第一章 绪论

Wang Shujia

Department of Statistics, School of Economics
Shenzhen University

Overview

Outline

1. 课程介绍
2. 贝叶斯定理
3. 贝叶斯推断
4. 频率学派推断与贝叶斯推断的差异
5. 教材及参考文献

贝叶斯是谁？

汤马斯・贝叶斯 (Thomas Bayes,1702-1761)，英国牧师。生前没有片纸只字的科学论著发表，但是 1742 年却当选为英国皇家学会会员（相当于今天的英国科学院院士)。

他的传世“遗作”是与朋友 Condon 的通信，被整理成论文“Essay Towards Solving a Problem in the Doctrine of Chances”，发表在 1764 年的学术杂志《Philosophical Transactions》。

“这个生性孤僻，哲学气味重于数学气味的学术怪杰，以一篇遗作的思想重大地影响了两个世纪以后的统计学术界，顶住了统计学的半边天”。

中国科学院院士陈希孺

课程目标

- 了解贝叶斯方法的基本理论
- 能够应用贝叶斯统计进行建模和推断
- 了解马尔科夫链蒙特卡洛（MCMC）算法
- 能够运用 R、WinBUGS/OpenBUGS、Stan 和 JAGS 等贝叶斯分析软件
- 应用贝叶斯方法于
  - 参数推断
  - 多元线性模型
  - 广义线性模型
  - 多层模型
  - 贝叶斯模型平均法等

课程内容

- 贝叶斯模型
- 正态分布与学生 t 分布的贝叶斯模型
- 先验分布的选择
- 贝叶斯 MCMC 软件运用（教材 Ch11）
- MCMC 收敛性（教材 Ch14）
- 贝叶斯线性回归模型
- 模型质量的评价
- 假设检验与贝叶斯因子
- 贝叶斯决策理论
- 贝叶斯多层模型（教材 Ch12）
- 学术论文讨论（贝叶斯模型平均）

课程评价

- 期末考试（闭卷，70%）
- 作业及项目（20%）
- 课堂表现（10%）

Outline

贝叶斯定理 (离散型)

定理 (贝叶斯定理 (离散型))

事件 B 发生了，其发生必来源于如下完备事件组：$A_1, \ldots, A_n$（两两独立，全部事件的并是必然事件），则对 $k = 1, 2, \ldots, n$，有

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^{n} P(A_i)P(B|A_i)}$$

- 应用于逆概率：妻子某天回家发现了一件新女式内衣 (事件 B)，出现这件新内衣可能原因是：
  - A_1：丈夫出轨了，A_2：其它情况
- 据统计，30% 的丈夫曾经出轨（称为**先验概率**，Prior probability)
- 妻子想知道的概率是：发现新内衣后，他出轨的概率 $P(A_1|B)$（称为**后验概率**，Posterior probability）
- 贝叶斯定理的逻辑：根据**观察结果**，对先验概率进行修正，修正系数为 $P(B|A_k)/P(B)$

贝叶斯定理：一个计算例子

某市新型冠状病毒感染者在总人口中的占比为 0.5% 左右。某人进行了新型冠状病毒的核酸检验，结果呈**阳性** (事件 B)，他想知道他确实感染了病毒的概率。

- A_1：他感染病毒了，A_2：他得的是一般感冒。
- 他感兴趣的概率：$P(A_1|B)$
- 必须知道如下三个概率：
  1. 先验概率 $P(A_1)$：该市新型冠状病毒感染者在总人口中的占比为 0.5% 左右
  2. $P(B|A_1)$：感染者中，核酸检验呈阳性的概率（检验准确率，95%）
  3. $P(B|A_2)$：非感染者中，检验呈阳性的概率（检验错误率，5%）

贝叶斯定理：一个计算例子（续）

他检验结果呈阳性，真正得病的概率

$$P(A_1|B) = \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$$
$$= \frac{0.5\% \times 95\%}{0.5\% \times 95\% + 99.5\% \times 5\%} = 8.72\%$$

表明：**检验之前**，他感染病毒的概率为 0.5%（先验），**检验之后**，他确实感染病毒的概率修正为 8.72%（后验）

- 该先生再检验一次，结果仍为阳性（B_2），则他确实为艾滋病人的概率又是多少？
  - 此时 $P(A_1) = 8.72\%, P(A_1|B_2) = 64.48\%$
  - 依次，$P(A_1|B_3) = 97.18\%$

贝叶斯定理 (连续型)

定理 (贝叶斯定理 (连续型))

设总体分布为 $X|\theta \sim f(x|\theta)$，未知参数服从 $\theta \sim \pi(\theta)$ 分布，则在给定数据 $X = x$ 下，参数 θ 的条件分布 *pdf* 为

$$\theta|x \sim p(\theta|x) = \frac{\pi(\theta)f(x|\theta)}{f_X(x)} \propto \pi(\theta)f(x|\theta)$$

- 参数 θ 在得到观察数据之前的分布为 $\theta \sim \pi(\theta)$，称为 θ 的**先验分布**
- 得到观察数据 $X = x$ 后，参数 θ 的条件件 *pdf* $p(\theta|x)$ 称为 θ 的**后验分布**
- 后验分布是根据观察数据对未知参数 θ 先验分布的修正

在连续型贝叶斯定理中，关于 X 的边缘分布为

$$X \sim f_X(x) = \int_\Theta f(x,\theta)\mathrm{d}\theta = \int_\Theta \pi(\theta) f(x|\theta)\mathrm{d}\theta$$

称为边缘似然函数 (marginal likelihood)，或先验预测分布 (prior predictive distribution)

Outline

Outline

3. 贝叶斯推断
   - 从女士品茶谈起
   - 贝叶斯推断

女士品茶，醉汉猜硬币，音乐家识谱

20 世纪 20 年代末，一位女士对一群绅士宣称，她可以分辨出奶茶中是**先放奶**还是**先放茶**。著名统计学家 R.A.Fisher 随机安排 10 杯奶茶（各 5 杯）进行检验，结果该女士说对了 9 杯。试问该女士是否确有辨别奶茶的能力？

- 一位喝醉了的朋友声称他闭着眼睛都能预测出抛硬币的结果，结果测试了 10 次，他说对了 9 次。试问该醉汉真有这个特异功能吗？

女士品茶，醉汉猜硬币，音乐家识谱

20 世纪 20 年代末，一位女士对一群绅士宣称，她可以分辨出奶茶中是**先放奶**还是**先放茶**。著名统计学家 R.A.Fisher 随机安排 10 杯奶茶（各 5 杯）进行检验，结果该女士说对了 9 杯。试问该女士是否确有辨别奶茶的能力？

- 一位喝醉了的朋友声称他闭着眼睛都能预测出抛硬币的结果，结果测试了 10 次，他说对了 9 次。试问该醉汉真有这个特异功能吗？
- 如果一位音乐家声称她能够辨别出一个乐谱是海登的还是莫扎特的。测试 10 份乐谱，他说对了 9 次。

频率学派的推断

频率学派推断：X 表示该女士正确辨识奶茶的次数，则

- 模型：$X|\theta \sim Bin(10, \theta)$，观察结果是 $x = 9$。
- 需要检验：$H_0 : \theta = 0.5$ $H_1 : \theta > 0.5$。
  - 假设检验的 p-值 $= P(X \geq 9|H_0) = 1 - \text{pbinom}(8, 10, 0.5) = 0.011$
  - 结论：支持该女士真有辨识奶茶的能力
- 对醉汉和音乐家的结论也是一样的（数据一样）。
- 你真的相信这些结论吗？

女士品茶的贝叶斯推断

贝叶斯学派推断：除了数据，还要加上个人对参数 θ 的分布进行先验判断。

- 模型：$X|\theta \sim Bin(10, \theta)$，即 $P(X = x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}(x = 0, 1, \ldots, 10)$

女士品茶的贝叶斯推断

贝叶斯学派推断：除了数据，还要加上个人对参数 θ 的分布进行先验判断。

- 模型：$X|\theta \sim Bin(10, \theta)$，即 $P(X = x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}(x = 0, 1, \ldots, 10)$
- 观察结果是 $x = 9$。

女士品茶的贝叶斯推断

贝叶斯学派推断：除了数据，还要加上个人对参数 θ 的分布进行先验判断。

- 模型：$X|\theta \sim Bin(10,\theta)$，即 $P(X=x|\theta)=\binom{10}{x}\theta^x(1-\theta)^{10-x}(x=0,1,\ldots,10)$
- 观察结果是 $x=9$。
- θ 的先验分布假设为 $\theta \sim U(0,1)$，即 $\pi(\theta)=1\ (0<\theta<1)$

女士品茶的贝叶斯推断

贝叶斯学派推断：除了数据，还要加上个人对参数 θ 的分布进行先验判断。

- 模型：$X|\theta \sim Bin(10, \theta)$，即 $P(X = x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}(x = 0, 1, \ldots, 10)$
- 观察结果是 $x = 9$。
- θ 的先验分布假设为 $\theta \sim U(0, 1)$，即 $\pi(\theta) = 1\ (0 < \theta < 1)$
- θ 的后验分布为：

女士品茶的贝叶斯推断

贝叶斯学派推断：除了数据，还要加上个人对参数 θ 的分布进行先验判断。

- 模型：$X|\theta \sim Bin(10, \theta)$，即 $P(X = x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}(x = 0, 1, \ldots, 10)$
- 观察结果是 $x = 9$。
- θ 的先验分布假设为 $\theta \sim U(0, 1)$，即 $\pi(\theta) = 1 \ (0 < \theta < 1)$
- θ 的后验分布为：

女士品茶的贝叶斯推断

贝叶斯学派推断：除了数据，还要加上个人对参数 θ 的分布进行先验判断。

- 模型：$X|\theta \sim Bin(10, \theta)$，即 $P(X = x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}(x = 0, 1, \ldots, 10)$
- 观察结果是 $x = 9$。
- θ 的先验分布假设为 $\theta \sim U(0, 1)$，即 $\pi(\theta) = 1\ (0 < \theta < 1)$
- θ 的后验分布为：

$$p(\theta|x) \propto \pi(\theta)f(x|\theta) \propto \theta^x(1-\theta)^{10-x}$$

即服从

女士品茶的贝叶斯推断

贝叶斯学派推断：除了数据，还要加上个人对参数 θ 的分布进行先验判断。

- 模型：$X|\theta \sim Bin(10, \theta)$，即 $P(X = x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}(x = 0, 1, \ldots, 10)$
- 观察结果是 $x = 9$。
- θ 的先验分布假设为 $\theta \sim U(0, 1)$，即 $\pi(\theta) = 1\ (0 < \theta < 1)$
- θ 的后验分布为：

$$p(\theta|x) \propto \pi(\theta)f(x|\theta) \propto \theta^x(1-\theta)^{10-x}$$

即服从 $\theta|x \sim Beta(x + 1, 11 - x)$

- 需要检验：$H_0 : \theta = 0.5\ H_1 : \theta > 0.5$。

女士品茶的贝叶斯推断

贝叶斯学派推断：除了数据，还要加上个人对参数 θ 的分布进行先验判断。

- 模型：$X|\theta \sim Bin(10, \theta)$，即 $P(X = x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}(x = 0, 1, \ldots, 10)$
- 观察结果是 $x = 9$。
- θ 的先验分布假设为 $\theta \sim U(0, 1)$，即 $\pi(\theta) = 1\ (0 < \theta < 1)$
- θ 的后验分布为：

$$p(\theta|x) \propto \pi(\theta)f(x|\theta) \propto \theta^x(1-\theta)^{10-x}$$

即服从 $\theta|x \sim Beta(x + 1, 11 - x)$

- 需要检验：$H_0 : \theta = 0.5\ H_1 : \theta > 0.5$。
- H_1 成立的后验概率 $P(\theta > 0.5|x = 9) = 1 - \text{pbeta}(0.5, 10, 2) = 0.994$

女士品茶的贝叶斯推断

贝叶斯学派推断：除了数据，还要加上个人对参数 θ 的分布进行先验判断。

- 模型：$X|\theta \sim Bin(10, \theta)$，即 $P(X = x|\theta) = \binom{10}{x}\theta^x(1-\theta)^{10-x}(x = 0, 1, \ldots, 10)$
- 观察结果是 $x = 9$。
- θ 的先验分布假设为 $\theta \sim U(0, 1)$，即 $\pi(\theta) = 1\ (0 < \theta < 1)$
- θ 的后验分布为：

$$p(\theta|x) \propto \pi(\theta)f(x|\theta) \propto \theta^x(1-\theta)^{10-x}$$

即服从 $\theta|x \sim Beta(x + 1, 11 - x)$

- 需要检验：$H_0 : \theta = 0.5\ H_1 : \theta > 0.5$。
- H_1 成立的后验概率 $P(\theta > 0.5|x = 9) = 1 - \text{pbeta}(0.5, 10, 2) = 0.994$
- 结论：强烈支持 H_1(该女士真有辨别奶茶的能力)，与频率学派结论相同。

醉汉猜硬币的贝叶斯推断

- 模型：$X|\theta \sim Bin(10, \theta)$，观察结果是 $x = 9$

醉汉猜硬币的贝叶斯推断

- 模型：$X|\theta \sim Bin(10, \theta)$，观察结果是 $x = 9$
- θ 的先验分布假设为 $\theta \sim \mathrm{Beta}(3, 30)$，

醉汉猜硬币的贝叶斯推断

- 模型：$X|\theta \sim Bin(10, \theta)$，观察结果是 $x = 9$
- θ 的先验分布假设为 $\theta \sim \mathrm{Beta}(3, 30)$，
- θ 的后验分布为：

醉汉猜硬币的贝叶斯推断

- 模型：$X|\theta \sim Bin(10, \theta)$，观察结果是 $x = 9$
- θ 的先验分布假设为 $\theta \sim \text{Beta}(3, 30)$，
- θ 的后验分布为：

醉汉猜硬币的贝叶斯推断

- 模型：$X|\theta \sim Bin(10, \theta)$，观察结果是 $x = 9$
- θ 的先验分布假设为 $\theta \sim \text{Beta}(3, 30)$，
- θ 的后验分布为：

$$
\begin{aligned}
p(\theta|x) &\propto \pi(\theta) f(x|\theta) \\
&\propto \theta^{3-1}(1-\theta)^{30-1}\theta^{x}(1-\theta)^{10-x} \\
&= \theta^{x+3-1}(1-\theta)^{40-x-1}
\end{aligned}
$$

即

醉汉猜硬币的贝叶斯推断

- 模型：$X|\theta \sim Bin(10, \theta)$，观察结果是 $x = 9$
- θ 的先验分布假设为 $\theta \sim \text{Beta}(3, 30)$，
- θ 的后验分布为：

$$
\begin{aligned}
p(\theta|x) &\propto \pi(\theta) f(x|\theta) \\
&\propto \theta^{3-1}(1-\theta)^{30-1}\theta^{x}(1-\theta)^{10-x} \\
&= \theta^{x+3-1}(1-\theta)^{40-x-1}
\end{aligned}
$$

即 $\theta|x \sim Beta(x+3, 40-x)$

- 需要检验：$H_0 : \theta \leq 0.5$ $H_1 : \theta > 0.5$。

醉汉猜硬币的贝叶斯推断

- 模型：$X|\theta \sim Bin(10, \theta)$，观察结果是 $x = 9$
- θ 的先验分布假设为 $\theta \sim \mathrm{Beta}(3, 30)$，
- θ 的后验分布为：

$$
\begin{aligned}
p(\theta|x) &\propto \pi(\theta) f(x|\theta) \\
&\propto \theta^{3-1}(1-\theta)^{30-1}\theta^{x}(1-\theta)^{10-x} \\
&= \theta^{x+3-1}(1-\theta)^{40-x-1}
\end{aligned}
$$

即 $\theta|x \sim Beta(x+3, 40-x)$

- 需要检验：$H_0 : \theta \le 0.5$ $H_1 : \theta > 0.5$。
- H_1 成立的后验概率
  $P(\theta > 0.5|x = 9) = 1 - \mathrm{pbeta}(0.5, 12, 31) = 0.001$

醉汉猜硬币的贝叶斯推断

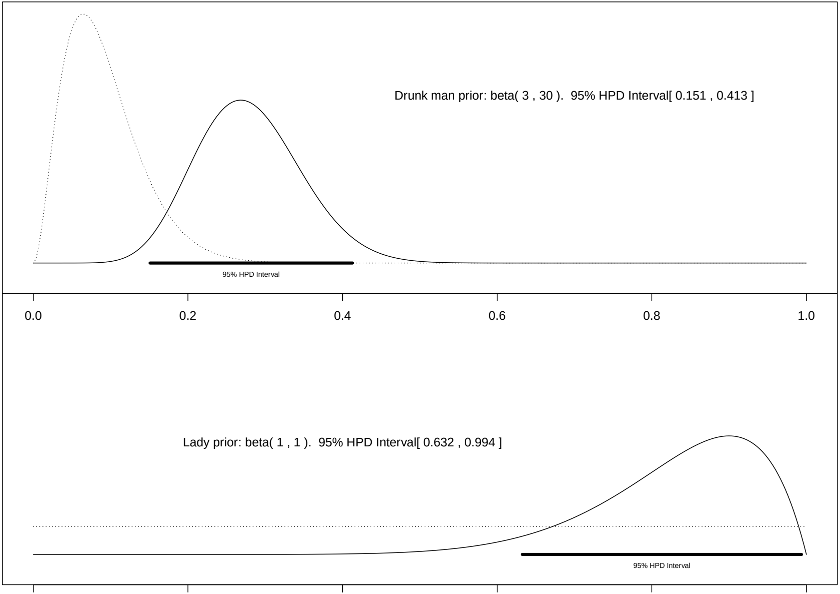
- 模型：$X|\theta \sim Bin(10, \theta)$，观察结果是 $x = 9$
- θ 的先验分布假设为 $\theta \sim \text{Beta}(3, 30)$，
- θ 的后验分布为：

$$
\begin{aligned}
p(\theta|x) &\propto \pi(\theta) f(x|\theta) \\
&\propto \theta^{3-1}(1-\theta)^{30-1}\theta^{x}(1-\theta)^{10-x} \\
&= \theta^{x+3-1}(1-\theta)^{40-x-1}
\end{aligned}
$$

即 $\theta|x \sim Beta(x + 3, 40 - x)$

- 需要检验：$H_0 : \theta \leq 0.5$ $H_1 : \theta > 0.5$。
- H_1 成立的后验概率
  $P(\theta > 0.5|x = 9) = 1 - \text{pbeta}(0.5, 12, 31) = 0.001$
- 结论：强烈支持 H_0(该醉汉没有预判硬币正反面的能力，与频率学派结论相反)

先验分布与后验分布



Drunk man prior: beta(3 , 30). 95% HPD Interval[0.151 , 0.413]
95% HPD Interval
0.0
0.2
0.4
0.6
0.8
1.0
Lady prior: beta(1 , 1). 95% HPD Interval[0.632 , 0.994]
95% HPD Interval

Outline

什么是贝叶斯统计推断？

- 贝叶斯方法是基于贝叶斯定理而发展起来用于系统地阐述和解决统计问题的方法。
- **贝叶斯推断**的基本方法是：
  1. 把总体的**未知参数**看作随机变量，在得到观察数据之前，根据经验和知识给出未知参数的概率分布，称为**先验分布** (Prior Distribution)
  2. 得到样本数据后，根据贝叶斯定理，基于给定的样本数据，得出未知参数的**后验分布** (Posterior Distribution)
  3. 根据后验分布进行统计推断
- 注意：贝叶斯统计的分析关注点是随机变量的**分布**！

贝叶斯推断的数学模型

贝叶斯模型包含如下三个部分：

1. 总体分布：$X|\theta \sim f(x|\theta)$，其中 θ 为总体的未知参数。
2. 先验分布：$\theta \sim \pi(\theta)$
3. 后验分布：$\theta|\boldsymbol{x} \sim p(\theta|\boldsymbol{x}) \propto \pi(\theta)f(\boldsymbol{x}|\theta)$

贝叶斯推断的方法：根据**后验分布**进行统计推断，包括点估计、区间估计、假设检验、预测等

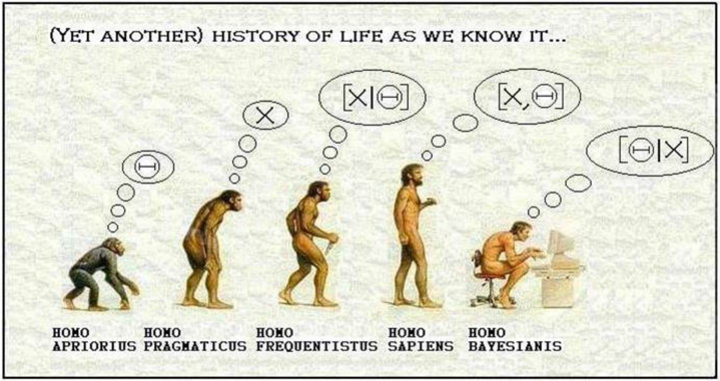
贝叶斯推断的特点：

总体信息 (总体分布) + 样本信息 (数据) + 先验信息 (参数)
= 推断结论

贝叶斯统计的发展

- 贝叶斯学派形成于 20 世纪 30 年代，但是发展缓慢。
  - 需要预先给出先验信息
  - 后验分布推算复杂
  - 受到 Fisher 等著名统计学家的压制
- 贝叶斯统计为什么在今天会如此流行？
  - 方法优势：贝叶斯统计能够把多种不同来源的信息结合起来分析，传统方法做不到
  - 算法革命：90 年代 Markov chain Monte Carlo (MCMC) 算法的发明，解决了贝叶斯统计发展的最大技术难题
  - 免费的统计软件（如 WinBUGS, JAGS, STAN, 以及 R 中各种各样的软件包）使得我们能够针对复杂现象建立复杂的统计模型，打开了广阔的应用前景
- “21 世纪将是贝叶斯统计一统天下的局面”

贝叶斯统计发展史



(YET ANOTHER) HISTORY OF LIFE AS WE KNOW IT...
[Θ]
[X]
[X|Θ]
[X,Θ]
[Θ|X]
HOMO APRIORIUS
HOMO PRAGMATICUS
HOMO FREQUENTISTUS
HOMO SAPIENS
HOMO BAYESIANIS

Outline

频率学派的统计推断

- 频率学派的推断：给定未知参数 θ，观察样本 $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)'$ 是随机变量，基于分布 $\boldsymbol{X}|\theta \sim f(\boldsymbol{x}|\theta)$ 进行推断
- 频率学派的推断方法：根据**样本的理论分布**进行点估计、区间估计、假设检验、预测等
- 频率学派推断的特点：
  总体信息 (总体分布) + 样本信息（数据）
  = 推断结论（少了先验信息）

似然函数

假设总体分布的密度函数（或概率函数）为 $f(x|\theta)$, 其中 θ 是总体未知参数。从总体中随机抽取样本 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)'$，样本观察值记为 $\boldsymbol{x}=(x_1,\ldots,x_n)^T$，随机样本 $\boldsymbol{X}$ 的联合密度函数在样本观察值 $\boldsymbol{x}$ 处取值为 $f(\boldsymbol{x}|\theta)$。
如果 $f(\boldsymbol{x}|\theta)$ 看作 θ 的函数，记为 $L(\theta|\boldsymbol{x})$，则称之为**似然函数** (Likelihood function)，记为

$$L(\theta|\boldsymbol{x})=f(\boldsymbol{x}|\theta)=\prod_{i=1}^{n}f(x_i|\theta)$$

贝叶斯定理也可以表示为：$p(\theta|\boldsymbol{x})\propto\pi(\theta)L(\theta|\boldsymbol{x})$
密度函数与似然函数的区别：

1. 密度函数：给定某一参数，求某一结果的可能性的函数
2. 似然函数：给定某一结果，求某一参数值的可能性的函数

似然函数

假设总体分布的密度函数（或概率函数）为 $f(x|\theta)$, 其中 θ 是总体未知参数。从总体中随机抽取样本 $\boldsymbol{X}=(X_1,X_2,\cdots,X_n)'$，样本观察值记为 $\boldsymbol{x}=(x_1,\ldots,x_n)^T$，随机样本 $\boldsymbol{X}$ 的联合密度函数在样本观察值 $\boldsymbol{x}$ 处取值为 $f(\boldsymbol{x}|\theta)$。

如果 $f(\boldsymbol{x}|\theta)$ 看作 θ 的函数，记为 $L(\theta|\boldsymbol{x})$，则称之为**似然函数** (Likelihood function)，记为

$$L(\theta|\boldsymbol{x})=f(\boldsymbol{x}|\theta)=\prod_{i=1}^{n}f(x_i|\theta)$$

贝叶斯定理也可以表示为：$p(\theta|\boldsymbol{x})\propto\pi(\theta)L(\theta|\boldsymbol{x})$

密度函数与似然函数的区别：

1. 密度函数：给定某一参数，求某一结果的可能性的函数
2. 似然函数：给定某一结果，求某一参数值的可能性的函数

- 问题：似然函数 $L(\theta|\boldsymbol{x})$ 是否 θ 的概率密度函数？

贝叶斯派 vs 频率学派

频率学派

- 概率：频率
- 参数：固定
- 数据：随机
- 推论：基于抽样分布
- 积分：对样本空间
- 机理：减少抽样误差

贝叶斯派

- 概率：信念程度
- 参数：随机
- 数据：固定
- 推论：基于后验分布
- 积分：对参数空间
- 机理：学习机制

对概率的不同理解

经典学派（Classical） 也称频率学派（Frequentist），认为事件的概率是在大量重复的独立试验中，事件发生的频率的极限值。

- 也叫客观概率。
- 缺点：大量事件不可重复

贝叶斯学派（Bayesian） 认为事件发生的概率就是它发生的可能性，是个人对事件发生的相信程度（degree of belief）。

- 也叫主观概率。
- 例子：看同一个人的表情举止，你判断他是小偷的概率为零，但在反扒专家眼中，他是小偷的概率很大。
- 看云判断明天下雨的概率

频率学派存在的主要问题

- 概率的频率解释不合理：需要假设大量重复观察（明天下雨概率？恐怖袭击的可能性?）

频率学派存在的主要问题

- 概率的频率解释不合理：需要假设大量重复观察（明天下雨概率？恐怖袭击的可能性?）
- 区间估计的解释有问题：置信区间的概率解释需要借助实际抽样之前的分布**假设**（该假设往往是错误的）

频率学派存在的主要问题

- 概率的频率解释不合理：需要假设大量重复观察（明天下雨概率？恐怖袭击的可能性？）
- 区间估计的解释有问题：置信区间的概率解释需要借助实际抽样之前的分布**假设**（该假设往往是错误的）
  - 该区间由一次抽样结果计算得到，但置信度的解释需要假设无穷多次重复抽样

频率学派存在的主要问题

- 概率的频率解释不合理：需要假设大量重复观察（明天下雨概率？恐怖袭击的可能性？）
- 区间估计的解释有问题：置信区间的概率解释需要借助实际抽样之前的分布**假设**（该假设往往是错误的）
  - 该区间由一次抽样结果计算得到，但置信度的解释需要假设无穷多次重复抽样
  - 99% 的 CI 可能更宽，而 90% 的 CI 可能更窄

频率学派存在的主要问题

- 概率的频率解释不合理：需要假设大量重复观察（明天下雨概率？恐怖袭击的可能性？）
- 区间估计的解释有问题：置信区间的概率解释需要借助实际抽样之前的分布**假设**（该假设往往是错误的）
  - 该区间由一次抽样结果计算得到，但置信度的解释需要假设无穷多次重复抽样
  - 99% 的 CI 可能更宽，而 90% 的 CI 可能更窄
- 假设检验及 p-值解释有问题：

频率学派存在的主要问题

- 概率的频率解释不合理：需要假设大量重复观察（明天下雨概率？恐怖袭击的可能性？）
- 区间估计的解释有问题：置信区间的概率解释需要借助实际抽样之前的分布**假设**（该假设往往是错误的）
  - 该区间由一次抽样结果计算得到，但置信度的解释需要假设无穷多次重复抽样
  - 99% 的 CI 可能更宽，而 90% 的 CI 可能更窄
- 假设检验及 p-值解释有问题：
  1. 假设检验的程序逻辑不清晰：实际差异很小也可能检验显著（如样本量很大时）

频率学派存在的主要问题

- 概率的频率解释不合理：需要假设大量重复观察（明天下雨概率？恐怖袭击的可能性？）
- 区间估计的解释有问题：置信区间的概率解释需要借助实际抽样之前的分布**假设**（该假设往往是错误的）
  - 该区间由一次抽样结果计算得到，但置信度的解释需要假设无穷多次重复抽样
  - 99% 的 CI 可能更宽，而 90% 的 CI 可能更窄
- 假设检验及 *p*-值解释有问题：
  1. 假设检验的程序逻辑不清晰：实际差异很小也可能检验显著（如样本量很大时）
  2. *p*-值不是 H_0 成立的概率

频率学派存在的主要问题

- 概率的频率解释不合理：需要假设大量重复观察（明天下雨概率？恐怖袭击的可能性？）
- 区间估计的解释有问题：置信区间的概率解释需要借助实际抽样之前的分布**假设**（该假设往往是错误的）
    - 该区间由一次抽样结果计算得到，但置信度的解释需要假设无穷多次重复抽样
    - 99% 的 CI 可能更宽，而 90% 的 CI 可能更窄
- 假设检验及 *p*-值解释有问题：
    1. 假设检验的程序逻辑不清晰：实际差异很小也可能检验显著（如样本量很大时）
    2. *p*-值不是 H_0 成立的概率
    3. *p*-值不表示数据对结论的支持程度：双样本检验中，*p*-值小不表示不同处理之间差距大

频率学派存在的主要问题

- 概率的频率解释不合理：需要假设大量重复观察（明天下雨概率？恐怖袭击的可能性？）
- 区间估计的解释有问题：置信区间的概率解释需要借助实际抽样之前的分布**假设**（该假设往往是错误的）
  - 该区间由一次抽样结果计算得到，但置信度的解释需要假设无穷多次重复抽样
  - 99% 的 CI 可能更宽，而 90% 的 CI 可能更窄
- 假设检验及 p-值解释有问题：
  1. 假设检验的程序逻辑不清晰：实际差异很小也可能检验显著（如样本量很大时）
  2. p-值不是 H_0 成立的概率
  3. p-值不表示数据对结论的支持程度：双样本检验中，p-值小不表示不同处理之间差距大

定义 (p-值)

在 H_0 成立条件下，检验统计量比实际观察值更极端的概率

频率学派存在的主要问题

- 频率学派违背**似然原理**：如果两个不同的抽样方案导致成比例的似然函数，则关于未知参数的推断结论应该一样

频率学派存在的主要问题

- 频率学派违背**似然原理**：如果两个不同的抽样方案导致成比例的似然函数，则关于未知参数的推断结论应该一样
- 例子：在 12 次试验中成功 9 次，假设每次成功概率为 θ，检验 $H_0 : \theta = 0.5,\ H_1 : \theta > 0.5$

频率学派存在的主要问题

- 频率学派违背**似然原理**：如果两个不同的抽样方案导致成比例的似然函数，则关于未知参数的推断结论应该一样
- 例子：在 12 次试验中成功 9 次，假设每次成功概率为 θ，检验 $H_0: \theta = 0.5$, $H_1: \theta > 0.5$
  - 设 X 表示 12 次试验中成功的次数，则 $X \sim B(12, \theta)$，观察结果为 $x = 9$，似然函数为

$$L_1(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = \binom{12}{9} \theta^9 (1-\theta)^3$$

频率学派存在的主要问题

- 频率学派违背**似然原理**：如果两个不同的抽样方案导致成比例的似然函数，则关于未知参数的推断结论应该一样
- 例子：在 12 次试验中成功 9 次，假设每次成功概率为 θ，检验 $H_0: \theta = 0.5,\ H_1: \theta > 0.5$
  - 设 X 表示 12 次试验中成功的次数，则 $X \sim B(12, \theta)$，观察结果为 $x = 9$，似然函数为
  $$L_1(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = \binom{12}{9} \theta^9 (1-\theta)^3$$
  - 设 Y 表示第 $r = 3$ 次失败时的试验成功的次数，则 $Y \sim NB(3, \theta)$，观察结果为 $y = 9$，似然函数为
  $$L_2(\theta) = \binom{r+y-1}{y} \theta^y (1-\theta)^r = \binom{11}{9} \theta^9 (1-\theta)^3$$

频率学派存在的主要问题

- 频率学派违背**似然原理**：如果两个不同的抽样方案导致成比例的似然函数，则关于未知参数的推断结论应该一样
- 例子：在 12 次试验中成功 9 次，假设每次成功概率为 θ，检验 $H_0: \theta = 0.5,\ H_1: \theta > 0.5$
  - 设 X 表示 12 次试验中成功的次数，则 $X \sim B(12, \theta)$，观察结果为 $x = 9$，似然函数为
  $$L_1(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = \binom{12}{9} \theta^9 (1-\theta)^3$$
  - 设 Y 表示第 $r = 3$ 次失败时的试验成功的次数，则 $Y \sim NB(3, \theta)$，观察结果为 $y = 9$，似然函数为
  $$L_2(\theta) = \binom{r+y-1}{y} \theta^y (1-\theta)^r = \binom{11}{9} \theta^9 (1-\theta)^3$$
  - 计算 p-值：结果得出矛盾结论

频率学派存在的主要问题

- 频率学派违背**似然原理**：如果两个不同的抽样方案导致成比例的似然函数，则关于未知参数的推断结论应该一样
- 例子：在 12 次试验中成功 9 次，假设每次成功概率为 θ，检验 $H_0: \theta = 0.5,\ H_1: \theta > 0.5$
  - 设 X 表示 12 次试验中成功的次数，则 $X \sim B(12, \theta)$，观察结果为 $x = 9$，似然函数为
$$L_1(\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x} = \binom{12}{9} \theta^9 (1-\theta)^3$$
  - 设 Y 表示第 $r = 3$ 次失败时的试验成功的次数，则 $Y \sim NB(3, \theta)$，观察结果为 $y = 9$，似然函数为
$$L_2(\theta) = \binom{r+y-1}{y} \theta^y (1-\theta)^r = \binom{11}{9} \theta^9 (1-\theta)^3$$
  - 计算 p-值：结果得出矛盾结论
    - ★ $p_1 = P(X \geq 9 | \theta = 0.5) = 0.075$，接受 H_0

频率学派存在的主要问题

- 频率学派违背**似然原理**：如果两个不同的抽样方案导致成比例的似然函数，则关于未知参数的推断结论应该一样
- 例子：在 12 次试验中成功 9 次，假设每次成功概率为 θ，检验 $H_0 : \theta = 0.5,\ H_1 : \theta > 0.5$
  - 设 X 表示 12 次试验中成功的次数，则 $X \sim B(12, \theta)$，观察结果为 $x = 9$，似然函数为
$$L_1(\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x} = \binom{12}{9}\theta^9(1-\theta)^3$$
  - 设 Y 表示第 $r = 3$ 次失败时的试验成功的次数，则 $Y \sim NB(3, \theta)$，观察结果为 $y = 9$，似然函数为
$$L_2(\theta) = \binom{r+y-1}{y}\theta^y(1-\theta)^r = \binom{11}{9}\theta^9(1-\theta)^3$$
  - 计算 p-值：结果得出矛盾结论
    - ★ $p_1 = P(X \geq 9|\theta = 0.5) = 0.075$，接受 H_0
    - ★ $p_2 = P(Y \geq 9|\theta = 0.5) = 0.0325$，拒绝 H_0

频率学派存在的主要问题

- 频率学派违背**似然原理**：如果两个不同的抽样方案导致成比例的似然函数，则关于未知参数的推断结论应该一样
- 例子：在 12 次试验中成功 9 次，假设每次成功概率为 θ，检验 $H_0:\theta=0.5,\ H_1:\theta>0.5$
  - 设 X 表示 12 次试验中成功的次数，则 $X\sim B(12,\theta)$，观察结果为 $x=9$，似然函数为
    $$L_1(\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x}=\binom{12}{9}\theta^9(1-\theta)^3$$
  - 设 Y 表示第 $r=3$ 次失败时的试验成功的次数，则 $Y\sim NB(3,\theta)$，观察结果为 $y=9$，似然函数为
    $$L_2(\theta)=\binom{r+y-1}{y}\theta^y(1-\theta)^r=\binom{11}{9}\theta^9(1-\theta)^3$$
  - 计算 p-值：结果得出矛盾结论
    - ★ $p_1=P(X\geq 9|\theta=0.5)=0.075$，接受 H_0
    - ★ $p_2=P(Y\geq 9|\theta=0.5)=0.0325$，拒绝 H_0
- 经济和社会研究的数据问题：非随机、内在不可重复数据；基于全部数据

贝叶斯方法的优势

- 方法具有统一性：不管什么问题，贝叶斯方法都是从贝叶斯定理出发，从而更方便应用和解释。而传统方法取决于特定问题、特定估计方法和特定模型，一类问题的推断方法不能直接应用于另一类问题或模型。

贝叶斯方法的优势

- 方法具有统一性：不管什么问题，贝叶斯方法都是从贝叶斯定理出发，从而更方便应用和解释。而传统方法取决于特定问题、特定估计方法和特定模型，一类问题的推断方法不能直接应用于另一类问题或模型。
- 更易于理解和解释：如可信区间的解释，$P(1.2 < \theta < 2.5|x) = 0.95$

贝叶斯方法的优势

- 方法具有统一性：不管什么问题，贝叶斯方法都是从贝叶斯定理出发，从而更方便应用和解释。而传统方法取决于特定问题、特定估计方法和特定模型，一类问题的推断方法不能直接应用于另一类问题或模型。
- 更易于理解和解释：如可信区间的解释，$P(1.2 < \theta < 2.5|x) = 0.95$
- 符合基本原理：满足似然原理；不会得出荒谬结论：假设 $X \sim Poisson(\lambda)$，参数 $\theta = \exp(-\lambda)$，可以证明传统的 UMVUE 为 $(-1)^x$ ！

贝叶斯方法的优势

- 方法具有统一性：不管什么问题，贝叶斯方法都是从贝叶斯定理出发，从而更方便应用和解释。而传统方法取决于特定问题、特定估计方法和特定模型，一类问题的推断方法不能直接应用于另一类问题或模型。
- 更易于理解和解释：如可信区间的解释，$P(1.2 < \theta < 2.5|x) = 0.95$
- 符合基本原理：满足似然原理；不会得出荒谬结论：假设 $X \sim Poisson(\lambda)$，参数 $\theta = \exp(-\lambda)$，可以证明传统的 UMVUE 为 $(-1)^x$！
- 符合人类认识规律（学习机制）：可以利用过去信息，不断从经验中学习。过去的认识（先验），观察新数据，先验信息更新为后验信息。。。
  —过去的认识（先验）- -观察新数据- -后验信息（修正先验认识）- -再观察新数据- -后验信息（先验再修正）- -。。。

贝叶斯方法的优势

- 方法具有统一性：不管什么问题，贝叶斯方法都是从贝叶斯定理出发，从而更方便应用和解释。而传统方法取决于特定问题、特定估计方法和特定模型，一类问题的推断方法不能直接应用于另一类问题或模型。
- 更易于理解和解释：如可信区间的解释，$P(1.2 < \theta < 2.5|x) = 0.95$
- 符合基本原理：满足似然原理；不会得出荒谬结论：假设 $X \sim Poisson(\lambda)$，参数 $\theta = \exp(-\lambda)$，可以证明传统的 UMVUE 为 $(-1)^x$!
- 符合人类认识规律（学习机制）：可以利用过去信息，不断从经验中学习。过去的认识（先验），观察新数据，先验信息更新为后验信息。。。
  —过去的认识（先验）- -观察新数据- -后验信息（修正先验认识）- -再观察新数据- -后验信息（先验再修正）- -。。。
- 方法更具弹性：不需要大样本；能处理高维甚至可变维数问题

贝叶斯方法的优势

- 方法具有统一性：不管什么问题，贝叶斯方法都是从贝叶斯定理出发，从而更方便应用和解释。而传统方法取决于特定问题、特定估计方法和特定模型，一类问题的推断方法不能直接应用于另一类问题或模型。
- 更易于理解和解释：如可信区间的解释，$P(1.2 < \theta < 2.5|x) = 0.95$
- 符合基本原理：满足似然原理；不会得出荒谬结论：假设 $X \sim Poisson(\lambda)$，参数 $\theta = \exp(-\lambda)$，可以证明传统的 UMVUE 为 $(-1)^x$!
- 符合人类认识规律（学习机制）：可以利用过去信息，不断从经验中学习。过去的认识（先验），观察新数据，先验信息更新为后验信息。。。
  —过去的认识（先验）- -观察新数据- -后验信息（修正先验认识）- -再观察新数据- -后验信息（先验再修正）- -。。。
- 方法更具弹性：不需要大样本；能处理高维甚至可变维数问题
- 广泛性：传统统计方法是贝叶斯的特例（无信息先验）

贝叶斯方法的短板

- 需要预先指定先验分布：贝叶斯明确允许主观判断（其实是优点？频率学派开了许多“主观”的后门，如重复抽样，显著性水平的选取等）

贝叶斯方法的短板

- 需要预先指定先验分布：贝叶斯明确允许主观判断（其实是优点？频率学派开了许多“主观”的后门，如重复抽样，显著性水平的选取等）
- 需要研究者丰富的经验才能建立可靠的贝叶斯模型。

贝叶斯方法的短板

- 需要预先指定先验分布：贝叶斯明确允许主观判断（其实是优点？频率学派开了许多“主观”的后门，如重复抽样，显著性水平的选取等）
- 需要研究者丰富的经验才能建立可靠的贝叶斯模型。
- 分析推导更复杂

贝叶斯方法的短板

- 需要预先指定先验分布：贝叶斯明确允许主观判断（其实是优点？频率学派开了许多“主观”的后门，如重复抽样，显著性水平的选取等）
- 需要研究者丰富的经验才能建立可靠的贝叶斯模型。
- 分析推导更复杂
- 计算成本高

贝叶斯方法的短板

- 需要预先指定先验分布：贝叶斯明确允许主观判断（其实是优点？频率学派开了许多“主观”的后门，如重复抽样，显著性水平的选取等）
- 需要研究者丰富的经验才能建立可靠的贝叶斯模型。
- 分析推导更复杂
- 计算成本高
- 目前还没有分析报告的标准

贝叶斯为什么直接用概率分布来推断？

1. 一个后验分布可以得到未知参数的各种推断：估计值、精度、区间等等

贝叶斯为什么直接用概率分布来推断？

1. 一个后验分布可以得到未知参数的各种推断：估计值、精度、区间等等
2. 不需要用 p-值：直接计算尾部概率

贝叶斯为什么直接用概率分布来推断？

1. 一个后验分布可以得到未知参数的各种推断：估计值、精度、区间等等
2. 不需要用 p-值：直接计算尾部概率
3. 不需要置信区间：直接计算后验分布的 95% 中间区域

贝叶斯为什么直接用概率分布来推断？

1. 一个后验分布可以得到未知参数的各种推断：估计值、精度、区间等等
2. 不需要用 p-值：直接计算尾部概率
3. 不需要置信区间：直接计算后验分布的 95% 中间区域
4. 很容易做预测

贝叶斯为什么直接用概率分布来推断？

1. 一个后验分布可以得到未知参数的各种推断：估计值、精度、区间等等
2. 不需要用 p-值：直接计算尾部概率
3. 不需要置信区间：直接计算后验分布的 95% 中间区域
4. 很容易做预测
5. 适配决策、风险等分析框架

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
    - 两派都要（尖峰厚尾分布？）

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$
   - 存在推导和计算问题。

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布？）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识？
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$
   - 存在推导和计算问题。
   - 计算问题已解决：马尔可夫链蒙特卡罗（MCMC）模拟

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$
   - 存在推导和计算问题。
   - 计算问题已解决：马尔可夫链蒙特卡罗（MCMC）模拟
4. 模型质量评估

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$
   - 存在推导和计算问题。
   - 计算问题已解决：马尔可夫链蒙特卡罗（MCMC）模拟
4. 模型质量评估
   - 模型检验：模型与实际数据是否符合?

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$
   - 存在推导和计算问题。
   - 计算问题已解决：马尔可夫链蒙特卡罗（MCMC）模拟
4. 模型质量评估
   - 模型检验：模型与实际数据是否符合?
   - 模型比较：哪个模型更好?

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$
   - 存在推导和计算问题。
   - 计算问题已解决：马尔可夫链蒙特卡罗（MCMC）模拟
4. 模型质量评估
   - 模型检验：模型与实际数据是否符合?
   - 模型比较：哪个模型更好?
   - 对先验分布的敏感性

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$
   - 存在推导和计算问题。
   - 计算问题已解决：马尔可夫链蒙特卡罗（MCMC）模拟
4. 模型质量评估
   - 模型检验：模型与实际数据是否符合?
   - 模型比较：哪个模型更好?
   - 对先验分布的敏感性
5. 推断：如何从后验分布中抽取有用信息?

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$
   - 存在推导和计算问题。
   - 计算问题已解决：马尔可夫链蒙特卡罗（MCMC）模拟
4. 模型质量评估
   - 模型检验：模型与实际数据是否符合?
   - 模型比较：哪个模型更好?
   - 对先验分布的敏感性
5. 推断：如何从后验分布中抽取有用信息?
   - 均值、中位数、标准差、概率等

贝叶斯分析的一般过程

1. 确定总体分布 (及似然函数):$L(\theta|\boldsymbol{x}) = f(\boldsymbol{x}|\theta)$
   - 两派都要（尖峰厚尾分布?）
2. 确定 θ 的先验分布:$\theta \sim \pi(\theta)$
   - 如何用分布来表示先验知识?
3. 确定后验分布：$p(\theta|\boldsymbol{x}) \propto \pi(\theta)L(\theta|\boldsymbol{x})$
   - 存在推导和计算问题。
   - 计算问题已解决：马尔可夫链蒙特卡罗（MCMC）模拟
4. 模型质量评估
   - 模型检验：模型与实际数据是否符合?
   - 模型比较：哪个模型更好?
   - 对先验分布的敏感性
5. 推断：如何从后验分布中抽取有用信息?
   - 均值、中位数、标准差、概率等
   - 区间估计、假设检验、预测等

Outline

1. 课程介绍
2. 贝叶斯定理
3. 贝叶斯推断
4. 频率学派推断与贝叶斯推断的差异
5. 教材及参考文献

Gill, Jeff
Bayesian Methods: A Social and Behavioral Sciences Approach, Third Edition, Chapman & Hall/CRC Press, 2015. (With R package BaM)

Gelman,A.,et al.
Bayesian Data Analysis (Third edition), Chapman and Hall, 2013.

Greenberg,E.
Introduction to Bayesian Econometrics (Second edition), Cambridge University Press, 2012.

Lunn,D.,et al.
The BUGS Book: A Practical Introduction to Bayesian Analysis. CRC Press, 2012.