

贝 叶 斯 统 计

韦 来 生 编 著

高 等 教 育 出 版 社

第一版, 2016

贝 叶 斯 统 计

韦 来 生 编 著

高 等 教 育 出 版 社

第一版, 2016

第2章 先验分布的选取

1 2.1 主观概率

- 2.1.1 主观概率的定义
- 2.1.2 确定主观概率的方法

2 2.2 利用先验信息确定先验分布

- 2.2.1 直方图法
- 2.2.2 相对似然法
- 2.2.3 选定先验密度函数的形式, 再估计超参数
- 2.2.4 给定参数 θ 分位数, 确定累计分布函数

3 2.3 利用边缘分布 $m(x)$ 确定先验分布

- 2.3.1 边缘分布的定义及例
- 2.3.2 选择先验分布的 ML-II 方法
- 2.3.3 选择先验分布的矩方法

4 2.4 无信息先验

- 2.4.1 Laplace 先验与广义先验分布
- 2.4.2 位置参数的无信息先验
- 2.4.3 刻度参数的无信息先验
- 2.4.4 一般情形下的无信息先验

5 2.5 共轭先验分布

引言和定义

- 在经典的统计学中概率是用公理化定义的,即用非负性、正则性和可加性三条公理定义的. 概率的确定有两种方法: 一是古典方法 (包括几何方法); 另一种是频率方法. 故经典统计研究的对象是能大量重复的随机现象, 不是这类随机现象就不能用频率方法去确定有关事件的概率.
- 在诸多社会现象、经济领域和决策问题中, “事件”常常是不能大量重复的. 如气象预报“明天是晴天的概率为 0.8”, 就不能用频率去解释, 因天气随时间变化而变化, 不可重复.
- 因此主观概率的创立, 使我们在频率解释不能适用的情形也能讨论概率. 从这个意义上说, 主观概率至少是确定概率的频率方法和古典方法的补充.
- **定义 7.2.1** 主观概率是人们根据经验对事件发生机会的个人信念.

引言和定义

- 在经典的统计学中概率是用公理化定义的,即用非负性、正则性和可加性三条公理定义的. 概率的确定有两种方法: 一是古典方法 (包括几何方法); 另一种是频率方法. 故经典统计研究的对象是能大量重复的随机现象, 不是这类随机现象就不能用频率方法去确定有关事件的概率.
- 在诸多社会现象、经济领域和决策问题中, “事件”常常是不能大量重复的. 如气象预报“明天是晴天的概率为 0.8”, 就不能用频率去解释, 因天气随时间变化而变化, 不可重复.
- 因此主观概率的创立, 使我们在频率解释不能适用的情形也能讨论概率. 从这个意义上说, 主观概率至少是确定概率的频率方法和古典方法的补充.
- **定义 7.2.1** 主观概率是人们根据经验对事件发生机会的个人信念.

引言和定义

- 在经典的统计学中概率是用公理化定义的,即用非负性、正则性和可加性三条公理定义的. 概率的确定有两种方法: 一是古典方法 (包括几何方法); 另一种是频率方法. 故经典统计研究的对象是能大量重复的随机现象, 不是这类随机现象就不能用频率方法去确定有关事件的概率.
- 在诸多社会现象、经济领域和决策问题中, “事件”常常是不能大量重复的. 如气象预报“明天是晴天的概率为 0.8”, 就不能用频率去解释, 因天气随时间变化而变化, 不可重复.
- 因此主观概率的创立, 使我们在频率解释不能适用的情形也能讨论概率. 从这个意义上说, 主观概率至少是确定概率的频率方法和古典方法的补充.
- **定义 7.2.1** 主观概率是人们根据经验对事件发生机会的个人信念.

引言和定义

- 在经典的统计学中概率是用公理化定义的,即用非负性、正则性和可加性三条公理定义的. 概率的确定有两种方法: 一是古典方法 (包括几何方法); 另一种是频率方法. 故经典统计研究的对象是能大量重复的随机现象, 不是这类随机现象就不能用频率方法去确定有关事件的概率.
- 在诸多社会现象、经济领域和决策问题中, “事件” 常常是不能大量重复的. 如气象预报 “明天是晴天的概率为 0.8”, 就不能用频率去解释, 因天气随时间变化而变化, 不可重复.
- 因此主观概率的创立, 使我们在频率解释不能适用的情形也能讨论概率. 从这个意义上说, 主观概率至少是确定概率的频率方法和古典方法的补充.
- **定义 7.2.1** 主观概率是人们根据经验对事件发生机会的个人信念.

确定主观概率的方法

- 对一些事件进行对比, 确定相对似然性是确定主观概率最简单的办法. 例如, 某工厂已设计好一种新玩具, 决策者要评估新玩具畅销的概率. 根据新玩具的特点和经验认为畅销 (A) 是不畅销 (\bar{A}) 的可能性的 2 倍, 即 $P(A) = 2/3$, $P(\bar{A}) = 1/3$.
- 利用专家意见来确定主观概率的方法是常用的. 例如, 有一项带有风险的生意, 欲估计成功的概率. 为此向专家请教这生意成功的可能性有多大. 专家回答大约 0.6, 若他的估计保守, 可将成功概率修改为 0.7 (也可以请教几位专家评估, 用成功概率的平均值代替更可靠).
- 利用历史资料, 做一些对比修正, 确定主观概率. 例如, 某公司经营玩具, 现设计一种新式玩具将投放市场, 要估计未来市场销售情况. 经理查阅了本公司生产的 37 种玩具的销售记录, 得销售状态如下: A_1 : 畅销, A_2 : 一般, A_3 : 滞销, 分别有 29, 6, 2 种, 由历史资料得到销售状态的概率. 根据新玩具的特点, 对上述概率作修正确定其销售状态的主观概率.

确定主观概率的方法

- 对一些事件进行对比, 确定相对似然性是确定主观概率最简单的办法. 例如, 某工厂已设计好一种新玩具, 决策者要评估新玩具畅销的概率. 根据新玩具的特点和经验认为畅销 (A) 是不畅销 (\bar{A}) 的可能性的 2 倍, 即 $P(A) = 2/3$, $P(\bar{A}) = 1/3$.
- 利用专家意见来确定主观概率的方法是常用的. 例如, 有一项带有风险的生意, 欲估计成功的概率. 为此向专家请教这生意成功的可能性有多大. 专家回答大约 0.6, 若他的估计保守, 可将成功概率修改为 0.7 (也可以请教几位专家评估, 用成功概率的平均值代替更可靠).
- 利用历史资料, 做一些对比修正, 确定主观概率. 例如, 某公司经营玩具, 现设计一种新式玩具将投放市场, 要估计未来市场销售情况. 经理查阅了本公司生产的 37 种玩具的销售记录, 得销售状态如下: A_1 : 畅销, A_2 : 一般, A_3 : 滞销, 分别有 29, 6, 2 种, 由历史资料得到销售状态的概率. 根据新玩具的特点, 对上述概率作修正确定其销售状态的主观概率.

确定主观概率的方法

- 对一些事件进行对比, 确定相对似然性是确定主观概率最简单的办法. 例如, 某工厂已设计好一种新玩具, 决策者要评估新玩具畅销的概率. 根据新玩具的特点和经验认为畅销 (A) 是不畅销 (\bar{A}) 的可能性的 2 倍, 即 $P(A) = 2/3$, $P(\bar{A}) = 1/3$.
- 利用专家意见来确定主观概率的方法是常用的. 例如, 有一项带有风险的生意, 欲估计成功的概率. 为此向专家请教这生意成功的可能性有多大. 专家回答大约 0.6, 若他的估计保守, 可将成功概率修改为 0.7 (也可以请教几位专家评估, 用成功概率的平均值代替更可靠).
- 利用历史资料, 做一些对比修正, 确定主观概率. 例如, 某公司经营玩具, 现设计一种新式玩具将投放市场, 要估计未来市场销售情况. 经理查阅了本公司生产的 37 种玩具的销售记录, 得销售状态如下: A_1 : 畅销, A_2 : 一般, A_3 : 滞销, 分别有 29, 6, 2 种, 由历史资料得到销售状态的概率. 根据新玩具的特点, 对上述概率作修正确定其销售状态的主观概率.

1 2.1 主观概率

- 2.1.1 主观概率的定义
- 2.1.2 确定主观概率的方法

2 2.2 利用先验信息确定先验分布

- 2.2.1 直方图法
- 2.2.2 相对似然法
- 2.2.3 选定先验密度函数的形式, 再估计超参数
- 2.2.4 给定参数 θ 分位数, 确定累计分布函数

3 2.3 利用边缘分布 $m(x)$ 确定先验分布

- 2.3.1 边缘分布的定义及例
- 2.3.2 选择先验分布的 ML-II 方法
- 2.3.3 选择先验分布的矩方法

4 2.4 无信息先验

- 2.4.1 Laplace 先验与广义先验分布
- 2.4.2 位置参数的无信息先验
- 2.4.3 刻度参数的无信息先验
- 2.4.4 一般情形下的无信息先验

5 2.5 共轭先验分布

直方图法

- 这个方法与一般直方图法类似, 可以通过下述步骤实现:
 - (1) 当参数空间 Θ 为实轴上的区间时, 先把 Θ 分成一些小小区间, 通常为等长的子区间.
 - (2) 在每个小区间上决定主观概率或按历史数据算出频率.
 - (3) 绘制直方图, 纵坐标为主观概率或“频率/(小区间长)”
 - (4) 在直方图上画一光滑曲线, 使下方与直方图面积相等. 此曲线即为先验密度 $\pi(\theta)$ (使曲线与横轴形成的面积为 1).
- 例 2.2.1 某医药公司销售人参, 记录了 102 周的销售量, 每周销售最多 35kg, 数据见下表. 要寻求每周平均销售量 θ 的概率分布.

表 2.2.1 每周平均销售量统计表

销售量(kg)	[0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
天 数	5	27	34	22	10	3	1
频 率	0.05	0.26	0.33	0.22	0.10	0.03	0.01

直方图法

- 这个方法与一般直方图法类似, 可以通过下述步骤实现:
 - (1) 当参数空间 Θ 为实轴上的区间时, 先把 Θ 分成一些小小区间, 通常为等长的子区间.
 - (2) 在每个小区间上决定主观概率或按历史数据算出频率.
 - (3) 绘制直方图, 纵坐标为主观概率或“频率/(小区间长)”
 - (4) 在直方图上画一光滑曲线, 使下方与直方图面积相等. 此曲线即为先验密度 $\pi(\theta)$ (使曲线与横轴形成的面积为 1).
- 例 2.2.1 某医药公司销售人参, 记录了 102 周的销售量, 每周销售最多 35kg, 数据见下表. 要寻求每周平均销售量 θ 的概率分布.

表 2.2.1 每周平均销售量统计表

销售量(kg)	[0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
天 数	5	27	34	22	10	3	1
频 率	0.05	0.26	0.33	0.22	0.10	0.03	0.01

直方图法

- 这个方法与一般直方图法类似, 可以通过下述步骤实现:
 - (1) 当参数空间 Θ 为实轴上的区间时, 先把 Θ 分成一些小区间, 通常为等长的子区间.
 - (2) 在每个小区间上决定主观概率或按历史数据算出频率.
 - (3) 绘制直方图, 纵坐标为主观概率或“频率/(小区间长)”
 - (4) 在直方图上画一光滑曲线, 使下方与直方图面积相等. 此曲线即为先验密度 $\pi(\theta)$ (使曲线与横轴形成的面积为 1).
- 例 2.2.1 某医药公司销售人参, 记录了 102 周的销售量, 每周销售最多 35kg, 数据见下表. 要寻求每周平均销售量 θ 的概率分布.

表 2.2.1 每周平均销售量统计表

销售量(kg)	[0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
天 数	5	27	34	22	10	3	1
频 率	0.05	0.26	0.33	0.22	0.10	0.03	0.01

直方图法

- 这个方法与一般直方图法类似, 可以通过下述步骤实现:
 - (1) 当参数空间 Θ 为实轴上的区间时, 先把 Θ 分成一些小小区间, 通常为等长的子区间.
 - (2) 在每个小区间上决定主观概率或按历史数据算出频率.
 - (3) 绘制直方图, 纵坐标为主观概率或“频率/(小区间长)”
 - (4) 在直方图上画一光滑曲线, 使下方与直方图面积相等. 此曲线即为先验密度 $\pi(\theta)$ (使曲线与横轴形成的面积为 1).
- 例 2.2.1 某医药公司销售人参, 记录了 102 周的销售量, 每周销售最多 35kg, 数据见下表. 要寻求每周平均销售量 θ 的概率分布.

表 2.2.1 每周平均销售量统计表

销售量(kg)	[0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
天 数	5	27	34	22	10	3	1
频 率	0.05	0.26	0.33	0.22	0.10	0.03	0.01

直方图法

- 这个方法与一般直方图法类似, 可以通过下述步骤实现:
 - (1) 当参数空间 Θ 为实轴上的区间时, 先把 Θ 分成一些小小区间, 通常为等长的子区间.
 - (2) 在每个小区间上决定主观概率或按历史数据算出频率.
 - (3) 绘制直方图, 纵坐标为主观概率或“频率/(小区间长)”
 - (4) 在直方图上画一光滑曲线, 使下方与直方图面积相等. 此曲线即为先验密度 $\pi(\theta)$ (使曲线与横轴形成的面积为 1).
- 例 2.2.1 某医药公司销售人参, 记录了 102 周的销售量, 每周销售最多 35kg, 数据见下表. 要寻求每周平均销售量 θ 的概率分布.

表 2.2.1 每周平均销售量统计表

销售量(kg)	[0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
天 数	5	27	34	22	10	3	1
频 率	0.05	0.26	0.33	0.22	0.10	0.03	0.01

直方图法

- 这个方法与一般直方图法类似, 可以通过下述步骤实现:
 - (1) 当参数空间 Θ 为实轴上的区间时, 先把 Θ 分成一些小小区间, 通常为等长的子区间.
 - (2) 在每个小区间上决定主观概率或按历史数据算出频率.
 - (3) 绘制直方图, 纵坐标为主观概率或“频率/(小区间长)”
 - (4) 在直方图上画一光滑曲线, 使下方与直方图面积相等. 此曲线即为先验密度 $\pi(\theta)$ (使曲线与横轴形成的面积为 1).
- 例 2.2.1** 某医药公司销售人参, 记录了 102 周的销售量, 每周销售最多 35kg, 数据见下表. 要寻求每周平均销售量 θ 的概率分布.

表 2.2.1 每周平均销售量统计表

销售量(kg)	[0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
天 数	5	27	34	22	10	3	1
频 率	0.05	0.26	0.33	0.22	0.10	0.03	0.01

直方图法

- 这个方法与一般直方图法类似, 可以通过下述步骤实现:
 - (1) 当参数空间 Θ 为实轴上的区间时, 先把 Θ 分成一些小小区间, 通常为等长的子区间.
 - (2) 在每个小区间上决定主观概率或按历史数据算出频率.
 - (3) 绘制直方图, 纵坐标为主观概率或“频率/(小区间长)”
 - (4) 在直方图上画一光滑曲线, 使下方与直方图面积相等. 此曲线即为先验密度 $\pi(\theta)$ (使曲线与横轴形成的面积为 1).
- 例 2.2.1** 某医药公司销售人参, 记录了 102 周的销售量, 每周销售最多 35kg, 数据见下表. 要寻求每周平均销售量 θ 的概率分布.

表 2.2.1 每周平均销售量统计表

销售量(kg)	[0,5]	(5,10]	(10,15]	(15,20]	(20,25]	(25,30]	(30,35]
天 数	5	27	34	22	10	3	1
频 率	0.05	0.26	0.33	0.22	0.10	0.03	0.01

周平均销售量直方图

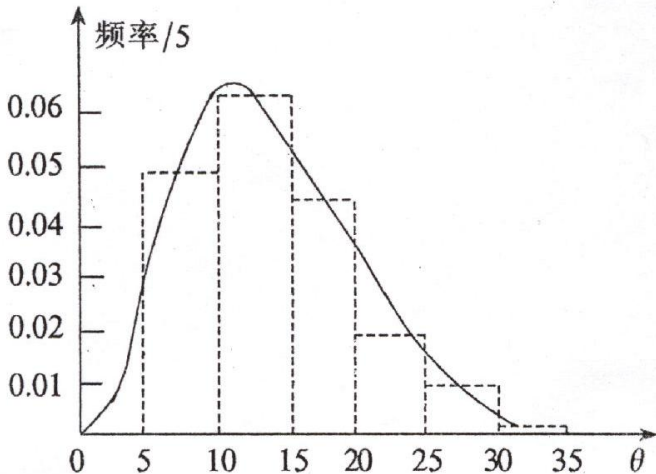


图 2.2.1 周平均销售量直方图

相对似然法

- 此法大多用于 Θ 为 $(-\infty, \infty)$ 的有限子区间的情形. 方法如下: 对 Θ 中的各种点的直观“似然”进行比较, 再按确定了的值画图, 即可得到先验密度草图, 用下例来作说明.
- 例如, 设 $\Theta = (0, 1)$, 从确定“最大可能”和“最小可能”的参数点的似然性入手. 设 $\theta = 3/4$ 为最大可能的点, $\theta = 0$ 为最小可能的点, 且 $\theta = 3/4$ 为 $\theta = 0$ 的似然性的 3 倍. 再确定 $\theta = 1/4$ 和 $\theta = 1/2$ 及 $\theta = 1$ 的相对似然性. 为简单计, 与 $\theta = 0$ 的可能性比较, $\theta = 1/2$ 和 $\theta = 1$ 的可能性 2 倍于 $\theta = 0$, $\theta = 1/4$ 的可能性为 $\theta = 0$ 的可能性的 1.5 倍. 令基本点 $\theta = 0$ 的先验密度为 1, 由此画出 $\tilde{\pi}(\theta)$.
- 但 $\int_0^1 \tilde{\pi}(\theta) d\theta \neq 1$. 记 $\pi(\theta) = c\tilde{\pi}(\theta)$, 使 $\int_0^1 \pi(\theta) d\theta = c \int_0^1 \tilde{\pi}(\theta) d\theta = 1$, 则 $\pi(\theta)$ 即为 θ 的先验密度.

相对似然法

- 此法大多用于 Θ 为 $(-\infty, \infty)$ 的有限子区间的情形. 方法如下: 对 Θ 中的各种点的直观“似然”进行比较, 再按确定了的值画图, 即可得到先验密度草图, 用下例来作说明.
- 例如, 设 $\Theta = (0, 1)$, 从确定“最大可能”和“最小可能”的参数点的似然性入手. 设 $\theta = 3/4$ 为最大可能的点, $\theta = 0$ 为最小可能的点, 且 $\theta = 3/4$ 为 $\theta = 0$ 的似然性的 3 倍. 再确定 $\theta = 1/4$ 和 $\theta = 1/2$ 及 $\theta = 1$ 的相对似然性. 为简单计, 与 $\theta = 0$ 的可能性比较, $\theta = 1/2$ 和 $\theta = 1$ 的可能性 2 倍于 $\theta = 0$, $\theta = 1/4$ 的可能性为 $\theta = 0$ 的可能性的 1.5 倍. 令基本点 $\theta = 0$ 的先验密度为 1, 由此画出 $\tilde{\pi}(\theta)$.
- 但 $\int_0^1 \tilde{\pi}(\theta) d\theta \neq 1$. 记 $\pi(\theta) = c\tilde{\pi}(\theta)$, 使 $\int_0^1 \pi(\theta) d\theta = c \int_0^1 \tilde{\pi}(\theta) d\theta = 1$, 则 $\pi(\theta)$ 即为 θ 的先验密度.

相对似然法

- 此法大多用于 Θ 为 $(-\infty, \infty)$ 的有限子区间的情形. 方法如下: 对 Θ 中的各种点的直观“似然”进行比较, 再按确定了的值画图, 即可得到先验密度草图, 用下例来作说明.
- 例如, 设 $\Theta = (0, 1)$, 从确定“最大可能”和“最小可能”的参数点的似然性入手. 设 $\theta = 3/4$ 为最大可能的点, $\theta = 0$ 为最小可能的点, 且 $\theta = 3/4$ 为 $\theta = 0$ 的似然性的 3 倍. 再确定 $\theta = 1/4$ 和 $\theta = 1/2$ 及 $\theta = 1$ 的相对似然性. 为简单计, 与 $\theta = 0$ 的可能性比较, $\theta = 1/2$ 和 $\theta = 1$ 的可能性 2 倍于 $\theta = 0$, $\theta = 1/4$ 的可能性为 $\theta = 0$ 的可能性的 1.5 倍. 令基本点 $\theta = 0$ 的先验密度为 1, 由此画出 $\tilde{\pi}(\theta)$.
- 但 $\int_0^1 \tilde{\pi}(\theta) d\theta \neq 1$. 记 $\pi(\theta) = c\tilde{\pi}(\theta)$, 使 $\int_0^1 \pi(\theta) d\theta = c \int_0^1 \tilde{\pi}(\theta) d\theta = 1$, 则 $\pi(\theta)$ 即为 θ 的先验密度.

相对似然性图

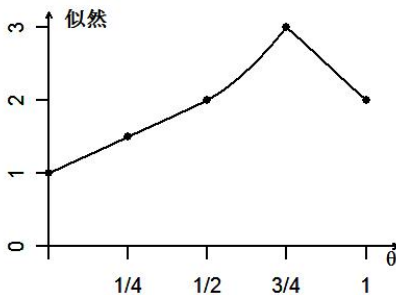


图 2.2.2 相对似然性

选定先验密度函数的形式, 再估计超参数

- **定义 2.2.1** 先验分布中的参数称为超参数.
- 例如, 根据先验信息, 选定 θ 的先验密度为 $\pi(\theta)$ 为 $N(\mu, \tau^2)$, 其中 μ 和 τ^2 称为超参数超参数.
- 一般方法: 设超参数为 α, β , 记先验密度为 $\pi(\theta; \alpha, \beta)$, 对超参数 α 和 β 作出估计, 其估计量为 $\hat{\alpha}$ 和 $\hat{\beta}$, 使 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 很接近于 $\pi(\theta; \alpha, \beta)$, 则 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 即为选定的先验密度函数.
- 这个方法的关键是 $\pi(\theta)$ 的形式的选定. 若选择不合适将导致失误.
- **例 2.2.2** 在例 2.2.1 中设参数 θ 为销售量, 选用 $N(\mu, \tau^2)$ 作为 θ 的先验分布 $\pi(\theta)$, 试确定这一先验分布.

选定先验密度函数的形式, 再估计超参数

- **定义 2.2.1** 先验分布中的参数称为超参数.
- 例如, 根据先验信息, 选定 θ 的先验密度为 $\pi(\theta)$ 为 $N(\mu, \tau^2)$, 其中 μ 和 τ^2 称为超参数超参数.
- 一般方法: 设超参数为 α, β , 记先验密度为 $\pi(\theta; \alpha, \beta)$, 对超参数 α 和 β 作出估计, 其估计量为 $\hat{\alpha}$ 和 $\hat{\beta}$, 使 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 很接近于 $\pi(\theta; \alpha, \beta)$, 则 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 即为选定的先验密度函数.
- 这个方法的关键是 $\pi(\theta)$ 的形式的选定. 若选择不合适将导致失误.
- **例 2.2.2** 在例 2.2.1 中设参数 θ 为销售量, 选用 $N(\mu, \tau^2)$ 作为 θ 的先验分布 $\pi(\theta)$, 试确定这一先验分布.

选定先验密度函数的形式, 再估计超参数

- **定义 2.2.1** 先验分布中的参数称为超参数.
- 例如, 根据先验信息, 选定 θ 的先验密度为 $\pi(\theta)$ 为 $N(\mu, \tau^2)$, 其中 μ 和 τ^2 称为超参数超参数.
- 一般方法: 设超参数为 α, β , 记先验密度为 $\pi(\theta; \alpha, \beta)$, 对超参数 α 和 β 作出估计, 其估计量为 $\hat{\alpha}$ 和 $\hat{\beta}$, 使 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 很接近于 $\pi(\theta; \alpha, \beta)$, 则 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 即为选定的先验密度函数.
- 这个方法的关键是 $\pi(\theta)$ 的形式的选定. 若选择不合适将导致失误.
- **例 2.2.2** 在例 2.2.1 中设参数 θ 为销售量, 选用 $N(\mu, \tau^2)$ 作为 θ 的先验分布 $\pi(\theta)$, 试确定这一先验分布.

选定先验密度函数的形式, 再估计超参数

- **定义 2.2.1** 先验分布中的参数称为超参数.
- 例如, 根据先验信息, 选定 θ 的先验密度为 $\pi(\theta)$ 为 $N(\mu, \tau^2)$, 其中 μ 和 τ^2 称为超参数超参数.
- 一般方法: 设超参数为 α, β , 记先验密度为 $\pi(\theta; \alpha, \beta)$, 对超参数 α 和 β 作出估计, 其估计量为 $\hat{\alpha}$ 和 $\hat{\beta}$, 使 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 很接近于 $\pi(\theta; \alpha, \beta)$, 则 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 即为选定的先验密度函数.
- 这个方法的关键是 $\pi(\theta)$ 的形式的选定. 若选择不合适将导致失误.
- **例 2.2.2** 在例 2.2.1 中设参数 θ 为销售量, 选用 $N(\mu, \tau^2)$ 作为 θ 的先验分布 $\pi(\theta)$, 试确定这一先验分布.

选定先验密度函数的形式, 再估计超参数

- **定义 2.2.1** 先验分布中的参数称为超参数.
- 例如, 根据先验信息, 选定 θ 的先验密度为 $\pi(\theta)$ 为 $N(\mu, \tau^2)$, 其中 μ 和 τ^2 称为超参数超参数.
- 一般方法: 设超参数为 α, β , 记先验密度为 $\pi(\theta; \alpha, \beta)$, 对超参数 α 和 β 作出估计, 其估计量为 $\hat{\alpha}$ 和 $\hat{\beta}$, 使 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 很接近于 $\pi(\theta; \alpha, \beta)$, 则 $\pi(\theta; \hat{\alpha}, \hat{\beta})$ 即为选定的先验密度函数.
- 这个方法的关键是 $\pi(\theta)$ 的形式的选定. 若选择不合适将导致失误.
- **例 2.2.2** 在例 2.2.1 中设参数 θ 为销售量, 选用 $N(\mu, \tau^2)$ 作为 θ 的先验分布 $\pi(\theta)$, 试确定这一先验分布.

选定先验密度函数的形式, 再估计超参数

- 在给定先验分布形式时, 决定超参数的另一方法是从先验信息中获得几个分位数的估计值, 然后通过这些分位数的值确定超参数 (即超参数的估计值), 请看下例.
- 例 2.2.3 设参数 θ 的取值范围为 $(-\infty, \infty)$, 先验分布为正态分布, 若从先验信息得知: (1) 先验的中位数为 0, (2) 其 0.25 和 0.75 的分位数分别为 -1 和 $+1$, 试求此先验分布.
- 又若在本例中假定 θ 不是正态分布, 而是柯西分布, 其余条件不变, 即 $\theta \sim \pi(\theta; \alpha, \beta) = \beta / \{\pi[\beta^2 + (\theta - \alpha)^2]\}$, $\theta \in R_1$, 确定先验分布的问题就转化为求 α, β 的估计.
- 因此, 同样的先验信息有 2 个先验分布可供选择. 若 2 个先验分布差别不大, 可任选其一. 本例中 $N(0, 1.481^2)$ 和 $C(0, 1)$ 密度函数形状上相似, 但柯西分布的尾部概率较大. 因此, 若 θ 的先验信息集中在中间, 则选择正态好些, 若先验信息较分散, 选择柯西分布更合适些.

选定先验密度函数的形式, 再估计超参数

- 在给定先验分布形式时, 决定超参数的另一方法是从先验信息中获得几个分位数的估计值, 然后通过这些分位数的值确定超参数 (即超参数的估计值), 请看下例.
- 例 2.2.3** 设参数 θ 的取值范围为 $(-\infty, \infty)$, 先验分布为正态分布, 若从先验信息得知: (1) 先验的中位数为 0, (2) 其 0.25 和 0.75 的分位数分别为 -1 和 $+1$, 试求此先验分布.
- 又若在本例中假定 θ 不是正态分布, 而是柯西分布, 其余条件不变, 即 $\theta \sim \pi(\theta; \alpha, \beta) = \beta / \{\pi[\beta^2 + (\theta - \alpha)^2]\}$, $\theta \in R_1$, 确定先验分布的问题就转化为求 α, β 的估计.
- 因此, 同样的先验信息有 2 个先验分布可供选择. 若 2 个先验分布差别不大, 可任选其一. 本例中 $N(0, 1.481^2)$ 和 $C(0, 1)$ 密度函数形状上相似, 但柯西分布的尾部概率较大. 因此, 若 θ 的先验信息集中在中间, 则选择正态好些, 若先验信息较分散, 选择柯西分布更合适些.

选定先验密度函数的形式, 再估计超参数

- 在给定先验分布形式时, 决定超参数的另一方法是从先验信息中获得几个分位数的估计值, 然后通过这些分位数的值确定超参数 (即超参数的估计值), 请看下例.
- 例 2.2.3** 设参数 θ 的取值范围为 $(-\infty, \infty)$, 先验分布为正态分布, 若从先验信息得知: (1) 先验的中位数为 0, (2) 其 0.25 和 0.75 的分位数分别为 -1 和 $+1$, 试求此先验分布.
- 又若在本例中假定 θ 不是正态分布, 而是柯西分布, 其余条件不变, 即 $\theta \sim \pi(\theta; \alpha, \beta) = \beta / \{\pi[\beta^2 + (\theta - \alpha)^2]\}$, $\theta \in R_1$, 确定先验分布的问题就转化为求 α, β 的估计.
- 因此, 同样的先验信息有 2 个先验分布可供选择. 若 2 个先验分布差别不大, 可任选其一. 本例中 $N(0, 1.481^2)$ 和 $C(0, 1)$ 密度函数形状上相似, 但柯西分布的尾部概率较大. 因此, 若 θ 的先验信息集中在中间, 则选择正态好些, 若先验信息较分散, 选择柯西分布更合适些.

选定先验密度函数的形式, 再估计超参数

- 在给定先验分布形式时, 决定超参数的另一方法是从先验信息中获得几个分位数的估计值, 然后通过这些分位数的值确定超参数 (即超参数的估计值), 请看下例.
- 例 2.2.3** 设参数 θ 的取值范围为 $(-\infty, \infty)$, 先验分布为正态分布, 若从先验信息得知: (1) 先验的中位数为 0, (2) 其 0.25 和 0.75 的分位数分别为 -1 和 $+1$, 试求此先验分布.
- 又若在本例中假定 θ 不是正态分布, 而是柯西分布, 其余条件不变, 即 $\theta \sim \pi(\theta; \alpha, \beta) = \beta / \{\pi[\beta^2 + (\theta - \alpha)^2]\}$, $\theta \in R_1$, 确定先验分布的问题就转化为求 α, β 的估计.
- 因此, 同样的先验信息有 2 个先验分布可供选择. 若 2 个先验分布差别不大, 可任选其一. 本例中 $N(0, 1.481^2)$ 和 $C(0, 1)$ 密度函数形状上相似, 但柯西分布的尾部概率较大. 因此, 若 θ 的先验信息集中在中间, 则选择正态好些, 若先验信息较分散, 选择柯西分布更合适些.

2.2 利用先验信息确定先验分布

2.2.4 给定参数 θ 分位数, 确定累计分布函数租金 θ 的分位数表

例 2.2.4 设某仓库租金 θ 在区间 $[2, 2.4]$ 中取值, 其几个分位数点 $(\alpha, z(\alpha))$ 见表 2.2.2, 求租金 θ 的 CDF 曲线和概率直方图.

表 2.2.2 租金 θ 的累积概率表

分位数 (α)	累积概率 ($z(\alpha)$)
2.00	0
2.10	0.125
2.20	0.500
2.30	0.875
2.40	1.000

2.2 利用先验信息确定先验分布

2.2.4 给定参数 θ 分位数, 确定累计分布函数租金 θ 的分位数表

例 2.2.4 设某仓库租金 θ 在区间 $[2, 2.4]$ 中取值, 其几个分位数点 $(\alpha, z(\alpha))$ 见表 2.2.2, 求租金 θ 的 CDF 曲线和概率直方图.

表 2.2.2 租金 θ 的累积概率表

分位数 (α)	累积概率 ($z(\alpha)$)
2.00	0
2.10	0.125
2.20	0.500
2.30	0.875
2.40	1.000

一般方法

- 这个方法是根据先验信息给出租金 θ 的几个分位数 α 的累积概率 $z(\alpha)$, 得到几个点 $(\alpha, z(\alpha))$;
- 在平面上将这几个点用一条光滑曲线连接起来得到累积分布函数, 简记为 CDF;
- 还可以通过先验分布的 CDF, 获得先验分布的概率直方图, 从而获得先验分布的概率密度函数.

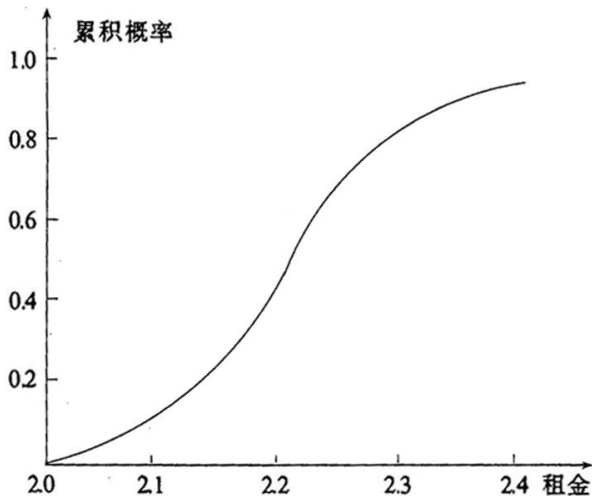
一般方法

- 这个方法是根据先验信息给出租金 θ 的几个分位数 α 的累积概率 $z(\alpha)$, 得到几个点 $(\alpha, z(\alpha))$;
- 在平面上将这几个点用一条光滑曲线连接起来得到累积分布函数, 简记为 CDF;
- 还可以通过先验分布的 CDF, 获得先验分布的概率直方图, 从而获得先验分布的概率密度函数.

一般方法

- 这个方法是根据先验信息给出租金 θ 的几个分位数 α 的累积概率 $z(\alpha)$, 得到几个点 $(\alpha, z(\alpha))$;
- 在平面上将这几个点用一条光滑曲线连接起来得到累积分布函数，简记为 CDF;
- 还可以通过先验分布的 CDF, 获得先验分布的概率直方图，从而获得先验分布的概率密度函数.

2.2 利用先验信息确定先验分布

2.2.4 给定参数 θ 分位数，确定累计分布函数租金 θ 累积概率曲线图图 2.2.4 租金 θ 累积概率曲线图

2.2 利用先验信息确定先验分布

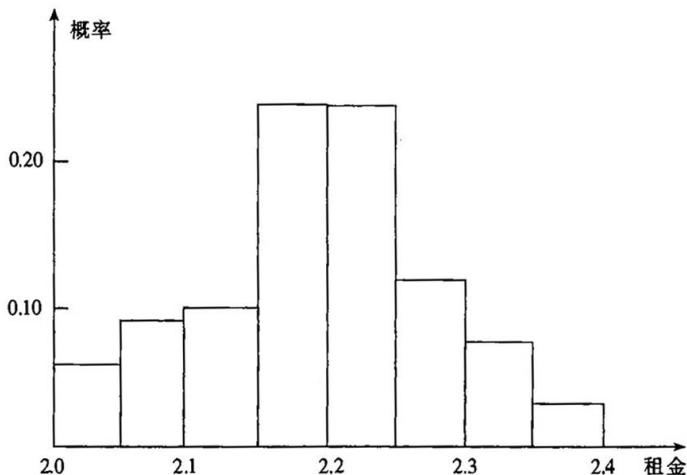
2.2.4 给定参数 θ 分位数, 确定累计分布函数租金 θ 的概率分布表表 2.2.3 租金 θ 的概率分布表

租金区间	分段中点	累积概率	分段概率
2.00-2.05	2.025	0.0575	0.0575
2.05-2.10	2.075	0.125	0.0675
2.10-2.15	2.125	0.250	0.125
2.15-2.20	2.175	0.500	0.250
2.20-2.25	2.225	0.750	0.250
2.25-2.30	2.275	0.875	0.125
2.30-2.35	2.325	0.9375	0.085
2.35-2.40	2.375	1.000	0.040

由表 2.2.3 中分段中点 x_i 和分段概率 p_i 计算 μ 和 σ^2 估计值为:

$$\hat{\mu} = \sum_{i=1}^8 x_i p_i = 2.198, \quad \hat{\sigma}^2 = \sum_{i=1}^8 x_i^2 p_i - \hat{\mu}^2 \approx 0.0824^2$$

2.2 利用先验信息确定先验分布

2.2.4 给定参数 θ 分位数, 确定累计分布函数租金 θ 的概率直方图图 2.2.5 租金 θ 的概率直方图

1 2.1 主观概率

- 2.1.1 主观概率的定义
- 2.1.2 确定主观概率的方法

2 2.2 利用先验信息确定先验分布

- 2.2.1 直方图法
- 2.2.2 相对似然法
- 2.2.3 选定先验密度函数的形式, 再估计超参数
- 2.2.4 给定参数 θ 分位数, 确定累计分布函数

3 2.3 利用边缘分布 $m(x)$ 确定先验分布

- 2.3.1 边缘分布的定义及例
- 2.3.2 选择先验分布的 ML-II 方法
- 2.3.3 选择先验分布的矩方法

4 2.4 无信息先验

- 2.4.1 Laplace 先验与广义先验分布
- 2.4.2 位置参数的无信息先验
- 2.4.3 刻度参数的无信息先验
- 2.4.4 一般情形下的无信息先验

5 2.5 共轭先验分布

缘分布的定义

- **定义 2.3.1** 设 r.v. X 有概率函数 $f(x|\theta)$, θ 有先验分布 $\pi(\theta)$, 则定义随机变量 X 的边缘分布为

$$\begin{aligned} m(x) &= \int_{\Theta} f(x|\theta) dF^{\pi}(\theta) \\ &= \begin{cases} \int_{\Theta} f(x|\theta) \pi(\theta) d\theta, & \text{当 } \theta \text{ 为连续随机变量,} \\ \sum_i f(x|\theta_i) \pi(\theta_i), & \text{当 } \theta \text{ 为离散随机变量.} \end{cases} \end{aligned}$$

- 当先验分布含有未知的超参数 λ , 如 $\pi(\theta) = \pi(\theta|\lambda)$, 则边缘分布 $m(x)$ 也依赖 λ , 此时可记 $m(x) = m(x|\lambda)$. 这种边缘分布在本节后面求后验分布时常用到.

缘分布的定义

- **定义 2.3.1** 设 r.v. X 有概率函数 $f(x|\theta)$, θ 有先验分布 $\pi(\theta)$, 则定义随机变量 X 的边缘分布为

$$\begin{aligned} m(x) &= \int_{\Theta} f(x|\theta) dF^{\pi}(\theta) \\ &= \begin{cases} \int_{\Theta} f(x|\theta) \pi(\theta) d\theta, & \text{当 } \theta \text{ 为连续随机变量,} \\ \sum_i f(x|\theta_i) \pi(\theta_i), & \text{当 } \theta \text{ 为离散随机变量.} \end{cases} \end{aligned}$$

- 当先验分布含有未知的超参数 λ , 如 $\pi(\theta) = \pi(\theta|\lambda)$, 则边缘分布 $m(x)$ 也依赖 λ , 此时可记 $m(x) = m(x|\lambda)$. 这种边缘分布在本节后面求后验分布时常用到.

混合分布及注释

- 为了说明边缘分布的统计意义，下面引入混合分布的概念. 设随机变量 X 以概率 π 在总体 F_1 中取值，以 $(1 - \pi)$ 的概率在总体 F_2 中取值，若 $F(x|\theta_1)$ 和 $F(x|\theta_2)$ 分别记这两个总体的分布函数，则 X 的混合分布函数为 $F(x) = \pi F(x|\theta_1) + (1 - \pi)F(x|\theta_2)$. 用概率函数表示，则混合分布为 $f(x) = \pi f(x|\theta_1) + (1 - \pi)f(x|\theta_2)$ ，则称 $F(x)$ 为 $F(x|\theta_1)$ 和 $F(x|\theta_2)$ 的混合分布. 而 π 和 $1 - \pi$ 可以视为随机变量 θ 的分布，即 $\pi(\theta_1) = \pi$, $\pi(\theta_2) = 1 - \pi$. 从混合分布中抽取容量为 n 的样本 X_1, \dots, X_n ，那么约有 $n\pi(\theta_1)$ 个抽自总体 $F(x|\theta_1)$ ，约有 $n\pi(\theta_2)$ 个抽自总体 $F(x|\theta_2)$ ，这样的样本称为混合样本.
- 从混合分布的定义可见，边缘分布 $m(x)$ 是混合分布的推广. 特别当 θ 为离散随机变量时，边缘分布 $m(x)$ 是有限或可列个概率函数混合而成，当 θ 为连续随机变量时， $m(x)$ 可视为无限不可数个密度函数混合而成.

混合分布及注释

- 为了说明边缘分布的统计意义，下面引入混合分布的概念. 设随机变量 X 以概率 π 在总体 F_1 中取值，以 $(1 - \pi)$ 的概率在总体 F_2 中取值，若 $F(x|\theta_1)$ 和 $F(x|\theta_2)$ 分别记这两个总体的分布函数，则 X 的混合分布函数为 $F(x) = \pi F(x|\theta_1) + (1 - \pi)F(x|\theta_2)$. 用概率函数表示，则混合分布为 $f(x) = \pi f(x|\theta_1) + (1 - \pi)f(x|\theta_2)$ ，则称 $F(x)$ 为 $F(x|\theta_1)$ 和 $F(x|\theta_2)$ 的混合分布. 而 π 和 $1 - \pi$ 可以视为随机变量 θ 的分布，即 $\pi(\theta_1) = \pi$, $\pi(\theta_2) = 1 - \pi$. 从混合分布中抽取容量为 n 的样本 X_1, \dots, X_n ，那么约有 $n\pi(\theta_1)$ 个抽自总体 $F(x|\theta_1)$ ，约有 $n\pi(\theta_2)$ 个抽自总体 $F(x|\theta_2)$ ，这样的样本称为混合样本.
- 从混合分布的定义可见，边缘分布 $m(x)$ 是混合分布的推广. 特别当 θ 为离散随机变量时，边缘分布 $m(x)$ 是有限或可列个概率函数混合而成，当 θ 为连续随机变量时， $m(x)$ 可视为无限不可数个密度函数混合而成.

2.3 利用边缘分布 $m(x)$ 确定先验分布

2.3.1 边缘分布的定义及例

一个例子

- **例 2.3.1** 设给定 θ 时样本 X 服从正态分布 $N(\theta, \sigma^2)$, σ^2 已知, 设 θ 有先验分布 $N(\mu_\pi, \sigma_\pi^2)$, 求 X 的边缘分布 $m(x)$.
- 为简化 $m(x)$ 计算, 令

$$A = \frac{1}{\sigma^2} + \frac{1}{\sigma_\pi^2}, \quad B = \frac{x}{\sigma^2} + \frac{\mu_\pi}{\sigma_\pi^2}, \quad C = \frac{x^2}{\sigma^2} + \frac{\mu_\pi^2}{\sigma_\pi^2}.$$

- 因此有

$$\begin{aligned} m(x) &= \int_{-\infty}^{\infty} f(x|\theta)\pi(\theta)d\theta \\ &= \frac{1}{2\pi\sigma\sigma_\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\mu_\pi-\theta)^2}{\sigma_\pi^2}\right]\right\}d\theta \\ &= \frac{1}{2\pi\sigma\sigma_\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{A}{2}\left(\theta - \frac{B}{A}\right)^2\right\} \cdot \exp\left\{-\frac{1}{2}\left(C - \frac{B^2}{A}\right)\right\}d\theta \\ &= \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_\pi^2)}} \exp\left\{-\frac{(x - \mu_\pi)^2}{2(\sigma^2 + \sigma_\pi^2)}\right\}, \end{aligned} \quad (3.1)$$

即边缘分布 $m(x)$ 是正态分布 $N(\mu_\pi, \sigma^2 + \sigma_\pi^2)$.

2.3 利用边缘分布 $m(x)$ 确定先验分布

2.3.1 边缘分布的定义及例

一个例子

- **例 2.3.1** 设给定 θ 时样本 X 服从正态分布 $N(\theta, \sigma^2)$, σ^2 已知, 设 θ 有先验分布 $N(\mu_\pi, \sigma_\pi^2)$, 求 X 的边缘分布 $m(x)$.
- 为简化 $m(x)$ 计算, 令

$$A = \frac{1}{\sigma^2} + \frac{1}{\sigma_\pi^2}, \quad B = \frac{x}{\sigma^2} + \frac{\mu_\pi}{\sigma_\pi^2}, \quad C = \frac{x^2}{\sigma^2} + \frac{\mu_\pi^2}{\sigma_\pi^2}.$$

- 因此有

$$\begin{aligned} m(x) &= \int_{-\infty}^{\infty} f(x|\theta)\pi(\theta)d\theta \\ &= \frac{1}{2\pi\sigma\sigma_\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\mu_\pi-\theta)^2}{\sigma_\pi^2}\right]\right\}d\theta \\ &= \frac{1}{2\pi\sigma\sigma_\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{A}{2}\left(\theta - \frac{B}{A}\right)^2\right\} \cdot \exp\left\{-\frac{1}{2}\left(C - \frac{B^2}{A}\right)\right\}d\theta \\ &= \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_\pi^2)}} \exp\left\{-\frac{(x - \mu_\pi)^2}{2(\sigma^2 + \sigma_\pi^2)}\right\}, \end{aligned} \quad (3.1)$$

即边缘分布 $m(x)$ 是正态分布 $N(\mu_\pi, \sigma^2 + \sigma_\pi^2)$.

2.3 利用边缘分布 $m(x)$ 确定先验分布

2.3.1 边缘分布的定义及例

一个例子

- **例 2.3.1** 设给定 θ 时样本 X 服从正态分布 $N(\theta, \sigma^2)$, σ^2 已知, 设 θ 有先验分布 $N(\mu_\pi, \sigma_\pi^2)$, 求 X 的边缘分布 $m(x)$.
- 为简化 $m(x)$ 计算, 令

$$A = \frac{1}{\sigma^2} + \frac{1}{\sigma_\pi^2}, \quad B = \frac{x}{\sigma^2} + \frac{\mu_\pi}{\sigma_\pi^2}, \quad C = \frac{x^2}{\sigma^2} + \frac{\mu_\pi^2}{\sigma_\pi^2}.$$

- 因此有

$$\begin{aligned} m(x) &= \int_{-\infty}^{\infty} f(x|\theta)\pi(\theta)d\theta \\ &= \frac{1}{2\pi\sigma\sigma_\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\frac{(x-\theta)^2}{\sigma^2} + \frac{(\mu_\pi-\theta)^2}{\sigma_\pi^2}\right]\right\}d\theta \\ &= \frac{1}{2\pi\sigma\sigma_\pi} \int_{-\infty}^{\infty} \exp\left\{-\frac{A}{2}\left(\theta - \frac{B}{A}\right)^2\right\} \cdot \exp\left\{-\frac{1}{2}\left(C - \frac{B^2}{A}\right)\right\}d\theta \\ &= \frac{1}{\sqrt{2\pi(\sigma^2 + \sigma_\pi^2)}} \exp\left\{-\frac{(x - \mu_\pi)^2}{2(\sigma^2 + \sigma_\pi^2)}\right\}, \end{aligned} \quad (3.1)$$

即边缘分布 $m(x)$ 是正态分布 $N(\mu_\pi, \sigma^2 + \sigma_\pi^2)$.

选择先验分布的 ML-II 方法

- 当观测到样本 X , 若有 2 个先验分布 π_1 和 π_2 , 使得

$$m(\mathbf{x}|\pi_1) > m(\mathbf{x}|\pi_2)$$

则认为当先验取 π_1 时样本 X 出现的似然性比先验取为 π_2 时的似然性更大, 即认为样本 X 由分布 $m(\mathbf{x}|\pi_1)$ 中产生. 这一思想与经典统计方法中的极大似然原理相类似, 这儿的 π 起似然函数中 θ 的作用, 这就引入了 ML-II 方法.

- 定义 2.3.2 设 Γ 为所考虑的先验类, 若存在 $\hat{\pi} \in \Gamma$, 有了样本 $\mathbf{x} = (x_1, \dots, x_n)$ 后, 使得

$$m(\mathbf{x}|\hat{\pi}) = \sup_{\pi \in \Gamma} m(\mathbf{x}|\pi) = \sup_{\pi \in \Gamma} \prod_{i=1}^n m(x_i|\pi)$$

则称 $\hat{\pi}$ 为类型 II 中最大似然先验, 或简称 ML-II 先验.

选择先验分布的 ML-II 方法

- 当观测到样本 X , 若有 2 个先验分布 π_1 和 π_2 , 使得

$$m(\mathbf{x}|\pi_1) > m(\mathbf{x}|\pi_2)$$

则认为当先验取 π_1 时样本 X 出现的似然性比先验取为 π_2 时的似然性更大, 即认为样本 X 由分布 $m(\mathbf{x}|\pi_1)$ 中产生. 这一思想与经典统计方法中的极大似然原理相类似, 这儿的 π 起似然函数中 θ 的作用, 这就引入了 ML-II 方法.

- 定义 2.3.2** 设 Γ 为所考虑的先验类, 若存在 $\hat{\pi} \in \Gamma$, 有了样本 $\mathbf{x} = (x_1, \dots, x_n)$ 后, 使得

$$m(\mathbf{x}|\hat{\pi}) = \sup_{\pi \in \Gamma} m(\mathbf{x}|\pi) = \sup_{\pi \in \Gamma} \prod_{i=1}^n m(x_i|\pi)$$

则称 $\hat{\pi}$ 为类型 II 中最大似然先验, 或简称 ML-II 先验.

2.3 利用边缘分布 $m(\mathbf{x})$ 确定先验分布

2.3.2 选择先验分布的 ML-II 方法

选择先验分布的 ML-II 方法

- 若样本 $\mathbf{X} = (X_1, \dots, X_n)$ 所涉及先验密度函数的形式已知, 未知的仅是其中的超参数, 如 $\theta \sim N(\mu_\pi, \sigma_\pi^2)$, μ_π 和 σ_π^2 称为超参数, 即先验密度的类 Γ 可表为

$$\Gamma = \{\pi(\theta|\lambda), \lambda \text{ 为超参数(或超参数向量)}, \lambda \in \Lambda\},$$

此处 Λ 为 λ 的参数空间.

- 这时寻求 ML-II 先验较为简单. 只要求出这样的 $\hat{\lambda}$, 使得

$$m(\mathbf{x}|\hat{\lambda}) = \sup_{\lambda \in \Lambda} m(\mathbf{x}|\lambda) = \sup_{\lambda \in \Lambda} \prod_{i=1}^n m(x_i|\lambda)$$

即通过使似然函数极大化方法求出 $\hat{\lambda}$, 则先验分布 $\pi(\theta|\hat{\lambda})$ 即为所确定的先验分布. 而 $\hat{\lambda}$ 称为 ML-II 超参数.

- 例 2.3.2 设随机变量 $X \sim N(\theta, \sigma^2)$, 其中 σ^2 已知, 又设 $\theta \sim N(\mu_\pi, \sigma_\pi^2)$. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从边缘分布 $m(\mathbf{x}|\lambda)$ 中抽取的 iid 样本, 试确定 θ 的先验分布.

选择先验分布的 ML-II 方法

- 若样本 $\mathbf{X} = (X_1, \dots, X_n)$ 所涉及的先验密度函数的形式已知, 未知的仅是其中的超参数, 如 $\theta \sim N(\mu_\pi, \sigma_\pi^2)$, μ_π 和 σ_π^2 称为超参数, 即先验密度的类 Γ 可表为

$$\Gamma = \{\pi(\theta|\lambda), \lambda \text{ 为超参数(或超参数向量)}, \lambda \in \Lambda\},$$

此处 Λ 为 λ 的参数空间.

- 这时寻求 ML-II 先验较为简单. 只要求出这样的 $\hat{\lambda}$, 使得

$$m(\mathbf{x}|\hat{\lambda}) = \sup_{\lambda \in \Lambda} m(\mathbf{x}|\lambda) = \sup_{\lambda \in \Lambda} \prod_{i=1}^n m(x_i|\lambda)$$

即通过使似然函数极大化方法求出 $\hat{\lambda}$, 则先验分布 $\pi(\theta|\hat{\lambda})$ 即为所确定的先验分布. 而 $\hat{\lambda}$ 称为 ML-II 超参数.

- 例 2.3.2 设随机变量 $X \sim N(\theta, \sigma^2)$, 其中 σ^2 已知, 又设 $\theta \sim N(\mu_\pi, \sigma_\pi^2)$. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从边缘分布 $m(\mathbf{x}|\lambda)$ 中抽取的 iid 样本, 试确定 θ 的先验分布.

选择先验分布的 ML-II 方法

- 若样本 $\mathbf{X} = (X_1, \dots, X_n)$ 所涉及的先验密度函数的形式已知, 未知的仅是其中的超参数, 如 $\theta \sim N(\mu_\pi, \sigma_\pi^2)$, μ_π 和 σ_π^2 称为超参数, 即先验密度的类 Γ 可表为

$$\Gamma = \{\pi(\theta|\lambda), \lambda \text{ 为超参数(或超参数向量)}, \lambda \in \Lambda\},$$

此处 Λ 为 λ 的参数空间.

- 这时寻求 ML-II 先验较为简单. 只要求出这样的 $\hat{\lambda}$, 使得

$$m(\mathbf{x}|\hat{\lambda}) = \sup_{\lambda \in \Lambda} m(\mathbf{x}|\lambda) = \sup_{\lambda \in \Lambda} \prod_{i=1}^n m(x_i|\lambda)$$

即通过使似然函数极大化方法求出 $\hat{\lambda}$, 则先验分布 $\pi(\theta|\hat{\lambda})$ 即为所确定的先验分布. 而 $\hat{\lambda}$ 称为 ML-II 超参数.

- 例 2.3.2** 设随机变量 $X \sim N(\theta, \sigma^2)$, 其中 σ^2 已知, 又设 $\theta \sim N(\mu_\pi, \sigma_\pi^2)$. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从边缘分布 $m(\mathbf{x}|\lambda)$ 中抽取的 iid 样本, 试确定 θ 的先验分布.

选择先验分布的矩方法

- 当先验分布 $\pi(\theta|\lambda)$ 的形式已知,但含有未知超参数 λ 时,可利用先验分布的矩与边缘分布的矩之间的关系寻求超参数 λ 的估计量 $\hat{\lambda}$, 从而获得先验分布 $\pi(\theta|\hat{\lambda})$.
- 这个方法的思想如下: 首先, 将边缘分布一些阶的矩表成超参数的函数, 得到一个方程组或方程; 将方程组或方程中的边缘分布的矩用相应的样本矩代替, 得到以超参数为变量的方程或方程组, 解方程或解方程组获得超参数的估计量, 从而确定先验分布.
- 下面给出具体的步骤
 - (1) 计算样本分布 $f(x|\theta)$ 的期望 $\mu(\theta)$ 和方差 $\sigma^2(\theta)$, 即

$$\mu(\theta) = E^{X|\theta}(X), \quad \sigma^2(\theta) = E^{X|\theta}[X - \mu(\theta)]^2$$

此处 $E^{X|\theta}$ 表示在给定 θ 的条件下关于条件分布 $f(x|\theta)$ 求期望, 以下类似符号亦作类似解释.

选择先验分布的矩方法

- 当先验分布 $\pi(\theta|\lambda)$ 的形式已知,但含有未知超参数 λ 时,可利用先验分布的矩与边缘分布的矩之间的关系寻求超参数 λ 的估计量 $\hat{\lambda}$, 从而获得先验分布 $\pi(\theta|\hat{\lambda})$.
- 这个方法的思想如下: 首先, 将边缘分布一些阶的矩表成超参数的函数, 得到一个方程组或方程; 将方程组或方程中的边缘分布的矩用相应的样本矩代替, 得到以超参数为变量的方程或方程组, 解方程或解方程组获得超参数的估计量, 从而确定先验分布.
- 下面给出具体的步骤
(1) 计算样本分布 $f(x|\theta)$ 的期望 $\mu(\theta)$ 和方差 $\sigma^2(\theta)$, 即

$$\mu(\theta) = E^{X|\theta}(X), \quad \sigma^2(\theta) = E^{X|\theta}[X - \mu(\theta)]^2$$

此处 $E^{X|\theta}$ 表示在给定 θ 的条件下关于条件分布 $f(x|\theta)$ 求期望, 以下类似符号亦作类似解释.

选择先验分布的矩方法

- 当先验分布 $\pi(\theta|\lambda)$ 的形式已知,但含有未知超参数 λ 时,可利用先验分布的矩与边缘分布的矩之间的关系寻求超参数 λ 的估计量 $\hat{\lambda}$, 从而获得先验分布 $\pi(\theta|\hat{\lambda})$.
- 这个方法的思想如下: 首先, 将边缘分布一些阶的矩表成超参数的函数, 得到一个方程组或方程; 将方程组或方程中的边缘分布的矩用相应的样本矩代替, 得到以超参数为变量的方程或方程组, 解方程或解方程组获得超参数的估计量, 从而确定先验分布.
- 下面给出具体的步骤
 - (1) 计算样本分布 $f(x|\theta)$ 的期望 $\mu(\theta)$ 和方差 $\sigma^2(\theta)$, 即

$$\mu(\theta) = E^{X|\theta}(X), \quad \sigma^2(\theta) = E^{X|\theta}[X - \mu(\theta)]^2$$

此处 $E^{X|\theta}$ 表示在给定 θ 的条件下关于条件分布 $f(x|\theta)$ 求期望, 以下类似符号亦作类似解释.

选择先验分布的矩方法

- (2) 计算边缘密度 $m(x) = m(x|\lambda)$ 的期望 $\mu_m(\lambda)$ 和方差 $\sigma_m^2(\lambda)$:

$$\begin{aligned}\mu_m(\lambda) &= E^{X|\lambda}(X) = \int_{\mathcal{X}} x m(x|\lambda) dx = \int_{\mathcal{X}} \int_{\Theta} x f(x|\theta) \pi(\theta|\lambda) d\theta dx \\ &= \int_{\Theta} \left[\int_{\mathcal{X}} x f(x|\theta) dx \right] \pi(\theta|\lambda) d\theta = \int_{\Theta} \mu(\theta) \pi(\theta|\lambda) d\theta \\ &= E^{\theta|\lambda}[\mu(\theta)],\end{aligned}\tag{2.3.5}$$

- 以及

$$\begin{aligned}\sigma_m^2(\lambda) &= E^{X|\lambda}[X - \mu_m(\lambda)]^2 = \int_{\mathcal{X}} (x - \mu_m(\lambda))^2 m(x|\lambda) dx \\ &= \int_{\mathcal{X}} \int_{\Theta} (x - \mu_m(\lambda))^2 f(x|\theta) \pi(\theta|\lambda) d\theta dx \\ &= \int_{\Theta} E^{X|\theta} [x - \mu_m(\lambda)]^2 \pi(\theta|\lambda) d\theta,\end{aligned}\tag{2.3.6}$$

选择先验分布的矩方法

- (2) 计算边缘密度 $m(x) = m(x|\lambda)$ 的期望 $\mu_m(\lambda)$ 和方差 $\sigma_m^2(\lambda)$:

$$\begin{aligned}\mu_m(\lambda) &= E^{X|\lambda}(X) = \int_{\mathcal{X}} x m(x|\lambda) dx = \int_{\mathcal{X}} \int_{\Theta} x f(x|\theta) \pi(\theta|\lambda) d\theta dx \\ &= \int_{\Theta} \left[\int_{\mathcal{X}} x f(x|\theta) dx \right] \pi(\theta|\lambda) d\theta = \int_{\Theta} \mu(\theta) \pi(\theta|\lambda) d\theta \\ &= E^{\theta|\lambda}[\mu(\theta)],\end{aligned}\tag{2.3.5}$$

- 以及

$$\begin{aligned}\sigma_m^2(\lambda) &= E^{X|\lambda}[X - \mu_m(\lambda)]^2 = \int_{\mathcal{X}} (x - \mu_m(\lambda))^2 m(x|\lambda) dx \\ &= \int_{\mathcal{X}} \int_{\Theta} (x - \mu_m(\lambda))^2 f(x|\theta) \pi(\theta|\lambda) d\theta dx \\ &= \int_{\Theta} E^{X|\theta} [x - \mu_m(\lambda)]^2 \pi(\theta|\lambda) d\theta,\end{aligned}\tag{2.3.6}$$

选择先验分布的矩方法

- 其中

$$\begin{aligned} E^{X|\theta} [x - \mu_m(\lambda)]^2 &= E^{X|\theta} [(x - \mu(\theta)) + (\mu(\theta) - \mu_m(\lambda))]^2 \\ &= E^{X|\theta} [x - \mu(\theta)]^2 + E^{X|\theta} [\mu(\theta) - \mu_m(\lambda)]^2 \\ &= \sigma^2(\theta) + [\mu(\theta) - \mu_m(\lambda)]^2 \end{aligned} \quad (2.3.7)$$

- 将 (2.3.7) 带入 (2.3.6) 得

$$\begin{aligned} \sigma_m^2(\lambda) &= \int_{\Theta} \sigma^2(\theta) \pi(\theta|\lambda) d\theta + \int [\mu(\theta) - \mu_m(\lambda)]^2 \pi(\theta|\lambda) d\theta \\ &= E^{\theta|\lambda} [\sigma^2(\theta)] + E^{\theta|\lambda} [\mu(\theta) - \mu_m(\lambda)]^2 \end{aligned} \quad (2.3.8)$$

- 由 (2.3.5) 和 (2.3.8) 可见边缘分布的期望 $\mu_m(\lambda)$ 和方差 $\sigma_m^2(\lambda)$ 皆与 $E^{\theta|\lambda}[\mu(\theta)]$, $E^{\theta|\lambda}[\sigma^2(\theta)]$, 和 $E^{\theta|\lambda}[\mu(\theta) - \mu_m(\lambda)]^2$ 有关。

选择先验分布的矩方法

- 其中

$$\begin{aligned} E^{X|\theta} [x - \mu_m(\lambda)]^2 &= E^{X|\theta} [(x - \mu(\theta)) + (\mu(\theta) - \mu_m(\lambda))]^2 \\ &= E^{X|\theta} [x - \mu(\theta)]^2 + E^{X|\theta} [\mu(\theta) - \mu_m(\lambda)]^2 \\ &= \sigma^2(\theta) + [\mu(\theta) - \mu_m(\lambda)]^2 \end{aligned} \quad (2.3.7)$$

- 将 (2.3.7) 带入 (2.3.6) 得

$$\begin{aligned} \sigma_m^2(\lambda) &= \int_{\Theta} \sigma^2(\theta) \pi(\theta|\lambda) d\theta + \int [\mu(\theta) - \mu_m(\lambda)]^2 \pi(\theta|\lambda) d\theta \\ &= E^{\theta|\lambda} [\sigma^2(\theta)] + E^{\theta|\lambda} [\mu(\theta) - \mu_m(\lambda)]^2 \end{aligned} \quad (2.3.8)$$

- 由 (2.3.5) 和 (2.3.8) 可见边缘分布的期望 $\mu_m(\lambda)$ 和方差 $\sigma_m^2(\lambda)$ 皆与 $E^{\theta|\lambda}[\mu(\theta)]$, $E^{\theta|\lambda}[\sigma^2(\theta)]$, 和 $E^{\theta|\lambda}[\mu(\theta) - \mu_m(\lambda)]^2$ 有关。

选择先验分布的矩方法

- 其中

$$\begin{aligned} E^{X|\theta} [x - \mu_m(\lambda)]^2 &= E^{X|\theta} [(x - \mu(\theta)) + (\mu(\theta) - \mu_m(\lambda))]^2 \\ &= E^{X|\theta} [x - \mu(\theta)]^2 + E^{X|\theta} [\mu(\theta) - \mu_m(\lambda)]^2 \\ &= \sigma^2(\theta) + [\mu(\theta) - \mu_m(\lambda)]^2 \end{aligned} \quad (2.3.7)$$

- 将 (2.3.7) 带入 (2.3.6) 得

$$\begin{aligned} \sigma_m^2(\lambda) &= \int_{\Theta} \sigma^2(\theta) \pi(\theta|\lambda) d\theta + \int [\mu(\theta) - \mu_m(\lambda)]^2 \pi(\theta|\lambda) d\theta \\ &= E^{\theta|\lambda} [\sigma^2(\theta)] + E^{\theta|\lambda} [\mu(\theta) - \mu_m(\lambda)]^2 \end{aligned} \quad (2.3.8)$$

- 由 (2.3.5) 和 (2.3.8) 可见边缘分布的期望 $\mu_m(\lambda)$ 和方差 $\sigma_m^2(\lambda)$ 皆与 $E^{\theta|\lambda}[\mu(\theta)]$, $E^{\theta|\lambda}[\sigma^2(\theta)]$, 和 $E^{\theta|\lambda}[\mu(\theta) - \mu_m(\lambda)]^2$ 有关。

选择先验分布的矩方法

- (3) 当先验分布只含有 2 个超参数 $\lambda_1, \lambda_2, \lambda = (\lambda_1, \lambda_2)$ 时, 用

$$\hat{\mu}_m = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{和} \quad \hat{\sigma}_m^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

分别作为 $\mu_m(\lambda)$ 和 $\sigma_m^2(\lambda)$ 的估计.

- 将 (2.3.5) 和 (2.3.8) 左边的 $\mu_m(\lambda)$ 和 $\sigma_m^2(\lambda)$ 分别用这两个估计量代替, 得方程组

$$\begin{cases} \hat{\mu}_m = E^{\theta|\lambda}[\mu(\theta)] \\ \hat{\sigma}_m^2 = E^{\theta|\lambda}[\sigma^2(\theta)] + E^{\theta|\lambda}[\mu(\theta) - \mu_m(\lambda)]^2 \end{cases}$$

解此方程组, 可得超参数 λ_1 和 λ_2 的估计量 $\hat{\lambda}_1$ 和 $\hat{\lambda}_2$ 以及 $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2)$. 故 $\pi(\theta|\lambda)$ 可用 $\pi(\theta|\hat{\lambda})$ 作为其估计而得到.

选择先验分布的矩方法

- (3) 当先验分布只含有 2 个超参数 $\lambda_1, \lambda_2, \lambda = (\lambda_1, \lambda_2)$ 时, 用

$$\hat{\mu}_m = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad \text{和} \quad \hat{\sigma}_m^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

分别作为 $\mu_m(\lambda)$ 和 $\sigma_m^2(\lambda)$ 的估计.

- 将 (2.3.5) 和 (2.3.8) 左边的 $\mu_m(\lambda)$ 和 $\sigma_m^2(\lambda)$ 分别用这两个估计量代替, 得方程组

$$\begin{cases} \hat{\mu}_m = E^{\theta|\lambda}[\mu(\theta)] \\ \hat{\sigma}_m^2 = E^{\theta|\lambda}[\sigma^2(\theta)] + E^{\theta|\lambda}[\mu(\theta) - \mu_m(\lambda)]^2 \end{cases}$$

解此方程组, 可得超参数 λ_1 和 λ_2 的估计量 $\hat{\lambda}_1$ 和 $\hat{\lambda}_2$ 以及 $\hat{\lambda} = (\hat{\lambda}_1, \hat{\lambda}_2)$. 故 $\pi(\theta|\lambda)$ 可用 $\pi(\theta|\hat{\lambda})$ 作为其估计而得到.

一个例子

- **例 2.3.3** 设 $X \sim N(\theta, 1)$, 参数 θ 的先验分布取为共轭先验 $N(\mu_\pi, \sigma_\pi^2)$, 其中 $\lambda = (\mu_\pi, \sigma_\pi^2)$ 未知. 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从边缘分布 $m(x|\lambda)$ 中抽取的 iid 样本, 由样本算出样本均值 $\bar{X} = 10$, $S^2 = 3$. 试确定 θ 的先验分布.

1 2.1 主观概率

- 2.1.1 主观概率的定义
- 2.1.2 确定主观概率的方法

2 2.2 利用先验信息确定先验分布

- 2.2.1 直方图法
- 2.2.2 相对似然法
- 2.2.3 选定先验密度函数的形式, 再估计超参数
- 2.2.4 给定参数 θ 分位数, 确定累计分布函数

3 2.3 利用边缘分布 $m(x)$ 确定先验分布

- 2.3.1 边缘分布的定义及例
- 2.3.2 选择先验分布的 ML-II 方法
- 2.3.3 选择先验分布的矩方法

4 2.4 无信息先验

- 2.4.1 Laplace 先验与广义先验分布
- 2.4.2 位置参数的无信息先验
- 2.4.3 刻度参数的无信息先验
- 2.4.4 一般情形下的无信息先验

5 2.5 共轭先验分布

1. 均匀分布与广义先验分布

贝叶斯分析的一个重要特点就是在统计推断时要利用先验信息. 但常常会出现这样的情况: 没有先验信息或者只有极少的先验信息可利用, 但仍想用贝叶斯方法. 此时所需要的是一种无信息先验, 称为 **Laplace** 先验, 即对参数空间 Θ 中的任何一点 θ 没有偏爱的先验信息.

- 若 Θ 为有限集, 即 θ 只可能取有限个值, 如 $\theta = \theta_i, i = 1, 2, \dots, n$, 无信息先验给 Θ 中的每个元素以概率 $1/n$, 即

$$P(\theta = \theta_i) = 1/n, \quad i = 1, 2, \dots, n.$$

- 若 Θ 为 R_1 上的有限区间 $[a, b]$, 则取无信息先验为区间 $[a, b]$ 上的均匀分布 $U(a, b)$ (有时也记为 $R(a, b)$).
- 问题是若参数空间 Θ 无界, 无信息先验如何选取? 例如样本分布为 $N(\theta, \sigma^2)$, σ^2 已知, 此时 $\Theta = (-\infty, \infty)$. 若无信息先验取为 $\pi(\theta) \equiv 1$, 则 $\pi(\theta)$ 不是通常的密度, 因为 $\int_{-\infty}^{\infty} \pi(\theta) d\theta = \infty$. 这就引出广义先验分布的概念.

1. 均匀分布与广义先验分布

贝叶斯分析的一个重要特点就是在统计推断时要利用先验信息. 但常常会出现这样的情况: 没有先验信息或者只有极少的先验信息可利用, 但仍想用贝叶斯方法. 此时所需要的是一种无信息先验, 称为 **Laplace 先验**, 即对参数空间 Θ 中的任何一点 θ 没有偏爱的先验信息.

- 若 Θ 为有限集, 即 θ 只可能取有限个值, 如 $\theta = \theta_i, i = 1, 2, \dots, n$, 无信息先验给 Θ 中的每个元素以概率 $1/n$, 即

$$P(\theta = \theta_i) = 1/n, \quad i = 1, 2, \dots, n.$$

- 若 Θ 为 R_1 上的有限区间 $[a, b]$, 则取无信息先验为区间 $[a, b]$ 上的均匀分布 $U(a, b)$ (有时也记为 $R(a, b)$).
- 问题是若参数空间 Θ 无界, 无信息先验如何选取? 例如样本分布为 $N(\theta, \sigma^2)$, σ^2 已知, 此时 $\Theta = (-\infty, \infty)$. 若无信息先验取为 $\pi(\theta) \equiv 1$, 则 $\pi(\theta)$ 不是通常的密度, 因为 $\int_{-\infty}^{\infty} \pi(\theta) d\theta = \infty$. 这就引出广义先验分布的概念.

1. 均匀分布与广义先验分布

贝叶斯分析的一个重要特点就是在统计推断时要利用先验信息. 但常常会出现这样的情况: 没有先验信息或者只有极少的先验信息可利用, 但仍想用贝叶斯方法. 此时所需要的是种无信息先验, 称为 **Laplace** 先验, 即对参数空间 Θ 中的任何一点 θ 没有偏爱的先验信息.

- 若 Θ 为有限集, 即 θ 只可能取有限个值, 如 $\theta = \theta_i, i = 1, 2, \dots, n$, 无信息先验给 Θ 中的每个元素以概率 $1/n$, 即

$$P(\theta = \theta_i) = 1/n, \quad i = 1, 2, \dots, n.$$

- 若 Θ 为 R_1 上的有限区间 $[a, b]$, 则取无信息先验为区间 $[a, b]$ 上的均匀分布 $U(a, b)$ (有时也记为 $R(a, b)$).
- 问题是若参数空间 Θ 无界, 无信息先验如何选取? 例如样本分布为 $N(\theta, \sigma^2)$, σ^2 已知, 此时 $\Theta = (-\infty, \infty)$. 若无信息先验取为 $\pi(\theta) \equiv 1$, 则 $\pi(\theta)$ 不是通常的密度, 因为 $\int_{-\infty}^{\infty} \pi(\theta) d\theta = \infty$. 这就引出广义先验分布的概念.

1. 均匀分布与广义先验分布

贝叶斯分析的一个重要特点就是在统计推断时要利用先验信息. 但常常会出现这样的情况: 没有先验信息或者只有极少的先验信息可利用, 但仍想用贝叶斯方法. 此时所需要的是一种无信息先验, 称为 **Laplace** 先验, 即对参数空间 Θ 中的任何一点 θ 没有偏爱的先验信息.

- 若 Θ 为有限集, 即 θ 只可能取有限个值, 如 $\theta = \theta_i, i = 1, 2, \dots, n$, 无信息先验给 Θ 中的每个元素以概率 $1/n$, 即

$$P(\theta = \theta_i) = 1/n, \quad i = 1, 2, \dots, n.$$

- 若 Θ 为 R_1 上的有限区间 $[a, b]$, 则取无信息先验为区间 $[a, b]$ 上的均匀分布 $U(a, b)$ (有时也记为 $R(a, b)$).
- 问题是若参数空间 Θ 无界, 无信息先验如何选取? 例如样本分布为 $N(\theta, \sigma^2)$, σ^2 已知, 此时 $\Theta = (-\infty, \infty)$. 若无信息先验取为 $\pi(\theta) \equiv 1$, 则 $\pi(\theta)$ 不是通常的密度, 因为 $\int_{-\infty}^{\infty} \pi(\theta) d\theta = \infty$. 这就引出广义先验分布的概念.

2. 广义先验密度定义及例

- **定义 2.4.1** 设随机变量 $X \sim f(x|\theta)$, $\theta \in \Theta$, 若 θ 的先验分布 $\pi(\theta)$ 满足下列条件: (i) $\pi(\theta) \geq 0$ 且 $\int_{\Theta} \pi(\theta) d\theta = \infty$, (ii) 后验密度 $\pi(\theta|x)$ 是正常密度, 则称 $\pi(\theta)$ 为广义先验密度.
- **例 2.4.1** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 $N(\theta, 1)$ 总体中抽取的随机样本, 设 θ 的先验密度 $\pi(\theta) \equiv 1$, 求 θ 的后验密度.

解 由公式 (1.2.1) 可知

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(\mathbf{x}|\theta)\pi(\theta)d\theta} = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}d\theta} \\ &= \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\theta - \bar{x})^2\right\}.\end{aligned}$$

这是正态分布 $N(\bar{x}, 1/n)$ 的密度函数, 后验分布 $\pi(\theta|\mathbf{x})$ 仍为正常的密度函数, 故 $\pi(\theta) \equiv 1$ 为广义无信息先验密度.

2. 广义先验密度定义及例

- **定义 2.4.1** 设随机变量 $X \sim f(x|\theta)$, $\theta \in \Theta$, 若 θ 的先验分布 $\pi(\theta)$ 满足下列条件: (i) $\pi(\theta) \geq 0$ 且 $\int_{\Theta} \pi(\theta) d\theta = \infty$, (ii) 后验密度 $\pi(\theta|x)$ 是正常密度, 则称 $\pi(\theta)$ 为广义先验密度.
- **例 2.4.1** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 $N(\theta, 1)$ 总体中抽取的随机样本, 设 θ 的先验密度 $\pi(\theta) \equiv 1$, 求 θ 的后验密度.

解 由公式 (1.2.1) 可知

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(\mathbf{x}|\theta)\pi(\theta)d\theta} = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}d\theta} \\ &= \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\theta - \bar{x})^2\right\}.\end{aligned}$$

这是正态分布 $N(\bar{x}, 1/n)$ 的密度函数, 后验分布 $\pi(\theta|\mathbf{x})$ 仍为正常的密度函数, 故 $\pi(\theta) \equiv 1$ 为广义无信息先验密度.

2. 广义先验密度定义及例

- **定义 2.4.1** 设随机变量 $X \sim f(x|\theta)$, $\theta \in \Theta$, 若 θ 的先验分布 $\pi(\theta)$ 满足下列条件: (i) $\pi(\theta) \geq 0$ 且 $\int_{\Theta} \pi(\theta) d\theta = \infty$, (ii) 后验密度 $\pi(\theta|x)$ 是正常密度, 则称 $\pi(\theta)$ 为广义先验密度.
- **例 2.4.1** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 $N(\theta, 1)$ 总体中抽取的随机样本, 设 θ 的先验密度 $\pi(\theta) \equiv 1$, 求 θ 的后验密度.

解 由公式 (1.2.1) 可知

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(\mathbf{x}|\theta)\pi(\theta)d\theta} = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}d\theta} \\ &= \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\theta - \bar{x})^2\right\}.\end{aligned}$$

这是正态分布 $N(\bar{x}, 1/n)$ 的密度函数, 后验分布 $\pi(\theta|\mathbf{x})$ 仍为正常的密度函数, 故 $\pi(\theta) \equiv 1$ 为广义无信息先验密度.

2. 广义先验密度定义及例

- **定义 2.4.1** 设随机变量 $X \sim f(x|\theta)$, $\theta \in \Theta$, 若 θ 的先验分布 $\pi(\theta)$ 满足下列条件: (i) $\pi(\theta) \geq 0$ 且 $\int_{\Theta} \pi(\theta) d\theta = \infty$, (ii) 后验密度 $\pi(\theta|x)$ 是正常密度, 则称 $\pi(\theta)$ 为广义先验密度.
- **例 2.4.1** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从 $N(\theta, 1)$ 总体中抽取的随机样本, 设 θ 的先验密度 $\pi(\theta) \equiv 1$, 求 θ 的后验密度.

解 由公式 (1.2.1) 可知

$$\begin{aligned}\pi(\theta|\mathbf{x}) &= \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{-\infty}^{\infty} f(\mathbf{x}|\theta)\pi(\theta)d\theta} = \frac{\exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}}{\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\sum_{i=1}^n(x_i - \theta)^2\right\}d\theta} \\ &= \sqrt{\frac{n}{2\pi}} \exp\left\{-\frac{n}{2}(\theta - \bar{x})^2\right\}.\end{aligned}$$

这是正态分布 $N(\bar{x}, 1/n)$ 的密度函数, 后验分布 $\pi(\theta|\mathbf{x})$ 仍为正常的密度函数, 故 $\pi(\theta) \equiv 1$ 为广义无信息先验密度.

1. 位置参数族

- **定义 2.4.2** 设总体 X 的密度函数有形式 $f(x - \theta)$, 其样本空间 \mathcal{X} 和参数空间 Θ 皆为实轴 R , 则此类密度函数构成的分布族称为位置参数族, $\theta \in \Theta$ 称为位置参数.

- 例如, 设 $X \sim N(\theta, \sigma^2)$, 其中 σ^2 已知, 则 X 的密度函数

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \theta)^2\right\} = f(x - \theta)$$

属于位置参数族, θ 是位置参数.

- 又如, X 服从柯西分布 $C(\mu, \lambda)$, λ 已知, 则 X 的密度函数

$$\frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x - \mu)^2} = f(x - \mu)$$

也属于位置参数族, μ 是位置参数.

1. 位置参数族

- **定义 2.4.2** 设总体 X 的密度函数有形式 $f(x - \theta)$, 其样本空间 \mathcal{X} 和参数空间 Θ 皆为实轴 R , 则此类密度函数构成的分布族称为位置参数族, $\theta \in \Theta$ 称为位置参数.

- 例如, 设 $X \sim N(\theta, \sigma^2)$, 其中 σ^2 已知, 则 X 的密度函数

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \theta)^2\right\} = f(x - \theta)$$

属于位置参数族, θ 是位置参数.

- 又如, X 服从柯西分布 $C(\mu, \lambda)$, λ 已知, 则 X 的密度函数

$$\frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x - \mu)^2} = f(x - \mu)$$

也属于位置参数族, μ 是位置参数.

1. 位置参数族

- **定义 2.4.2** 设总体 X 的密度函数有形式 $f(x - \theta)$, 其样本空间 \mathcal{X} 和参数空间 Θ 皆为实轴 R , 则此类密度函数构成的分布族称为位置参数族, $\theta \in \Theta$ 称为位置参数.

- 例如, 设 $X \sim N(\theta, \sigma^2)$, 其中 σ^2 已知, 则 X 的密度函数

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \theta)^2\right\} = f(x - \theta)$$

属于位置参数族, θ 是位置参数.

- 又如, X 服从柯西分布 $C(\mu, \lambda)$, λ 已知, 则 X 的密度函数

$$\frac{1}{\pi} \cdot \frac{\lambda}{\lambda^2 + (x - \mu)^2} = f(x - \mu)$$

也属于位置参数族, μ 是位置参数.

2. 位置参数的无信息先验

- 位置参数族具有在平移变换群下的不变性. 对 \mathbf{X} 作平移变换得到 $\mathbf{Y} = \mathbf{X} + \mathbf{c}$, 同时对 θ 也作平移变换得到 $\eta = \theta + \mathbf{c}$. 显然 \mathbf{Y} 的密度函数有形式 $f(\mathbf{y} - \eta)$, η 仍为位置参数. 所以 (\mathbf{X}, θ) 与 (\mathbf{Y}, η) 的统计问题结构相同. 因此主张它们有相同的无信息先验是合理的.
- 理解这一点的另一方法: \mathbf{X} 和 \mathbf{Y} 的测量原点不同, 由于测量原点的选择是非常任意的, 所以无信息先验应当与这种选择无关. 如果无信息先验不依赖于原点的选择, 则它在等长区间内的先验概率应当一样. 即取 θ 的无信息先验密度 $\pi(\theta) \equiv 1$. 它是一个广义先验密度. 可以证明当 θ 为位置参数时, 其无信息先验密度取为常数 c 或者 1.
- 例 2.4.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, σ^2 已知. θ 的先验为无信息先验, 求 θ 的后验分布.

2. 位置参数的无信息先验

- 位置参数族具有在平移变换群下的不变性. 对 X 作平移变换得到 $Y = X + c$, 同时对 θ 也作平移变换得到 $\eta = \theta + c$. 显然 Y 的密度函数有形式 $f(y - \eta)$, η 仍为位置参数. 所以 (X, θ) 与 (Y, η) 的统计问题结构相同. 因此主张它们有相同的无信息先验是合理的.
- 理解这一点的另一方法: X 和 Y 的测量原点不同, 由于测量原点的选择是非常任意的, 所以无信息先验应当与这种选择无关. 如果无信息先验不依赖于原点的选择, 则它在等长区间内的先验概率应当一样. 即取 θ 的无信息先验密度 $\pi(\theta) \equiv 1$. 它是一个广义先验密度. 可以证明当 θ 为位置参数时, 其无信息先验密度取为常数 c 或者 1.
- 例 2.4.2 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, σ^2 已知. θ 的先验为无信息先验, 求 θ 的后验分布.

2. 位置参数的无信息先验

- 位置参数族具有在平移变换群下的不变性. 对 \mathbf{X} 作平移变换得到 $\mathbf{Y} = \mathbf{X} + \mathbf{c}$, 同时对 θ 也作平移变换得到 $\eta = \theta + \mathbf{c}$. 显然 \mathbf{Y} 的密度函数有形式 $f(\mathbf{y} - \eta)$, η 仍为位置参数. 所以 (\mathbf{X}, θ) 与 (\mathbf{Y}, η) 的统计问题结构相同. 因此主张它们有相同的无信息先验是合理的.
- 理解这一点的另一方法: \mathbf{X} 和 \mathbf{Y} 的测量原点不同, 由于测量原点的选择是非常任意的, 所以无信息先验应当与这种选择无关. 如果无信息先验不依赖于原点的选择, 则它在等长区间内的先验概率应当一样. 即取 θ 的无信息先验密度 $\pi(\theta) \equiv 1$. 它是一个广义先验密度. 可以证明当 θ 为位置参数时, 其无信息先验密度取为常数 c 或者 1.
- **例 2.4.2** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 为从总体 $N(\theta, \sigma^2)$ 中抽取的简单样本, σ^2 已知. θ 的先验为无信息先验, 求 θ 的后验分布.

1. 刻度参数族

- **定义 2.4.3** 设总体 X 的密度函数有形式 $\sigma^{-1}\varphi(x/\sigma)$, 其中 $\sigma > 0$ 为刻度参数, 参数空间为 $R^+ = (0, \infty)$, 则此类密度函数构成的分布族称为刻度参数族.

- 例如, $X \sim N(0, \sigma^2)$, X 的密度函数

$$\begin{aligned} f(x|\sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \\ &= \sigma^{-1} \left[\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\} \right] = \sigma^{-1} \varphi\left(\frac{x}{\sigma}\right) \end{aligned}$$

符合上述定义, 故属于刻度参数族, σ 为刻度参数.

- 又如, X 服从伽玛分布 $\Gamma(r, \lambda^{-1})$, 其密度函数

$$f(x|\lambda) = \frac{\lambda^{-r}}{\Gamma(r)} x^{r-1} e^{-x/\lambda} = \lambda^{-1} \left[\frac{1}{\Gamma(r)} \left(\frac{x}{\lambda}\right)^{r-1} e^{-x/\lambda} \right] = \lambda^{-1} \varphi\left(\frac{x}{\lambda}\right)$$

其中 $r > 0$ 已知, 它也符合上述定义, 因此也属于刻度参数, λ 为刻度参数.

1. 刻度参数族

- **定义 2.4.3** 设总体 X 的密度函数有形式 $\sigma^{-1}\varphi(x/\sigma)$, 其中 $\sigma > 0$ 为刻度参数, 参数空间为 $R^+ = (0, \infty)$, 则此类密度函数构成的分布族称为刻度参数族.

- 例如, $X \sim N(0, \sigma^2)$, X 的密度函数

$$\begin{aligned} f(x|\sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \\ &= \sigma^{-1} \left[\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\} \right] = \sigma^{-1} \varphi\left(\frac{x}{\sigma}\right) \end{aligned}$$

符合上述定义, 故属于刻度参数族, σ 为刻度参数.

- 又如, X 服从伽玛分布 $\Gamma(r, \lambda^{-1})$, 其密度函数

$$f(x|\lambda) = \frac{\lambda^{-r}}{\Gamma(r)} x^{r-1} e^{-x/\lambda} = \lambda^{-1} \left[\frac{1}{\Gamma(r)} \left(\frac{x}{\lambda}\right)^{r-1} e^{-x/\lambda} \right] = \lambda^{-1} \varphi\left(\frac{x}{\lambda}\right)$$

其中 $r > 0$ 已知, 它也符合上述定义, 因此也属于刻度参数, λ 为刻度参数.

1. 刻度参数族

- **定义 2.4.3** 设总体 X 的密度函数有形式 $\sigma^{-1}\varphi(x/\sigma)$, 其中 $\sigma > 0$ 为刻度参数, 参数空间为 $R^+ = (0, \infty)$, 则此类密度函数构成的分布族称为刻度参数族.
- 例如, $X \sim N(0, \sigma^2)$, X 的密度函数

$$\begin{aligned} f(x|\sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \\ &= \sigma^{-1} \left[\frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right\} \right] = \sigma^{-1} \varphi\left(\frac{x}{\sigma}\right) \end{aligned}$$

符合上述定义, 故属于刻度参数族, σ 为刻度参数.

- 又如, X 服从伽玛分布 $\Gamma(r, \lambda^{-1})$, 其密度函数

$$f(x|\lambda) = \frac{\lambda^{-r}}{\Gamma(r)} x^{r-1} e^{-x/\lambda} = \lambda^{-1} \left[\frac{1}{\Gamma(r)} \left(\frac{x}{\lambda}\right)^{r-1} e^{-x/\lambda} \right] = \lambda^{-1} \varphi\left(\frac{x}{\lambda}\right)$$

其中 $r > 0$ 已知, 它也符合上述定义, 因此也属于刻度参数, λ 为刻度参数.

2. 刻度参数的无信息先验

- 刻度参数族具有在刻度变换群下的不变性. 对 X 作变换 $Y = cX$, $c > 0$, 同时对 σ 作相应的变换 $\eta = c\sigma$. 不难算出 Y 的密度仍为 $\eta^{-1}\varphi(y/\eta)$. 可见 (X, σ) 和 (Y, η) 统计问题的结构相同, 故主张 σ 与 η 的无信息先验相同.
- 理解这一点的另一方法: X 和 Y 的度量单位不同, 先验分布应当不依赖于度量单位的选择, 则对任何 $a, b, 0 < a < b, c > 0$, σ 落在 $[a, b]$ 内的先验概率, 应当等于 η 落在 $[ca, cb]$ 内的先验概率, 可以证明, 这只有在先验密度为 $1/\sigma$ 时才可能, 即取 σ 的无信息先验为 $\pi(\sigma) = 1/\sigma, \sigma > 0$.
- 例 2.4.3 设总体 X 密度为 $f(x|\lambda) = \lambda^{-1} \exp\{-x/\lambda\} I_{(0,\infty)}(x)$, $\lambda > 0$ 为刻度参数. 令 $\mathbf{X} = (X_1, \dots, X_n)$ 是从上述分布中抽取的简单样本, λ 的先验密度为无信息先验, 求其后验密度.

2. 刻度参数的无信息先验

- 刻度参数族具有在刻度变换群下的不变性. 对 X 作变换 $Y = cX$, $c > 0$, 同时对 σ 作相应的变换 $\eta = c\sigma$. 不难算出 Y 的密度仍为 $\eta^{-1}\varphi(y/\eta)$. 可见 (X, σ) 和 (Y, η) 统计问题的结构相同, 故主张 σ 与 η 的无信息先验相同.
- 理解这一点的另一方法: X 和 Y 的度量单位不同, 先验分布应当不依赖于度量单位的选择, 则对任何 $a, b, 0 < a < b, c > 0$, σ 落在 $[a, b]$ 内的先验概率, 应当等于 η 落在 $[ca, cb]$ 内的先验概率, 可以证明, 这只有在先验密度为 $1/\sigma$ 时才可能, 即取 σ 的无信息先验为 $\pi(\sigma) = 1/\sigma, \sigma > 0$.
- 例 2.4.3 设总体 X 密度为 $f(x|\lambda) = \lambda^{-1} \exp\{-x/\lambda\} I_{(0,\infty)}(x)$, $\lambda > 0$ 为刻度参数. 令 $\mathbf{X} = (X_1, \dots, X_n)$ 是从上述分布中抽取的简单样本, λ 的先验密度为无信息先验, 求其后验密度.

2. 刻度参数的无信息先验

- 刻度参数族具有在刻度变换群下的不变性. 对 X 作变换 $Y = cX$, $c > 0$, 同时对 σ 作相应的变换 $\eta = c\sigma$. 不难算出 Y 的密度仍为 $\eta^{-1}\varphi(y/\eta)$. 可见 (X, σ) 和 (Y, η) 统计问题的结构相同, 故主张 σ 与 η 的无信息先验相同.
- 理解这一点的另一方法: X 和 Y 的度量单位不同, 先验分布应当不依赖于度量单位的选择, 则对任何 $a, b, 0 < a < b, c > 0$, σ 落在 $[a, b]$ 内的先验概率, 应当等于 η 落在 $[ca, cb]$ 内的先验概率, 可以证明, 这只有在先验密度为 $1/\sigma$ 时才可能, 即取 σ 的无信息先验为 $\pi(\sigma) = 1/\sigma, \sigma > 0$.
- **例 2.4.3** 设总体 X 密度为 $f(x|\lambda) = \lambda^{-1} \exp\{-x/\lambda\} I_{(0,\infty)}(x)$, $\lambda > 0$ 为刻度参数. 令 $\mathbf{X} = (X_1, \dots, X_n)$ 是从上述分布中抽取的简单样本, λ 的先验密度为无信息先验, 求其后验密度.

一般情形下的无信息先验*-Jeffreys 先验

- 对非位置参数族和刻度参数族的无信息先验如何求, 被广泛采用的是 Jeffreys (1961) 的方法, 由于推导涉及到变换群和 Harr 测度, 这里只给出结果, 不推导结果是如何得来的.
- 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是来自总体 $f(x|\theta)$ 的简单样本, 这里 $\theta = (\theta_1, \dots, \theta_p)$ 为 p 维参数向量. 在对 θ 无先验信息可用时, Jeffreys 用 Fisher 信息阵行列式的平方根作为 θ 的无信息先验, 此先验称为 Jeffreys 无信息先验. 其求解步骤如下:
- 写出样本的对数似然函数

$$l(\theta|\mathbf{x}) = \ln \left[\prod_{i=1}^n f(x_i|\theta) \right] = \sum_{i=1}^n \ln f(x_i|\theta).$$

一般情形下的无信息先验*-Jeffreys 先验

- 对非位置参数族和刻度参数族的无信息先验如何求, 被广泛采用的是 Jeffreys (1961) 的方法, 由于推导涉及到变换群和 Harr 测度, 这里只给出结果, 不推导结果是如何得来的.
- 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是来自总体 $f(x|\theta)$ 的简单样本, 这里 $\theta = (\theta_1, \dots, \theta_p)$ 为 p 维参数向量. 在对 θ 无先验信息可用时, Jeffreys 用 Fisher 信息阵行列式的平方根作为 θ 的无信息先验, 此先验称为 Jeffreys 无信息先验. 其求解步骤如下:
- 写出样本的对数似然函数

$$l(\theta|\mathbf{x}) = \ln \left[\prod_{i=1}^n f(x_i|\theta) \right] = \sum_{i=1}^n \ln f(x_i|\theta).$$

一般情形下的无信息先验*-Jeffreys 先验

- 对非位置参数族和刻度参数族的无信息先验如何求, 被广泛采用的是 Jeffreys (1961) 的方法, 由于推导涉及到变换群和 Harr 测度, 这里只给出结果, 不推导结果是如何得来的.
- 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是来自总体 $f(x|\theta)$ 的简单样本, 这里 $\theta = (\theta_1, \dots, \theta_p)$ 为 p 维参数向量. 在对 θ 无先验信息可用时, Jeffreys 用 Fisher 信息阵行列式的平方根作为 θ 的无信息先验, 此先验称为 Jeffreys 无信息先验. 其求解步骤如下:
- 写出样本的对数似然函数

$$l(\theta|\mathbf{x}) = \ln \left[\prod_{i=1}^n f(x_i|\theta) \right] = \sum_{i=1}^n \ln f(x_i|\theta).$$

Jeffreys 无信息先验

- 求样本的信息阵

$$I(\theta) = (I_{ij}(\theta))_{p \times p}, \quad I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} \quad i, j = 1, \dots, p.$$

易证在 C-R 正则条件下

$$I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} = E_{\mathbf{x}|\theta} \left\{ \frac{\partial l}{\partial \theta_i} \cdot \frac{\partial l}{\partial \theta_j} \right\}.$$

- 特别对 $p = 1$, 即 θ 为单参数的情形 $I(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta^2} \right\}$.

- θ 的无信息先验的密度为

$$\pi(\theta) = [\det I(\theta)]^{1/2},$$

其中 $\det I(\theta)$ 表示 p 阶方阵 $I(\theta)$ 的行列式.

- 特别 $p = 1$, 即单参数场合有 $\pi(\theta) = [I(\theta)]^{1/2}$.

Jeffreys 无信息先验

- 求样本的信息阵

$$I(\theta) = (I_{ij}(\theta))_{p \times p}, \quad I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} \quad i, j = 1, \dots, p.$$

易证在 C-R 正则条件下

$$I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} = E_{\mathbf{x}|\theta} \left\{ \frac{\partial l}{\partial \theta_i} \cdot \frac{\partial l}{\partial \theta_j} \right\}.$$

- 特别对 $p = 1$, 即 θ 为单参数的情形 $I(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta^2} \right\}$.

- θ 的无信息先验的密度为

$$\pi(\theta) = [\det I(\theta)]^{1/2},$$

其中 $\det I(\theta)$ 表示 p 阶方阵 $I(\theta)$ 的行列式.

- 特别 $p = 1$, 即单参数场合有 $\pi(\theta) = [I(\theta)]^{1/2}$.

Jeffreys 无信息先验

- 求样本的信息阵

$$I(\theta) = (I_{ij}(\theta))_{p \times p}, \quad I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} \quad i, j = 1, \dots, p.$$

易证在 C-R 正则条件下

$$I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} = E_{\mathbf{x}|\theta} \left\{ \frac{\partial l}{\partial \theta_i} \cdot \frac{\partial l}{\partial \theta_j} \right\}.$$

- 特别对 $p = 1$, 即 θ 为单参数的情形 $I(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta^2} \right\}$.
- θ 的无信息先验的密度为

$$\pi(\theta) = [\det I(\theta)]^{1/2},$$

其中 $\det I(\theta)$ 表示 p 阶方阵 $I(\theta)$ 的行列式.

- 特别 $p = 1$, 即单参数场合有 $\pi(\theta) = [I(\theta)]^{1/2}$.

Jeffreys 无信息先验

- 求样本的信息阵

$$I(\theta) = (I_{ij}(\theta))_{p \times p}, \quad I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} \quad i, j = 1, \dots, p.$$

易证在 C-R 正则条件下

$$I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} = E_{\mathbf{x}|\theta} \left\{ \frac{\partial l}{\partial \theta_i} \cdot \frac{\partial l}{\partial \theta_j} \right\}.$$

- 特别对 $p = 1$, 即 θ 为单参数的情形 $I(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta^2} \right\}$.
- θ 的无信息先验的密度为

$$\pi(\theta) = [\det I(\theta)]^{1/2},$$

其中 $\det I(\theta)$ 表示 p 阶方阵 $I(\theta)$ 的行列式.

- 特别 $p = 1$, 即单参数场合有 $\pi(\theta) = [I(\theta)]^{1/2}$.

Jeffreys 无信息先验

- 求样本的信息阵

$$I(\theta) = (I_{ij}(\theta))_{p \times p}, \quad I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} \quad i, j = 1, \dots, p.$$

易证在 C-R 正则条件下

$$I_{ij}(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta_i \partial \theta_j} \right\} = E_{\mathbf{x}|\theta} \left\{ \frac{\partial l}{\partial \theta_i} \cdot \frac{\partial l}{\partial \theta_j} \right\}.$$

- 特别对 $p = 1$, 即 θ 为单参数的情形 $I(\theta) = E_{\mathbf{x}|\theta} \left\{ -\frac{\partial^2 l}{\partial \theta^2} \right\}$.
- θ 的无信息先验的密度为

$$\pi(\theta) = [\det I(\theta)]^{1/2},$$

其中 $\det I(\theta)$ 表示 p 阶方阵 $I(\theta)$ 的行列式.

- 特别 $p = 1$, 即单参数场合有 $\pi(\theta) = [I(\theta)]^{1/2}$.

Jeffreyes 先验*的例子

- **例 2.4.4** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的简单样本, 记 $\theta = (\mu, \sigma)$, 求 (μ, σ) 的联合无信息先验.
- 按上述步骤易求 (μ, σ) 的联合无信息先验为 $1/\sigma^2$. 它的几个特例为:
 - (1) 当 σ 已知时, $I(\mu) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu^2}\right\} = n/\sigma^2$, 故取 $\pi(\mu) \equiv 1$.
 - (2) 当 μ 已知时, $I(\sigma) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \sigma^2}\right\} = 2n/\sigma^2$, 故取 $\pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
 - (3) 当 μ 和 σ 独立时, $\pi(\mu, \sigma) = \pi(\mu) \cdot \pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
- 由此可见, 当 μ 和 σ 的先验不独立时, 他们的联合无信息先验为 $1/\sigma^2$; 而当 μ 和 σ 的先验独立时, 其联合无信息先验为 $1/\sigma$. Jeffreyes 最终推荐联合无信息先验用 $\pi(\mu, \sigma) = 1/\sigma$.

Jeffreyes 先验*的例子

- **例 2.4.4** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的简单样本, 记 $\theta = (\mu, \sigma)$, 求 (μ, σ) 的联合无信息先验.
- 按上述步骤易求 (μ, σ) 的联合无信息先验为 $1/\sigma^2$. 它的几个特例为:
 - (1) 当 σ 已知时, $I(\mu) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu^2}\right\} = n/\sigma^2$, 故取 $\pi(\mu) \equiv 1$.
 - (2) 当 μ 已知时, $I(\sigma) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \sigma^2}\right\} = 2n/\sigma^2$, 故取 $\pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
 - (3) 当 μ 和 σ 独立时, $\pi(\mu, \sigma) = \pi(\mu) \cdot \pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
- 由此可见, 当 μ 和 σ 的先验不独立时, 他们的联合无信息先验为 $1/\sigma^2$; 而当 μ 和 σ 的先验独立时, 其联合无信息先验为 $1/\sigma$. Jeffreyes 最终推荐联合无信息先验用 $\pi(\mu, \sigma) = 1/\sigma$.

Jeffreyes 先验*的例子

- **例 2.4.4** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的简单样本, 记 $\theta = (\mu, \sigma)$, 求 (μ, σ) 的联合无信息先验.
- 按上述步骤易求 (μ, σ) 的联合无信息先验为 $1/\sigma^2$. 它的几个特例为:
 - (1) 当 σ 已知时, $I(\mu) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu^2}\right\} = n/\sigma^2$, 故取 $\pi(\mu) \equiv 1$.
 - (2) 当 μ 已知时, $I(\sigma) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \sigma^2}\right\} = 2n/\sigma^2$, 故取 $\pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
 - (3) 当 μ 和 σ 独立时, $\pi(\mu, \sigma) = \pi(\mu) \cdot \pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
- 由此可见, 当 μ 和 σ 的先验不独立时, 他们的联合无信息先验为 $1/\sigma^2$; 而当 μ 和 σ 的先验独立时, 其联合无信息先验为 $1/\sigma$. Jeffreyes 最终推荐联合无信息先验用 $\pi(\mu, \sigma) = 1/\sigma$.

Jeffreyes 先验*的例子

- **例 2.4.4** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的简单样本, 记 $\theta = (\mu, \sigma)$, 求 (μ, σ) 的联合无信息先验.
- 按上述步骤易求 (μ, σ) 的联合无信息先验为 $1/\sigma^2$. 它的几个特例为:
 - (1) 当 σ 已知时, $I(\mu) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu^2}\right\} = n/\sigma^2$, 故取 $\pi(\mu) \equiv 1$.
 - (2) 当 μ 已知时, $I(\sigma) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \sigma^2}\right\} = 2n/\sigma^2$, 故取 $\pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
 - (3) 当 μ 和 σ 独立时, $\pi(\mu, \sigma) = \pi(\mu) \cdot \pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
- 由此可见, 当 μ 和 σ 的先验不独立时, 他们的联合无信息先验为 $1/\sigma^2$; 而当 μ 和 σ 的先验独立时, 其联合无信息先验为 $1/\sigma$. Jeffreyes 最终推荐联合无信息先验用 $\pi(\mu, \sigma) = 1/\sigma$.

Jeffreyes 先验*的例子

- **例 2.4.4** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的简单样本, 记 $\theta = (\mu, \sigma)$, 求 (μ, σ) 的联合无信息先验.
- 按上述步骤易求 (μ, σ) 的联合无信息先验为 $1/\sigma^2$. 它的几个特例为:
 - (1) 当 σ 已知时, $I(\mu) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu^2}\right\} = n/\sigma^2$, 故取 $\pi(\mu) \equiv 1$.
 - (2) 当 μ 已知时, $I(\sigma) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \sigma^2}\right\} = 2n/\sigma^2$, 故取 $\pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
 - (3) 当 μ 和 σ 独立时, $\pi(\mu, \sigma) = \pi(\mu) \cdot \pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
- 由此可见, 当 μ 和 σ 的先验不独立时, 他们的联合无信息先验为 $1/\sigma^2$; 而当 μ 和 σ 的先验独立时, 其联合无信息先验为 $1/\sigma$. Jeffreyes 最终推荐联合无信息先验用 $\pi(\mu, \sigma) = 1/\sigma$.

Jeffreyes 先验*的例子

- **例 2.4.4** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的简单样本, 记 $\theta = (\mu, \sigma)$, 求 (μ, σ) 的联合无信息先验.
- 按上述步骤易求 (μ, σ) 的联合无信息先验为 $1/\sigma^2$. 它的几个特例为:
 - (1) 当 σ 已知时, $I(\mu) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu^2}\right\} = n/\sigma^2$, 故取 $\pi(\mu) \equiv 1$.
 - (2) 当 μ 已知时, $I(\sigma) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \sigma^2}\right\} = 2n/\sigma^2$, 故取 $\pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
 - (3) 当 μ 和 σ 独立时, $\pi(\mu, \sigma) = \pi(\mu) \cdot \pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
- 由此可见, 当 μ 和 σ 的先验不独立时, 他们的联合无信息先验为 $1/\sigma^2$; 而当 μ 和 σ 的先验独立时, 其联合无信息先验为 $1/\sigma$. Jeffreyes 最终推荐联合无信息先验用 $\pi(\mu, \sigma) = 1/\sigma$.

Jeffreyes 先验*的例子

- **例 2.4.4** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从总体 $N(\mu, \sigma^2)$ 中抽取的简单样本, 记 $\theta = (\mu, \sigma)$, 求 (μ, σ) 的联合无信息先验.
- 按上述步骤易求 (μ, σ) 的联合无信息先验为 $1/\sigma^2$. 它的几个特例为:
 - (1) 当 σ 已知时, $I(\mu) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \mu^2}\right\} = n/\sigma^2$, 故取 $\pi(\mu) \equiv 1$.
 - (2) 当 μ 已知时, $I(\sigma) = E\left\{-\frac{\partial^2 l(\theta|\mathbf{x})}{\partial \sigma^2}\right\} = 2n/\sigma^2$, 故取 $\pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
 - (3) 当 μ 和 σ 独立时, $\pi(\mu, \sigma) = \pi(\mu) \cdot \pi(\sigma) = 1/\sigma$, $\sigma \in (0, \infty)$.
- 由此可见, 当 μ 和 σ 的先验不独立时, 他们的联合无信息先验为 $1/\sigma^2$; 而当 μ 和 σ 的先验独立时, 其联合无信息先验为 $1/\sigma$. Jeffreyes 最终推荐联合无信息先验用 $\pi(\mu, \sigma) = 1/\sigma$.

Jeffreyes 先验*的例子

- 例 2.4.5 设 θ 为 Benoulli 试验中成功概率, 则在 n 次独立的 Benoulli 试验中, 成功次数 $X \sim b(n, \theta)$. 即

$$P(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n.$$

求 θ 的 Jeffreyes 先验.

- 注 2.4.2 本例中求出 θ 的 Jeffreyes 先验是 $Be(1/2, 1/2)$, 我们还可以取 θ 的无信息先验为 $(0, 1)$ 上的均匀分布. 一般说来无信息先验不唯一, 它们对贝叶斯推断影响都很小, 所以任何无信息先验都可以接受. 当今无信息先验采用越来越多, 就连经典统计学者也认为无信息先验是客观的, 是可以接受的. 这是近几十年中贝叶斯学派研究中最成功的部分.

Jeffreyes 先验*的例子

- **例 2.4.5** 设 θ 为 Benoulli 试验中成功概率, 则在 n 次独立的 Benoulli 试验中, 成功次数 $X \sim b(n, \theta)$. 即

$$P(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}, \quad x = 0, 1, \dots, n.$$

求 θ 的 Jeffreyes 先验.

- **注 2.4.2** 本例中求出 θ 的 Jeffreyes 先验是 $Be(1/2, 1/2)$, 我们还可以取 θ 的无信息先验为 $(0, 1)$ 上的均匀分布. 一般说来无信息先验不唯一, 它们对贝叶斯推断影响都很小, 所以任何无信息先验都可以接受. 当今无信息先验采用越来越多, 就连经典统计学者也认为无信息先验是客观的, 是可以接受的. 这是近几十年中贝叶斯学派研究中最成功的部分.

1 2.1 主观概率

- 2.1.1 主观概率的定义
- 2.1.2 确定主观概率的方法

2 2.2 利用先验信息确定先验分布

- 2.2.1 直方图法
- 2.2.2 相对似然法
- 2.2.3 选定先验密度函数的形式, 再估计超参数
- 2.2.4 给定参数 θ 分位数, 确定累计分布函数

3 2.3 利用边缘分布 $m(x)$ 确定先验分布

- 2.3.1 边缘分布的定义及例
- 2.3.2 选择先验分布的 ML-II 方法
- 2.3.3 选择先验分布的矩方法

4 2.4 无信息先验

- 2.4.1 Laplace 先验与广义先验分布
- 2.4.2 位置参数的无信息先验
- 2.4.3 刻度参数的无信息先验
- 2.4.4 一般情形下的无信息先验

5 2.5 共轭先验分布

共轭先验分布的定义及例

- 另外一种选择先验的方法是在已知样本分布的情形下, 为了理论上的需要常常选参数的先验分布为共轭先验分布.
- **定义 2.5.1** 设 \mathcal{F} 为 θ 的先验分布 π 构成的分布族, 样本 X 的分布为 $f(x|\theta)$, 如果对任意取的 $\pi \in \mathcal{F}$ 及样本 x , 后验分布 $\pi(\theta|x)$ 仍属于 \mathcal{F} , 则称 \mathcal{F} 是一个共轭先验分布族.
- **注 2.5.1** 上述定义中的后验分布 $\pi(\theta|x)$ 的形式不仅依赖先验分布 π , 还依赖于样本分布族的形式. 离开指定参数及其所在的样本分布族去谈论共轭先验分布是没有意义的. 因此, 指定的先验分布族是否是共轭, 要视样本分布族而定.
- **例 2.5.1** 设 $X \sim b(n, \theta)$. (1) 设 $\theta \sim$ 均匀分布 $U(0, 1)$, 证明 θ 的后验分布为贝塔分布; (2) 若取 θ 的先验分布为贝塔分布 $Be(a, b)$, 证明 θ 的后验分布仍为贝塔分布. 即样本分布如果为二项分布, 则共轭先验分布为贝塔分布.

共轭先验分布的定义及例

- 另外一种选择先验的方法是在已知样本分布的情形下, 为了理论上的需要常常选参数的先验分布为共轭先验分布.
- **定义 2.5.1** 设 \mathcal{F} 为 θ 的先验分布 π 构成的分布族, 样本 X 的分布为 $f(x|\theta)$, 如果对任意取的 $\pi \in \mathcal{F}$ 及样本 x , 后验分布 $\pi(\theta|x)$ 仍属于 \mathcal{F} , 则称 \mathcal{F} 是一个共轭先验分布族.
- **注 2.5.1** 上述定义中的后验分布 $\pi(\theta|x)$ 的形式不仅依赖先验分布 π , 还依赖于样本分布族的形式. 离开指定参数及其所在的样本分布族去谈论共轭先验分布是没有意义的. 因此, 指定的先验分布族是否是共轭, 要视样本分布族而定.
- **例 2.5.1** 设 $X \sim b(n, \theta)$. (1) 设 $\theta \sim$ 均匀分布 $U(0, 1)$, 证明 θ 的后验分布为贝塔分布; (2) 若取 θ 的先验分布为贝塔分布 $Be(a, b)$, 证明 θ 的后验分布仍为贝塔分布. 即样本分布如果为二项分布, 则共轭先验分布为贝塔分布.

共轭先验分布的定义及例

- 另外一种选择先验的方法是在已知样本分布的情形下, 为了理论上的需要常常选参数的先验分布为共轭先验分布.
- **定义 2.5.1** 设 \mathcal{F} 为 θ 的先验分布 π 构成的分布族, 样本 X 的分布为 $f(x|\theta)$, 如果对任意取的 $\pi \in \mathcal{F}$ 及样本 x , 后验分布 $\pi(\theta|x)$ 仍属于 \mathcal{F} , 则称 \mathcal{F} 是一个共轭先验分布族.
- **注 2.5.1** 上述定义中的后验分布 $\pi(\theta|x)$ 的形式不仅依赖先验分布 π , 还依赖于样本分布族的形式. 离开指定参数及其所在的样本分布族去谈论共轭先验分布是没有意义的. 因此, 指定的先验分布族是否是共轭, 要视样本分布族而定.
- **例 2.5.1** 设 $X \sim b(n, \theta)$. (1) 设 $\theta \sim$ 均匀分布 $U(0, 1)$, 证明 θ 的后验分布为贝塔分布; (2) 若取 θ 的先验分布为贝塔分布 $Be(a, b)$, 证明 θ 的后验分布仍为贝塔分布. 即样本分布如果为二项分布, 则共轭先验分布为贝塔分布.

共轭先验分布的定义及例

- 另外一种选择先验的方法是在已知样本分布的情形下, 为了理论上的需要常常选参数的先验分布为共轭先验分布.
- **定义 2.5.1** 设 \mathcal{F} 为 θ 的先验分布 π 构成的分布族, 样本 X 的分布为 $f(x|\theta)$, 如果对任意取的 $\pi \in \mathcal{F}$ 及样本 x , 后验分布 $\pi(\theta|x)$ 仍属于 \mathcal{F} , 则称 \mathcal{F} 是一个共轭先验分布族.
- **注 2.5.1** 上述定义中的后验分布 $\pi(\theta|x)$ 的形式不仅依赖先验分布 π , 还依赖于样本分布族的形式. 离开指定参数及其所在的样本分布族去谈论共轭先验分布是没有意义的. 因此, 指定的先验分布族是否是共轭, 要视样本分布族而定.
- **例 2.5.1** 设 $X \sim b(n, \theta)$. (1) 设 $\theta \sim$ 均匀分布 $U(0, 1)$, 证明 θ 的后验分布为贝塔分布; (2) 若取 θ 的先验分布为贝塔分布 $Be(a, b)$, 证明 θ 的后验分布仍为贝塔分布. 即样本分布如果为二项分布, 则共轭先验分布为贝塔分布.

共轭先验分布的定义及例

- 另外一种选择先验的方法是在已知样本分布的情形下, 为了理论上的需要常常选参数的先验分布为共轭先验分布.
- **定义 2.5.1** 设 \mathcal{F} 为 θ 的先验分布 π 构成的分布族, 样本 X 的分布为 $f(x|\theta)$, 如果对任意取的 $\pi \in \mathcal{F}$ 及样本 x , 后验分布 $\pi(\theta|x)$ 仍属于 \mathcal{F} , 则称 \mathcal{F} 是一个共轭先验分布族.
- **注 2.5.1** 上述定义中的后验分布 $\pi(\theta|x)$ 的形式不仅依赖先验分布 π , 还依赖于样本分布族的形式. 离开指定参数及其所在的样本分布族去谈论共轭先验分布是没有意义的. 因此, 指定的先验分布族是否是共轭, 要视样本分布族而定.
- **例 2.5.1** 设 $X \sim b(n, \theta)$. (1) 设 $\theta \sim$ 均匀分布 $U(0, 1)$, 证明 θ 的后验分布为贝塔分布; (2) 若取 θ 的先验分布为贝塔分布 $Be(a, b)$, 证明 θ 的后验分布仍为贝塔分布. 即样本分布如果为二项分布, 则共轭先验分布为贝塔分布.

后验分布的计算

- 后验密度的计算公式由 (1.2.1) 式给出, 其中 $f(x|\theta)$ 是样本的密度函数, 它可以用似然函数 $l(\theta|x)$ 代替.
- 在公式 (1.2.1) 式中, $m(x)$ 是 X 的边缘密度. 由于 $m(x)$ 与 θ 无关, 故可将 $1/m(x)$ 看成与 θ 无关的常数, 因此有

$$\pi(\theta|x) = f(x|\theta)\pi(\theta)/m(x) \propto f(x|\theta)\pi(\theta). \quad (2.5.3)$$

- 对共轭先验分布情形, 求后验密度可按下列步骤:
 - (1) 写出似然函数 $l(\theta|x)$ 的核, 即 $l(\theta|x)$ 中与 θ 有关的因子. 再写出先验密度 $\pi(\theta)$ 的核, 即 $\pi(\theta)$ 中与 θ 有关的因子.
 - (2) 类似公式 (2.5.3), 写出后验密度的核, 即
$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \{l(\theta|x) \text{ 的核}\} \cdot \{\pi(\theta) \text{ 的核}\}. \quad (2.5.4)$$
即“后验密度核”是“似然函数核与先验密度核之积”.
 - (3) 将公式 (2.5.4) 的右边添加一个正则化常数因子 (可以与 x 有关), 即可得到后验密度.

后验分布的计算

- 后验密度的计算公式由 (1.2.1) 式给出, 其中 $f(x|\theta)$ 是样本的密度函数, 它可以用似然函数 $l(\theta|x)$ 代替.
- 在公式 (1.2.1) 式中, $m(x)$ 是 X 的边缘密度. 由于 $m(x)$ 与 θ 无关, 故可将 $1/m(x)$ 看成与 θ 无关的常数, 因此有

$$\pi(\theta|x) = f(x|\theta)\pi(\theta)/m(x) \propto f(x|\theta)\pi(\theta). \quad (2.5.3)$$

- 对共轭先验分布情形, 求后验密度可按下列步骤:
 - (1) 写出似然函数 $l(\theta|x)$ 的核, 即 $l(\theta|x)$ 中与 θ 有关的因子. 再写出先验密度 $\pi(\theta)$ 的核, 即 $\pi(\theta)$ 中与 θ 有关的因子.
 - (2) 类似公式 (2.5.3), 写出后验密度的核, 即
$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \{l(\theta|x) \text{ 的核} \} \cdot \{\pi(\theta) \text{ 的核} \}. \quad (2.5.4)$$
即“后验密度核”是“似然函数核与先验密度核之积”.
 - (3) 将公式 (2.5.4) 的右边添加一个正则化常数因子 (可以与 x 有关), 即可得到后验密度.

后验分布的计算

- 后验密度的计算公式由 (1.2.1) 式给出, 其中 $f(x|\theta)$ 是样本的密度函数, 它可以用似然函数 $l(\theta|x)$ 代替.
- 在公式 (1.2.1) 式中, $m(x)$ 是 X 的边缘密度. 由于 $m(x)$ 与 θ 无关, 故可将 $1/m(x)$ 看成与 θ 无关的常数, 因此有

$$\pi(\theta|x) = f(x|\theta)\pi(\theta)/m(x) \propto f(x|\theta)\pi(\theta). \quad (2.5.3)$$

- 对共轭先验分布情形, 求后验密度可按下列步骤:
 - (1) 写出似然函数 $l(\theta|x)$ 的核, 即 $l(\theta|x)$ 中与 θ 有关的因子. 再写出先验密度 $\pi(\theta)$ 的核, 即 $\pi(\theta)$ 中与 θ 有关的因子.
 - (2) 类似公式 (2.5.3), 写出后验密度的核, 即
$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \{l(\theta|x) \text{ 的核}\} \cdot \{\pi(\theta) \text{ 的核}\}. \quad (2.5.4)$$
即“后验密度核”是“似然函数核与先验密度核之积”.
 - (3) 将公式 (2.5.4) 的右边添加一个正则化常数因子 (可以与 x 有关), 即可得到后验密度.

后验分布的计算

- 后验密度的计算公式由 (1.2.1) 式给出, 其中 $f(x|\theta)$ 是样本的密度函数, 它可以用似然函数 $l(\theta|x)$ 代替.
- 在公式 (1.2.1) 式中, $m(x)$ 是 X 的边缘密度. 由于 $m(x)$ 与 θ 无关, 故可将 $1/m(x)$ 看成与 θ 无关的常数, 因此有

$$\pi(\theta|x) = f(x|\theta)\pi(\theta)/m(x) \propto f(x|\theta)\pi(\theta). \quad (2.5.3)$$

- 对共轭先验分布情形, 求后验密度可按下列步骤:
 - (1) 写出似然函数 $l(\theta|x)$ 的核, 即 $l(\theta|x)$ 中与 θ 有关的因子. 再写出先验密度 $\pi(\theta)$ 的核, 即 $\pi(\theta)$ 中与 θ 有关的因子.
 - (2) 类似公式 (2.5.3), 写出后验密度的核, 即
$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \{l(\theta|x) \text{ 的核}\} \cdot \{\pi(\theta) \text{ 的核}\}. \quad (2.5.4)$$
即“后验密度核”是“似然函数核与先验密度核之积”.
 - (3) 将公式 (2.5.4) 的右边添加一个正则化常数因子 (可以与 x 有关), 即可得到后验密度.

后验分布的计算

- 后验密度的计算公式由 (1.2.1) 式给出, 其中 $f(x|\theta)$ 是样本的密度函数, 它可以用似然函数 $l(\theta|x)$ 代替.
- 在公式 (1.2.1) 式中, $m(x)$ 是 X 的边缘密度. 由于 $m(x)$ 与 θ 无关, 故可将 $1/m(x)$ 看成与 θ 无关的常数, 因此有

$$\pi(\theta|x) = f(x|\theta)\pi(\theta)/m(x) \propto f(x|\theta)\pi(\theta). \quad (2.5.3)$$

- 对共轭先验分布情形, 求后验密度可按下列步骤:
 - 写出似然函数 $l(\theta|x)$ 的核, 即 $l(\theta|x)$ 中与 θ 有关的因子. 再写出先验密度 $\pi(\theta)$ 的核, 即 $\pi(\theta)$ 中与 θ 有关的因子.
 - 类似公式 (2.5.3), 写出后验密度的核, 即
$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \{l(\theta|x) \text{ 的核}\} \cdot \{\pi(\theta) \text{ 的核}\}. \quad (2.5.4)$$
即“后验密度核”是“似然函数核与先验密度核之积”.
 - 将公式 (2.5.4) 的右边添加一个正则化常数因子 (可以与 x 有关), 即可得到后验密度.

后验分布的计算

- 后验密度的计算公式由 (1.2.1) 式给出, 其中 $f(x|\theta)$ 是样本的密度函数, 它可以用似然函数 $l(\theta|x)$ 代替.
- 在公式 (1.2.1) 式中, $m(x)$ 是 X 的边缘密度. 由于 $m(x)$ 与 θ 无关, 故可将 $1/m(x)$ 看成与 θ 无关的常数, 因此有

$$\pi(\theta|x) = f(x|\theta)\pi(\theta)/m(x) \propto f(x|\theta)\pi(\theta). \quad (2.5.3)$$

- 对共轭先验分布情形, 求后验密度可按下列步骤:
 - 写出似然函数 $l(\theta|x)$ 的核, 即 $l(\theta|x)$ 中与 θ 有关的因子. 再写出先验密度 $\pi(\theta)$ 的核, 即 $\pi(\theta)$ 中与 θ 有关的因子.
 - 类似公式 (2.5.3), 写出后验密度的核, 即
$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \{l(\theta|x) \text{ 的核}\} \cdot \{\pi(\theta) \text{ 的核}\}. \quad (2.5.4)$$
即“后验密度核”是“似然函数核与先验密度核之积”.
 - 将公式 (2.5.4) 的右边添加一个正则化常数因子 (可以与 x 有关), 即可得到后验密度.

注释及例子

- **注 2.5.2** 上述计算后验分布的简化方法, 只对先验分布为共轭或无信息先验的情形有效. 当先验分布非上述情形, 获得后验分布的核之后, 如果不能判断后验分布的类型, 就不知道如何添加正则化常数因子将“后验密度的核”变成“后验密度”. 此时只有老老实实按公式 (1.2.1) 计算后验密度.
- **续例 2.5.1** 设 $X \sim b(n, \theta)$, 若取 θ 的先验分布为 $Be(a, b)$, 用上面介绍的方法求 θ 的后验分布.
- **例 2.5.2** 设 $X \sim N(\theta, \sigma^2)$, σ^2 已知而 θ 未知. 令 θ 的先验分布 $\pi(\theta)$ 是 $N(\mu, \tau^2)$, 其中超参数 μ 和 τ^2 已知, 求 θ 的后验分布 $\pi(\theta|x)$.
- 在此例中, 若进一步假定 X_1, \dots, X_n i.i.d. $\sim N(\theta, \sigma^2)$, σ^2 已知, $\theta \sim N(\mu, \tau^2)$, 求 θ 的后验密度.

注释及例子

- **注 2.5.2** 上述计算后验分布的简化方法, 只对先验分布为共轭或无信息先验的情形有效. 当先验分布非上述情形, 获得后验分布的核之后, 如果不能判断后验分布的类型, 就不知道如何添加正则化常数因子将“后验密度的核”变成“后验密度”. 此时只有老老实实按公式 (1.2.1) 计算后验密度.
- **续例 2.5.1** 设 $X \sim b(n, \theta)$, 若取 θ 的先验分布为 $Be(a, b)$, 用上面介绍的方法求 θ 的后验分布.
- **例 2.5.2** 设 $X \sim N(\theta, \sigma^2)$, σ^2 已知而 θ 未知. 令 θ 的先验分布 $\pi(\theta)$ 是 $N(\mu, \tau^2)$, 其中超参数 μ 和 τ^2 已知, 求 θ 的后验分布 $\pi(\theta|x)$.
- 在此例中, 若进一步假定 X_1, \dots, X_n i.i.d. $\sim N(\theta, \sigma^2)$, σ^2 已知, $\theta \sim N(\mu, \tau^2)$, 求 θ 的后验密度.

注释及例子

- **注 2.5.2** 上述计算后验分布的简化方法, 只对先验分布为共轭或无信息先验的情形有效. 当先验分布非上述情形, 获得后验分布的核之后, 如果不能判断后验分布的类型, 就不知道如何添加正则化常数因子将“后验密度的核”变成“后验密度”. 此时只有老老实实按公式 (1.2.1) 计算后验密度.
- **续例 2.5.1** 设 $X \sim b(n, \theta)$, 若取 θ 的先验分布为 $Be(a, b)$, 用上面介绍的方法求 θ 的后验分布.
- **例 2.5.2** 设 $X \sim N(\theta, \sigma^2)$, σ^2 已知而 θ 未知. 令 θ 的先验分布 $\pi(\theta)$ 是 $N(\mu, \tau^2)$, 其中超参数 μ 和 τ^2 已知, 求 θ 的后验分布 $\pi(\theta|x)$.
- 在此例中, 若进一步假定 X_1, \dots, X_n i.i.d. $\sim N(\theta, \sigma^2)$, σ^2 已知, $\theta \sim N(\mu, \tau^2)$, 求 θ 的后验密度.

注释及例子

- **注 2.5.2** 上述计算后验分布的简化方法, 只对先验分布为共轭或无信息先验的情形有效. 当先验分布非上述情形, 获得后验分布的核之后, 如果不能判断后验分布的类型, 就不知道如何添加正则化常数因子将“后验密度的核”变成“后验密度”. 此时只有老老实实按公式 (1.2.1) 计算后验密度.
- **续例 2.5.1** 设 $X \sim b(n, \theta)$, 若取 θ 的先验分布为 $Be(a, b)$, 用上面介绍的方法求 θ 的后验分布.
- **例 2.5.2** 设 $X \sim N(\theta, \sigma^2)$, σ^2 已知而 θ 未知. 令 θ 的先验分布 $\pi(\theta)$ 是 $N(\mu, \tau^2)$, 其中超参数 μ 和 τ^2 已知, 求 θ 的后验分布 $\pi(\theta|x)$.
- 在此例中, 若进一步假定 X_1, \dots, X_n i.i.d. $\sim N(\theta, \sigma^2)$, σ^2 已知, $\theta \sim N(\mu, \tau^2)$, 求 θ 的后验密度.

逆伽玛分布的定义及例

- **定义 2.5.3 (逆伽玛分布)** 若样本 X 的密度函数为

$$f(x|\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\lambda}{x}}, \quad x > 0,$$

则称 r.v. X 的分布为逆伽玛分布, 记为 $X \sim \Gamma^{-1}(\alpha, \lambda)$. 显然, 若 $X \sim \Gamma(\alpha, \lambda)$, 则 $Y = 1/X \sim \Gamma^{-1}(\alpha, \lambda)$.

- **例 2.5.5** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态分布 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 θ 已知, 求 σ^2 的共轭先验分布.

逆伽玛分布的定义及例

- **定义 2.5.3 (逆伽玛分布)** 若样本 X 的密度函数为

$$f(x|\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-\frac{\lambda}{x}}, \quad x > 0,$$

则称 r.v. X 的分布为逆伽玛分布, 记为 $X \sim \Gamma^{-1}(\alpha, \lambda)$. 显然, 若 $X \sim \Gamma(\alpha, \lambda)$, 则 $Y = 1/X \sim \Gamma^{-1}(\alpha, \lambda)$.

- **例 2.5.5** 设 $\mathbf{X} = (X_1, \dots, X_n)$ 是从正态分布 $N(\theta, \sigma^2)$ 中抽取的简单样本, 其中 θ 已知, 求 σ^2 的共轭先验分布.

2.5 共轭先验分布

2.5.2 后验分布的计算及例

共轭先验分布表

样本分布 $f(x \theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta x)$
正态 $N(\theta, \sigma^2)$	正态 $N(\mu, \tau^2)$	正态 $N\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
伽玛 $\Gamma(r, \theta)$ ($r=1$ 为指数分布)	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(r + \alpha, x + \beta)$
均匀 $U(0, \theta)$	帕雷托 $Pa(\theta_0, \alpha)$	帕雷托 $Pa(\theta_1, \alpha + 1)$ $\theta_1 = \max(x, \theta_0)$
伽玛 $\Gamma(r, 1/\theta)$	逆伽玛 $\Gamma^{-1}(\alpha, \beta)$	逆伽玛 $\Gamma^{-1}(r + \alpha, x + \beta)$
泊松 $P(\theta)$	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(x + \alpha, 1 + \beta)$
二项 $B(n, \theta)$ ($n=1$ 为两点分布)	贝塔 $Be(a, b)$	贝塔 $Be(x + a, n - x + b)$
负二项 $Nb(r, \theta)$ ($r=1$ 为几何分布)	贝塔 $Be(a, b)$	贝塔 $Be(r + a, x - r + b)$
多项 $M(n, \theta)$ $\theta = (\theta_1, \dots, \theta_k)$	$D(\alpha_1, \dots, \alpha_k)$ 狄利克雷分布	$D(x_1 + \alpha_1, \dots, x_k + \alpha_k)$ $x_1 + \dots + x_k = n$

2.5 共轭先验分布

2.5.2 后验分布的计算及例

共轭先验分布表

样本分布 $f(x \theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta x)$
正态 $N(\theta, \sigma^2)$	正态 $N(\mu, \tau^2)$	正态 $N\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
伽玛 $\Gamma(r, \theta)$ ($r=1$ 为指数分布)	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(r + \alpha, x + \beta)$
均匀 $U(0, \theta)$	帕雷托 $Pa(\theta_0, \alpha)$	帕雷托 $Pa(\theta_1, \alpha + 1)$ $\theta_1 = \max(x, \theta_0)$
伽玛 $\Gamma(r, 1/\theta)$	逆伽玛 $\Gamma^{-1}(\alpha, \beta)$	逆伽玛 $\Gamma^{-1}(r + \alpha, x + \beta)$
泊松 $P(\theta)$	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(x + \alpha, 1 + \beta)$
二项 $B(n, \theta)$ ($n=1$ 为两点分布)	贝塔 $Be(a, b)$	贝塔 $Be(x + a, n - x + b)$
负二项 $Nb(r, \theta)$ ($r=1$ 为几何分布)	贝塔 $Be(a, b)$	贝塔 $Be(r + a, x - r + b)$
多项 $M(n, \theta)$ $\theta = (\theta_1, \dots, \theta_k)$	$D(\alpha_1, \dots, \alpha_k)$ 狄利克雷分布	$D(x_1 + \alpha_1, \dots, x_k + \alpha_k)$ $x_1 + \dots + x_k = n$

2.5 共轭先验分布

2.5.2 后验分布的计算及例

共轭先验分布表

样本分布 $f(x \theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta x)$
正态 $N(\theta, \sigma^2)$	正态 $N(\mu, \tau^2)$	正态 $N\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
伽玛 $\Gamma(r, \theta)$ ($r=1$ 为指数分布)	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(r + \alpha, x + \beta)$
均匀 $U(0, \theta)$	帕雷托 $Pa(\theta_0, \alpha)$	帕雷托 $Pa(\theta_1, \alpha + 1)$ $\theta_1 = \max(x, \theta_0)$
伽玛 $\Gamma(r, 1/\theta)$	逆伽玛 $\Gamma^{-1}(\alpha, \beta)$	逆伽玛 $\Gamma^{-1}(r + \alpha, x + \beta)$
泊松 $P(\theta)$	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(x + \alpha, 1 + \beta)$
二项 $B(n, \theta)$ ($n=1$ 为两点分布)	贝塔 $Be(a, b)$	贝塔 $Be(x + a, n - x + b)$
负二项 $Nb(r, \theta)$ ($r=1$ 为几何分布)	贝塔 $Be(a, b)$	贝塔 $Be(r + a, x - r + b)$
多项 $M(n, \theta)$ $\theta = (\theta_1, \dots, \theta_k)$	$D(\alpha_1, \dots, \alpha_k)$ 狄利克雷分布	$D(x_1 + \alpha_1, \dots, x_k + \alpha_k)$ $x_1 + \dots + x_k = n$

2.5 共轭先验分布

2.5.2 后验分布的计算及例

共轭先验分布表

样本分布 $f(x \theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta x)$
正态 $N(\theta, \sigma^2)$	正态 $N(\mu, \tau^2)$	正态 $N\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
伽玛 $\Gamma(r, \theta)$ ($r=1$ 为指数分布)	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(r + \alpha, x + \beta)$
均匀 $U(0, \theta)$	帕雷托 $Pa(\theta_0, \alpha)$	帕雷托 $Pa(\theta_1, \alpha + 1)$ $\theta_1 = \max(x, \theta_0)$
伽玛 $\Gamma(r, 1/\theta)$	逆伽玛 $\Gamma^{-1}(\alpha, \beta)$	逆伽玛 $\Gamma^{-1}(r + \alpha, x + \beta)$
泊松 $P(\theta)$	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(x + \alpha, 1 + \beta)$
二项 $B(n, \theta)$ ($n=1$ 为两点分布)	贝塔 $Be(a, b)$	贝塔 $Be(x + a, n - x + b)$
负二项 $Nb(r, \theta)$ ($r=1$ 为几何分布)	贝塔 $Be(a, b)$	贝塔 $Be(r + a, x - r + b)$
多项 $M(n, \theta)$ $\theta = (\theta_1, \dots, \theta_k)$	$D(\alpha_1, \dots, \alpha_k)$ 狄利克雷分布	$D(x_1 + \alpha_1, \dots, x_k + \alpha_k)$ $x_1 + \dots + x_k = n$

共轭先验分布表

样本分布 $f(x \theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta x)$
正态 $N(\theta, \sigma^2)$	正态 $N(\mu, \tau^2)$	正态 $N\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
伽玛 $\Gamma(r, \theta)$ ($r=1$ 为指数分布)	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(r + \alpha, x + \beta)$
均匀 $U(0, \theta)$	帕雷托 $Pa(\theta_0, \alpha)$	帕雷托 $Pa(\theta_1, \alpha + 1)$ $\theta_1 = \max(x, \theta_0)$
伽玛 $\Gamma(r, 1/\theta)$	逆伽玛 $\Gamma^{-1}(\alpha, \beta)$	逆伽玛 $\Gamma^{-1}(r + \alpha, x + \beta)$
泊松 $P(\theta)$	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(x + \alpha, 1 + \beta)$
二项 $B(n, \theta)$ ($n=1$ 为两点分布)	贝塔 $Be(a, b)$	贝塔 $Be(x + a, n - x + b)$
负二项 $Nb(r, \theta)$ ($r=1$ 为几何分布)	贝塔 $Be(a, b)$	贝塔 $Be(r + a, x - r + b)$
多项 $M(n, \theta)$ $\theta = (\theta_1, \dots, \theta_k)$	$D(\alpha_1, \dots, \alpha_k)$ 狄利克雷分布	$D(x_1 + \alpha_1, \dots, x_k + \alpha_k)$ $x_1 + \dots + x_k = n$

共轭先验分布表

样本分布 $f(x \theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta x)$
正态 $N(\theta, \sigma^2)$	正态 $N(\mu, \tau^2)$	正态 $N\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
伽玛 $\Gamma(r, \theta)$ ($r=1$ 为指数分布)	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(r + \alpha, x + \beta)$
均匀 $U(0, \theta)$	帕雷托 $Pa(\theta_0, \alpha)$	帕雷托 $Pa(\theta_1, \alpha + 1)$ $\theta_1 = \max(x, \theta_0)$
伽玛 $\Gamma(r, 1/\theta)$	逆伽玛 $\Gamma^{-1}(\alpha, \beta)$	逆伽玛 $\Gamma^{-1}(r + \alpha, x + \beta)$
泊松 $P(\theta)$	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(x + \alpha, 1 + \beta)$
二项 $B(n, \theta)$ ($n=1$ 为两点分布)	贝塔 $Be(a, b)$	贝塔 $Be(x + a, n - x + b)$
负二项 $Nb(r, \theta)$ ($r=1$ 为几何分布)	贝塔 $Be(a, b)$	贝塔 $Be(r + a, x - r + b)$
多项 $M(n, \theta)$ $\theta = (\theta_1, \dots, \theta_k)$	$D(\alpha_1, \dots, \alpha_k)$ 狄利克雷分布	$D(x_1 + \alpha_1, \dots, x_k + \alpha_k)$ $x_1 + \dots + x_k = n$

共轭先验分布表

样本分布 $f(x \theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta x)$
正态 $N(\theta, \sigma^2)$	正态 $N(\mu, \tau^2)$	正态 $N\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
伽玛 $\Gamma(r, \theta)$ ($r=1$ 为指数分布)	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(r + \alpha, x + \beta)$
均匀 $U(0, \theta)$	帕雷托 $Pa(\theta_0, \alpha)$	帕雷托 $Pa(\theta_1, \alpha + 1)$ $\theta_1 = \max(x, \theta_0)$
伽玛 $\Gamma(r, 1/\theta)$	逆伽玛 $\Gamma^{-1}(\alpha, \beta)$	逆伽玛 $\Gamma^{-1}(r + \alpha, x + \beta)$
泊松 $P(\theta)$	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(x + \alpha, 1 + \beta)$
二项 $B(n, \theta)$ ($n=1$ 为两点分布)	贝塔 $Be(a, b)$	贝塔 $Be(x + a, n - x + b)$
负二项 $Nb(r, \theta)$ ($r=1$ 为几何分布)	贝塔 $Be(a, b)$	贝塔 $Be(r + a, x - r + b)$
多项 $M(n, \theta)$ $\theta = (\theta_1, \dots, \theta_k)$	$D(\alpha_1, \dots, \alpha_k)$ 狄利克雷分布	$D(x_1 + \alpha_1, \dots, x_k + \alpha_k)$ $x_1 + \dots + x_k = n$

共轭先验分布表

样本分布 $f(x \theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta x)$
正态 $N(\theta, \sigma^2)$	正态 $N(\mu, \tau^2)$	正态 $N\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
伽玛 $\Gamma(r, \theta)$ ($r=1$ 为指数分布)	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(r + \alpha, x + \beta)$
均匀 $U(0, \theta)$	帕雷托 $Pa(\theta_0, \alpha)$	帕雷托 $Pa(\theta_1, \alpha + 1)$ $\theta_1 = \max(x, \theta_0)$
伽玛 $\Gamma(r, 1/\theta)$	逆伽玛 $\Gamma^{-1}(\alpha, \beta)$	逆伽玛 $\Gamma^{-1}(r + \alpha, x + \beta)$
泊松 $P(\theta)$	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(x + \alpha, 1 + \beta)$
二项 $B(n, \theta)$ ($n=1$ 为两点分布)	贝塔 $Be(a, b)$	贝塔 $Be(x + a, n - x + b)$
负二项 $Nb(r, \theta)$ ($r=1$ 为几何分布)	贝塔 $Be(a, b)$	贝塔 $Be(r + a, x - r + b)$
多项 $M(n, \theta)$ $\theta = (\theta_1, \dots, \theta_k)$	$D(\alpha_1, \dots, \alpha_k)$ 狄利克雷分布	$D(x_1 + \alpha_1, \dots, x_k + \alpha_k)$ $x_1 + \dots + x_k = n$

共轭先验分布表

样本分布 $f(x \theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta x)$
正态 $N(\theta, \sigma^2)$	正态 $N(\mu, \tau^2)$	正态 $N\left(\frac{\sigma^2\mu + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}\right)$
伽玛 $\Gamma(r, \theta)$ ($r=1$ 为指数分布)	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(r + \alpha, x + \beta)$
均匀 $U(0, \theta)$	帕雷托 $Pa(\theta_0, \alpha)$	帕雷托 $Pa(\theta_1, \alpha + 1)$ $\theta_1 = \max(x, \theta_0)$
伽玛 $\Gamma(r, 1/\theta)$	逆伽玛 $\Gamma^{-1}(\alpha, \beta)$	逆伽玛 $\Gamma^{-1}(r + \alpha, x + \beta)$
泊松 $P(\theta)$	伽玛 $\Gamma(\alpha, \beta)$	伽玛 $\Gamma(x + \alpha, 1 + \beta)$
二项 $B(n, \theta)$ ($n=1$ 为两点分布)	贝塔 $Be(a, b)$	贝塔 $Be(x + a, n - x + b)$
负二项 $Nb(r, \theta)$ ($r=1$ 为几何分布)	贝塔 $Be(a, b)$	贝塔 $Be(r + a, x - r + b)$
多项 $M(n, \theta)$ $\theta = (\theta_1, \dots, \theta_k)$	$D(\alpha_1, \dots, \alpha_k)$ 狄利克雷分布	$D(x_1 + \alpha_1, \dots, x_k + \alpha_k)$ $x_1 + \dots + x_k = n$

共轭先验分布表的注释

- 上式共轭先验分布表中样本分布是对样本大小为 1 的情形来讨论的.
- 当从总体 X 中抽取大小为 n 的样本 X_1, \dots, X_n 时, 则样本分布应当用样本 X_1, \dots, X_n 的联合分布代替, 后验分布应作相应的调整.
- 若充分统计量存在, 则也可以用充分统计量的分布代替样本 X_1, \dots, X_n 的联合分布, 此时后验分布也应作相应改动.

共轭先验分布表的注释

- 上式共轭先验分布表中样本分布是对样本大小为 1 的情形来讨论的.
- 当从总体 X 中抽取大小为 n 的样本 X_1, \dots, X_n 时, 则样本分布应当用样本 X_1, \dots, X_n 的联合分布代替, 后验分布应作相应的调整.
- 若充分统计量存在, 则也可以用充分统计量的分布代替样本 X_1, \dots, X_n 的联合分布, 此时后验分布也应作相应改动.

共轭先验分布表的注释

- 上式共轭先验分布表中样本分布是对样本大小为 1 的情形来讨论的.
- 当从总体 X 中抽取大小为 n 的样本 X_1, \dots, X_n 时, 则样本分布应当用样本 X_1, \dots, X_n 的联合分布代替, 后验分布应作相应的调整.
- 若充分统计量存在, 则也可以用充分统计量的分布代替样本 X_1, \dots, X_n 的联合分布, 此时后验分布也应作相应改动.

共轭先验的优点

- 共轭先验分布具有下列优点: (1) 计算方便; (2) 后验分布的某些参数常可以得到很好的解释.
- 如例 2.5.2 中, 后验分布为 $N(\mu_n(\mathbf{x}), \eta_n^2)$, 其中

$$\begin{aligned}\mu_n(\mathbf{x}) &= \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \mu + \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x}, \\ &= r_n \mu + (1 - r_n) \bar{x},\end{aligned}$$

此处 $r_n = \frac{\tau^2}{\sigma^2/n + \tau^2}$. 后验均值 $\mu_n(\mathbf{x})$ 是样本均值 \bar{x} 和先验均值 μ 的加权平均. 若 σ^2/n 很小, 即样本信息量相对于先验信息量很大, 则后验均值主要由 \bar{x} 决定, 反之, 则后验均值主要由先验均值 μ 来决定. 故后验均值是样本均值和先验均值的一个折衷.

共轭先验的优点

- 共轭先验分布具有下列优点: (1) 计算方便; (2) 后验分布的某些参数常可以得到很好的解释.
- 如例 2.5.2 中, 后验分布为 $N(\mu_n(\mathbf{x}), \eta_n^2)$, 其中

$$\begin{aligned}\mu_n(\mathbf{x}) &= \frac{\sigma^2/n}{\sigma^2/n + \tau^2} \mu + \frac{\tau^2}{\sigma^2/n + \tau^2} \bar{x}, \\ &= r_n \mu + (1 - r_n) \bar{x},\end{aligned}$$

此处 $r_n = \frac{\tau^2}{\sigma^2/n + \tau^2}$. 后验均值 $\mu_n(\mathbf{x})$ 是样本均值 \bar{x} 和先验均值 μ 的加权平均. 若 σ^2/n 很小, 即样本信息量相对于先验信息量很大, 则后验均值主要由 \bar{x} 决定, 反之, 则后验均值主要由先验均值 μ 来决定. 故后验均值是样本均值和先验均值的一个折衷.

1 2.1 主观概率

- 2.1.1 主观概率的定义
- 2.1.2 确定主观概率的方法

2 2.2 利用先验信息确定先验分布

- 2.2.1 直方图法
- 2.2.2 相对似然法
- 2.2.3 选定先验密度函数的形式, 再估计超参数
- 2.2.4 给定参数 θ 分位数, 确定累计分布函数

3 2.3 利用边缘分布 $m(x)$ 确定先验分布

- 2.3.1 边缘分布的定义及例
- 2.3.2 选择先验分布的 ML-II 方法
- 2.3.3 选择先验分布的矩方法

4 2.4 无信息先验

- 2.4.1 Laplace 先验与广义先验分布
- 2.4.2 位置参数的无信息先验
- 2.4.3 刻度参数的无信息先验
- 2.4.4 一般情形下的无信息先验

5 2.5 共轭先验分布

多层先验分布的定义

- 当所给先验分布中的超参数难于确定时, 可以对超参数再给出一个先验, 第二个先验称为超先验; 若超先验中的超参数还是难以确定时, 还可以再给出第三个先验, 等等. 由先验和超先验决定的一个新先验就称为多层先验, 这是由 Good (1983) 命名的, 也叫做分阶段先验. 其想法大致如下:
- 第一步: 设 $\Gamma_1 = \{\pi_1(\theta|\lambda) : \pi_1 \text{ 的函数形式已知, } \lambda \in \Lambda\}$; 其中 Λ 为超参数 λ 的取值范围, 且 λ 未知.
- 第二步: 设 λ 为随机变量
$$\Gamma_2 = \{\pi_2(\lambda|\delta) : \pi_2 \text{ 的函数形式已知, } \delta \in \Delta\};$$
其中 Δ 为 δ 的取值范围.
- 第三步: 设 δ 也是随机变量, 它有先验分布 $\pi_3(\delta)$.

多层先验分布的定义

- 当所给先验分布中的超参数难于确定时, 可以对超参数再给出一个先验, 第二个先验称为超先验; 若超先验中的超参数还是难以确定时, 还可以再给出第三个先验, 等等. 由先验和超先验决定的一个新先验就称为多层先验, 这是由 Good (1983) 命名的, 也叫做分阶段先验. 其想法大致如下:
- 第一步: 设 $\Gamma_1 = \{\pi_1(\theta|\lambda) : \pi_1 \text{ 的函数形式已知, } \lambda \in \Lambda\}$; 其中 Λ 为超参数 λ 的取值范围, 且 λ 未知.
- 第二步: 设 λ 为随机变量
$$\Gamma_2 = \{\pi_2(\lambda|\delta) : \pi_2 \text{ 的函数形式已知, } \delta \in \Delta\};$$
其中 Δ 为 δ 的取值范围.
- 第三步: 设 δ 也是随机变量, 它有先验分布 $\pi_3(\delta)$.

多层先验分布的定义

- 当所给先验分布中的超参数难于确定时, 可以对超参数再给出一个先验, 第二个先验称为超先验; 若超先验中的超参数还是难以确定时, 还可以再给出第三个先验, 等等. 由先验和超先验决定的一个新先验就称为多层先验, 这是由 Good (1983) 命名的, 也叫做分阶段先验. 其想法大致如下:
- 第一步: 设 $\Gamma_1 = \{\pi_1(\theta|\lambda) : \pi_1 \text{ 的函数形式已知, } \lambda \in \Lambda\}$; 其中 Λ 为超参数 λ 的取值范围, 且 λ 未知.
- 第二步: 设 λ 为随机变量
$$\Gamma_2 = \{\pi_2(\lambda|\delta) : \pi_2 \text{ 的函数形式已知, } \delta \in \Delta\};$$
其中 Δ 为 δ 的取值范围.
- 第三步: 设 δ 也是随机变量, 它有先验分布 $\pi_3(\delta)$.

多层先验分布的定义

- 当所给先验分布中的超参数难于确定时, 可以对超参数再给出一个先验, 第二个先验称为超先验; 若超先验中的超参数还是难以确定时, 还可以再给出第三个先验, 等等. 由先验和超先验决定的一个新先验就称为多层先验, 这是由 Good (1983) 命名的, 也叫做分阶段先验. 其想法大致如下:
- 第一步: 设 $\Gamma_1 = \{\pi_1(\theta|\lambda) : \pi_1 \text{ 的函数形式已知, } \lambda \in \Lambda\}$; 其中 Λ 为超参数 λ 的取值范围, 且 λ 未知.
- 第二步: 设 λ 为随机变量
$$\Gamma_2 = \{\pi_2(\lambda|\delta) : \pi_2 \text{ 的函数形式已知, } \delta \in \Delta\};$$
其中 Δ 为 δ 的取值范围.
- 第三步: 设 δ 也是随机变量, 它有先验分布 $\pi_3(\delta)$.

多层先验分布的定义

- 多层结构与其说是完全新的内容,不如说是先验的一种方便表现方法,任何一个多层先验都可写成一个一般规范的先验.
- 以二层先验为例, 这个规范的先验是

$$\pi(\theta) = \int_{\Lambda} \pi_1(\theta|\lambda)\pi_2(\lambda)d\lambda = \int_{\Lambda} \pi(\theta, \lambda)d\lambda, \quad (2.7.1)$$

此处 $\theta \sim \pi_1(\theta|\lambda)$, $\lambda \sim \pi_2(\lambda)$, θ 和 λ 的联合密度为 $\pi(\theta, \lambda) = \pi_1(\theta|\lambda)\pi_2(\lambda)$, 故 $\pi(\theta)$ 作为联合密度 $\pi(\theta, \lambda)$ 的边缘密度就是 θ 的先验密度.

- 对于三层的先验, 这个规范的先验是

$$\begin{aligned} \pi(\theta) &= \int_{\Lambda} \int_{\Delta} \pi_1(\theta|\lambda)\pi_2(\lambda|\delta)\pi_3(\delta)d\lambda d\delta \\ &= \int_{\Lambda} \pi_1(\theta|\lambda) \left[\int_{\Delta} \pi(\lambda, \delta)d\delta \right] d\lambda. \end{aligned} \quad (2.7.2)$$

其中 $\pi(\lambda, \delta) = \pi_2(\lambda|\delta)\pi_3(\delta)$. 对于更多层的先验, 可用类似方法求得规范的先验 $\pi(\theta)$.

多层先验分布的定义

- 多层结构与其说是完全新的内容,不如说是先验的一种方便表现方法,任何一个多层先验都可写成一个一般规范的先验.
- 以二层先验为例, 这个规范的先验是

$$\pi(\theta) = \int_{\Lambda} \pi_1(\theta|\lambda)\pi_2(\lambda)d\lambda = \int_{\Lambda} \pi(\theta, \lambda)d\lambda, \quad (2.7.1)$$

此处 $\theta \sim \pi_1(\theta|\lambda)$, $\lambda \sim \pi_2(\lambda)$, θ 和 λ 的联合密度为 $\pi(\theta, \lambda) = \pi_1(\theta|\lambda)\pi_2(\lambda)$, 故 $\pi(\theta)$ 作为联合密度 $\pi(\theta, \lambda)$ 的边缘密度就是 θ 的先验密度.

- 对于三层的先验, 这个规范的先验是

$$\begin{aligned} \pi(\theta) &= \int_{\Lambda} \int_{\Delta} \pi_1(\theta|\lambda)\pi_2(\lambda|\delta)\pi_3(\delta)d\lambda d\delta \\ &= \int_{\Lambda} \pi_1(\theta|\lambda) \left[\int_{\Delta} \pi(\lambda, \delta)d\delta \right] d\lambda. \end{aligned} \quad (2.7.2)$$

其中 $\pi(\lambda, \delta) = \pi_2(\lambda|\delta)\pi_3(\delta)$. 对于更多层的先验, 可用类似方法求得规范的先验 $\pi(\theta)$.

多层先验分布的定义

- 多层结构与其说是完全新的内容,不如说是先验的一种方便表现方法,任何一个多层先验都可写成一个一般规范的先验.
- 以二层先验为例, 这个规范的先验是

$$\pi(\theta) = \int_{\Lambda} \pi_1(\theta|\lambda)\pi_2(\lambda)d\lambda = \int_{\Lambda} \pi(\theta, \lambda)d\lambda, \quad (2.7.1)$$

此处 $\theta \sim \pi_1(\theta|\lambda)$, $\lambda \sim \pi_2(\lambda)$, θ 和 λ 的联合密度为 $\pi(\theta, \lambda) = \pi_1(\theta|\lambda)\pi_2(\lambda)$, 故 $\pi(\theta)$ 作为联合密度 $\pi(\theta, \lambda)$ 的边缘密度就是 θ 的先验密度.

- 对于三层的先验, 这个规范的先验是

$$\begin{aligned} \pi(\theta) &= \int_{\Lambda} \int_{\Delta} \pi_1(\theta|\lambda)\pi_2(\lambda|\delta)\pi_3(\delta)d\lambda d\delta \\ &= \int_{\Lambda} \pi_1(\theta|\lambda) \left[\int_{\Delta} \pi(\lambda, \delta)d\delta \right] d\lambda. \end{aligned} \quad (2.7.2)$$

其中 $\pi(\lambda, \delta) = \pi_2(\lambda|\delta)\pi_3(\delta)$. 对于更多层的先验, 可用类似方法求得规范的先验 $\pi(\theta)$.

一个例子

- 由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后, 视 $\pi(\theta)$ 为 θ 的先验分布进行贝叶斯分析, 这就与通常先验分布下的贝叶斯分析无本质差别.
- 例 2.6.1 设对某产品的不合格率 θ 了解较少, 只知道 θ 较小. 现需确定 θ 的先验分布. 决策人经反复思考, 决定考虑用多层先验. 他的思路是这样的:
 - (1) 开始考虑用 $(0, 1)$ 上的均匀分布作为 θ 的先验分布.
 - (2) 后来他觉得不妥, 因为产品的不合格率 θ 较小, 不会超过 0.5, 于是他改用区间 $(0, 0.5)$ 上均匀分布作为先验分布.
 - (3) 有人对上限 0.5 提出意见, 问为什么不可以把上限改成 0.4 呢? 也有人建议上限用 0.1 他也无把握. 这些促使他考虑用分层先验.
- 最后提出如下方法是: (i) θ 的先验分布 $\pi_1(\theta|\lambda)$ 为 $U(0, \lambda)$, (ii) λ 的先验分布为 $\pi_2(\lambda)$ 为 $U(0.1, 0.5)$.

一个例子

- 由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后, 视 $\pi(\theta)$ 为 θ 的先验分布进行贝叶斯分析, 这就与通常先验分布下的贝叶斯分析无本质差别.
- **例 2.6.1** 设对某产品的不合格率 θ 了解较少, 只知道 θ 较小. 现需确定 θ 的先验分布. 决策人经反复思考, 决定考虑用多层先验. 他的思路是这样的:
 - (1) 开始考虑用 $(0, 1)$ 上的均匀分布作为 θ 的先验分布.
 - (2) 后来他觉得不妥, 因为产品的不合格率 θ 较小, 不会超过 0.5, 于是他改用区间 $(0, 0.5)$ 上均匀分布作为先验分布.
 - (3) 有人对上限 0.5 提出意见, 问为什么不可以把上限改成 0.4 呢? 也有人建议上限用 0.1 他也无把握. 这些促使他考虑用分层先验.
- 最后提出如下方法是: (i) θ 的先验分布 $\pi_1(\theta|\lambda)$ 为 $U(0, \lambda)$, (ii) λ 的先验分布为 $\pi_2(\lambda)$ 为 $U(0.1, 0.5)$.

一个例子

- 由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后, 视 $\pi(\theta)$ 为 θ 的先验分布进行贝叶斯分析, 这就与通常先验分布下的贝叶斯分析无本质差别.
- **例 2.6.1** 设对某产品的不合格率 θ 了解较少, 只知道 θ 较小. 现需确定 θ 的先验分布. 决策人经反复思考, 决定考虑用多层先验. 他的思路是这样的:
 - (1) 开始考虑用 $(0, 1)$ 上的均匀分布作为 θ 的先验分布.
 - (2) 后来他觉得不妥, 因为产品的不合格率 θ 较小, 不会超过 0.5, 于是他改用区间 $(0, 0.5)$ 上均匀分布作为先验分布.
 - (3) 有人对上限 0.5 提出意见, 问为什么不可以把上限改成 0.4 呢? 也有人建议上限用 0.1 他也无把握. 这些促使他考虑用分层先验.
- 最后提出如下方法是: (i) θ 的先验分布 $\pi_1(\theta|\lambda)$ 为 $U(0, \lambda)$, (ii) λ 的先验分布为 $\pi_2(\lambda)$ 为 $U(0.1, 0.5)$.

一个例子

- 由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后, 视 $\pi(\theta)$ 为 θ 的先验分布进行贝叶斯分析, 这就与通常先验分布下的贝叶斯分析无本质差别.
- **例 2.6.1** 设对某产品的不合格率 θ 了解较少, 只知道 θ 较小. 现需确定 θ 的先验分布. 决策人经反复思考, 决定考虑用多层先验. 他的思路是这样的:
 - (1) 开始考虑用 $(0, 1)$ 上的均匀分布作为 θ 的先验分布.
 - (2) 后来他觉得不妥, 因为产品的不合格率 θ 较小, 不会超过 0.5, 于是他改用区间 $(0, 0.5)$ 上均匀分布作为先验分布.
 - (3) 有人对上限 0.5 提出意见, 问为什么不可以把上限改成 0.4 呢? 也有人建议上限用 0.1 他也无把握. 这些促使他考虑用分层先验.
- 最后提出如下方法是: (i) θ 的先验分布 $\pi_1(\theta|\lambda)$ 为 $U(0, \lambda)$, (ii) λ 的先验分布为 $\pi_2(\lambda)$ 为 $U(0.1, 0.5)$.

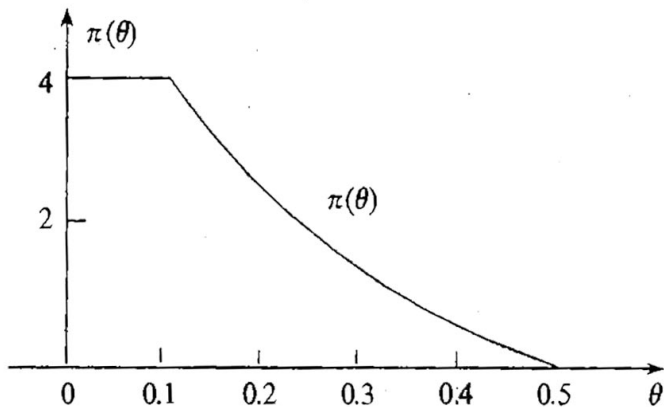
一个例子

- 由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后, 视 $\pi(\theta)$ 为 θ 的先验分布进行贝叶斯分析, 这就与通常先验分布下的贝叶斯分析无本质差别.
- 例 2.6.1** 设对某产品的不合格率 θ 了解较少, 只知道 θ 较小. 现需确定 θ 的先验分布. 决策人经反复思考, 决定考虑用多层先验. 他的思路是这样的:
 - (1) 开始考虑用 $(0, 1)$ 上的均匀分布作为 θ 的先验分布.
 - (2) 后来他觉得不妥, 因为产品的不合格率 θ 较小, 不会超过 0.5, 于是他改用区间 $(0, 0.5)$ 上均匀分布作为先验分布.
 - (3) 有人对上限 0.5 提出意见, 问为什么不可以把上限改成 0.4 呢? 也有人建议上限用 0.1 他也无把握. 这些促使他考虑用分层先验.
- 最后提出如下方法是: (i) θ 的先验分布 $\pi_1(\theta|\lambda)$ 为 $U(0, \lambda)$, (ii) λ 的先验分布为 $\pi_2(\lambda)$ 为 $U(0.1, 0.5)$.

一个例子

- 由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后, 视 $\pi(\theta)$ 为 θ 的先验分布进行贝叶斯分析, 这就与通常先验分布下的贝叶斯分析无本质差别.
- **例 2.6.1** 设对某产品的不合格率 θ 了解较少, 只知道 θ 较小. 现需确定 θ 的先验分布. 决策人经反复思考, 决定考虑用多层先验. 他的思路是这样的:
 - (1) 开始考虑用 $(0, 1)$ 上的均匀分布作为 θ 的先验分布.
 - (2) 后来他觉得不妥, 因为产品的不合格率 θ 较小, 不会超过 0.5, 于是他改用区间 $(0, 0.5)$ 上均匀分布作为先验分布.
 - (3) 有人对上限 0.5 提出意见, 问为什么不可以把上限改成 0.4 呢? 也有人建议上限用 0.1 他也无把握. 这些促使他考虑用分层先验.
- 最后提出如下方法是: (i) θ 的先验分布 $\pi_1(\theta|\lambda)$ 为 $U(0, \lambda)$, (ii) λ 的先验分布为 $\pi_2(\lambda)$ 为 $U(0.1, 0.5)$.

图 2.6.1

图 2.6.1 θ 的多层先验

确定多层先验的方法和步骤

- 首先对未知参数 θ 给出一个形式已知的密度函数作为先验分布, 即 $\theta \sim \pi_1(\theta|\lambda)$, 其中 λ 为超参数, $\lambda \in \Lambda$.
- 第二步是对超参数 λ 再给出一个先验分布 $\pi_2(\lambda)$.
- 对两层先验, 按公式 (2.7.1) 求得规范的先验 $\pi(\theta)$. 对于三层先验, 可按公式 (2.7.2) 求得规范的先验; 对更多层的先验, 可用类似方法求得规范的先验 $\pi(\theta)$.
- 在由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后, 任一个与多层先验有关的贝叶斯分析问题都可从 $\pi(\theta)$ 出发进行, 它与通常的贝叶斯分析无本质差别.

确定多层先验的方法和步骤

- 首先对未知参数 θ 给出一个形式已知的密度函数作为先验分布, 即 $\theta \sim \pi_1(\theta|\lambda)$, 其中 λ 为超参数, $\lambda \in \Lambda$.
- 第二步是对超参数 λ 再给出一个先验分布 $\pi_2(\lambda)$.
- 对两层先验, 按公式 (2.7.1) 求得规范的先验 $\pi(\theta)$. 对于三层先验, 可按公式 (2.7.2) 求得规范的先验; 对更多层的先验, 可用类似方法求得规范的先验 $\pi(\theta)$.
- 在由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后, 任一个与多层先验有关的贝叶斯分析问题都可从 $\pi(\theta)$ 出发进行, 它与通常的贝叶斯分析无本质差别.

确定多层先验的方法和步骤

- 首先对未知参数 θ 给出一个形式已知的密度函数作为先验分布, 即 $\theta \sim \pi_1(\theta|\lambda)$, 其中 λ 为超参数, $\lambda \in \Lambda$.
- 第二步是对超参数 λ 再给出一个先验分布 $\pi_2(\lambda)$.
- 对两层先验, 按公式 (2.7.1) 求得规范的先验 $\pi(\theta)$. 对于三层先验, 可按公式 (2.7.2) 求得规范的先验; 对更多层的先验, 可用类似方法求得规范的先验 $\pi(\theta)$.
- 在由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后, 任一个与多层先验有关的贝叶斯分析问题都可从 $\pi(\theta)$ 出发进行, 它与通常的贝叶斯分析无本质差别.

确定多层先验的方法和步骤

- 首先对未知参数 θ 给出一个形式已知的密度函数作为先验分布，即 $\theta \sim \pi_1(\theta|\lambda)$ ，其中 λ 为超参数， $\lambda \in \Lambda$.
- 第二步是对超参数 λ 再给出一个先验分布 $\pi_2(\lambda)$.
- 对两层先验，按公式 (2.7.1) 求得规范的先验 $\pi(\theta)$. 对于三层先验，可按公式 (2.7.2) 求得规范的先验；对更多层的先验，可用类似方法求得规范的先验 $\pi(\theta)$.
- 在由多层先验按上述方法获得规范的先验 $\pi(\theta)$ 后，任一个与多层先验有关的贝叶斯分析问题都可从 $\pi(\theta)$ 出发进行，它与通常的贝叶斯分析无本质差别.

注释

- 既然可以把一个分层贝叶斯模型转化为规范形式的先验 (即一个单层先验) 模型, 为什么我们还要研究多层贝叶斯模型呢?
- 因为多层贝叶斯模型允许我们在建模时, 可以把相对复杂的情况分解为一系列简单的情形. 以两层先验为例, $\pi_1(\theta|\lambda)$ 和 $\pi_2(\lambda)$ 都可以是简单的函数形式 (共轭先验或无信息先验等), 但式 (2.7.1) 表示的 $\pi(\theta)$ 可能非常的复杂.
- 分层先验模型的另一个重要特点, 是便于计算. 以两层先验为例, 如果用通常的贝叶斯模型计算 θ 的后验分布 $\pi(\theta|\mathbf{x})$ 和它的某些数字特征, 由于 θ 的联合后验密度没有显式表达而通过积分表示, 计算非常困难. 为了克服计算上的难题, 首先需要建立容易使用的后验分布的表达式. 即由多层结构各阶段的后验来表达 θ 的联合后验, 这些各阶段的后验大多有显式表达, 且容易计算.

注释

- 既然可以把一个分层贝叶斯模型转化为规范形式的先验 (即一个单层先验) 模型, 为什么我们还要研究多层贝叶斯模型呢?
- 因为多层贝叶斯模型允许我们在建模时, 可以把相对复杂的情况分解为一系列简单的情形. 以两层先验为例, $\pi_1(\theta|\lambda)$ 和 $\pi_2(\lambda)$ 都可以是简单的函数形式 (共轭先验或无信息先验等), 但式 (2.7.1) 表示的 $\pi(\theta)$ 可能非常的复杂.
- 分层先验模型的另一个重要特点, 是便于计算. 以两层先验为例, 如果用通常的贝叶斯模型计算 θ 的后验分布 $\pi(\theta|\mathbf{x})$ 和它的某些数字特征, 由于 θ 的联合后验密度没有显式表达而通过积分表示, 计算非常困难. 为了克服计算上的难题, 首先需要建立容易使用的后验分布的表达式. 即由多层结构各阶段的后验来表达 θ 的联合后验, 这些各阶段的后验大多有显式表达, 且容易计算.

注释

- 既然可以把一个分层贝叶斯模型转化为规范形式的先验 (即一个单层先验) 模型, 为什么我们还要研究多层贝叶斯模型呢?
- 因为多层贝叶斯模型允许我们在建模时, 可以把相对复杂的情况分解为一系列简单的情形. 以两层先验为例, $\pi_1(\theta|\lambda)$ 和 $\pi_2(\lambda)$ 都可以是简单的函数形式 (共轭先验或无信息先验等), 但式 (2.7.1) 表示的 $\pi(\theta)$ 可能非常的复杂.
- 分层先验模型的另一个重要特点, 是便于计算. 以两层先验为例, 如果用通常的贝叶斯模型计算 θ 的后验分布 $\pi(\theta|\mathbf{x})$ 和它的某些数字特征, 由于 θ 的联合后验密度没有显式表达而通过积分表示, 计算非常困难. 为了克服计算上的难题, 首先需要建立容易使用的后验分布的表达式. 即由多层结构各阶段的后验来表达 θ 的联合后验, 这些各阶段的后验大多有显式表达, 且容易计算.