

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

第一章：绪论

1

1.1 引言

- 1.1.1 从贝叶斯公式说起
- 1.1.2 三种信息
- 1.1.3 历史
- 1.1.4 古典学派和贝叶斯学派的论争

2

1.2 贝叶斯统计推断的若干基本概念

- 1.2.1 先验分布与后验分布
- 1.2.2 点估计问题
- 1.2.3 假设检验问题
- 1.2.4 区间估计问题及例

3

1.3 贝叶斯统计决策的若干基本概念

- 1.3.1 统计判决三要素
- 1.3.2 风险函数和一致最优决策函数
- 1.3.3 贝叶斯风险和贝叶斯解

1 1.1 引言

- 1.1.1 从贝叶斯公式说起
- 1.1.2 三种信息
- 1.1.3 历史
- 1.1.4 古典学派和贝叶斯学派的论争

2 1.2 贝叶斯统计推断的若干基本概念

- 1.2.1 先验分布与后验分布
- 1.2.2 点估计问题
- 1.2.3 假设检验问题
- 1.2.4 区间估计问题及例

3 1.3 贝叶斯统计决策的若干基本概念

- 1.3.1 统计判决三要素
- 1.3.2 风险函数和一致最优决策函数
- 1.3.3 贝叶斯风险和贝叶斯解

全概率公式、贝叶斯公式

- 设 B_1, B_2, \dots, B_n (n 为有限或无穷) 是样本空间 Ω 中的一个完备事件群 (又称为 Ω 的一个分划), 令 A 为 Ω 中的一个事件, 则全概率公式为:

$$P(A) = P\left(\sum_{i=1}^n AB_i\right) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

- 这个公式将整个事件 A 分解成一些两两不相交的事件之并, 直接计算 $P(A)$ 不容易, 但分解后的那些事件的概率容易计算, 从而使 $P(A)$ 的计算变得容易了.
- 在全概率公式条件下, 设 A 为 Ω 中的一个事件, $P(B_i) > 0$, $i = 1, 2, \dots, n$, $P(A) > 0$, 则按条件概率计算方法有

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}.$$

这个公式称为贝叶斯 (Bayes) 公式.

全概率公式、贝叶斯公式

- 设 B_1, B_2, \dots, B_n (n 为有限或无穷) 是样本空间 Ω 中的一个完备事件群 (又称为 Ω 的一个分划), 令 A 为 Ω 中的一个事件, 则全概率公式为:

$$P(A) = P\left(\sum_{i=1}^n AB_i\right) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

- 这个公式将整个事件 A 分解成一些两两不相交的事件之并, 直接计算 $P(A)$ 不容易, 但分解后的那些事件的概率容易计算, 从而使 $P(A)$ 的计算变得容易了.
- 在全概率公式条件下, 设 A 为 Ω 中的一个事件, $P(B_i) > 0$, $i = 1, 2, \dots, n$, $P(A) > 0$, 则按条件概率计算方法有

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}.$$

这个公式称为贝叶斯 (Bayes) 公式.

全概率公式、贝叶斯公式

- 设 B_1, B_2, \dots, B_n (n 为有限或无穷) 是样本空间 Ω 中的一个完备事件群 (又称为 Ω 的一个分划), 令 A 为 Ω 中的一个事件, 则全概率公式为:

$$P(A) = P\left(\sum_{i=1}^n AB_i\right) = \sum_{i=1}^n P(A|B_i)P(B_i).$$

- 这个公式将整个事件 A 分解成一些两两不相交的事件之并, 直接计算 $P(A)$ 不容易, 但分解后的那些事件的概率容易计算, 从而使 $P(A)$ 的计算变得容易了.
- 在全概率公式条件下, 设 A 为 Ω 中的一个事件, $P(B_i) > 0$, $i = 1, 2, \dots, n$, $P(A) > 0$, 则按条件概率计算方法有

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}.$$

这个公式称为贝叶斯 (Bayes) 公式.

贝叶斯公式的意义

- 从形式上看这个公式不过是条件概率定义与全概率公式的简单推论. 之所以著名, 是在这个公式的哲理意义上: 在没有进一步信息 (不知道 A 发生) 下, 人们对 B_1, B_2, \dots, B_n 发生可能性大小的认识通过 $P(B_1), P(B_2), \dots, P(B_n)$ 来体现; 在有了新信息 (知道 A 发生) 后, 人们对事件 B_1, B_2, \dots, B_n 发生可能性大小新认识体现在 $P(B_1|A), P(B_2|A), \dots, P(B_n|A)$.
- 如果把事件 A 看成“结果”, 把诸事件 B_1, B_2, \dots, B_n 看成导致这一结果的可能“原因”, 则可以形象地把全概率公式看成由“原因”推“结果”. 而贝叶斯公式正好相反, 其作用在于由“结果”推“原因”. 现在有了结果 A , 在导致 A 发生的众多原因中, 到底是哪个原因导致了 A 发生 (或曰: 到底是哪个原因导致 A 发生的可能性最大)? 这是日常生活和科学技术研究中常见到的问题. 请看下例:

贝叶斯公式的意义

- 从形式上看这个公式不过是条件概率定义与全概率公式的简单推论. 之所以著名, 是在这个公式的哲理意义上: 在没有进一步信息 (不知道 A 发生) 下, 人们对 B_1, B_2, \dots, B_n 发生可能性大小的认识通过 $P(B_1), P(B_2), \dots, P(B_n)$ 来体现; 在有了新信息 (知道 A 发生) 后, 人们对事件 B_1, B_2, \dots, B_n 发生可能性大小新认识体现在 $P(B_1|A), P(B_2|A), \dots, P(B_n|A)$.
- 如果把事件 A 看成“结果”, 把诸事件 B_1, B_2, \dots, B_n 看成导致这一结果的可能“原因”, 则可以形象地把全概率公式看成由“原因”推“结果”. 而贝叶斯公式正好相反, 其作用在于由“结果”推“原因”. 现在有了结果 A , 在导致 A 发生的众多原因中, 到底是哪个原因导致了 A 发生 (或曰: 到底是哪个原因导致 A 发生的可能性最大)? 这是日常生活和科学技术研究中常见到的问题. 请看下例:

一个例子

- **例 1.1.1** 一种诊断某癌症的试剂, 经临床试验有如下记录: 癌症病人试验结果是阳性的概率为 95%, 非癌症病人试验结果是阴性的概率为 95%. 现用这种试剂在某社区进行癌症普查, 设该社区癌症发病率为 0.5%, 问某人反应为阳性时该如何判断他是否患有癌症.

- 设 A 表示“反应为阳性”的事件, B 表示“被诊断者患癌症”的事件, 则 $B_1 = B$ 和 $B_2 = \bar{B}$ 构成完备事件群. 由题意

$$P(A|B_1) = 0.95, \quad P(A|B_2) = 1 - 0.95 = 0.05,$$

$$P(B_1) = 0.005, \quad P(B_2) = 0.995.$$

- 现在要算的是 $P(B_1|A)$ 和 $P(B_2|A)$. 由贝叶斯公式易得

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = 0.087 = 8.7\%$$

$$P(B_2|A) = 1 - 0.087 = 0.913 = 91.3\%.$$

一个例子

- **例 1.1.1** 一种诊断某癌症的试剂, 经临床试验有如下记录: 癌症病人试验结果是阳性的概率为 95%, 非癌症病人试验结果是阴性的概率为 95%. 现用这种试剂在某社区进行癌症普查, 设该社区癌症发病率为 0.5%, 问某人反应为阳性时该如何判断他是否患有癌症.

- 设 A 表示“反应为阳性”的事件, B 表示“被诊断者患癌症”的事件, 则 $B_1 = B$ 和 $B_2 = \bar{B}$ 构成完备事件群. 由题意

$$P(A|B_1) = 0.95, \quad P(A|B_2) = 1 - 0.95 = 0.05,$$

$$P(B_1) = 0.005, \quad P(B_2) = 0.995.$$

- 现在要算的是 $P(B_1|A)$ 和 $P(B_2|A)$. 由贝叶斯公式易得

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = 0.087 = 8.7\%$$

$$P(B_2|A) = 1 - 0.087 = 0.913 = 91.3\%.$$

一个例子

- **例 1.1.1** 一种诊断某癌症的试剂, 经临床试验有如下记录: 癌症病人试验结果是阳性的概率为 95%, 非癌症病人试验结果是阴性的概率为 95%. 现用这种试剂在某社区进行癌症普查, 设该社区癌症发病率为 0.5%, 问某人反应为阳性时该如何判断他是否患有癌症.

- 设 A 表示“反应为阳性”的事件, B 表示“被诊断者患癌症”的事件, 则 $B_1 = B$ 和 $B_2 = \bar{B}$ 构成完备事件群. 由题意

$$P(A|B_1) = 0.95, \quad P(A|B_2) = 1 - 0.95 = 0.05,$$

$$P(B_1) = 0.005, \quad P(B_2) = 0.995.$$

- 现在要算的是 $P(B_1|A)$ 和 $P(B_2|A)$. 由贝叶斯公式易得

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} = 0.087 = 8.7\%$$

$$P(B_2|A) = 1 - 0.087 = 0.913 = 91.3\%.$$

贝叶斯方法

- 导致某人试验结果为阳性，可能原因有两个，一个是他患有癌症使试验结果呈阳性；另一个原因是他根本就没有患有癌症，但也可以使试验结果呈阳性。某人真正患癌症的可能性很小，只有 8.7%，告诉他不必紧张，可以到医院去做进一步的检查，以便排除这一疑点。
- 贝叶斯方法的基本观点是由贝叶斯公式引伸而来的。从形式上看这一公式不过是条件概率定义的一个简单推论，但它却包含了归纳推理的一种思想，后来学者把它发展成一种关于统计推断的系统理论和方法，称为贝叶斯方法。
- 由这种方法获得的统计推断的全部结果，构成了贝叶斯统计学的内容。信奉贝叶斯统计，乃至鼓吹贝叶斯观点是关于统计推断的唯一正确方法的那些学者形成了数理统计学中的贝叶斯学派。经典（频率）学派与贝叶斯学派近几十年来的争论推动了数理统计学的发展。

贝叶斯方法

- 导致某人试验结果为阳性，可能原因有两个，一个是他患有癌症使试验结果呈阳性；另一个原因是他根本就没有患有癌症，但也可以使试验结果呈阳性。某人真正患癌症的可能性很小，只有 8.7%，告诉他不必紧张，可以到医院去做进一步的检查，以便排除这一疑点。
- 贝叶斯方法的基本观点是由贝叶斯公式引伸而来的。从形式上看这一公式不过是条件概率定义的一个简单推论，但它却包含了归纳推理的一种思想，后来学者把它发展成一种关于统计推断的系统理论和方法，称为贝叶斯方法。
- 由这种方法获得的统计推断的全部结果，构成了贝叶斯统计学的内容。信奉贝叶斯统计，乃至鼓吹贝叶斯观点是关于统计推断的唯一正确方法的那些学者形成了数理统计学中的贝叶斯学派。经典（频率）学派与贝叶斯学派近几十年来的争论推动了数理统计学的发展。

贝叶斯方法

- 导致某人试验结果为阳性，可能原因有两个，一个是他患有癌症使试验结果呈阳性；另一个原因是他根本就没有患有癌症，但也可以使试验结果呈阳性。某人真正患癌症的可能性很小，只有 8.7%，告诉他不必紧张，可以到医院去做进一步的检查，以便排除这一疑点。
- 贝叶斯方法的基本观点是由贝叶斯公式引伸而来的。从形式上看这一公式不过是条件概率定义的一个简单推论，但它却包含了归纳推理的一种思想，后来学者把它发展成一种关于统计推断的系统理论和方法，称为**贝叶斯方法**。
- 由这种方法获得的统计推断的全部结果，构成了**贝叶斯统计学**的内容。信奉贝叶斯统计，乃至鼓吹贝叶斯观点是关于统计推断的唯一正确方法的那些学者形成了数理统计学中的**贝叶斯学派**。经典 (频率) 学派与贝叶斯学派近几十年来的争论推动了数理统计学的发展。

三种信息和贝叶斯统计学

- 当把样本视为随机变量，它有概率分布，称为总体分布. 如果我们已经知道总体的分布形式，这就给了我们一种信息，称为总体信息.
- 另外一种信息是样本本身的信息，就是从总体中抽取的样本所提供的信息. 这是最“鲜活”的信息. 样本越多，提供的信息越多. 总体信息和样本信息放在一起,称为抽样信息.
- 基于总体信息和样本信息 (或统称为抽样信息) 进行统计推断的理论和方法称为经典（古典）统计学.
- 另外一种信息称为先验信息，就是在抽样之前，有关问题中未知参数的信息. 一般先验信息来自经验和历史资料.
- 基于上述三种信息进行统计推断的方法和理论称为贝叶斯统计学. 它与经典统计学的主要区别在于是否利用先验信息. 在使用样本上也是存在差别的. 贝叶斯方法重视已出现的样本, 对尚未发生的样本值不予考虑.

三种信息和贝叶斯统计学

- 当把样本视为随机变量，它有概率分布，称为总体分布. 如果我们已经知道总体的分布形式，这就给了我们一种信息，称为总体信息.
- 另外一种信息是样本本身的信息，就是从总体中抽取的样本所提供的信息. 这是最“鲜活”的信息. 样本越多，提供的信息越多. 总体信息和样本信息放在一起,称为抽样信息.
- 基于总体信息和样本信息 (或统称为抽样信息) 进行统计推断的理论和方法称为经典 (古典) 统计学.
- 另外一种信息称为先验信息，就是在抽样之前，有关问题中未知参数的信息. 一般先验信息来自经验和历史资料.
- 基于上述三种信息进行统计推断的方法和理论称为贝叶斯统计学. 它与经典统计学的主要区别在于是否利用先验信息. 在使用样本上也是存在差别的. 贝叶斯方法重视已出现的样本, 对尚未发生的样本值不予考虑.

三种信息和贝叶斯统计学

- 当把样本视为随机变量，它有概率分布，称为总体分布. 如果我们已经知道总体的分布形式，这就给了我们一种信息，称为总体信息.
- 另外一种信息是样本本身的信息，就是从总体中抽取的样本所提供的信息. 这是最“鲜活”的信息. 样本越多，提供的信息越多. 总体信息和样本信息放在一起,称为抽样信息.
- 基于总体信息和样本信息 (或统称为抽样信息) 进行统计推断的理论和方法称为**经典（古典）统计学**.
- 另外一种信息称为先验信息，就是在抽样之前，有关问题中未知参数的信息. 一般先验信息来自经验和历史资料.
- 基于上述三种信息进行统计推断的方法和理论称为**贝叶斯统计学**. 它与经典统计学的主要区别在于是否利用先验信息. 在使用样本上也是存在差别的. 贝叶斯方法重视已出现的样本, 对尚未发生的样本值不予考虑.

三种信息和贝叶斯统计学

- 当把样本视为随机变量，它有概率分布，称为总体分布. 如果我们已经知道总体的分布形式，这就给了我们一种信息，称为总体信息.
- 另外一种信息是样本本身的信息，就是从总体中抽取的样本所提供的信息. 这是最“鲜活”的信息. 样本越多，提供的信息越多. 总体信息和样本信息放在一起,称为抽样信息.
- 基于总体信息和样本信息 (或统称为抽样信息) 进行统计推断的理论和方法称为**经典（古典）统计学**.
- 另外一种信息称为先验信息，就是在抽样之前，有关问题中未知参数的信息. 一般先验信息来自经验和历史资料.
- 基于上述三种信息进行统计推断的方法和理论称为**贝叶斯统计学**. 它与经典统计学的主要区别在于是否利用先验信息. 在使用样本上也是存在差别的. 贝叶斯方法重视已出现的样本, 对尚未发生的样本值不予考虑.

三种信息和贝叶斯统计学

- 当把样本视为随机变量，它有概率分布，称为总体分布. 如果我们已经知道总体的分布形式，这就给了我们一种信息，称为总体信息.
- 另外一种信息是样本本身的信息，就是从总体中抽取的样本所提供的信息. 这是最“鲜活”的信息. 样本越多，提供的信息越多. 总体信息和样本信息放在一起,称为抽样信息.
- 基于总体信息和样本信息 (或统称为抽样信息) 进行统计推断的理论和方法称为**经典（古典）统计学**.
- 另外一种信息称为先验信息，就是在抽样之前，有关问题中未知参数的信息. 一般先验信息来自经验和历史资料.
- 基于上述三种信息进行统计推断的方法和理论称为**贝叶斯统计学**. 它与经典统计学的主要区别在于是否利用先验信息. 在使用样本上也是存在差别的. 贝叶斯方法重视已出现的样本, 对尚未发生的样本值不予考虑.

先验信息存在的例子

- **例 1.1.2** 英国统计学家 L. J. Savage (1961) 提出一个令人信服的例子说明先验信息有时是很重要的，且看下面两试验：
(1) 一位常饮牛奶和茶的女士说她能辨别先倒进杯子里的是茶还是牛奶. 对此做了 10 次试验她多说对了.
(2) 一位音乐家说他能够从一页乐谱辨别是海顿 (Haydn) 还是莫扎特 (Mozart) 的作品，在 10 次试验中他都说对了.
- 在上面两次试验中，如果认为试验者是猜对的，每次成功概率为 0.5，则 10 次都猜中的概率为 $(1/2)^{10} = 0.0009766$ ，这是一个很小的概率，几乎不可能发生.
- **例 1.1.3** 某工厂生产一种产品，每日抽查一部分产品以检查废品率 θ ，经过一段时间后获得大量数据，对 θ 做出估计. 就当日被抽查的那批产品的废品率 θ 而言，它只是一个固定的数，并无随机性可言；但逐日的废品率 θ 受随机因素的影响会有些波动. 从长期看，将“一日废品率 θ ”视为随机变量，而要估计的某日的废品率是这个随机变量的一个观察值.

先验信息存在的例子

- **例 1.1.2** 英国统计学家 L. J. Savage (1961) 提出一个令人信服的例子说明先验信息有时是很重要的，且看下面两试验：
(1) 一位常饮牛奶和茶的女士说她能辨别先倒进杯子里的是茶还是牛奶. 对此做了 10 次试验她多说对了.
(2) 一位音乐家说他能够从一页乐谱辨别是海顿 (Haydn) 还是莫扎特 (Mozart) 的作品，在 10 次试验中他都说对了.
- 在上面两次试验中，如果认为试验者是猜对的，每次成功概率为 0.5，则 10 次都猜中的概率为 $(1/2)^{10} = 0.0009766$ ，这是一个很小的概率，几乎不可能发生.
- **例 1.1.3** 某工厂生产一种产品，每日抽查一部分产品以检查废品率 θ ，经过一段时间后获得大量数据，对 θ 做出估计. 就当日被抽查的那批产品的废品率 θ 而言，它只是一个固定的数，并无随机性可言；但逐日的废品率 θ 受随机因素的影响会有些波动. 从长期看，将“一日废品率 θ ”视为随机变量，而要估计的某日的废品率是这个随机变量的一个观察值.

先验信息存在的例子

- **例 1.1.2** 英国统计学家 L. J. Savage (1961) 提出一个令人信服的例子说明先验信息有时是很重要的，且看下面两试验：
(1) 一位常饮牛奶和茶的女士说她能辨别先倒进杯子里的是茶还是牛奶. 对此做了 10 次试验她多说对了.
(2) 一位音乐家说他能够从一页乐谱辨别是海顿 (Haydn) 还是莫扎特 (Mozart) 的作品，在 10 次试验中他都说对了.
- 在上面两次试验中，如果认为试验者是猜对的，每次成功概率为 0.5，则 10 次都猜中的概率为 $(1/2)^{10} = 0.0009766$ ，这是一个很小的概率，几乎不可能发生.
- **例 1.1.3** 某工厂生产一种产品，每日抽查一部分产品以检查废品率 θ ，经过一段时间后获得大量数据，对 θ 做出估计. 就当日被抽查的那批产品的废品率 θ 而言，它只是一个固定的数，并无随机性可言；但逐日的废品率 θ 受随机因素的影响会有些波动. 从长期看，将“一日废品率 θ ”视为随机变量，而要估计的某日的废品率是这个随机变量的一个观察值.

经验贝叶斯方法

- 根据历史资料可构造一个分布

$$P\left(\theta = \frac{i}{n}\right) = \pi_i, \quad i = 0, 1, 2, \dots, n; \quad \sum_{i=1}^n \pi_i = 1.$$

这个对先验信息进行整理加工而得到的分布称为先验分布。该分布总结了工厂过去产品质量情况。

- 贝叶斯方法的一个主要问题是如何确定先验分布, 先验分布的确定有很大的主观随意性. 当先验分布完全未知或部分未知时, 如果人为地给出的先验分布与实际情形偏离较大时, 贝叶斯解的性质就较差. 针对这一问题, Robbins (1955) 首先提出了经验贝叶斯 (简称 EB) 方法. 它的实质是利用历史样本对先验分布或先验分布的某些数字特征作出直接或间接的估计, 因此 EB 方法是对贝叶斯方法的改进和推广. 它是介于经典统计学和贝叶斯统计学之间的一种统计推断方法.

经验贝叶斯方法

- 根据历史资料可构造一个分布

$$P\left(\theta = \frac{i}{n}\right) = \pi_i, \quad i = 0, 1, 2, \dots, n; \quad \sum_{i=1}^n \pi_i = 1.$$

这个对先验信息进行整理加工而得到的分布称为先验分布. 该分布总结了工厂过去产品质量情况.

- 贝叶斯方法的一个主要问题是如何确定先验分布, 先验分布的确定有很大的主观随意性. 当先验分布完全未知或部分未知时, 如果人为地给出的先验分布与实际情形偏离较大时, 贝叶斯解的性质就较差. 针对这一问题, **Robbins (1955)** 首先提出了经验贝叶斯 (简称 **EB**) 方法. 它的实质是利用历史样本对先验分布或先验分布的某些数字特征作出直接或间接的估计, 因此 **EB** 方法是对贝叶斯方法的改进和推广. 它是介于经典统计学和贝叶斯统计学之间的一种统计推断方法.

发展史

- 贝叶斯统计起源于英国学者 Thomas 贝叶斯 (1701-1761) 在他去世两年后 (1763 年) 发表的一篇论文. 在此论文中他提出著名的 贝叶斯公式和一种归纳推理的方法.
- 20 世纪上半叶正是经典统计学大发展的一个时期, 发现了一些具有普遍应用意义的有力的统计方法, 如创建假设检验理论 (Neyman-Pearson 引理)、似然比检验、拟合优度检验、列联表检验和估计的最优性理论等. 自 20 世纪中叶以来, 经典统计学的发展遇到一些问题, 如数学化程度越来越高, 小样本方法缺乏进展从而越来越转向大样本理论的研究, 在应用工作中产生了不满. 在这种背景下贝叶斯统计以其操作简单, 加之在解释上的某些合理性吸引了不少应用统计学者.
- 贝叶斯学派自 20 世纪下半叶进入全盛时期, 在这中间起过重要作用的有 H. Jeffreys, L.J. Savage, 还有 D.V. Lindley、G.P. Box & G.C. Tiao 和 J.O. Berger 等.

发展史

- 贝叶斯统计起源于英国学者 Thomas 贝叶斯 (1701-1761) 在他去世两年后 (1763 年) 发表的一篇论文. 在此论文中他提出著名的 贝叶斯公式和一种归纳推理的方法.
- 20 世纪上半叶正是经典统计学大发展的一个时期, 发现了一些具有普遍应用意义的有力的统计方法, 如创建假设检验理论 (Neyman-Pearson 引理)、似然比检验、拟合优度检验、列联表检验和估计的最优性理论等. 自 20 世纪中叶以来, 经典统计学的发展遇到一些问题, 如数学化程度越来越高, 小样本方法缺乏进展从而越来越转向大样本理论的研究, 在应用工作中产生了不满. 在这种背景下贝叶斯统计以其操作简单, 加之在解释上的某些合理性吸引了不少应用统计学者.
- 贝叶斯学派自 20 世纪下半叶进入全盛时期, 在这中间起过重要作用的有 H. Jeffreys, L.J. Savage, 还有 D.V. Lindley、G.P. Box & G.C. Tiao 和 J.O. Berger 等.

发展史

- 贝叶斯统计起源于英国学者 Thomas 贝叶斯 (1701-1761) 在他去世两年后 (1763 年) 发表的一篇论文. 在此论文中他提出著名的 贝叶斯公式和一种归纳推理的方法.
- 20 世纪上半叶正是经典统计学大发展的一个时期, 发现了一些具有普遍应用意义的有力的统计方法, 如创建假设检验理论 (Neyman-Pearson 引理)、似然比检验、拟合优度检验、列联表检验和估计的最优性理论等. 自 20 世纪中叶以来, 经典统计学的发展遇到一些问题, 如数学化程度越来越高, 小样本方法缺乏进展从而越来越转向大样本理论的研究, 在应用工作中产生了不满. 在这种背景下贝叶斯统计以其操作简单, 加之在解释上的某些合理性吸引了不少应用统计学者.
- 贝叶斯学派自 20 世纪下半叶进入全盛时期, 在这中间起过重要作用的有 H. Jeffreys, L.J. Savage, 还有 D.V. Lindley、G.P. Box & G.C. Tiao 和 J.O. Berger 等.

1. 频率学派对贝叶斯学派的批评

- 凡是坚持概率的频率解释, 因而对数理统计学中的概念、结果和方法性能的评价等, 必须在大量重复的意义上理解的, 都属于频率学派. 频率学派对贝叶斯学派的批评, 主要集中在主观概率以及相关的先验分布的确定上.
- 贝叶斯学派常使用主观概率. 主观概率定义为认识主体对事情发生机会的个人信念, 即不同的人对同一事件的概率可以得到不同的结果. 坚持频率解释的人认为这不仅难以捉摸, 是主观随意性的产物, 没有客观性, 因而也就没有科学性, 当然也不能接受建立在这个基础上的统计方法.
- 频率学派还认为在许多情况下将 θ 视为随机变量是不合理的. 例如, 估计某矿体内一种金属的含量, 很难把这一问题纳入贝叶斯观点的模式中. 此外, 就算在某些问题中将 θ 看作随机变量有一定的合理性, 但关于 θ 的先验知识, 往往不是确切到可以提出一定的先验分布来. 在这种情况下, 指定一种先验分布不免带有人为性, 缺乏合理性.

1. 频率学派对贝叶斯学派的批评

- 凡是坚持概率的频率解释, 因而对数理统计学中的概念、结果和方法性能的评价等, 必须在大量重复的意义上理解的, 都属于频率学派. 频率学派对贝叶斯学派的批评, 主要集中在主观概率以及相关的先验分布的确定上.
- 贝叶斯学派常使用主观概率. 主观概率定义为认识主体对事情发生机会的个人信念, 即不同的人对同一事件的概率可以得到不同的结果. 坚持频率解释的人认为这不仅难以捉摸, 是主观随意性的产物, 没有客观性, 因而也就没有科学性, 当然也不能接受建立在这个基础上的统计方法.
- 频率学派还认为在许多情况下将 θ 视为随机变量是不合理的. 例如, 估计某矿体内一种金属的含量, 很难把这一问题纳入贝叶斯观点的模式中. 此外, 就算在某些问题中将 θ 看作随机变量有一定的合理性, 但关于 θ 的先验知识, 往往不是确切到可以提出一定的先验分布来. 在这种情况下, 指定一种先验分布不免带有人为性, 缺乏合理性.

1. 频率学派对贝叶斯学派的批评

- 凡是坚持概率的频率解释, 因而对数理统计学中的概念、结果和方法性能的评价等, 必须在大量重复的意义上理解的, 都属于频率学派. 频率学派对贝叶斯学派的批评, 主要集中在主观概率以及相关的先验分布的确定上.
- 贝叶斯学派常使用主观概率. 主观概率定义为认识主体对事情发生机会的个人信念, 即不同的人对同一事件的概率可以得到不同的结果. 坚持频率解释的人认为这不仅难以捉摸, 是主观随意性的产物, 没有客观性, 因而也就没有科学性, 当然也不能接受建立在这个基础上的统计方法.
- 频率学派还认为在许多情况下将 θ 视为随机变量是不合理的. 例如, 估计某矿体内一种金属的含量, 很难把这一问题纳入贝叶斯观点的模式中. 此外, 就算在某些问题中将 θ 看作随机变量有一定的合理性, 但关于 θ 的先验知识, 往往不是确切到可以提出一定的先验分布来. 在这种情况下, 指定一种先验分布不免带有人为性, 缺乏合理性.

2. 贝叶斯学派对频率学派的批评

- 贝叶斯学派对频率学派的批评，主要在以下方面：首先涉及“频率解释”本身。贝叶斯学派认为许多应用问题是一次性的，在诸多自然现象、社会现象和经济决策问题中，事件常常是不能大量重复的。例如灾害预报中，像地震、水灾等都不可在相同条件下重复。
- 贝叶斯学派认为，虽然频率学派没有明白使用先验分布，但事实上，经典统计中的一些重要统计推断方法之所以能站住脚，只是它暗合于某个合理的贝叶斯解。如 $N(\theta, 1)$ 中的 θ 的无偏估计 \bar{X} ，恰好是当 θ 有广义先验密度 $\pi(\theta) \equiv 1$ ，即所谓无信息先验分布之下的贝叶斯解。
- 频率学派承认贝叶斯方法在一些情况下可用，但限于先验分布可给予某种频率解释的时候。至于两派涉及的基本哲学观点，看来是无法调和的。

2. 贝叶斯学派对频率学派的批评

- 贝叶斯学派对频率学派的批评，主要在以下方面：首先涉及“频率解释”本身。贝叶斯学派认为许多应用问题是一次性的，在诸多自然现象、社会现象和经济决策问题中，事件常常是不能大量重复的。例如灾害预报中，像地震、水灾等都不可在相同条件下重复。
- 贝叶斯学派认为，虽然频率学派没有明白使用先验分布，但事实上，经典统计中的一些重要统计推断方法之所以能站住脚，只是它暗合于某个合理的贝叶斯解。如 $N(\theta, 1)$ 中的 θ 的无偏估计 \bar{X} ，恰好是当 θ 有广义先验密度 $\pi(\theta) \equiv 1$ ，即所谓无信息先验分布之下的贝叶斯解。
- 频率学派承认贝叶斯方法在一些情况下可用，但限于先验分布可给予某种频率解释的时候。至于两派涉及的基本哲学观点，看来是无法调和的。

2. 贝叶斯学派对频率学派的批评

- 贝叶斯学派对频率学派的批评，主要在以下方面：首先涉及“频率解释”本身。贝叶斯学派认为许多应用问题是一次性的，在诸多自然现象、社会现象和经济决策问题中，事件常常是不能大量重复的。例如灾害预报中，像地震、水灾等都不可在相同条件下重复。
- 贝叶斯学派认为，虽然频率学派没有明白使用先验分布，但事实上，经典统计中的一些重要统计推断方法之所以能站住脚，只是它暗合于某个合理的贝叶斯解。如 $N(\theta, 1)$ 中的 θ 的无偏估计 \bar{X} ，恰好是当 θ 有广义先验密度 $\pi(\theta) \equiv 1$ ，即所谓无信息先验分布之下的贝叶斯解。
- 频率学派承认贝叶斯方法在一些情况下可用，但限于先验分布可给予某种频率解释的时候。至于两派涉及的基本哲学观点，看来是无法调和的。

评论

- 频率学派与贝叶斯学派有不少共同点, 如都承认样本有概率分布, 概率的计算遵循共同的准则. 分歧在于把未知参数 θ 看成一个未知的固定量呢, 还是看成一个随机变量, 其余的分歧多少都是由此派生而来.
- 对上述争论有一个至高无上的裁判者, 即应用的结果. 统计方法无论在理论上如何精细高明, 总要用实践来检验. 迄今为止, 实践显示这两派得分都不低, 也正是因为它们在上表现不错, 才能各自聚合了一些追随者形成各自的学派.
- 作为一个统计学者, 可以不执著于任何一派的观点, 而是各取其所长为我所用.

评论

- 频率学派与贝叶斯学派有不少共同点, 如都承认样本有概率分布, 概率的计算遵循共同的准则. 分歧在于把未知参数 θ 看成一个未知的固定量呢, 还是看成一个随机变量, 其余的分歧多少都是由此派生而来.
- 对上述争论有一个至高无上的裁判者, 即应用的结果. 统计方法无论在理论上如何精细高明, 总要用实践来检验. 迄今为止, 实践显示这两派得分都不低, 也正是因为它们在实际应用上表现不错, 才能各自聚合了一些追随者形成各自的学派.
- 作为一个统计学者, 可以不执著于任何一派的观点, 而是各取其所长为我所用.

评论

- 频率学派与贝叶斯学派有不少共同点, 如都承认样本有概率分布, 概率的计算遵循共同的准则. 分歧在于把未知参数 θ 看成一个未知的固定量呢, 还是看成一个随机变量, 其余的分歧多少都是由此派生而来.
- 对上述争论有一个至高无上的裁判者, 即应用的结果. 统计方法无论在理论上如何精细高明, 总要用实践来检验. 迄今为止, 实践显示这两派得分都不低, 也正是因为它们在上表现不错, 才能各自聚合了一些追随者形成各自的学派.
- 作为一个统计学者, 可以不执著于任何一派的观点, 而是各取其所长为我所用.

1 1.1 引言

- 1.1.1 从贝叶斯公式说起
- 1.1.2 三种信息
- 1.1.3 历史
- 1.1.4 古典学派和贝叶斯学派的论争

2 1.2 贝叶斯统计推断的若干基本概念

- 1.2.1 先验分布与后验分布
- 1.2.2 点估计问题
- 1.2.3 假设检验问题
- 1.2.4 区间估计问题及例

3 1.3 贝叶斯统计决策的若干基本概念

- 1.3.1 统计判决三要素
- 1.3.2 风险函数和一致最优决策函数
- 1.3.3 贝叶斯风险和贝叶斯解

先验分布与后验分布的定义

- **定义 1.2.1** (先验分布) 参数空间 Θ 上的任一概率分布都称先验分布.

本教材中用 $\pi(\theta)$ 表示 θ 的先验分布. 这里 $\pi(\theta)$ 是随机变量 θ 的概率函数 (即当 θ 为离散型随机变量时, $\pi(\theta_i)$, $i = 1, 2, \dots$ 表示事件 $\{\theta = \theta_i\}$ 的概率分布. 当 θ 为连续型随机变量时, $\pi(\theta)$ 表示的密度函数). $F^\pi(\theta)$ 表示分布函数.

- **定义 1.2.2** (后验分布) 在获得样本 x 后, θ 的后验分布就是给定 $X = x$ 条件下 θ 的条件分布, 记为 $\pi(\theta|x)$. 在有密度的情形, 它的密度函数为

$$\pi(\theta|x) = \frac{f(x, \theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}, \quad (1.2.1)$$

其中 $f(x, \theta) = f(x|\theta)\pi(\theta)$ 为 X 和 θ 的联合分布. 而

$$m(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta$$

为 X 的边缘分布.

先验分布与后验分布的定义

- **定义 1.2.1** (先验分布) 参数空间 Θ 上的任一概率分布都称先验分布.

本教材中用 $\pi(\theta)$ 表示 θ 的先验分布. 这里 $\pi(\theta)$ 是随机变量 θ 的概率函数 (即当 θ 为离散型随机变量时, $\pi(\theta_i)$, $i = 1, 2, \dots$ 表示事件 $\{\theta = \theta_i\}$ 的概率分布. 当 θ 为连续型随机变量时, $\pi(\theta)$ 表示的密度函数). $F^\pi(\theta)$ 表示分布函数.

- **定义 1.2.2** (后验分布) 在获得样本 x 后, θ 的后验分布就是给定 $X = x$ 条件下 θ 的条件分布, 记为 $\pi(\theta|x)$. 在有密度的情形, 它的密度函数为

$$\pi(\theta|x) = \frac{f(x, \theta)}{m(x)} = \frac{f(x|\theta)\pi(\theta)}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}, \quad (1.2.1)$$

其中 $f(x, \theta) = f(x|\theta)\pi(\theta)$ 为 X 和 θ 的联合分布. 而

$$m(x) = \int_{\Theta} f(x|\theta)\pi(\theta)d\theta$$

为 X 的边缘分布.

后验分布的定义

- 先验分布 $\pi(\theta)$ 是在抽取样本 X 之前对参数 θ 的认识. 由于样本也包含 θ 的信息, 一旦获得抽样信息后, 人们对 θ 的认识发生了变化和调整, 调整的结果获得对 θ 的新认识称为后验分布, 故后验分布是人们用总体信息和样本信息对先验分布作调整的结果, 是三种信息的综合.
- (1.2.1) 就是贝叶斯公式的密度形式, 它集中了总体、样本和先验三种信息中有关 θ 的一切信息, 从贝叶斯学派的观点看, 获取后验分布 $\pi(\theta|x)$ 后, 一切统计推断都须从 $\pi(\theta|x)$ 出发.
- 当 θ 是离散随机变量时, 先验分布可用先验分布列 $\{\pi(\theta_i), i = 1, 2, \dots\}$ 表示, 这时后验分布是如下离散形式

$$\pi(\theta_i|x) = f(x|\theta_i)\pi(\theta_i) / \sum_i f(x|\theta_i)\pi(\theta_i), \quad (1.2.2)$$

若 X 为离散时, 把 (1.2.2) 中密度 $f(x|\theta_i)$ 看着事件 $\{X = x|\theta_i\}$ 的概率 $P(X = x|\theta_i)$, 此时, (1.2.2) 就是贝叶斯公式.

后验分布的定义

- 先验分布 $\pi(\theta)$ 是在抽取样本 X 之前对参数 θ 的认识. 由于样本也包含 θ 的信息, 一旦获得抽样信息后, 人们对 θ 的认识发生了变化和调整, 调整的结果获得对 θ 的新认识称为后验分布, 故后验分布是人们用总体信息和样本信息对先验分布作调整的结果, 是三种信息的综合.
- (1.2.1) 就是贝叶斯公式的密度形式, 它集中了总体、样本和先验三种信息中有关 θ 的一切信息, 从贝叶斯学派的观点看, 获取后验分布 $\pi(\theta|x)$ 后, 一切统计推断都须从 $\pi(\theta|x)$ 出发.
- 当 θ 是离散随机变量时, 先验分布可用先验分布列 $\{\pi(\theta_i), i = 1, 2, \dots\}$ 表示, 这时后验分布是如下离散形式

$$\pi(\theta_i|x) = f(x|\theta_i)\pi(\theta_i) / \sum_i f(x|\theta_i)\pi(\theta_i), \quad (1.2.2)$$

若 X 为离散时, 把 (1.2.2) 中密度 $f(x|\theta_i)$ 看着事件 $\{X = x|\theta_i\}$ 的概率 $P(X = x|\theta_i)$, 此时, (1.2.2) 就是贝叶斯公式.

后验分布的定义

- 先验分布 $\pi(\theta)$ 是在抽取样本 X 之前对参数 θ 的认识. 由于样本也包含 θ 的信息, 一旦获得抽样信息后, 人们对 θ 的认识发生了变化和调整, 调整的结果获得对 θ 的新认识称为后验分布, 故后验分布是人们用总体信息和样本信息对先验分布作调整的结果, 是三种信息的综合.
- (1.2.1) 就是贝叶斯公式的密度形式, 它集中了总体、样本和先验三种信息中有关 θ 的一切信息, 从贝叶斯学派的观点看, 获取后验分布 $\pi(\theta|x)$ 后, 一切统计推断都须从 $\pi(\theta|x)$ 出发.
- 当 θ 是离散随机变量时, 先验分布可用先验分布列 $\{\pi(\theta_i), i = 1, 2, \dots\}$ 表示, 这时后验分布是如下离散形式

$$\pi(\theta_i|x) = f(x|\theta_i)\pi(\theta_i) / \sum_i f(x|\theta_i)\pi(\theta_i), \quad (1.2.2)$$

若 X 为离散时, 把 (1.2.2) 中密度 $f(x|\theta_i)$ 看着事件 $\{X = x|\theta_i\}$ 的概率 $P(X = x|\theta_i)$, 此时, (1.2.2) 就是贝叶斯公式.

点估计问题

- 在获得参数 θ 的后验分布后, θ 的估计可以用后验均值

$$\hat{\theta}_B = E(\theta|x) = \int_{\Theta} \theta \pi(\theta|x) d\theta = \frac{\int_{\Theta} \theta f(x|\theta) \pi(\theta) d\theta}{m(x)}.$$

- 也可以用后验分布的中位数或后验众数作为 θ 的估计量.

点估计问题

- 在获得参数 θ 的后验分布后, θ 的估计可以用后验均值

$$\hat{\theta}_B = E(\theta|x) = \int_{\Theta} \theta \pi(\theta|x) d\theta = \frac{\int_{\Theta} \theta f(x|\theta) \pi(\theta) d\theta}{m(x)}.$$

- 也可以用后验分布的中位数或后验众数作为 θ 的估计量.

假设检验问题

- 设假设检验问题的一般形式是：

$$H : \theta \in \Theta_H \longleftrightarrow K : \theta \in \Theta_K$$

此处 $\Theta_H \cup \Theta_K = \Theta$, 其中 Θ 是参数空间, Θ_H 是 Θ 的非空真子集

- 获得参数 θ 的后验分布后, 计算 Θ_H 和 Θ_K 的后验概率

$$p_H(x) = P(\theta \in \Theta_H | x), \quad p_K(x) = P(\theta \in \Theta_K | x).$$

若 $p_H(x) > 1/2$, 则接受 H , 否则拒绝 H .

假设检验问题

- 设假设检验问题的一般形式是：

$$H : \theta \in \Theta_H \longleftrightarrow K : \theta \in \Theta_K$$

此处 $\Theta_H \cup \Theta_K = \Theta$, 其中 Θ 是参数空间, Θ_H 是 Θ 的非空真子集

- 获得参数 θ 的后验分布后, 计算 Θ_H 和 Θ_K 的后验概率

$$p_H(x) = P(\theta \in \Theta_H | x), \quad p_K(x) = P(\theta \in \Theta_K | x).$$

若 $p_H(x) > 1/2$, 则接受 H , 否则拒绝 H .

区间估计问题

- 在求得 θ 的后验密度 $\pi(\theta|x)$ 后, 求统计量 $A(x)$ 和 $B(x)$, 使得

$$P(A(x) \leq \theta \leq B(x)|x) = 1 - \alpha,$$

其中 $\alpha \in (0, 1)$ 为给定的常数, 则称 $[A(x), B(x)]$ 为 θ 的可信度为 $1 - \alpha$ 的可信区间.

- 例 1.2.1 设随机变量 (r.v.) X 服从二项分布 $B(n, \theta)$, θ 的先验分布为 $(0, 1)$ 上的均匀分布 $U(0, 1)$, 求 θ 的贝叶斯点估计.

区间估计问题

- 在求得 θ 的后验密度 $\pi(\theta|x)$ 后, 求统计量 $A(x)$ 和 $B(x)$, 使得

$$P(A(x) \leq \theta \leq B(x)|x) = 1 - \alpha,$$

其中 $\alpha \in (0, 1)$ 为给定的常数, 则称 $[A(x), B(x)]$ 为 θ 的可信度为 $1 - \alpha$ 的可信区间.

- 例 1.2.1** 设随机变量 (r.v.) X 服从二项分布 $B(n, \theta)$, θ 的先验分布为 $(0, 1)$ 上的均匀分布 $U(0, 1)$, 求 θ 的贝叶斯点估计.

1 1.1 引言

- 1.1.1 从贝叶斯公式说起
- 1.1.2 三种信息
- 1.1.3 历史
- 1.1.4 古典学派和贝叶斯学派的论争

2 1.2 贝叶斯统计推断的若干基本概念

- 1.2.1 先验分布与后验分布
- 1.2.2 点估计问题
- 1.2.3 假设检验问题
- 1.2.4 区间估计问题及例

3 1.3 贝叶斯统计决策的若干基本概念

- 1.3.1 统计判决三要素
- 1.3.2 风险函数和一致最优决策函数
- 1.3.3 贝叶斯风险和贝叶斯解

统计决策三要素

- **样本空间和样本分布族：** 取值于样本空间 \mathcal{X} 内的随机变量 X 及其分布族 $\{F_\theta(x), \theta \in \Theta\}$ 是构成统计决策问题的第一个要素. 这里 $F_\theta(x)$ 是 X 的分布函数, θ 是未知参数, Θ 为参数空间, $\mathbf{X} = (X_1, \dots, X_n)$ 是总体 X 的简单样本.
- **行动空间：** 决策者或统计工作者对某个统计决策问题可能采取的行动所构成的非空集合, 被称为行动空间, 记为 \mathcal{D} . 在估计问题中, \mathcal{D} 由一切估计量 $d(x)$ 构成, 常取 $\mathcal{D} = \Theta$. 在检验问题中 \mathcal{D} 只有两个行动组成, 即 $\mathcal{D} = \{d_0, d_1\}$, 其中 d_0 表示接受原假设 H_0 , d_1 表示拒绝 H_0 .
- **损失函数：** 损失函数是定义于 $\Theta \times \mathcal{D}$ 上的非负函数, 记为 $L(\theta, d)$. 它表示参数为 θ 时采取行动 $d \in \mathcal{D}$ 所蒙受的损失. 常用的有平方损失、绝对值损失和线性损失等.
- **统计决策问题就是研究如何根据样本 X 的值恰当地选取行动 d 使得引起的平均损失最小.**

统计决策三要素

- **样本空间和样本分布族：** 取值于样本空间 \mathcal{X} 内的随机变量 X 及其分布族 $\{F_\theta(x), \theta \in \Theta\}$ 是构成统计决策问题的第一个要素. 这里 $F_\theta(x)$ 是 X 的分布函数, θ 是未知参数, Θ 为参数空间, $\mathbf{X} = (X_1, \dots, X_n)$ 是总体 X 的简单样本.
- **行动空间：** 决策者或统计工作者对某个统计决策问题可能采取的行动所构成的非空集合, 被称为行动空间, 记为 \mathcal{D} . 在估计问题中, \mathcal{D} 由一切估计量 $d(x)$ 构成, 常取 $\mathcal{D} = \Theta$. 在检验问题中 \mathcal{D} 只有两个行动组成, 即 $\mathcal{D} = \{d_0, d_1\}$, 其中 d_0 表示接受原假设 H_0 , d_1 表示拒绝 H_0 .
- **损失函数：** 损失函数是定义于 $\Theta \times \mathcal{D}$ 上的非负函数, 记为 $L(\theta, d)$. 它表示参数为 θ 时采取行动 $d \in \mathcal{D}$ 所蒙受的损失. 常用的有平方损失、绝对值损失和线性损失等.
- **统计决策问题就是研究如何根据样本 X 的值恰当地选取行动 d 使得引起的平均损失最小.**

统计决策三要素

- **样本空间和样本分布族：** 取值于样本空间 \mathcal{X} 内的随机变量 X 及其分布族 $\{F_\theta(x), \theta \in \Theta\}$ 是构成统计决策问题的第一个要素. 这里 $F_\theta(x)$ 是 X 的分布函数, θ 是未知参数, Θ 为参数空间, $\mathbf{X} = (X_1, \dots, X_n)$ 是总体 X 的简单样本.
- **行动空间：** 决策者或统计工作者对某个统计决策问题可能采取的行动所构成的非空集合, 被称为行动空间, 记为 \mathcal{D} . 在估计问题中, \mathcal{D} 由一切估计量 $d(x)$ 构成, 常取 $\mathcal{D} = \Theta$. 在检验问题中 \mathcal{D} 只有两个行动组成, 即 $\mathcal{D} = \{d_0, d_1\}$, 其中 d_0 表示接受原假设 H_0 , d_1 表示拒绝 H_0 .
- **损失函数：** 损失函数是定义于 $\Theta \times \mathcal{D}$ 上的非负函数, 记为 $L(\theta, d)$. 它表示参数为 θ 时采取行动 $d \in \mathcal{D}$ 所蒙受的损失. 常用的有平方损失、绝对值损失和线性损失等.
- 统计决策问题就是研究如何根据样本 X 的值恰当地选取行动 d 使得引起的平均损失最小.

统计决策三要素

- **样本空间和样本分布族：** 取值于样本空间 \mathcal{X} 内的随机变量 X 及其分布族 $\{F_\theta(x), \theta \in \Theta\}$ 是构成统计决策问题的第一个要素. 这里 $F_\theta(x)$ 是 X 的分布函数, θ 是未知参数, Θ 为参数空间, $\mathbf{X} = (X_1, \dots, X_n)$ 是总体 X 的简单样本.
- **行动空间：** 决策者或统计工作者对某个统计决策问题可能采取的行动所构成的非空集合, 被称为行动空间, 记为 \mathcal{D} . 在估计问题中, \mathcal{D} 由一切估计量 $d(x)$ 构成, 常取 $\mathcal{D} = \Theta$. 在检验问题中 \mathcal{D} 只有两个行动组成, 即 $\mathcal{D} = \{d_0, d_1\}$, 其中 d_0 表示接受原假设 H_0 , d_1 表示拒绝 H_0 .
- **损失函数：** 损失函数是定义于 $\Theta \times \mathcal{D}$ 上的非负函数, 记为 $L(\theta, d)$. 它表示参数为 θ 时采取行动 $d \in \mathcal{D}$ 所蒙受的损失. 常用的有平方损失、绝对值损失和线性损失等.
- 统计决策问题就是研究如何根据样本 \mathbf{X} 的值恰当地选取行动 d 使得引起的平均损失最小.

1. 风险函数

- **定义 1.3.1** 定义于样本空间 \mathcal{X} 内而取值于行动空间 \mathcal{D} 内的函数 $\delta = \delta(\mathbf{x})$ 称为决策函数或判决函数.
- 设 δ 是决策行动, 参数为 θ , 损失函数是 $L(\theta, \delta)$, 这个量与样本 \mathbf{X} 有关, 因而是随机的. 故采取行动 δ 的效果用平均损失去度量是相对合理的. 这就引入如下风险函数的概念.
- **定义 1.3.2** 设 $\delta(\mathbf{x})$ 是一个决策函数, 称平均损失

$$\begin{aligned} R(\theta, \delta) &= E[L(\theta, \delta(\mathbf{X}))] = \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) dF(\mathbf{x}|\theta) \\ &= \begin{cases} \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) f(\mathbf{x}|\theta) d\mathbf{x}, & \text{当 } X \text{ 为连续随机变量,} \\ \sum_{\mathbf{x} \in \mathcal{X}} L(\theta, \delta(\mathbf{x})) f(\mathbf{x}|\theta), & \text{当 } X \text{ 为离散随机变量.} \end{cases} \end{aligned}$$

为 δ 的风险函数, 此处 $F(\mathbf{x}|\theta)$ 是给定 θ 时 \mathbf{X} 的分布函数.

1. 风险函数

- **定义 1.3.1** 定义于样本空间 \mathcal{X} 内而取值于行动空间 \mathcal{D} 内的函数 $\delta = \delta(\mathbf{x})$ 称为决策函数或判决函数.
- 设 δ 是决策行动, 参数为 θ , 损失函数是 $L(\theta, \delta)$, 这个量与样本 \mathbf{X} 有关, 因而是随机的. 故采取行动 δ 的效果用平均损失去度量是相对合理的. 这就引入如下风险函数的概念.
- **定义 1.3.2** 设 $\delta(\mathbf{x})$ 是一个决策函数, 称平均损失

$$\begin{aligned} R(\theta, \delta) &= E[L(\theta, \delta(\mathbf{X}))] = \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) dF(\mathbf{x}|\theta) \\ &= \begin{cases} \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) f(\mathbf{x}|\theta) d\mathbf{x}, & \text{当 } X \text{ 为连续随机变量,} \\ \sum_{\mathbf{x} \in \mathcal{X}} L(\theta, \delta(\mathbf{x})) f(\mathbf{x}|\theta), & \text{当 } X \text{ 为离散随机变量.} \end{cases} \end{aligned}$$

为 δ 的风险函数, 此处 $F(\mathbf{x}|\theta)$ 是给定 θ 时 \mathbf{X} 的分布函数.

1. 风险函数

- **定义 1.3.1** 定义于样本空间 \mathcal{X} 内而取值于行动空间 \mathcal{D} 内的函数 $\delta = \delta(\mathbf{x})$ 称为决策函数或判决函数.
- 设 δ 是决策行动, 参数为 θ , 损失函数是 $L(\theta, \delta)$, 这个量与样本 \mathbf{X} 有关, 因而是随机的. 故采取行动 δ 的效果用平均损失去度量是相对合理的. 这就引入如下风险函数的概念.
- **定义 1.3.2** 设 $\delta(\mathbf{x})$ 是一个决策函数, 称平均损失

$$\begin{aligned} R(\theta, \delta) &= E[L(\theta, \delta(\mathbf{X}))] = \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) dF(\mathbf{x}|\theta) \\ &= \begin{cases} \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) f(\mathbf{x}|\theta) d\mathbf{x}, & \text{当 } X \text{ 为连续随机变量,} \\ \sum_{\mathbf{x} \in \mathcal{X}} L(\theta, \delta(\mathbf{x})) f(\mathbf{x}|\theta), & \text{当 } X \text{ 为离散随机变量.} \end{cases} \end{aligned}$$

为 δ 的风险函数, 此处 $F(\mathbf{x}|\theta)$ 是给定 θ 时 \mathbf{X} 的分布函数.

2. 一致最优决策函数

- 按照 Wald 的统计决策理论, 评价一个决策函数的唯一依据, 就是其风险函数. 风险函数愈小愈好.
- 定义 1.3.3 设 $\delta_1(\mathbf{x})$ 和 $\delta_2(\mathbf{x})$ 为 θ 的两个不同的决策函数, 若 $R(\theta, \delta_1(\mathbf{x})) \leq R(\theta, \delta_2(\mathbf{x}))$, 对一切 $\theta \in \Theta$, 且至少存在一个 $\theta_0 \in \Theta$ 使严格不等号成立, 则称 $\delta_1(\mathbf{x})$ 优于 $\delta_2(\mathbf{x})$. 若存在 $\delta^*(\mathbf{x})$, 使得对任一决策函数 $\delta(\mathbf{x})$ 有

$$R(\theta, \delta^*(\mathbf{x})) \leq R(\theta, \delta(\mathbf{x})), \text{ 对一切 } \theta \in \Theta,$$

则称 $\delta^*(\mathbf{x})$ 为一致最优解或一致最优决策函数.

- 对决策函数 $\delta(\mathbf{x})$, 若不存在一致优于它的决策函数, 则称 $\delta(\mathbf{x})$ 为可容许的决策函数.

2. 一致最优决策函数

- 按照 Wald 的统计决策理论, 评价一个决策函数的唯一依据, 就是其风险函数. 风险函数愈小愈好.
- 定义 1.3.3** 设 $\delta_1(\mathbf{x})$ 和 $\delta_2(\mathbf{x})$ 为 θ 的两个不同的决策函数, 若 $R(\theta, \delta_1(\mathbf{x})) \leq R(\theta, \delta_2(\mathbf{x}))$, 对一切 $\theta \in \Theta$, 且至少存在一个 $\theta_0 \in \Theta$ 使严格不等号成立, 则称 $\delta_1(\mathbf{x})$ 优于 $\delta_2(\mathbf{x})$. 若存在 $\delta^*(\mathbf{x})$, 使得对任一决策函数 $\delta(\mathbf{x})$ 有

$$R(\theta, \delta^*(\mathbf{x})) \leq R(\theta, \delta(\mathbf{x})), \text{ 对一切 } \theta \in \Theta,$$

则称 $\delta^*(\mathbf{x})$ 为一致最优解或一致最优决策函数.

- 对决策函数 $\delta(\mathbf{x})$, 若不存在一致优于它的决策函数, 则称 $\delta(\mathbf{x})$ 为可容许的决策函数.

2. 一致最优决策函数

- 按照 Wald 的统计决策理论, 评价一个决策函数的唯一依据, 就是其风险函数. 风险函数愈小愈好.
- 定义 1.3.3** 设 $\delta_1(\mathbf{x})$ 和 $\delta_2(\mathbf{x})$ 为 θ 的两个不同的决策函数, 若 $R(\theta, \delta_1(\mathbf{x})) \leq R(\theta, \delta_2(\mathbf{x}))$, 对一切 $\theta \in \Theta$, 且至少存在一个 $\theta_0 \in \Theta$ 使严格不等号成立, 则称 $\delta_1(\mathbf{x})$ 优于 $\delta_2(\mathbf{x})$. 若存在 $\delta^*(\mathbf{x})$, 使得对任一决策函数 $\delta(\mathbf{x})$ 有

$$R(\theta, \delta^*(\mathbf{x})) \leq R(\theta, \delta(\mathbf{x})), \text{ 对一切 } \theta \in \Theta,$$

则称 $\delta^*(\mathbf{x})$ 为一致最优解或一致最优决策函数.

- 对决策函数 $\delta(\mathbf{x})$, 若不存在一致优于它的决策函数, 则称 $\delta(\mathbf{x})$ 为可容许的决策函数.

贝叶斯风险和贝叶斯解

- **定义 1.3.5.** 设 $R(\theta, \delta(\mathbf{x}))$ 为风险函数, $\pi(\theta)$ 为 θ 的先验分布, 则称

$$\begin{aligned} R_{\pi}(\delta(\mathbf{x})) &= \int_{\Theta} R(\theta, \delta(\mathbf{x})) dF^{\pi}(\theta) = E^{\pi}[R(\theta, \delta(\mathbf{X}))] \\ &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) dF(\mathbf{x}|\theta) dF^{\pi}(\theta) \end{aligned}$$

是 $\delta(\mathbf{x})$ 的贝叶斯风险, 即它是将风险函数对 θ 的先验分布再求一次均值.

- **定义 1.3.6** 设 $\delta_1(\mathbf{x})$ 和 $\delta_2(\mathbf{x})$ 为 θ 的两个决策函数, $\pi(\theta)$ 为 θ 的先验分布, 若 $R_{\pi}(\delta_1(\mathbf{x})) \leq R_{\pi}(\delta_2(\mathbf{x}))$, 则称 $\delta_1(\mathbf{x})$ 在贝叶斯风险下优于 $\delta_2(\mathbf{x})$. 若存在 $\delta^*(\mathbf{x})$, 使得对任一决策函数 $\delta(\mathbf{x})$ 有

$$R_{\pi}(\delta^*(\mathbf{x})) \leq R_{\pi}(\delta(\mathbf{x})),$$

则称 $\delta^*(\mathbf{x})$ 为所考虑的统计判决问题的贝叶斯解.

贝叶斯风险和贝叶斯解

- **定义 1.3.5.** 设 $R(\theta, \delta(\mathbf{x}))$ 为风险函数, $\pi(\theta)$ 为 θ 的先验分布, 则称

$$\begin{aligned} R_{\pi}(\delta(\mathbf{x})) &= \int_{\Theta} R(\theta, \delta(\mathbf{x})) dF^{\pi}(\theta) = E^{\pi}[R(\theta, \delta(\mathbf{X}))] \\ &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(\mathbf{x})) dF(\mathbf{x}|\theta) dF^{\pi}(\theta) \end{aligned}$$

是 $\delta(\mathbf{x})$ 的贝叶斯风险, 即它是将风险函数对 θ 的先验分布再求一次均值.

- **定义 1.3.6** 设 $\delta_1(\mathbf{x})$ 和 $\delta_2(\mathbf{x})$ 为 θ 的两个决策函数, $\pi(\theta)$ 为 θ 的先验分布, 若 $R_{\pi}(\delta_1(\mathbf{x})) \leq R_{\pi}(\delta_2(\mathbf{x}))$, 则称 $\delta_1(\mathbf{x})$ 在贝叶斯风险下优于 $\delta_2(\mathbf{x})$. 若存在 $\delta^*(\mathbf{x})$, 使得对任一决策函数 $\delta(\mathbf{x})$ 有

$$R_{\pi}(\delta^*(\mathbf{x})) \leq R_{\pi}(\delta(\mathbf{x})),$$

则称 $\delta^*(\mathbf{x})$ 为所考虑的统计判决问题的贝叶斯解.