_{-j}}\left(z_j | x_1, \ldots, x_{j-1}, z_{j+1}, \ldots, z_d\right)}$$

证明...

Gibbs Sampling

记$x^{(t)} := \left(x_1^{(t)}, ..., x_d^{(t)}\right)$，systematic scan Gibbs sampler 的转移概率为

$$
\begin{aligned}
P\left(x^{(t-1)}, x^{(t)}\right) = \pi_{X_1|X_{-1}} & \left(x_1^{(t)} \middle| x_2^{(t-1)}, ..., x_d^{(t-1)}\right) \times \\
& \pi_{X_2|X_{-2}} \left(x_2^{(t)} \middle| x_1^{(t)}, x_3^{(t-1)}, ..., x_d^{(t-1)}\right) \times \cdots \\
& \times \pi_{X_d|X_{-d}} \left(x_d^{(t)} | x_1^{(t)}, ..., x_{d-1}^{(t)}\right).
\end{aligned}
$$

random scan Gibbs sampler 的转移概率为

$$
P\left(x^{(t-1)}, x^{(t)}\right) = \frac{1}{d} \sum_{j=1}^{d} \pi_{X_j|X_{-j}} \left(x_j^{(t)} | x_{-j}^{(t-1)}\right) \delta_{x_{-j}^{(t-1)}} \left(x_{-j}^{(t)}\right)
$$

其中$\delta_{x_{-j}^{(t-1)}}$为$x_{-j}^{(t-1)}$的 Dirac 测度。

Gibbs Sampling

命题 1

Systematic scan Gibbs sampler 的不变分布为 π.

证明...

Remark 1

Systematic scan Gibbs sampler 是不可逆的。

Gibbs Sampling

命题 2

Random scan Gibbs sampler 的不变分布为 π.

Remark 2

Random scan Gibbs sampler 是可逆的。

Gibbs sampler for Ising model

设

$$\pi = \frac{1}{Z}\mathrm{e}^{-\beta H(\sigma)},$$

其中$H(\sigma) = -J\sum_{(ij)\in B}\sigma(i)\sigma(j) - h\sum_i \sigma(i)$. 考虑转移核

$$P(\sigma,\sigma^i) = \frac{1}{N}\cdot\frac{e^{-\beta H(\sigma^i)}}{e^{-\beta H(\sigma)} + e^{-\beta H(\sigma^i)}}$$

其中$\frac{e^{-\beta H(\sigma^i)}}{e^{-\beta H(\sigma)}+e^{-\beta H(\sigma^i)}}$ 即为

$$\pi_{\sigma_i'|\sigma_{-i}} = \frac{\pi(\sigma_i',\sigma_{-i})}{\pi(\sigma_{-i})} = \frac{\pi(\sigma^i)}{\int_{\sigma_i=\{\pm 1\}}\pi(\sigma)}$$

Gibbs Sampling

定义 5 (π-不可约)

*Markov*链被称为π-不可约的，如果对于任何满足

$$\pi(A) \triangleq \int_A \pi\left(x_1, \ldots, x_d\right) d x_1 \ldots d x_d>0$$

的集合A，从任意初始状态x出发，存在某个正整数t，使得链在t步后到达A的概率为正。

这一条件确保了链在π的支撑上是“连通的”，即无法将状态空间分解为两个或多个 π 正测度的不相交子集，使得链无法从一个子集到达另一个。

Gibbs Sampling

命题 3

假设 π 满足正定性条件，那么 *Gibbs sampler* $X^{(t)}$ 是一个π-不可约、正常返的 *Markov* 链。

定理 2

假设 π 满足正定性条件，那么对于任意可积函数 $\varphi : S \to \mathbb{R}$

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \varphi\left(X^{(i)}\right) = \int_S \varphi(x)\,\pi(x)\,dx$$

Metropolis-Hastings 算法

设目标分布为π.

Metropolis-Hastings

1. 选择初始状态 $x^{(0)}$。
2. 对于每一步 $t = 1, 2, \ldots, N$:
   - 从提议分布 $q(x^\star|x^{(t-1)})$ 中生成候选状态 $x^\star$。
   - 计算接受概率：
   $$\alpha(x^\star|x^{(t-1)}) = \min\left(1, \frac{\pi(x^\star)q(x^{(t-1)}|x^\star)}{\pi(x^{(t-1)})q(x^\star|x^{(t-1)})}\right)$$
   - 以概率 α 接受候选状态 $x^\star$，否则保持状态 $x^{(t-1)}$。

     具体操作为：采样$U \sim$ Uniform$[0,1]$，若$\alpha \leq U$，则接受新状态，否则保持不变。

Metropolis-Hastings 算法

- Metropolis-Hastings 算法可以处理非归一化的概率密度，即，对于任意$\widetilde{\pi}\left(x\right) \propto \pi\left(x\right)$，有

$$\frac{\pi\left(x^{\star}\right) q\left(x^{(t-1)} \middle| x^{\star}\right)}{\pi\left(x^{(t-1)}\right) q\left(x^{\star} \middle| x^{(t-1)}\right)} = \frac{\widetilde{\pi}\left(x^{\star}\right) q\left(x^{(t-1)} \middle| x^{\star}\right)}{\widetilde{\pi}\left(x^{(t-1)}\right) q\left(x^{\star} \middle| x^{(t-1)}\right)}.$$

Metropolis-Hastings 算法

命题 4

MH 算法的转移概率为

$$P\left(x^{(t-1)}, x^{(t)}\right) = \alpha\left(x^{(t)} \middle| x^{(t-1)}\right) q\left(x^{(t)} \middle| x^{(t-1)}\right) + \left(1 - a\left(x^{(t-1)}\right)\right) \delta_{x^{(t-1)}}\left(x^{(t)}\right)$$

其中$\delta_{x^{(t-1)}}$是在$x^{(t-1)}$处的*Dirac*测度，

$$a\left(x^{(t-1)}\right) \triangleq \int_S \alpha\left(x \middle| x^{(t-1)}\right) q\left(x \middle| x^{(t-1)}\right) dx$$

表示在给定当前状态$x^{(t-1)}$时，会接受并更新状态的概率。

证明...

Metropolis-Hastings 算法

命题 5

*MH*算法是关于π可逆的，即

$$\pi\left(x^{(t-1)}\right) P\left(x^{(t-1)}, x^{(t)}\right)=\pi\left(x^{(t)}\right) P\left(x^{(t)}, x^{(t-1)}\right)$$

因此π也是*MH*算法的不变分布。

证明...

Metropolis-Hastings 算法的不可约性[7]

命题 6 (Informal)

如果存在$\delta,\ \epsilon > 0$，使得对任意$x \in S$，有

$$|x - x'| < \delta \Rightarrow q\left(x|x'\right) \geq \varepsilon,$$

那么*MH*算法是π-不可约的。

一个更强的条件是：如果对任意$x, x^\star \in \mathrm{supp}(\pi)$，有$q\left(x^\star|\, x\right) > 0$，那么MH链$\pi$-不可约。

[7] Gareth O. Roberts. Jeffrey S. Rosenthal. "General state space Markov chains and MCMC algorithms." Probab. Surveys 1 20 - 71, 2004. https://doi.org/10.1214/154957804100000024

Metropolis-Hastings 算法

定理 3

假设 *MH* 链是π-不可约的，那么对于任意可积函数 $\varphi : S \to \mathbb{R}$

$$\lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} \varphi\left(X^{(i)}\right) = \int_S \varphi(x)\, \pi(x)\, dx$$

Proposal 的选择

Symmetric Proposals

考虑 proposal 分布为

$$q\left(x|x^{(t-1)}\right) = q\left(x^{(t-1)}|x\right)$$

那么Metropolis-Hastings接受率为

$$\frac{\pi\left(x\right)q\left(x^{(t-1)}|x\right)}{\pi\left(x^{(t-1)}\right)q\left(x|x^{(t-1)}\right)} = \frac{\pi\left(x\right)}{\pi\left(x^{(t-1)}\right)}.$$

一种具体的实现：Random Walk Proposals

$$X = X^{(t-1)} + W$$

其中W服从d为标准高斯分布，$W \sim \mathcal{N}\left(0, I_d\right)$。

Proposal 的选择

Independent Proposals

考虑 proposal 分布为

$$q\left(x|x^{(t-1)}\right) = q(x)$$

那么Metropolis-Hastings接受率为

$$\frac{\pi(x)\, q\left(x^{(t-1)}|x\right)}{\pi\left(x^{(t-1)}\right) q\left(x|x^{(t-1)}\right)} = \frac{\pi(x)}{q(x)} \frac{q\left(x^{(t-1)}\right)}{\pi\left(x^{(t-1)}\right)}.$$

Proposal 的选择

定义 6 (Langevin扩散(Langevin Diffusion))

设$V:\mathbb{R}^d\to\mathbb{R}$是一个位势，我们称随机微分方程*(Stochastic Differential Equations, SDEs)*

$$\mathrm{d}X_t=-\nabla V(X_t)\mathrm{d}t+\sqrt{2}\mathrm{d}B_t.$$

的解为V对应的*Langevin扩散*/*Langevin Dynamics*, 其中B_t为一个标准布朗运动。*Langevin*扩散的不变测度为

$$\pi\propto e^{-V}$$

Langevin Proposals

$$X=X^{(t-1)}-\sigma\nabla\log\pi|_{X^{(t-1)}}+2\sigma W$$

其中$W\sim\mathcal{N}(0,I_d)$，$\nabla\log\pi|_{X^{(t-1)}}$表示在$X^{(t-1)}$处的目标密度函数的对数梯度，$\sigma$为可调整超参数。

Proposal 的选择

定义 6 (Langevin扩散(Langevin Diffusion))

设$V : \mathbb{R}^d \to \mathbb{R}$是一个位势，我们称随机微分方程*(Stochastic Differential Equations, SDEs)*

$$\mathrm{d}X_t = -\nabla V(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}B_t.$$

的解为V对应的*Langevin扩散*/*Langevin Dynamics*, 其中B_t为一个标准布朗运动。*Langevin*扩散的不变测度为

$$\pi \propto e^{-V}$$

Langevin Proposals

$$X = X^{(t-1)} - \sigma\nabla \log \pi|_{X^{(t-1)}} + 2\sigma W$$

其中$W \sim \mathcal{N}(0, I_d)$，$\nabla \log \pi|_{X^{(t-1)}}$表示在$X^{(t-1)}$处的目标密度函数的对数梯度，$\sigma$为可调整超参数。

Proposal 的选择

Informed Proposals

- 2020 JASA - Informed Proposals for Local MCMC in Discrete Spaces
- 2021 ICML - Oops i took a gradient: Scalable sampling for discrete distributions
- 2022 ICLR - Path auxiliary proposal for MCMC in discrete space
- 2022 ICML - A Langevin-like Sampler for Discrete Distributions
- 2022 NIPS - Optimal scaling for locally balanced proposals in discrete spaces
- 2023 ICLR - Any-scale balanced samplers for discrete space
- 2024 NIPS - Gradient-based Discrete Sampling with Automatic Cyclical Scheduling

总结

- Monte Carlo 方法：计算

$$I=\int_S f(x)\,dx=\int_S \varphi(x)\pi(x)dx=\mathbb{E}_\pi[\varphi].$$

转化为 simulation method to sample π

- 采样问题：设 π 是一个概率分布，如何获得随机样本 x，使得 x 的分布为 π。

- Markov chain Monte Carlo

总结：Gibbs sampler

- Systematic scan Gibbs sampler
  1. 选择初始状态 $X^{(0)} = (X_1^{(0)}, ..., X_d^{(0)})$。
  2. 对于每一步 $t = 1, 2, \ldots, N$:
     - 从提议分布 $\pi_{X_1|X_{-1}}\left(\cdot|X_2^{(t-1)}, ..., X_d^{(t-1)}\right)$ 中生成候选状态 $X_1^{(t)}$。
     - ...
     - 采样$X_j^{(t)} \sim \pi_{X_j|X_{-j}}\left(\cdot|X_1^{(t)}, ..., X_{j-1}^{(t)}, X_{j+1}^{(t-1)}, ..., X_d^{(t-1)}\right)$
     - ...
     - 采样$X_d^{(t)} \sim \pi_{X_d|X_{-d}}\left(\cdot|X_1^{(t)}, ..., X_{d-1}^{(t)}\right)$
- Random scan Gibbs sampler
  1. 选择初始状态 $X^{(0)} = (X_1^{(0)}, ..., X_d^{(0)})$。
  2. 对于每一步 $t = 1, 2, \ldots, N$:
     - 从维度指标集中$\{1, 2, \ldots, d\}$采样指标J，一般可以均匀的采样；
     - 采样$X_J^{(t)} \sim \pi_{X_J|X_{-J}}\left(\cdot|X_1^{(t-1)}, ..., X_{J-1}^{(t-1)}, X_{J+1}^{(t-1)}, ..., X_d^{(t-1)}\right)$

总结：Gibbs sampler

命题 7

Systematic/Random scan Gibbs sampler 的不变分布为 π，且在π 满足正定性条件下，对于任意可积函数 $\varphi : S \to \mathbb{R}$

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} \varphi\left(X^{(i)}\right) = \int_S \varphi(x)\,\pi(x)\,dx$$

总结：Metropolis-Hastings 算法

1. 选择初始状态 $x^{(0)}$。
2. 对于每一步 $t = 1, 2, \ldots, N$:
   - 从提议分布 $q(x^\star | x^{(t-1)})$ 中生成候选状态 $x^\star$。
   - 计算接受概率：

$$\alpha(x^\star | x^{(t-1)}) = \min\left(1, \frac{\pi(x^\star) q(x^{(t-1)} | x^\star)}{\pi(x^{(t-1)}) q(x^\star | x^{(t-1)})}\right)$$

   - 以概率 α 接受候选状态 $x^\star$，否则保持状态 $x^{(t-1)}$。

     具体操作为：采样$U \sim \text{Uniform}[0, 1]$，若$\alpha \leq U$，则接受新状态，否则保持不变。

总结：Metropolis-Hastings 算法

命题 8

*MH*算法是关于π可逆的，因此π也是*MH*算法的不变分布，而且若*MH* 链是π-不可约的，那么对于任意可积函数 $\varphi : S \to \mathbb{R}$

$$\lim_{t\to\infty}\frac{1}{t}\sum_{i=1}^{t}\varphi\left(X^{(i)}\right)=\int_S\varphi\left(x\right)\pi\left(x\right)dx$$

目录