博 学　慎 思　明 辨　笃 行

贝叶斯计算及其挑战

Bayesian Computation and some Challenges

何志坚

hezhijian@scut.edu.cn

www.hezhijian.com.cn

目　录

contents

贝叶斯与贝叶斯公式

贝叶斯学派起源于英国学者贝叶斯在1763 年发表的著名论文《论有关机遇问题的求解》

似然函数

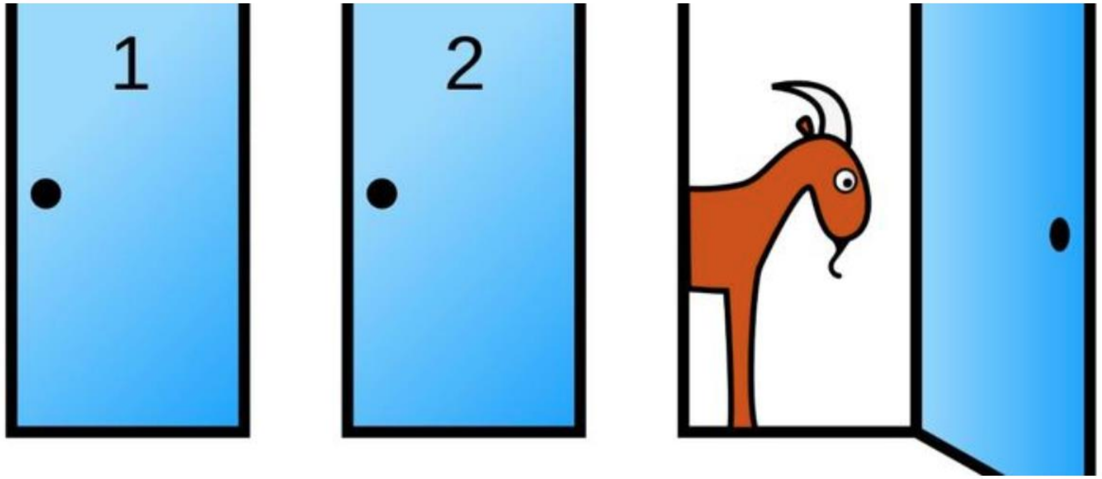
$$P(A|D) = P(A)\frac{P(D|A)}{P(D)}$$

后验概率　先验概率　证据

Thomas Bayes 1702-1761

贝叶斯公式应用1：三门问题（Monty Hall problem）

有三扇门，其中一扇有奖品，两扇有山羊。选手选完后，主持人打开其中一扇有山羊的门，主持人问选手是否换另外一扇门?



A: 选手一开始选中奖品（不换门中奖）

D: 主持人打开其中一扇有山羊的门

$$P(A|D) = P(A)\frac{P(D|A)}{P(D)}$$

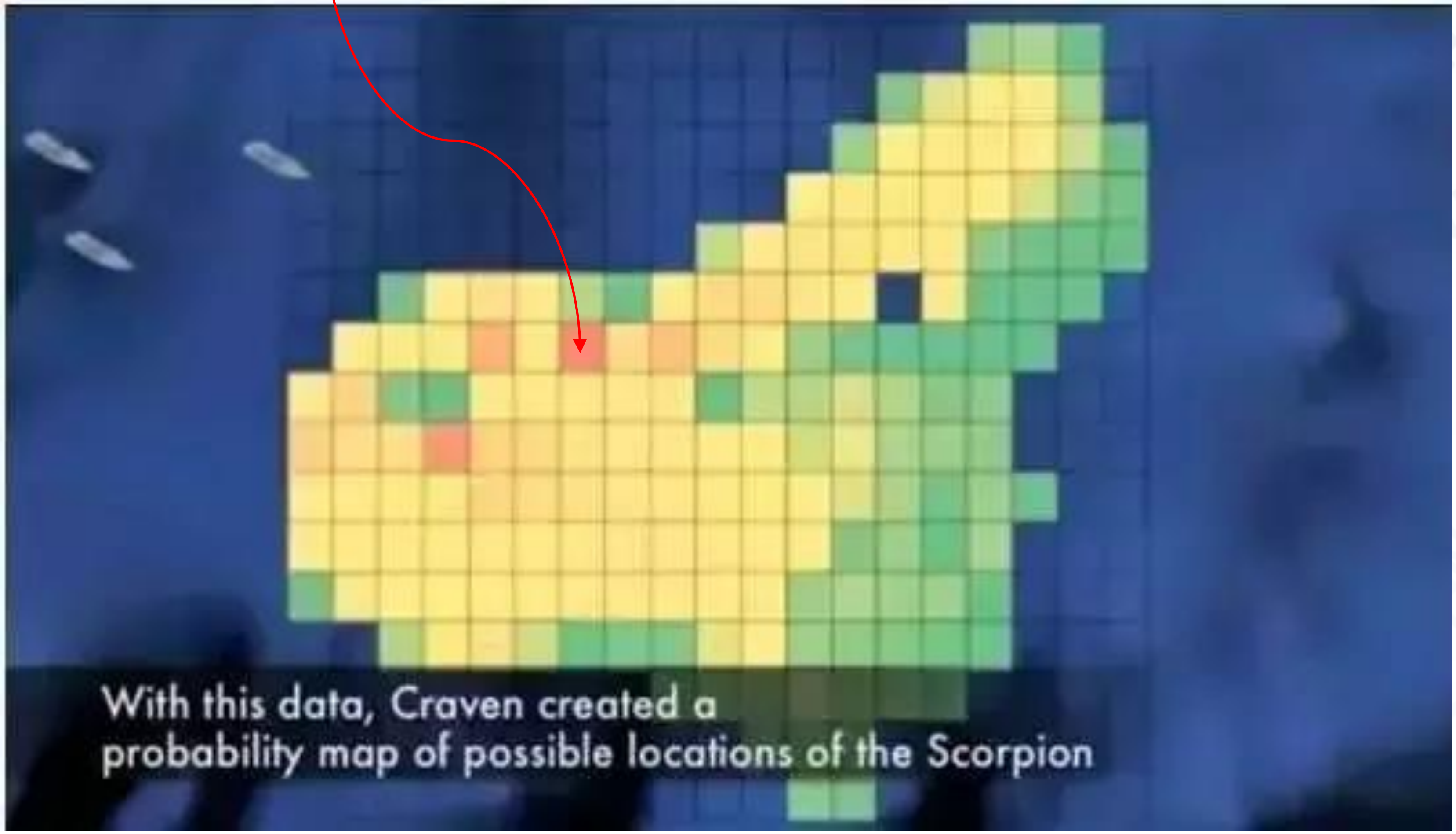
贝叶斯公式应用2：天蝎号核潜艇搜救

Part I

Part II

Part III

$$P(A_i|D) = P(A_i)\frac{P(D|A_i)}{P(D)}$$



贝叶斯计算的动机

密度形式的贝叶斯公式:

$$p(\theta|\boldsymbol{y}) = p(\theta)\frac{L(\boldsymbol{y}|\theta)}{p(\boldsymbol{y})} \propto p(\theta)L(\boldsymbol{y}|\theta)$$

后验分布　先验分布　似然函数　证据

$\theta = (\theta_1, \dots, \theta_d)$：

模型参数

$\boldsymbol{y} = (y_1, \dots, y_n)$:

观测数据

$$p(\boldsymbol{y}) = \int p(\theta)L(\boldsymbol{y}|\theta)\mathrm{d}\theta$$

一般情况下，$p(\boldsymbol{y})$不可计算

感兴趣的问题

$$\mathrm{E}[f(\theta)|\boldsymbol{y}] = \int f(\theta)p(\theta|\boldsymbol{y})d\theta$$

- 后验均值
- 后验概率

需要引入数值算法!

$\because p(\theta|\boldsymbol{y})$未知

$\therefore \mathrm{E}[f(\theta)|\boldsymbol{y}]$未知

贝叶斯计算：从1763年到21世纪

$$\mathrm{E}[f(\theta)|\boldsymbol{y}] = \int f(\theta)p(\theta|\boldsymbol{y})d\theta$$

- 1774年，Laplace积分逼近方法
- 1953年，蒙特卡洛(Monte Carlo)方法和Metropolis算法
- 1964年，重要性抽样(Importance Sampling)方法
- 1970年，推广的Metropolis算法——Metropolis-Hastings算法
- 1980s， Gibbs抽样方法
- 21世纪：第二次计算革命（面临一些新挑战）

现代贝叶斯计算新挑战

$$p(\theta|\boldsymbol{y}) \quad = \quad p(\theta)\frac{L(\boldsymbol{y}|\theta)}{p(\boldsymbol{y})} \propto p(\theta)L(\boldsymbol{y}|\theta)$$

挑战一：似然函数$L(\boldsymbol{y}|\theta)$没有显示表达式

- 分位数模型，生成模型 $\boldsymbol{y} = G(\boldsymbol{v}|\theta)$, $\boldsymbol{v}$是随机向量
- 一些深度学习模型
- 隐马尔科夫模型(HMM)

现代贝叶斯计算新挑战

$$p(\theta|\boldsymbol{y}) \quad = \quad p(\theta)\frac{L(\boldsymbol{y}|\theta)}{p(\boldsymbol{y})} \propto p(\theta)L(\boldsymbol{y}|\theta)$$

挑战二：似然函数$L(\boldsymbol{y}|\theta)$计算复杂度高

- $\boldsymbol{y} = (y_1, \dots, y_n)$数据量为$n$, 似然函数复杂度为$O(n)$
- 对大数据问题不可扩展(not scalable)

现代贝叶斯计算新挑战

$$p(\theta|\boldsymbol{y}) \quad = \quad p(\theta)\frac{L(\boldsymbol{y}|\theta)}{p(\boldsymbol{y})} \propto p(\theta)L(\boldsymbol{y}|\theta)$$

挑战三：高维参数$\theta = (\theta_1, \dots, \theta_d)$

- 潜(冗余)变量的存在, $d > n$
- 对大数据问题不可扩展(not scalable)

例子：状态空间模型(隐马尔科夫模型)

两个随机过程$\{X_t\}_{t\geq 0}, \{Y_t\}_{t\geq 0}$

- $\{X_t\}_{t\geq 0}$是马尔科夫过程(不可观测, 潜变量)

$$X_0 \sim \mu_\theta(x_0), \qquad X_t | X_0, \dots, X_{t-1} \sim f_\theta(x_t | x_{t-1})$$

- $\{Y_t\}_{t\geq 0}$是可观测的随机过程, 满足

$$Y_t | X_0, \dots, X_t, Y_0, \dots, Y_{t-1} \sim g_\theta(y_t | x_t)$$

- θ是模型参数

例子：状态空间模型(隐马尔科夫模型)

- 观测数据$y_{0:T} = (y_0, y_1, \dots, y_T)$
- 如何推断马尔科夫过程$x_{0:T} = (x_0, x_1, \dots, x_T)$?
- 如何估计模型参数θ

似然函数

高维积分，通常无解析解!

$$p_\theta(y_{0:T}) = \int p_\theta(x_{0:T}, y_{0:T}) dx_{0:T}$$

$$= \int \mu_\theta(x_0) \prod_{t=1}^{T} f_\theta(x_t | x_{t-1}) \prod_{t=0}^{T} g_\theta(y_t | x_t) \, dx_{0:T}$$

若数据不断累积(T变大)，如何设计算法? Off-line or on-line?

现代贝叶斯计算新方法

$$p(\theta|\boldsymbol{y}) \quad = \quad p(\theta)\frac{L(\boldsymbol{y}|\theta)}{p(\boldsymbol{y})} \propto p(\theta)L(\boldsymbol{y}|\theta)$$

方法一：伪边际方法(Pseudo-Marginal Methods)

- 目标分布$p(\theta, \boldsymbol{x}|\boldsymbol{y})$
- $\boldsymbol{z}$：潜变量或者人工变量（用于估计似然函数$L(\boldsymbol{y}|\theta)$）

现代贝叶斯计算新方法

$$p(\theta|\boldsymbol{y}) \quad = \quad p(\theta)\frac{L(\boldsymbol{y}|\theta)}{p(\boldsymbol{y})} \propto p(\theta)L(\boldsymbol{y}|\theta)$$

方法二：近似贝叶斯计算
(Approximate Bayesian Computation, ABC)

- 目标分布$p_h(\theta,\boldsymbol{z}|\boldsymbol{y}) \to p(\theta,\boldsymbol{z}|\boldsymbol{y})$当$h \to 0$时。
- 这是一种近似算法，得到不是准确的后验分布

现代贝叶斯计算新方法

$$p(\theta|\mathbf{y}) \quad = \quad p(\theta)\frac{L(\mathbf{y}|\theta)}{p(\mathbf{y})} \propto p(\theta)L(\mathbf{y}|\theta)$$

方法三：变分贝叶斯(Variational Bayes)

- 在给定分布族中找一种“最优”分布来近似后验分布$p(\theta|\mathbf{y})$

$$q^*(\boldsymbol{\theta}) := \arg\min_{q(\boldsymbol{\theta})\in\mathcal{Q}} \mathrm{KL}\left[q(\boldsymbol{\theta})|p(\boldsymbol{\theta}|\mathbf{y})\right]$$

- Kullback-Leibler (KL) 散度, $KL(q|p) = E_q[\log(\frac{q(\theta)}{p(\theta|y)})]$

现代贝叶斯计算新方法

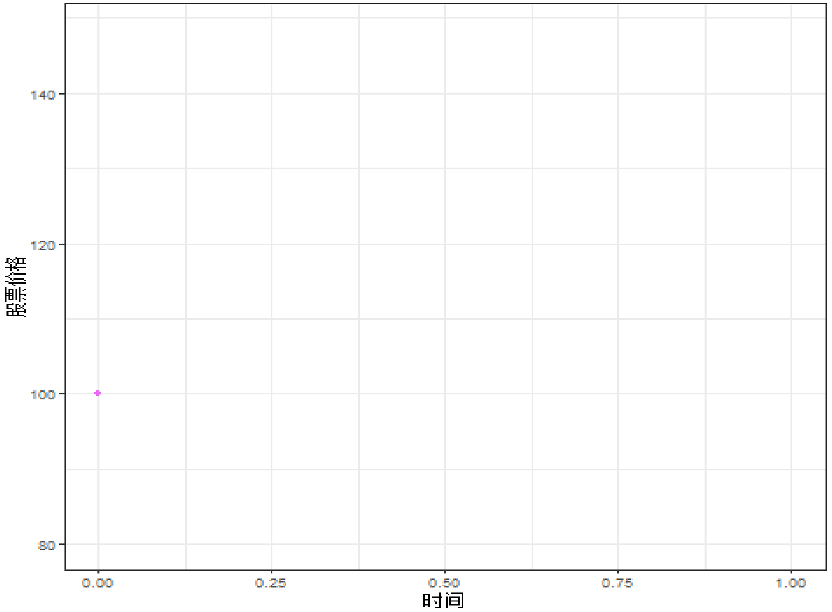
方法四：粒子滤波算法(Particle Filtering), 也叫序贯蒙特卡罗方法(sequential Monte Carlo)

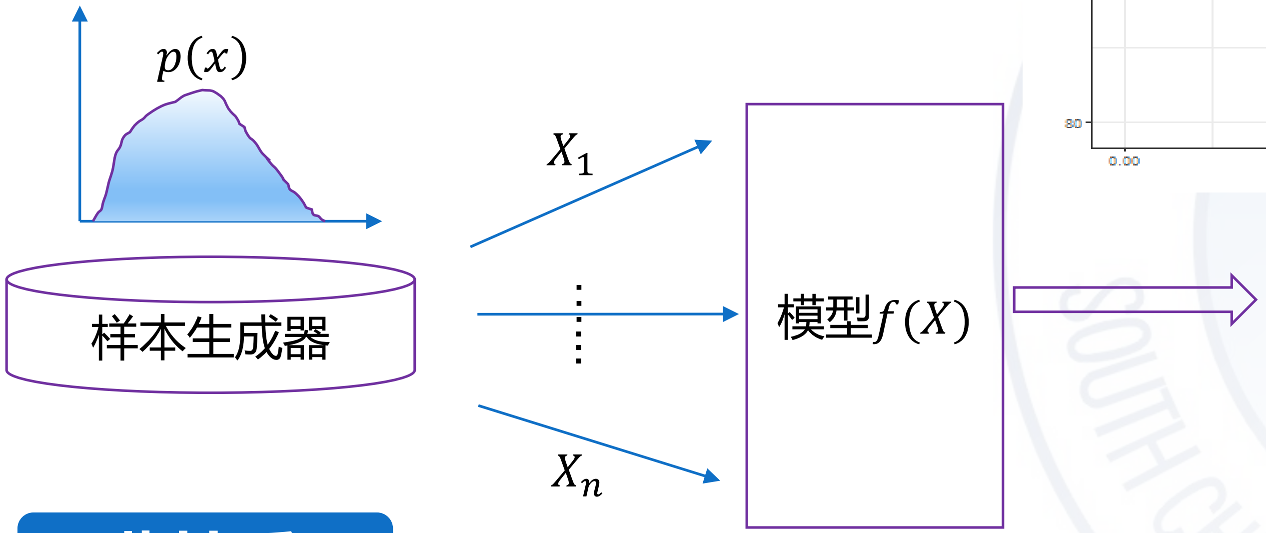
- 通过采样, 用离散分布近似

$$\hat{p}_\theta(y_{0:T}) = \hat{p}_\theta(y_0)\prod_{t=1}^{T}\hat{p}_\theta(y_t|y_{0:t-1})$$

基于随机模拟的计算方法：蒙特卡罗方法

工作机制：





$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n} f(X_i) \approx E[f(X)]$$

一些性质：

- 收敛阶为 $O(n^{-0.5})$，与维数 d 无关，与 $f(X)$ 光滑性无关。
- 收敛阶太低，往往不满足实际需求。

Part I

Part II

Part III

拟蒙特卡罗方法(QMC)

$$\int_{(0,1)^d} f(\boldsymbol{u})d\boldsymbol{u} \approx \frac{1}{n}\sum_{i=1}^{n} f(\boldsymbol{u}_i)$$

蒙特卡罗方法

Monte Carlo (MC)

- 使用独立同分布点列（伪随机数）
- 收敛阶$O(n^{-1/2})$

拟蒙特卡罗方法

Quasi-Monte Carlo (QMC)

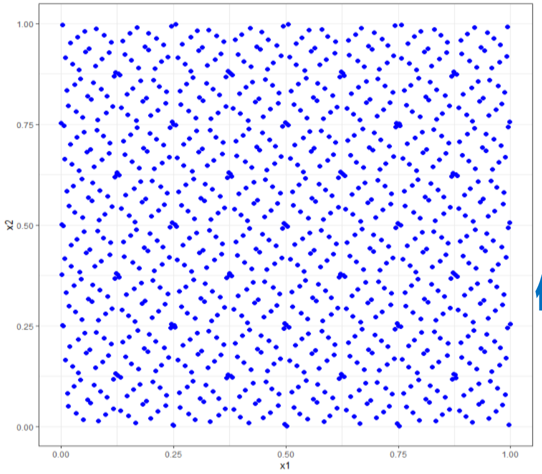
- 使用**确定性的超均匀（低偏差）序列**
- **高阶**算法

- 华罗庚和王元提出被国际上誉为“**华—王方法**”
- Halton序列， Sobol序列，Faure序列，Niederreiter序列，好格子点列
- 基于数论工具，其机理与蒙特卡罗不同



随机化拟蒙特卡罗方法(RQMC)

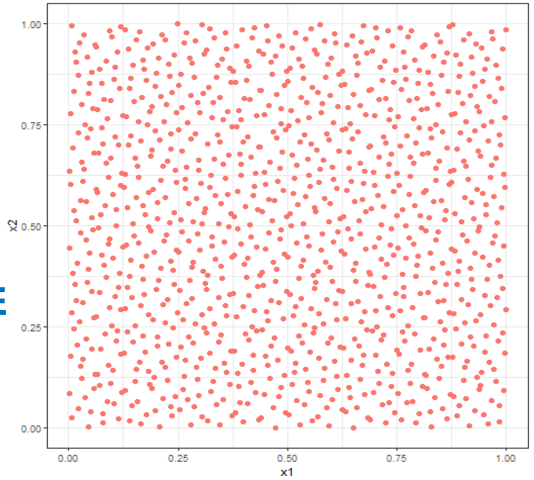
实际应用中，使用**随机化拟蒙特卡罗(RQMC)**：无偏估计、容易估计误差



1024个QMC点

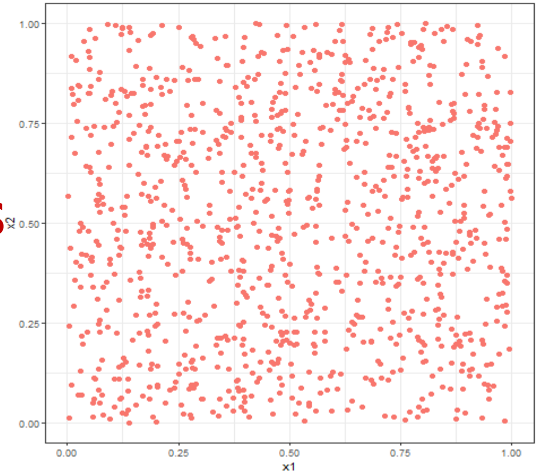
随机化

保持超均匀性



1024个RQMC点

VS



1024个MC点

为什么使用RQMC?

① 对于一般的函数，RQMC渐近快于MC：$o\left(n^{-1/2}\right)$ vs. $O\left(n^{-1/2}\right)$ **不会变差**

② 函数越光滑，RQMC的收敛阶越高

- 有界变差的函数：$O\left(n^{-1}(\log n)^d\right)$ **有可能得到更高收敛阶**
- 偏导平方可积的函数：$O\left(n^{-3/2}(\log n)^{(d-1)/2}\right)$

Thanks for your attention

何志坚

hezhijian@scut.edu.cn