统计计算 张楠 2019秋: Introduction to Markov Chain Monte Carlo Methods

(返回 统计计算 张楠 2019秋)

Introduction

MCMC(Markov Chain Monte Carlo) 方法的一般理论框架可以参看 Metropolis et al. (1953)以及 Hastings (1970), 以及其他各种介绍MCMC的专著. 本节我们介绍这种方法的基本思想和应用.

注意在前面介绍Monte Carlo方法估计积分 $\int_A g(t)dt$, 是把此积分表示成一个对某个概率密度$f(t)$下的期望. 从而积分估计问题转化成从目标概率密度$f(t)$中产生 随机样本.

在MCMC方法中, 首先建立一个马尔科夫链, 使得$f(t)$为其平稳分布. 则可以运行此马尔科夫链充分长时间直至 收敛到平稳分布, 那么从目标分布$f(t)$中产生随机样本, 就是从达到平稳状态的马尔科夫链中产生其样本路径. 我们将介绍几种建立这样的马尔科夫链的方法: Metropolis算法, Metropolis-Hastings算法, 以及Gibbs抽样方法. 一个好的链应该具有快速混合(rapid mixing)性质—从任意位置出发很快达到平稳分布.

Integration problems in Bayesian inference

Bayesian推断中的许多问题都是MCMC方法的应用. 从Bayesian的观点来看, 模型中的观测变量和参数都是随机变量. 因此, 样本$x = (x_1, \cdots, x_n)$和参数θ联合分布可以表示为 $f_{x,\theta}(x, \theta) = f_{x|\theta}(x_1, \cdots, x_n)\pi(\theta)$. 从而根据Bayes定理, 可以通过样本$x = (x_1, \cdots, x_n)$的信息对$\theta$的分布进行更新: $f_{\theta|x}(\theta|x) == \frac{f_{x|\theta}(x)\pi(\theta)}{\int f_{x|\theta}(x)\pi(\theta)d\theta}$. 则在后验分布下, $g(\theta)$的期望为 $Eg(\theta|x) = \int g(\theta)f_{\theta|x}(\theta|x)d\theta = \frac{\int g(\theta)f_{x|\theta}(x)\pi(\theta)d\theta}{\int f_{x|\theta}(x)\pi(\theta)d\theta}$. 此积分为值为样本$x$的函数. 因此可以对$g(\theta)$进行推断. 比如 $g(\theta) = \theta$时, 则$Eg(\theta|x) = E[\theta|x]$ 可以作为θ的估计.

对此类问题我们考虑一般形式: $Eg(Y) = \frac{\int g(t)\pi(t)dt}{\int \pi(t)dt}$. 这里$\pi(\cdot)$为一个密度或者似然. 若$\pi$为一密度, 则此期望即为常见的期望定义: $Eg(Y) = \int g(t)f_Y(t)dt$. 若$\pi$为一似然, 则需要一个正则化常数才可以成为密度. 在贝叶斯分析中, $\pi(\cdot)$为一后验密度. 当$\pi(\cdot)$ 的正则化常数未知时, 此积分 也可以计算. (分子分母中的正则化常数抵消了). 这是非常好的一个性质, 因为在 很多问题下, 正则化常数很难计算.

另一方面, 在很多具体问题中此积分根本没有显示表示, 数值方法也很难计算, 特别是在高维时, 而MCMC方法对此类积分提供了 一个很好的计算方法.

Markov Chain Monte Carlo Integration

积分 $E[g(\theta|x)] = \int g(\theta)f_{\theta|x}(\theta|x)d\theta$ 的Monte Carlo估计为样本均值 $\bar{g} = \frac{1}{m}\sum_{i=1}^{m} g(x_i)$, 其中$x_1, \cdots, x_m$ 为从分布$f_{\theta|x}(\theta|x)$中抽取的样本. 当$x_1, \cdots, x_m$ 独立时, 则可以根据大数律知当样本量n趋向于无穷时, $\bar{g}$收敛到$E[g(\theta|x)$.

但是在一些问题中, 从分布$f_{\theta|x}(\theta|x)$中抽取样本是非常困难的, MCMC方法就是为此目的而诞生的, 其第一个“MC”, Markov Chain, 就表示从目标分布中抽样, 第二个“MC”, Monte Carlo, 则表示在抽取到的样本下, 应用Monte Carlo积分方法对积分进行计算.

MCMC方法的理论依据在于下述的极限定理:

定理1

设$\{X_t, t \geq 0\}$为一不可约的,非周期的状态空间为Ω的马氏链, π为平稳分布, π_0是初始分布, 则 $\pi_t \rightarrow \pi, \quad t \rightarrow \infty.$

这里π_t表示马氏链在t时刻的边际分布. 此定理说明了当此马氏链运行充分长时间后($t = n, n$很大), 则在时刻n, X_n的分布为π, 且和初始分布无关.

定理2 (Markov chain Law of Large Numbers)

若$X_1, X_2, \cdots$为一遍历的平稳分布为π的马氏链值(其中每个X_t可以是多维的), 则X_n依分布收敛到 分布为π的随机变量X, 且对任意函数f, 当$E_\pi|f(X)|$存在时, 且$n \rightarrow \infty$时, 则 $\bar{f}_n = \frac{1}{n}\sum_{t=1}^n f(X_t) \rightarrow E_\pi f(X), \quad a.s.$

下面我们计算遍历均值$\bar{f}_n$的方差. 记$f^t = f(X_t), \gamma_k = cov(f^t, f^{t+k})$, 则$f^t$的方差为$\sigma^2 = \gamma_0$, k步相关系数$\rho_k = \gamma_k/\sigma^2$. 从而

$$
\begin{aligned}
\tau_n^2 &= nVar_\pi(\bar{f}_n) = nVar(\frac{1}{n}\sum_{t=1}^n f^t) = \frac{1}{n}Var_\pi(\sum_{t=1}^n f^t) \\
&= \frac{1}{n}\sum_{t=1}^n Var_\pi(f^t) + 2\frac{1}{n}\sum_{t=1}^{n-1}\sum_{k=1}^{n-t} cov(f^t, f^{t+k}) \\
&= \sigma^2 + 2\sum_{k=1}^{n-1}\frac{n-k}{n}\gamma_k = \sigma^2[1 + 2\sum_{k=1}^{n-1}\frac{n-k}{n}\rho_k].
\end{aligned}
$$

可以证明 $\tau_n^2 \rightarrow \tau^2 = \sigma^2[1 + 2\sum_{k=1}^\infty \rho_k]$. 因此

遍历均值$\bar{f}_n$的方差为$Var_\pi(\bar{f}_n) = \frac{\sigma^2}{n}[1 + 2\sum_{k=1}^{n-1}\frac{n-k}{n}\rho_k]$.

定理3

若一个链是一致的几何遍历(转移概率以速度$\lambda^t, 0 < \lambda < 1$收敛)的, 以及 f相对于平稳分布π是平方可积的, 则有 $\sqrt{n}\frac{\bar{f}_n - E_\pi f(x)}{\tau} \rightarrow N(0,1), \quad n \rightarrow \infty$

这个定理对构造平稳分布π的参数的置信区间提供了理论依据.

这些定理说明了:
- 不可约+非周期+正常返$\Longrightarrow$ 唯一的平稳分布.
- LLN: 马氏链的实值函数的遍历均值以概率1收敛到极限分布下的均值.
- CLT: 遍历均值作合适的变化依分布收敛到标准正态分布.

因此, 我们可以建立一个以π为平稳分布的马氏链, 则在运行此链足够长时间后, 该马氏链就会达到 平稳状态, 此时马氏链的值就相当于从分布π中抽取样本. 从而可以根据遍历定理对积分进行估计. MCMC方法的优点在于:

1. 适用于广泛且困难的问题;

2. 问题维数的增加通常不会降低其收敛速度或者使其更复杂.
